aws

Guía de administración

Amazon EMR

Copyright © 2023 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon EMR: Guía de administración

Copyright © 2023 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon de ninguna manera que pueda causar confusión entre los clientes y que menosprecie o desacredite a Amazon. Todas las demás marcas comerciales que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, conectados o patrocinados por Amazon.

Table of Contents

¿Qué es Amazon EMR? 1
Información general 1
Descripción de los clústeres y los nodos 2
Envío de trabajo a un clúster 3
Procesamiento de datos 3
Descripción del ciclo de vida del clúster 5
Ventajas 6
Ahorro de costos 7
AWSIntegración de 8
Implementación 8
Escalabilidad y flexibilidad 9
Fiabilidad 9
Seguridad 10
Supervisión 12
Interfaces de administración 12
Arquitectura 13
Almacenamiento 13
Administración de recursos de clúster 14
Marcos de procesamiento de datos 15
Aplicaciones y programas 16
Configuración de Amazon EMR 17
Registro en una Cuenta de AWS 17
Crear un usuario administrativo 17
Crear un par de claves de Amazon EC2 para SSH 18
Pasos siguientes 19
Explicación introductoria 20
Información general 20
Paso 1: planificar y configurar 21
Prepare el almacenamiento para Amazon EMR 21
Preparar una aplicación con datos de entrada para Amazon EMR 22
Lanzar un clúster de Amazon EMR 24
Paso 2: administrar 28
Enviar un trabajo a Amazon EMR 28
Ver los resultados 33

Paso 3: Limpieza 38
Terminar su clúster 38
Eliminar recursos de S3 41
Pasos siguientes 41
Explorar las aplicaciones de macrodatos para Amazon EMR 41
Planificar el hardware, las redes y la seguridad de los clústeres 42
Administrar clústeres 42
Utilizar una interfaz diferente 42
Consultar el blog técnico de EMR 42
Novedades de la consola 43
¿En qué consola estoy? 43
Uso de la consola antigua 44
Resumen de las diferencias 44
Compatibilidad de clústeres entre la consola antigua y la nueva 45
Diferencias a la hora de crear clústeres 45
Diferencias a la hora de enumerar y buscar clústeres 47
Diferencias al ver o editar los detalles del clúster 48
Diferencias al trabajar con configuraciones de seguridad 49
Amazon EMR Studio 51
Características principales 51
Historial de características 52
Cómo funciona 53
Autenticación e inicio de sesión de los usuarios 54
Control de acceso 57
Workspaces 58
Almacenamiento de cuadernos 59
Consideraciones 60
Consideraciones 60
Problemas conocidos 61
Limitaciones de características 63
Límites de los servicios 63
Prácticas recomendadas sobre VPC y subredes 63
Requisitos del clúster 64
Configuración de EMR Studio 66
Permisos de administrador para crear un EMR Studio 67
Configuración de un Amazon EMR Studio 72

Administrar un estudio ... 134
Controlar el tráfico de red de EMR Studio ... 142
Crear plantillas de clúster ... 145
Acceso y permisos para los repositorios basados en Git ... 151
Optimizar los trabajos de Spark ... 155
Utilizar un EMR Studio ... 156
Conceptos básicos de los espacios de trabajo ... 157
Colaboración en el espacio de trabajo ... 165
Ejecutar un espacio de trabajo con un rol de tiempo de ejecución ... 168
Ejecutar los cuadernos del espacio de trabajo mediante programación ... 174
Examinar los datos con SQL Explorer ... 174
Asociar computación a un espacio de trabajo ... 175
Vincular repositorios de Git ... 182
Depurar aplicaciones y trabajos ... 186
Instalar kernels y bibliotecas ... 191
Comandos magic ... 192
Usar cuadernos multilingües con kernels de Spark ... 202
EMR Notebooks ... 205
Cuadernos en la nueva consola ... 206
Acerca de la transición ... 206
¿Qué necesita hacer? ... 207
Ventajas del espacio de trabajo ... 207
Permisos necesarios ... 208
Consideraciones ... 209
Requisitos del clúster ... 209
Diferencias en capacidades por versión de clúster ... 210
Límites para cuadernos asociados de forma simultánea ... 212
Versiones de cuaderno de Jupyter y Python ... 212
Consideraciones en torno a la seguridad ... 212
Creación de un bloc de notas ... 213
Uso de los Cuadernos de EMR ... 216
Descripción de los estados del cuaderno ... 217
Uso del editor de cuadernos ... 218
Cambio de clústeres ... 220
Eliminación de cuadernos y archivos de cuadernos ... 220
Uso compartido de archivos de cuadernos ... 221

Ejecución programática 222
Información general 223
Permisos 223
Limitaciones 224
Ejemplos 225
Ejemplos de comandos de la CLI 225
Ejemplo de script del SDK de Boto3 231
Script de ejemplo de Ruby 234
Suplantación de usuarios para Spark 236
Configuración de la suplantación de usuarios de Spark 237
Uso del widget de supervisión de trabajos de Spark 238
Seguridad 239
Instalación y uso de kernels y bibliotecas 240
..... 241
Instalación de kernels y bibliotecas de Python en un nodo principal del clúster 241
Consideraciones y limitaciones de las bibliotecas adaptadas al cuaderno 244
Uso de bibliotecas adaptadas al cuaderno 244
Asociación de repositorios basados en Git con Cuadernos de EMR 246
Requisitos previos y consideraciones 247
Agregar un repositorio basado en Git a Amazon EMR 250
Actualizar o eliminar un repositorio basado en Git 253
Vincular o desvincular un repositorio basado en Git 255
Crear un nuevo cuaderno con un repositorio de Git asociado 258
Uso de repositorios de Git en un cuaderno 259
Planificar y configurar clústeres 260
Lanzamiento rápido de un clúster 260
Configurar la ubicación del clúster y el almacenamiento de datos 262
Elegir una región de AWS 262
Utilizar almacenamiento y sistemas de archivos 264
Preparar datos de entrada 268
Configurar una ubicación de salida 284
Planificación y configuración de nodos principales 291
Aplicaciones y características compatibles 292
Lanzar un clúster de Amazon EMR con varios nodos principales 302
Integración de Amazon EMR con grupos de ubicación de EC2 305
Consideraciones y prácticas recomendadas 311

Clústeres de EMR en AWS Outposts 313
Requisitos previos 313
Limitaciones 314
Consideraciones sobre la conectividad de red 315
Creación de un clúster de Amazon EMR en AWS Outposts 315
Clústeres de EMR en Zonas locales de AWS 317
Tipos de instancias admitidas 318
Creación de un clúster de Amazon EMR en zonas locales 318
Configuración de Docker 320
Registros de Docker 321
Configuración de registros de Docker 322
Configuración de YARN para acceder a Amazon ECR en EMR 6.0.0 y versiones anteriores 323
Control de la terminación de los clústeres 325
Configuración de un clúster para que continúe o termine después de la ejecución de pasos 326
Uso de una política de terminación automática 329
Uso de la protección de terminación 336
Uso de las AMI 344
Información general 344
Uso de la AMI predeterminada 345
Uso de una AMI personalizada 378
Cambio de versión de Amazon Linux al crear un clúster 392
Especificación del tamaño de volumen de dispositivo raíz de Amazon EBS 393
Configuración de software del clúster 396
Creación de acciones de arranque 396
Configuración del hardware y las redes de los clústeres 402
Comprender los tipos de nodos 403
Configuración de instancias de Amazon EC2 406
Configurar el registro y la depuración de un clúster 856
Archivos de registro predeterminados 856
Archivar archivos de registro en Amazon S3 857
Localización de registros 863
Habilitar la herramienta de depuración 864
Información sobre la opción de depuración 866
Etiquetado de clústeres 867

Restricciones de las etiquetas ... 869
Recursos de etiquetas para facturación ... 869
Agregar etiquetas a un clúster ... 870
Ver etiquetas en un clúster ... 873
Eliminar etiquetas de un clúster ... 874
Integración de controladores y aplicaciones de terceros ... 876
Utilizar herramientas de inteligencia empresarial con Amazon EMR ... 876
Seguridad ... 878
Configuraciones de seguridad ... 878
Protección de los datos ... 879
AWS Identity and Access Management con Amazon EMR ... 879
Kerberos ... 879
Lake Formation ... 880
Secure Socket Shell (SSH) ... 880
Grupos de seguridad de Amazon EC2 ... 880
Actualizaciones de la AMI de Amazon Linux ... 880
Uso de configuraciones de seguridad para configurar la seguridad del clúster ... 881
Creación de una configuración de seguridad ... 881
Especificación de una configuración de seguridad para un clúster ... 912
Protección de los datos ... 913
Cifrado de datos en reposo y en tránsito ... 914
IAM con Amazon EMR ... 929
Público ... 930
Autenticación con identidades ... 930
Administración de acceso mediante políticas ... 934
Cómo funciona Amazon EMR con IAM ... 937
Roles en tiempo de ejecución para los pasos de Amazon EMR ... 945
Configuración de roles de servicio para Amazon EMR ... 953
Ejemplos de políticas basadas en identidad ... 1009
Autenticación en nodos de clúster ... 1048
Uso de un par de claves de EC2 para credenciales de SSH ... 1049
Uso de la autenticación Kerberos ... 1049
Uso de la autenticación LDAP ... 1089
Integración de Amazon EMR con Lake Formation ... 1101
Cómo funciona Amazon EMR con Lake Formation ... 1101
Requisitos previos ... 1102

Habilitación de Lake Formation con Amazon EMR ... 1103
Hudi y Lake Formation ... 1106
Consideraciones ... 1109
Integración de Amazon EMR con Apache Ranger ... 1109
Información general de Ranger ... 1110
Compatibilidad y limitaciones de la aplicación ... 1113
Configuración de Amazon EMR para Apache Ranger ... 1115
Complementos de Apache Ranger ... 1134
Solución de problemas con Apache Ranger ... 1162
Control del tráfico de red con grupos de seguridad ... 1166
Uso de grupos de seguridad administrados por Amazon EMR ... 1168
Uso de grupos de seguridad adicionales ... 1179
Especificación de los grupos de seguridad ... 1180
Grupos de seguridad para Cuadernos de Amazon EMR ... 1184
Bloqueo de acceso público ... 1186
Validación de conformidad ... 1192
Resiliencia ... 1193
Seguridad de infraestructuras ... 1194
Conexión a Amazon EMR mediante un punto de conexión de VPC de tipo interfaz ... 1194
Administrar clústeres ... 1200
Conectarse a un clúster ... 1200
Antes de conectarse ... 1201
Conectarse al nodo principal mediante SSH ... 1204
Enviar trabajo a un clúster ... 1231
Agregar pasos con la consola ... 1232
Adición de pasos con la CLI ... 1236
Ejecutar varios pasos ... 1238
Visualización de pasos ... 1239
Cancelación de pasos ... 1240
Ver y monitorizar un clúster ... 1242
Ver el estado y los detalles del clúster ... 1243
Depuración de pasos mejorada ... 1250
Ver el historial de aplicaciones ... 1253
Ver archivos de registro de ... 1263
Ver instancias del clúster en Amazon EC2 ... 1269
Eventos y métricas de CloudWatch ... 1270

Ver métricas de aplicaciones de clúster con Ganglia 1338
Registro de llamadas a la API de Amazon EMR en AWS CloudTrail 1338
Usar el escalado de clústeres 1342
Consideraciones 1343
Escalado administrado 1343
Escalado automático con una política personalizada 1371
Cambiar el tamaño de un clúster en ejecución 1384
Tiempos de espera de aprovisionamiento 1393
Reducción vertical del clúster 1398
Terminar un clúster 1402
Terminación desde la consola 1403
Terminación desde la CLI 1405
Terminación desde la API 1406
Clonar un clúster 1406
Automatizar clústeres periódicos con AWS Data Pipeline 1408
Solucionar problemas de clústeres 1410
Herramientas de solución de problemas 1410
Ver detalles del clúster 1411
Ver detalles de errores 1411
Ejecutar scripts y configurar procesos 1412
Ver archivos de registro de 1412
Supervisar el rendimiento del clúster 1413
Ver y reiniciar procesos 1413
Ver procesos en ejecución 1414
Detener y reiniciar procesos 1415
Errores comunes 1418
Códigos de error 1419
Errores de recursos 1433
Errores de entrada y salida 1444
Errores de permisos 1447
Errores de clúster de Hive 1449
Errores de VPC 1450
Errores de clúster de streaming 1455
Errores de clúster JAR personalizados 1456
Errores de AWS GovCloud (Oeste de EE. UU.) 1457
Buscar un clúster que falta 1457

resolución de problemas de clústeres con errores 1458
Paso 1: recopilar datos sobre el problema 1458
Paso 2: comprobar el entorno 1459
Paso 3: comprobar el último cambio de estado 1461
Paso 4: examinar los archivos de registro 1461
Paso 5: comprobar el clúster paso a paso 1462
Solucionar problemas de clústeres lentos 1463
Paso 1: recopilar datos sobre el problema 1464
Paso 2: comprobar el entorno 1465
Paso 3: examinar los archivos de registro 1466
Paso 4: comprobar el clúster y el estado de la instancia 1468
Paso 5: comprobar si hay grupos suspendidos 1470
Paso 6: revisar los ajustes de configuración 1470
Paso 7: examinar los datos de entrada 1473
Solucionar problemas de un clúster de Lake Formation 1473
No se permite el acceso al lago de datos 1474
Vencimiento de la sesión 1474
No hay permisos para el usuario en la tabla solicitada 1474
Consulta de datos entre cuentas compartidos con Lake Formation 1475
Inserción, creación y alteración de tablas 1476
Escribir aplicaciones que lanzan y administran clústeres 1477
Muestra completa de código fuente de Java de Amazon EMR 1477
Conceptos comunes para las llamadas a la API 1481
Puntos de conexión para Amazon EMR 1482
Especificar parámetros de clúster en Amazon EMR 1482
Zonas de disponibilidad en Amazon EMR 1483
Cómo utilizar archivos y bibliotecas adicionales en clústeres de Amazon EMR 1483
Utilizar los SDK para llamar a las API de Amazon EMR 1484
Usar el AWS SDK for Java para crear un clúster de Amazon EMR 1484
Administrar las Service Quotas de Amazon EMR 1487
¿Qué son las Service Quotas de Amazon EMR? 1487
Cómo administrar Service Quotas para Amazon EMR 1488
Cuándo configurar los eventos de EMR en CloudWatch 1488
Glosario de AWS 1492

¿Qué es Amazon EMR?

Amazon EMR (conocido antes como Amazon Elastic MapReduce) es una plataforma de clúster administrada que simplifica la ejecución de los marcos de trabajo de macrodatos, tales como Apache Hadoop y Apache Spark, en AWS para procesar y analizar grandes cantidades de datos. Mediante el uso de estos marcos de trabajo y proyectos de código abierto relacionados, puede procesar datos para fines de análisis y cargas de trabajo de inteligencia empresarial. Además, Amazon EMR le permite transformar y trasladar grandes cantidades de datos hacia y desde otros almacenes de datos y bases de datos de AWS, tales como Amazon Simple Storage Service (Amazon S3) y Amazon DynamoDB.

Si está utilizando Amazon EMR por primera vez, le recomendamos que comience leyendo lo siguiente, además de esta sección:

- Amazon EMR: esta página de servicio ofrece los puntos destacados de Amazon EMR, los detalles del producto y la información de precios.
- Tutorial: Introducción a Amazon EMR: este tutorial le ayudará a empezar a utilizar Amazon EMR rápidamente.

En esta sección

- Información general de Amazon EMR
- Beneficios de usar Amazon EMR
- Descripción general de la arquitectura de Amazon EMR

Información general de Amazon EMR

En este tema, se ofrece información general de los clústeres de Amazon EMR, lo que incluye cómo enviar trabajos a un clúster, la forma en que se procesan los datos y los distintos estados por los que pasa el clúster durante el procesamiento.

En este tema

- Descripción de los clústeres y los nodos
- Envío de trabajo a un clúster
- Procesamiento de datos
- Descripción del ciclo de vida del clúster

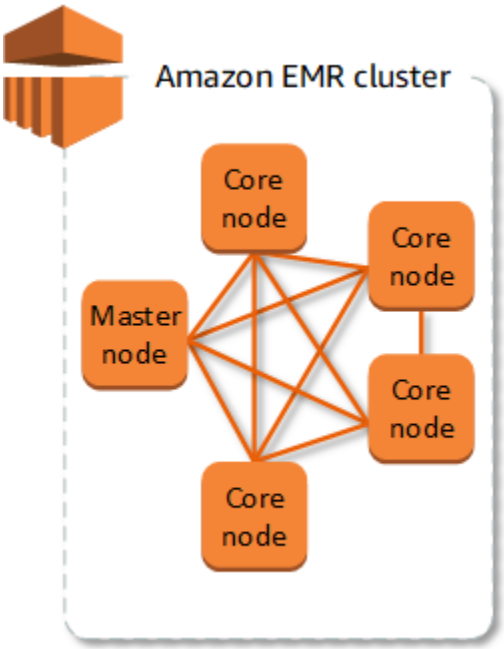
Descripción de los clústeres y los nodos

El componente central de Amazon EMR es el clúster. Un clúster es una colección de instancias de Amazon Elastic Compute Cloud (Amazon EC2). Cada instancia del clúster se denomina nodo. Cada nodo tiene un rol dentro del clúster, conocido como el tipo de nodo. Amazon EMR también instala distintos componentes de software en cada tipo de nodo, lo que proporciona a cada nodo un rol en una aplicación distribuida, como Apache Hadoop.

Los tipos de nodos en Amazon EMR son los siguientes:

- Nodo principal: un nodo que administra el clúster mediante la ejecución de componentes de software para coordinar la distribución de datos y las tareas entre los demás nodos para su procesamiento. El nodo principal hacer un seguimiento del estado de las tareas y supervisa el estado del clúster. Cada clúster tiene un nodo principal y se puede crear un clúster de un solo nodo con solo el nodo principal.
- Nodo secundario: un nodo con componentes de software que ejecutan tareas y almacenan datos en el Hadoop Distributed File System (HDFS) del clúster. Los clústeres de varios nodos tienen al menos un nodo secundario.
- Nodo de tareas: un nodo con componentes de software que solo ejecuta tareas y no almacena datos en HDFS. Los nodos de tareas son opcionales.

En el siguiente diagrama, se representa un clúster con un nodo principal y cuatro nodos secundarios.



Envío de trabajo a un clúster

Cuando se ejecuta un clúster en Amazon EMR, dispone de varias opciones sobre cómo especificar el trabajo que hay que llevar a cabo.

- Proporcionar toda la definición del trabajo que hay que realizar en funciones que debe especificar como pasos al crear un clúster. Esto se realiza normalmente para clústeres que procesan una cantidad definida de datos y, a continuación, terminan cuando se completa el procesamiento.
- Cree un clúster de ejecución prolongada y utilice la consola de Amazon EMR, la API de Amazon EMR o la AWS CLI para enviar pasos, que pueden contener uno o varios trabajos. Para obtener más información, consulte Enviar trabajo a un clúster.
- Cree un clúster, conéctese al nodo principal y a los demás nodos según sea necesario mediante SSH y utilice las interfaces que proporcionan las aplicaciones instaladas para llevar a cabo tareas y enviar consultas, ya sea de forma interactiva o con scripts. Para obtener más información, consulte la Guía de publicación de Amazon EMR.

Procesamiento de datos

Al lanzar el clúster, puede elegir los marcos de trabajo y las aplicaciones que desea instalar para sus necesidades de procesamiento de datos. Para procesar datos en el clúster de Amazon EMR, puede enviar los trabajos o las consultas directamente a las aplicaciones instaladas, o puede ejecutar pasos en el clúster.

Envío de trabajos directamente a las aplicaciones

Puede enviar trabajos e interactuar directamente con el software que está instalado en el clúster de Amazon EMR. Para ello, normalmente se conecta al nodo principal a través de una conexión segura y accede a las interfaces y herramientas que están disponibles para que el software se ejecute directamente en el clúster. Para obtener más información, consulte Conectarse a un clúster.

Ejecución de pasos para procesar datos

Puede enviar uno o varios pasos ordenados a un clúster de Amazon EMR. Cada paso es una unidad de trabajo que contiene instrucciones para manipular los datos para su procesamiento por el software instalado en el clúster.

A continuación se muestra un proceso de ejemplo que utiliza cuatro pasos:

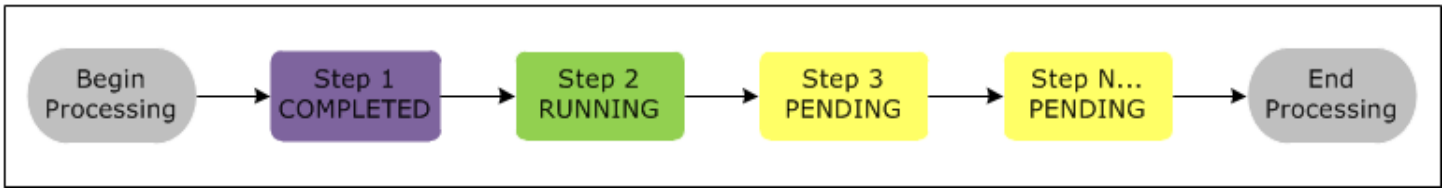
1. Enviar un conjunto de datos de entrada para procesamiento.
2. Procesar la salida del primer paso mediante un programa de Pig.
3. Procesar un segundo conjunto de datos de entrada mediante un programa de Hive.
4. Escribir un conjunto de datos de salida.

Por lo general, cuando se procesan datos en Amazon EMR, la entrada son datos almacenados como archivos en el sistema de archivos subyacente elegido, como Amazon S3 o HDFS. Estos datos se transfieren de un paso al siguiente en la secuencia de procesamiento. El último paso escribe los datos de salida en una ubicación especificada, como un bucket de Amazon S3.

Los pasos se ejecutan en la siguiente secuencia:

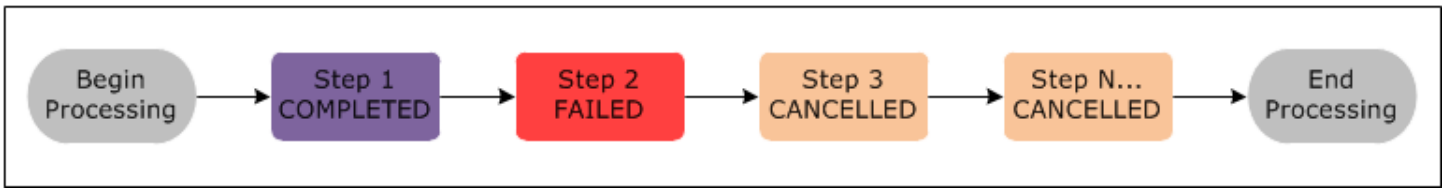
1. Se envía una solicitud para empezar los pasos de procesamiento.
2. El estado de todos los pasos se establece en PENDING (Pendiente).
3. Cuando se inicia el primer paso de la secuencia, su estado cambia a RUNNING (En ejecución). Los demás pasos permanecen en el estado PENDING (Pendiente).
4. Una vez que finaliza el primer paso, su estado cambia a COMPLETED (Completado).
5. El siguiente paso de la secuencia se inicia y su estado cambia a RUNNING (En ejecución). Una vez finalizado, su estado cambia a COMPLETED (Completado).
6. Este patrón se repite para cada paso hasta que se completen todos y el procesamiento finaliza.

El siguiente diagrama representa la secuencia de pasos y cambios de estado de los pasos a medida que se procesan.



Si un paso falla durante el procesamiento, su estado cambia a ERROR. Puede determinar lo que ocurre para cada paso. De forma predeterminada, los pasos restantes de la secuencia se establecen en CANCELADO y no se ejecutan si falla un paso anterior. También puede elegir omitir el error y permitir que los pasos restantes continúen o terminar el clúster inmediatamente.

El siguiente diagrama representa la secuencia de pasos y el cambio de estado predeterminado cuando un paso produce un error durante el procesamiento.



Descripción del ciclo de vida del clúster

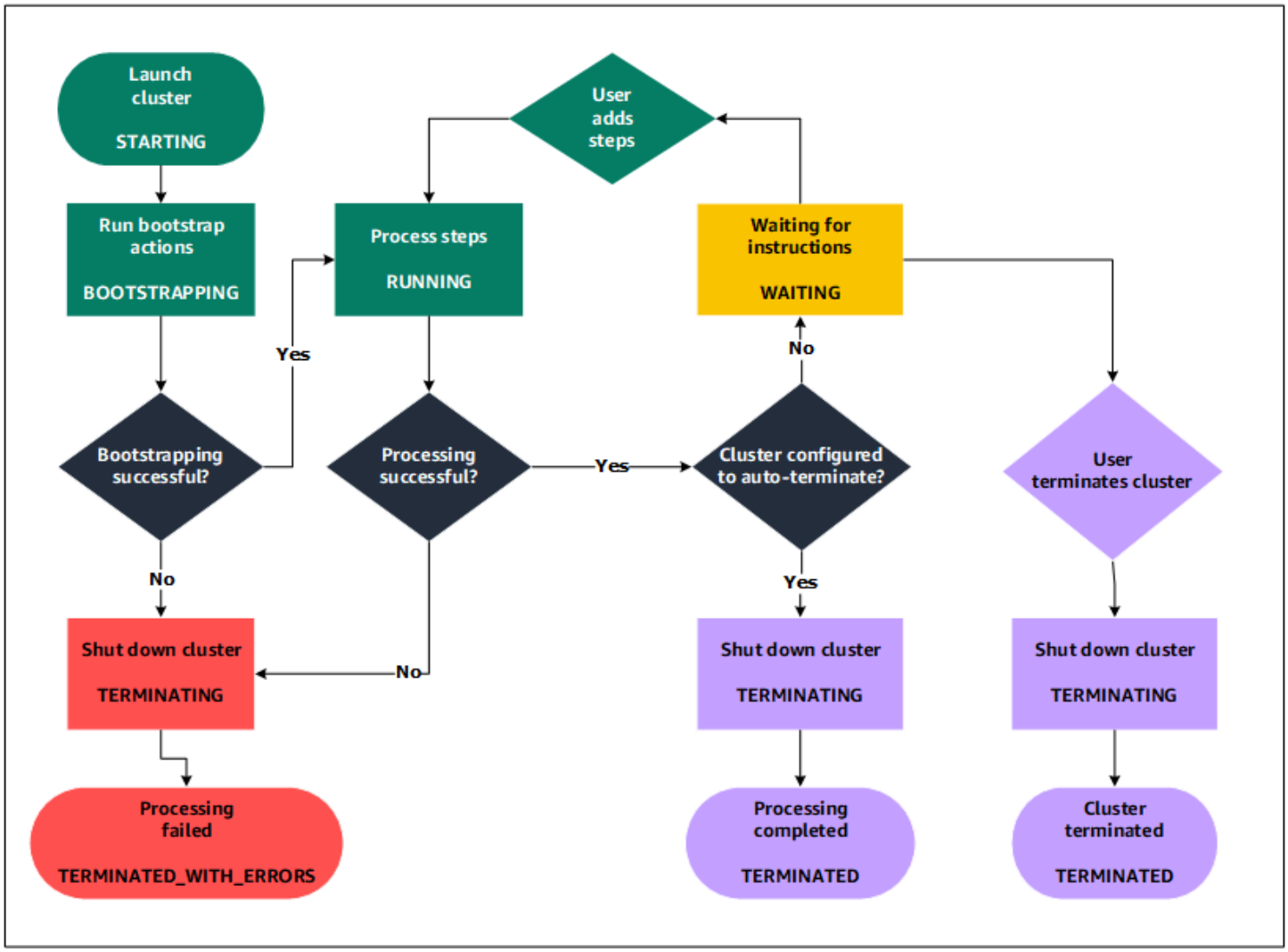
Un clúster de Amazon EMR correcto sigue este proceso:

1. Amazon EMR aprovisiona primero instancias de EC2 en el clúster para cada instancia de acuerdo con sus especificaciones. Para obtener más información, consulte Configuración del hardware y las redes de los clústeres. Para todas las instancias, Amazon EMR utiliza la AMI predeterminada de Amazon EMR o la AMI de Amazon Linux personalizada que se especifique. Para obtener más información, consulte Uso de una AMI personalizada. Durante esta fase, el estado del clúster es `STARTING`.
2. Amazon EMR ejecuta las acciones de arranque que especifique en cada instancia. Puede utilizar acciones de arranque para instalar aplicaciones personalizadas y realizar las personalizaciones que necesite. Para obtener más información, consulte Crear acciones de arranque para instalar software adicional. Durante esta fase, el estado del clúster es `BOOTSTRAPPING`.
3. Amazon EMR instala las aplicaciones nativas que especifique al crear el clúster, tales como Hive, Hadoop, Spark, etc.
4. Cuando las acciones de arranque se han completado correctamente y las aplicaciones nativas se han instalado correctamente, el estado del clúster es `RUNNING`. En este punto, puede conectarse a las instancias del clúster y el clúster ejecutará por orden todos los pasos que haya especificado al crear el clúster. Puede enviar pasos adicionales, que se ejecutarán después de los pasos anteriores. Para obtener más información, consulte Enviar trabajo a un clúster.
5. Una vez que los pasos se han ejecutado correctamente, el clúster pasa al estado `WAITING`. Si un clúster está configurado para que se termine automáticamente después de que se haya completado el último paso, pasa a un estado `TERMINATING` y luego al estado `TERMINATED`. Si el clúster está configurado para esperar, debe apagarlo manualmente cuando ya no lo necesite. Después de terminar manualmente el clúster, pasa al estado `TERMINATING` y, a continuación, al estado `TERMINATED`.

Un error durante el ciclo de vida del clúster hace que Amazon EMR termine el clúster y todas sus instancias, a menos que habilite la protección de terminación. Si se termina un clúster debido a un error, todos los datos almacenados en el clúster se eliminan y el estado del clúster se establece en

`TERMINATED_WITH_ERRORS`. Si ha habilitado la protección de terminación, puede recuperar los datos del clúster y después eliminar la protección de terminación y terminar el clúster. Para obtener más información, consulte Uso de la protección de terminación.

El siguiente diagrama representa el ciclo de vida de un clúster y cómo cada etapa del ciclo de vida se asocia a un determinado estado del clúster.



Beneficios de usar Amazon EMR

El uso de Amazon EMR conlleva muchos beneficios. En esta sección se ofrece información general sobre estos beneficios y enlaces a información adicional que le ayudan a seguir explorando.

Temas

- Ahorro de costos

- AWSIntegración de
- Implementación
- Escalabilidad y flexibilidad
- Fiabilidad
- Seguridad
- Supervisión
- Interfaces de administración

Ahorro de costos

Los precios de Amazon EMR dependen del tipo de instancia y del número de instancias de Amazon EC2 que implemente y de la región en la que lance el clúster. Los precios bajo demanda ofrecen tarifas reducidas, pero puede reducir aún más el costo mediante la adquisición de instancias reservadas o instancias de spot. Las instancias de spot pueden ofrecer ahorros significativos. En algunos casos, de solo una décima parte de los precios bajo demanda.

> **Note**
>
> Si utiliza Amazon S3, Amazon Kinesis o DynamoDB con el clúster de EMR, hay cargos adicionales por estos servicios que se facturan por separado de su uso de Amazon EMR.

> **Note**
>
> Al configurar un clúster de Amazon EMR en una subred privada, se recomienda que también configure puntos de conexión de VPC para Amazon S3. Si su clúster de EMR se encuentra en una subred privada sin puntos de conexión de VPC para Amazon S3, incurrirá en cargos adicionales de puerta de enlace de NAT asociados al tráfico de S3, ya que el tráfico entre su clúster de EMR y S3 no permanecerá dentro de su VPC.

Para obtener más información sobre las opciones de precios y los detalles, consulte Precios de Amazon EMR.

AWSIntegración de

Amazon EMR se integra con otros servicios de AWS para proporcionar capacidades y funcionalidades relacionadas con las redes, el almacenamiento, la seguridad, etc. para el clúster. La siguiente lista proporciona diversos ejemplos de esta integración:

- Amazon EC2 para las instancias que componen los nodos del clúster
- Amazon Virtual Private Cloud (Amazon VPC) para configurar la red virtual en la que lanzar sus instancias
- Amazon S3 para almacenar los datos de entrada y de salida
- Amazon CloudWatch para supervisar el rendimiento del clúster y configurar alarmas
- AWS Identity and Access Management (IAM) para configurar los permisos
- AWS CloudTrail para auditar solicitudes realizadas al servicio
- AWS Data Pipeline para programar e iniciar los clústeres
- AWS Lake Formation para descubrir, catalogar y proteger datos en un lago de datos de Amazon S3

Implementación

El clúster de EMR consta de instancias EC2, que realizan el trabajo que envía a su clúster. Al lanzar el clúster, Amazon EMR configura las instancias con las aplicaciones que elija, como Apache Hadoop o Spark. Elija el tamaño y el tipo de instancia que mejor se adapte a las necesidades de procesamiento de su clúster: procesamiento por lotes, consultas de baja latencia, streaming de datos o almacenamiento de datos de gran tamaño. Para obtener más información sobre los tipos de instancias disponibles para Amazon EMR, consulte Configuración del hardware y las redes de los clústeres.

Amazon EMR ofrece distintas formas de configurar el software en su clúster. Por ejemplo, puede instalar una versión de Amazon EMR con un conjunto seleccionado de aplicaciones que puede incluir marcos de trabajo versátiles como, por ejemplo, Hadoop y aplicaciones como Hive, Pig o Spark. También puede instalar una de las diversas distribuciones de MapR. Amazon EMR utiliza Amazon Linux, por lo que también puede instalar software en su clúster de forma manual o mediante el administrador de paquetes yum o desde el origen. Para obtener más información, consulte Configuración de software del clúster.

Escalabilidad y flexibilidad

Amazon EMR proporciona flexibilidad para aumentar o reducir el escalado del clúster a medida que cambien sus necesidades informáticas. Puede cambiar el tamaño del clúster para añadir instancias durante los picos de cargas de trabajo y eliminar instancias para controlar los costos cuando desaparezcan los picos de cargas de trabajo. Para obtener más información, consulte Cambiar manualmente el tamaño de un clúster en ejecución.

Amazon EMR también ofrece la opción de ejecutar varios grupos de instancias, para que puede utilizar instancias bajo demanda en un grupo para garantizar la potencia de procesamiento junto con las instancias de spot en otro grupo para completar los trabajos con mayor rapidez y para reducir costos. También puede combinar diversos tipos de instancias para aprovechar mejor los precios de un tipo de instancia de spot sobre otros. Para obtener más información, consulte ¿Cuándo se deben utilizar las instancias de spot?.

Además, Amazon EMR proporciona la flexibilidad necesaria para utilizar diversos sistemas de archivos para sus datos de entrada, de salida e intermedios. Por ejemplo, podría elegir el Sistema de archivos distribuido de Hadoop (HDFS), que se ejecuta en los nodos principal y secundarios del clúster para el procesamiento de datos que no es necesario almacenar más allá del ciclo de vida del clúster. Podría elegir el sistema de archivos de EMR (EMRFS) para utilizar Amazon S3 como capa de datos para aplicaciones que se ejecutan en su clúster para que pueda separar la computación y el almacenamiento, así como para conservar los datos cuando finaliza el ciclo de vida de su clúster. EMRFS ofrece el beneficio añadido de permitirle aumentar o reducir el escalado de sus necesidades de computación y almacenamiento de manera independiente. Para escalar sus necesidades de computación, puede cambiar el tamaño de su clúster y puede escalar sus necesidades de almacenamiento mediante el uso de Amazon S3. Para obtener más información, consulte Utilizar almacenamiento y sistemas de archivos.

Fiabilidad

Amazon EMR supervisa los nodos del clúster y termina y sustituye automáticamente una instancia en caso de que se produzca algún error.

Amazon EMR proporciona opciones de configuración que controlan cómo se termina el clúster: de forma automática o manual. Si configura el clúster para terminarlo de forma automática, se termina después de que se completen todos los pasos. Esto se conoce como clúster transitorio. Sin embargo, puede configurar el clúster para que se siga ejecutando una vez completado el procesamiento, para que pueda elegir terminarlo manualmente cuando ya no lo necesite. O bien,

puede crear un clúster, interactuar con las aplicaciones instaladas directamente y, a continuación, terminar el clúster manualmente cuando deje de necesitarlo. Los clústeres de estos ejemplos se denominan clústeres de ejecución prolongada.

Además, puede configurar la protección de terminación para evitar que las instancias del clúster se terminen por errores o problemas durante el procesamiento. Cuando se habilita la protección de terminación, puede recuperar los datos desde instancias antes de la terminación. La configuración predeterminada de estas opciones varía en función de si lanza el clúster a través de la consola, la CLI o la API. Para obtener más información, consulte Uso de la protección de terminación.

Seguridad

Amazon EMR utiliza otros servicios de AWS, como IAM y Amazon VPC, así como características tales como los pares de claves de Amazon EC2, para ayudarle a proteger los clústeres y los datos.

IAM

Amazon EMR se integra con IAM para administrar los permisos. Puede definir permisos mediante las políticas de IAM, que se asocian a usuarios o grupos de IAM. Los permisos que defina en la política determinan las acciones que dichos usuarios o miembros del grupo pueden realizar y los recursos a los que pueden acceder. Para obtener más información, consulte Cómo funciona Amazon EMR con IAM.

Además, Amazon EMR utiliza roles de IAM para el propio servicio de Amazon EMR y el perfil de instancia de EC2 para las instancias. Estos roles conceden permisos al servicio y a las instancias para obtener acceso a otros servicios de AWS en su nombre. Hay un rol predeterminado para el servicio de Amazon EMR y un rol predeterminado para el perfil de instancia de EC2. Los roles predeterminados utilizan políticas administradas de AWS, que se crean automáticamente la primera vez que lanza un clúster de EMR desde la consola y elige los permisos de forma predeterminada. También puede crear los roles de IAM predeterminados desde la AWS CLI. Si desea administrar los permisos en lugar de AWS, puede elegir roles personalizados para el servicio y el perfil de instancia. Para obtener más información, consulte Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS.

Grupos de seguridad

Amazon EMR utiliza grupos de seguridad para controlar el tráfico de entrada y salida a sus instancias de EC2. Cuando se lanza el clúster, Amazon EMR utiliza un grupo de seguridad para la instancia principal y un grupo de seguridad compartido por las instancias principales/de tarea. Amazon EMR

configura las reglas de grupo de seguridad para garantizar la comunicación entre las instancias del clúster. De forma opcional, puede configurar grupos de seguridad adicionales y asignarlos a sus instancias principales y secundarias o de tareas si necesita reglas más avanzadas. Para obtener más información, consulte Control del tráfico de red con grupos de seguridad.

Encryption (Cifrado)

Amazon EMR admite cifrado del cliente y del servidor opcional de Amazon S3 con EMRFS para ayudar a proteger los datos que se almacenan en Amazon S3. Con el cifrado del lado del servidor, Amazon S3 cifra sus datos después de cargarlos.

Con el cifrado del lado cliente, el proceso de cifrado y descifrado se produce en el cliente EMRFS en su clúster de EMR. Puede administrar la clave raíz para el cifrado del cliente mediante AWS Key Management Service (AWS KMS) o con su propio sistema de administración de claves.

Para obtener más información, consulte Especificación del cifrado de Amazon S3 con propiedades de EMRFS.

Amazon VPC

Amazon EMR admite el lanzamiento de clústeres en una nube privada virtual (VPC) en Amazon VPC. Una VPC es una red virtual aislada en AWS que ofrece la posibilidad de controlar aspectos avanzados de la configuración de la red y del acceso. Para obtener más información, consulte Configurar redes.

AWS CloudTrail

Amazon EMR se integra con CloudTrail para registrar información sobre las solicitudes hechas desde la cuenta de AWS o en su nombre. Con esta información, puede realizar un seguimiento de quién accede en cada momento a su clúster y la dirección IP desde la que se ha realizado la solicitud. Para obtener más información, consulte Registro de llamadas a la API de Amazon EMR en AWS CloudTrail.

Pares de claves de Amazon EC2

Para supervisar e interactuar con el clúster, puede formar una conexión segura entre el equipo remoto y el nodo principal. Puede utilizar el protocolo de red Secure Shell (SSH) para esta conexión o Kerberos para la autenticación. Si utiliza SSH, se requiere un par de claves de Amazon EC2. Para obtener más información, consulte Uso de un par de claves de EC2 para credenciales de SSH.

Supervisión

Puede utilizar las interfaces de administración de Amazon EMR y los archivos de registro para solucionar problemas de clúster como, por ejemplo, averías o errores. Amazon EMR ofrece la posibilidad de archivar los archivos de registro en Amazon S3 para que pueda almacenar registros y cuestiones de solución de problemas incluso después de que el clúster termine. Amazon EMR también proporciona una herramienta de depuración opcional en la consola de Amazon EMR para examinar los archivos de registro basada en pasos, trabajos y tareas. Para obtener más información, consulte Configurar el registro y la depuración de un clúster.

Amazon EMR se integra con CloudWatch para hacer un seguimiento de las métricas de rendimiento del clúster y los trabajos del clúster. Puede configurar alarmas basadas en diversas métricas como, por ejemplo, si el clúster está inactivo o el porcentaje de almacenamiento utilizado. Para obtener más información, consulte Monitorización de métricas de Amazon EMR con CloudWatch.

Interfaces de administración

Existen varias formas en las que puede interactuar con Amazon EMR:

- Consola: una interfaz gráfica de usuario que puede utilizar para lanzar y administrar clústeres. Con ella, puede rellenar formularios web para especificar los detalles de los clústeres que lanzar, ver los detalles de clústeres existentes, depurar y terminar clústeres. El uso de la consola es la manera más sencilla de empezar a utilizar Amazon EMR; no se requieren conocimientos de programación. La consola está disponible online en https://console.aws.amazon.com/elasticmapreduce/home.
- AWS Command Line Interface (AWS CLI): una aplicación cliente que se ejecuta en su equipo local para conectarse a Amazon EMR y crear y administrar clústeres. La AWS CLI contiene un conjunto de comandos con una amplia gama de características específicas de Amazon EMR. Con ella, puede escribir scripts que automatizan el proceso de lanzamiento y administración de clústeres. Si prefiere trabajar desde una línea de comandos, usar la AWS CLI es la mejor opción. Para obtener más información, consulte Amazon EMR en la Referencia de los comandos de la AWS CLI.
- Kit de desarrollo de software (SDK): los SDK proporcionan funciones que llaman a Amazon EMR para crear y administrar clústeres. Con ellos, puede escribir aplicaciones que automatizan el proceso de creación y administración de clústeres. Utilizar el SDK es la mejor opción para ampliar o personalizar la funcionalidad de Amazon EMR. Amazon EMR está disponible actualmente en los SDK siguientes: Go, Java, .NET (C# y VB.NET), Node.js, PHP, Python y Ruby. Para obtener más información sobre estos SDK, consulte Herramientas para crear en AWS y Código de muestra y bibliotecas de Amazon EMR.

- API de servicios web: un interfaz de bajo nivel que puede utilizar para llamar al servicio web directamente, utilizando JSON. El uso de la API es la mejor opción para crear un SDK personalizado que llame a Amazon EMR. Para obtener más información, consulte la Referencia de las API de Amazon EMR.

Descripción general de la arquitectura de Amazon EMR

La arquitectura de servicio de Amazon EMR se compone de varias capas, cada una de las cuales proporciona determinadas capacidades y funcionalidad al clúster. Esta sección proporciona información general sobre las capas y los componentes de cada una de ellas.

En este tema

- Almacenamiento
- Administración de recursos de clúster
- Marcos de procesamiento de datos
- Aplicaciones y programas

Almacenamiento

La capa de almacenamiento incluye los diferentes sistemas de archivos que se utilizan con el clúster. Existen varios tipos distintos de opciones de almacenamiento como se indica a continuación.

Hadoop Distributed File System (HDFS)

Hadoop Distributed File System (HDFS) es un sistema de archivos distribuido y escalable para Hadoop. HDFS distribuye los datos que almacena en instancias en el clúster, almacenando varias copias en datos en distintas instancias para garantizar que no se pierdan datos si una instancia individual falla. HDFS es un almacenamiento efímero que se reclama cuando se termina un clúster. HDFS resulta útil para almacenar en caché los resultados intermedios durante el procesamiento de MapReduce o para cargas de trabajo con una E/S aleatoria notable.

Para obtener más información, consulte Almacenamiento de la instancia en esta guía o vaya a la Guía del usuario de HDFS en el sitio web de Apache Hadoop.

Sistema de archivos de EMR (EMRFS)

Con el sistema de archivos de EMR (EMRFS), Amazon EMR amplía Hadoop para agregar la posibilidad de tener acceso directamente a los datos almacenados en Amazon S3 como si fueran un sistema de archivos similar a HDFS. Puede usar HDFS o Amazon S3 como sistema de archivos en su clúster. A menudo, Amazon S3 se usa para almacenar los datos de entrada y de salida y los resultados intermedios se almacenan en HDFS.

Sistema de archivos local

El sistema de archivos local se refiere a un disco conectado a nivel local. Cuando se crea un clúster de Hadoop, cada nodo se crea a partir de una instancia de Amazon EC2 que viene con un bloque preconfigurado de almacenamiento en disco preasociado que se denomina almacén de instancias. Los datos en volúmenes del almacén de instancias se conservan solo durante el ciclo de vida de su instancia de Amazon EC2.

Administración de recursos de clúster

La capa de administración de recursos es responsable de la administración de los recursos de clúster y de la programación de trabajos para procesamiento de datos.

De forma predeterminada, Amazon EMR utiliza YARN (Yet Another Resource Negotiator), que es un componente introducido en Apache Hadoop 2.0 para administrar de forma centralizada recursos de clúster para varios marcos de procesamiento de datos. Sin embargo, hay otros marcos y aplicaciones que se ofrecen en Amazon EMR que no utilizan YARN como administrador de recursos. Amazon EMR también dispone de un agente en cada nodo que administra los componentes de YARN, mantiene el clúster en buen estado y se comunica con el servicio de Amazon EMR.

Dado que las instancias de spot se utilizan a menudo para ejecutar nodos de tarea, Amazon EMR tiene una funcionalidad predeterminada para programar trabajos de YARN, de modo que los trabajos en ejecución no presenten errores cuando los nodos de tarea que se ejecutan en las instancias de spot se terminen. Para ello, Amazon EMR permite que los procesos maestros de la aplicación se ejecuten únicamente en los nodos principales. El proceso maestro de la aplicación controla los trabajos en ejecución y debe mantenerse activo durante toda la vida del trabajo.

La versión 5.19.0 y posteriores de Amazon EMR utilizan la característica integrada de etiquetas de nodo YARN para lograrlo. (Las versiones anteriores utilizaban una revisión de código). Las propiedades en las clasificaciones de configuración `yarn-site` y `capacity-scheduler` se ajustan de forma predeterminada para que capacity-scheduler y fair-scheduler de YARN utilicen las

etiquetas de nodo. Amazon EMR etiqueta automáticamente los nodos principales con la etiqueta `CORE` y establece las propiedades para que los maestros de la aplicación se programen únicamente en los nodos con la etiqueta CORE. La modificación manual de las propiedades relacionadas en las clasificaciones de configuración yarn-site y capacity-scheduler o directamente en los archivos XML asociados podría interrumpir esta característica o modificar esta funcionalidad.

Marcos de procesamiento de datos

La capa de marco de trabajo de procesamiento de datos es el motor que se utiliza para procesar y analizar datos. Existen muchos marcos de trabajo disponibles que se ejecutan en YARN o que tienen su propia administración de recursos. Los distintos marcos están disponibles para los diferentes tipos de necesidades de procesamiento tales como lotes, interactivo, en memoria, streaming, etc. El marco de trabajo que elija depende de su caso de uso. Esto afecta a los lenguajes y a los interfaces disponibles desde la capa de aplicación, que es la capa que se utiliza para interactuar con los datos que desea procesar. Los principales marcos de procesamiento disponibles para Amazon EMR son Hadoop MapReduce y Spark.

Hadoop MapReduce

Hadoop MapReduce es un modelo de programación de código abierto para informática distribuida. Simplifica el proceso de escritura de aplicaciones distribuidas en paralelo mediante el tratamiento de toda la lógica, mientras proporciona las funciones Map y Reduce. La función Map asigna datos a conjuntos de pares clave-valor denominados resultados intermedios. La función Reduce combina los resultados intermedios, aplica algoritmos adicionales y genera la salida final. Existen varios marcos de trabajo disponibles para MapReduce, como Hive, que generan automáticamente programas Map y Reduce.

Para obtener más información, consulte Cómo asignar y reducir operaciones que se están llevando a cabo en el sitio web de la wiki de Apache Hadoop.

Apache Spark

Spark es un marco de trabajo de clúster y un modelo de programación para el procesamiento de cargas de trabajo de big data. Al igual que Hadoop MapReduce, Spark es un sistema de procesamiento distribuido de código abierto, pero utiliza gráficos acíclicos dirigidos para planes de ejecución y el almacenamiento de caché en memoria para conjuntos de datos. Cuando se ejecuta Spark en Amazon EMR, es posible utilizar EMRFS para tener acceso directamente a los datos en Amazon S3. Spark admite diversos módulos de consulta interactivos como, por ejemplo, SparkSQL.

Para obtener más información, consulte Apache Spark en clústeres de Amazon EMR en la Guía de publicación de Amazon EMR.

Aplicaciones y programas

Amazon EMR es compatible con muchas aplicaciones, tales como Hive, Pig y la biblioteca Spark Streaming, para ofrecer diversas capacidades, como el uso de lenguajes de nivel superior para crear cargas de trabajo de procesamiento, el uso de algoritmos de machine learning, el desarrollo de aplicaciones de procesamiento de flujos y la creación de almacenamientos de datos. Además, Amazon EMR también admite proyectos de código abierto que tienen su propia funcionalidad de administración de clústeres en lugar de utilizar YARN.

Puede utilizar diversas bibliotecas y lenguajes para interactuar con las aplicaciones que se ejecutan en Amazon EMR. Por ejemplo, puede utilizar Java, Hive o Pig con MapReduce o Spark Streaming, Spark SQL, MLlib y GraphX con Spark.

Para obtener más información, consulte la Guía de publicación de Amazon EMR.

Configuración de Amazon EMR

Complete las tareas de esta sección antes de lanzar un clúster de Amazon EMR por primera vez:

Antes de usar Amazon EMR por primera vez, complete las siguientes tareas:

Registro en una Cuenta de AWS

Si no dispone de una Cuenta de AWS, siga los pasos que figuran a continuación para crear una.

Para registrarse en Cuenta de AWS

1. Abra https://portal.aws.amazon.com/billing/signup.
2. Siga las instrucciones en línea.

   Parte del procedimiento de inscripción consiste en recibir una llamada telefónica e indicar un código de verificación en el teclado del teléfono.

   Al registrase en una Cuenta de AWS, se crea un Usuario raíz de la cuenta de AWS. El usuario raíz tiene acceso a todos los recursos y Servicios de AWS de esa cuenta. Como práctica recomendada de seguridad, asigne acceso administrativo a un usuario administrativo y utilice únicamente el usuario raíz para realizar la ejecución tareas que requieren acceso de usuario raíz.

AWS le enviará un email de confirmación luego de completar el proceso de registro. Puede ver la actividad de la cuenta y administrar la cuenta en cualquier momento entrando en https://aws.amazon.com/ y seleccionando My Account (Mi cuenta).

Crear un usuario administrativo

Después de registrarse para obtener una Cuenta de AWS, cree un usuario administrativo para que no utilice el usuario raíz en las tareas cotidianas.

Proteger su Usuario raíz de la cuenta de AWS

1. Inicie sesión en AWS Management Console como propietario de cuenta eligiendo Usuario raíz e ingrese el email de su Cuenta de AWS. En la siguiente página, escriba su contraseña.

Para obtener ayuda para iniciar sesión con el usuario raíz, consulte Signing in as the root user (Iniciar sesión como usuario raíz) en la Guía del usuario de AWS Sign-In.

2. Active la autenticación multifactor (MFA) para el usuario raíz.

   Para obtener instrucciones, consulte Habilitar un dispositivo MFA virtual para el usuario raíz Cuenta de AWS (consola) en la Guía del usuario de IAM.

Crear un usuario administrativo

- Para las tareas administrativas diarias, conceda acceso administrativo a un usuario administrativo en AWS IAM Identity Center.

  Para obtener instrucciones, consulte Introducción en la Guía del usuario de AWS IAM Identity Center.

Iniciar sesión como usuario administrativo

- Para iniciar sesión con el usuario del Centro de identidades de IAM, utilice la URL de inicio de sesión que se envió a la dirección de correo electrónico cuando creó el usuario del Centro de identidades de IAM.

  Para obtener ayuda para iniciar sesión con un usuario del Centro de identidades de IAM, consulte Iniciar sesión en el portal de acceso de AWS en la Guía del usuario de AWS Sign-In.

Crear un par de claves de Amazon EC2 para SSH

> **Note**
>
> Con la versión 5.10.0 o las versiones posteriores de Amazon EMR, puede configurar Kerberos para autenticar a los usuarios y las conexiones SSH con el clúster. Para obtener más información, consulte Uso de Kerberos para la autenticación con Amazon EMR.

Para autenticarse y conectarse a los nodos de un clúster a través de un canal seguro mediante el protocolo Secure Shell (SSH), cree un par de claves de Amazon Elastic Compute Cloud (Amazon

EC2) antes de lanzar el clúster. También puede crear un clúster sin un par de claves. Esto se hace con clústeres transitorios que se inician, ejecutan pasos, y luego se terminan de forma automática.

Si…	Entonces...
Ya tiene un par de claves de Amazon EC2 que quiere usar o no necesita autenticarse en su clúster.	Omita este paso.
Debe crear un par de claves.	Consulte Crear un par de claves con Amazon EC2.

Pasos siguientes

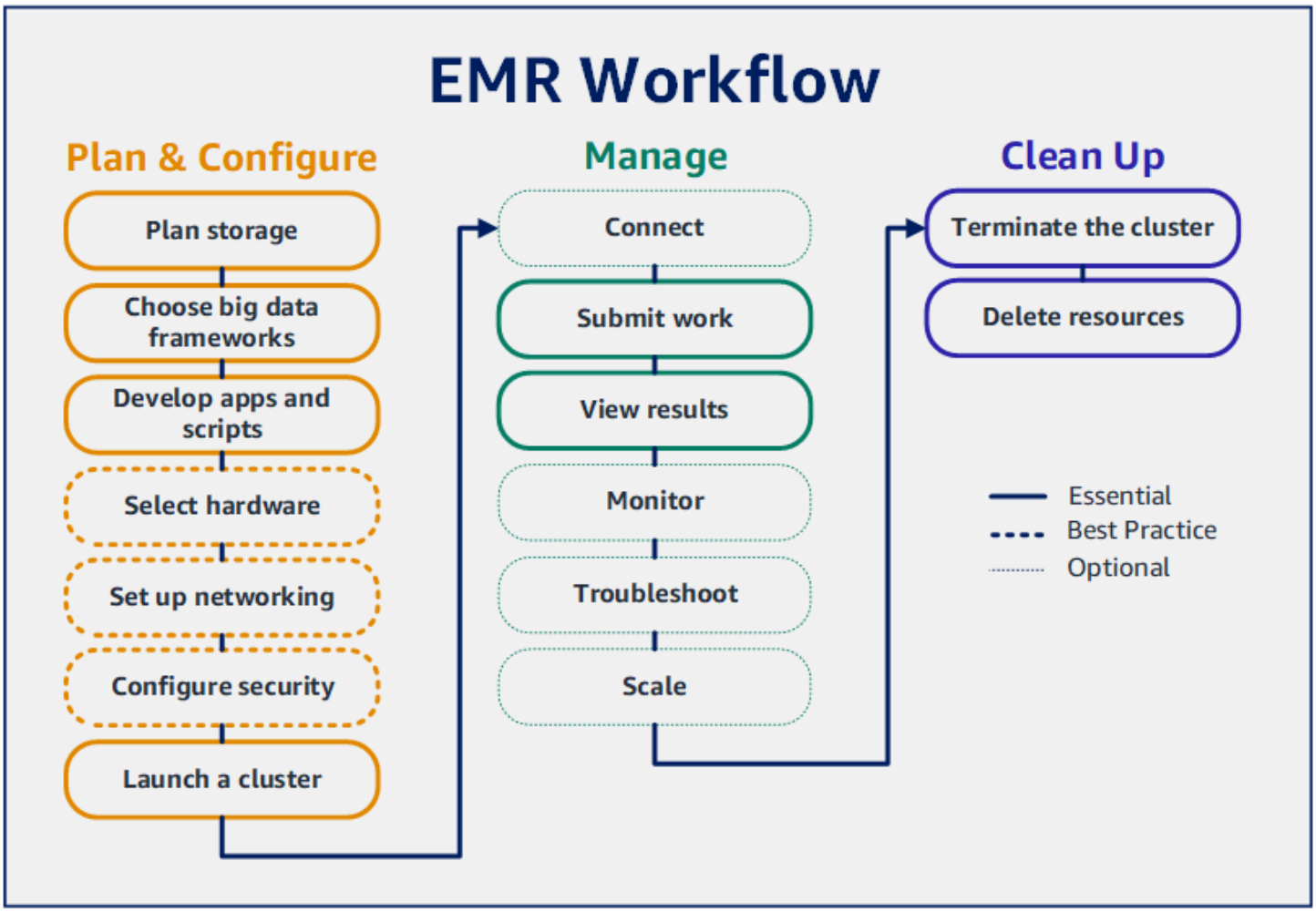
- Para obtener información sobre cómo crear un clúster de muestra, consulte Tutorial: Introducción a Amazon EMR.
- Para obtener más información sobre cómo configurar un clúster personalizado y controlar el acceso al mismo, consulte Planificar y configurar clústeres y Seguridad en Amazon EMR.

Tutorial: Introducción a Amazon EMR

Información general

Con Amazon EMR, puede configurar un clúster para procesar y analizar datos con marcos de macrodatos en solo unos minutos. En este tutorial, se muestra cómo lanzar un clúster de muestra con Spark y cómo ejecutar un script de PySpark sencillo almacenado en un bucket de Amazon S3. Cubre tareas esenciales de Amazon EMR en tres categorías principales de flujo de trabajo: planificación y configuración, administración y limpieza.

Encontrará enlaces a temas más detallados a medida que avance en el tutorial e ideas para pasos adicionales en la sección Pasos siguientes. Si tiene alguna pregunta o no sabe cómo continuar, contacte con el equipo de Amazon EMR en nuestro foro de discusión.



Requisitos previos

- Antes de lanzar un clúster de Amazon EMR, asegúrese de completar las tareas en Configuración de Amazon EMR.

Costo

- El clúster de ejemplo que cree se ejecuta en un entorno real. El clúster genera unos gastos mínimos. Para evitar cargos adicionales, asegúrese de completar las tareas de limpieza del último paso de este tutorial. Los cargos se acumulan en virtud la tarifa por segundo según los precios de Amazon EMR. Los cargos también varían según la región. Para obtener más información, consulte Precios de Amazon EMR.
- Es posible que se acumulen cargos mínimos por los archivos pequeños que almacene en Amazon S3. Es posible que no se apliquen algunos o todos los cargos de Amazon S3 si se encuentra dentro de los límites de uso del nivel gratuito de AWS. Para obtener más información, consulte Precios de Amazon S3 y Nivel gratuito de AWS.

Paso 1: planificar y configurar un clúster de Amazon EMR

Prepare el almacenamiento para Amazon EMR

Cuando utiliza Amazon EMR, puede elegir entre una variedad de sistemas de archivos para almacenar los datos de entrada, los datos de salida y los archivos de registro. En este tutorial, utilizará EMRFS para almacenar datos en un bucket de S3. EMRFS es una implementación del sistema de archivos de Hadoop que le permite leer y escribir archivos normales en Amazon S3. Para obtener más información, consulte Utilizar almacenamiento y sistemas de archivos.

Para crear un bucket para este tutorial, consulte Crear un bucket de S3 en la Guía del usuario de la consola de Amazon Simple Storage Service. Cree el bucket en la misma región de AWS en la que desee lanzar el clúster de Amazon EMR. Por ejemplo, Oeste de EE. UU. (Oregón) us-west-2.

Los buckets y las carpetas que utilice con Amazon EMR tienen las siguientes limitaciones:

- Los nombres pueden contener letras minúsculas, números, puntos (.) y guiones (-).
- Los nombres no pueden terminar en números.
- El nombre de un bucket debe ser único en todas las cuentas de AWS.
- La carpeta de salida debe estar vacía.

Preparar una aplicación con datos de entrada para Amazon EMR

La forma más común de preparar una aplicación para Amazon EMR consiste en cargar la aplicación y sus datos de entrada en Amazon S3. A continuación, cuando envíe el trabajo a su clúster, especifique las ubicaciones de Amazon S3 para el script y los datos.

En este paso, cargue un ejemplo del script de PySpark a su bucket de Amazon S3. Le hemos proporcionado un script de PySpark para que lo utilice. El script procesa los datos de inspección del establecimiento de alimentos y devuelve un archivo de resultados en su bucket de S3. En el archivo de resultados se enumeran los diez establecimientos con más infracciones de tipo “rojo”.

También debe cargar datos de entrada de muestra en Amazon S3 para que el script de PySpark los procese. Los datos de entrada son una versión modificada de los resultados de las inspecciones del Departamento de Salud en el condado de King (Washington) entre 2006 y 2020. Para obtener más información, consulte Datos abiertos del condado de King: datos de inspección de establecimientos alimentarios. A continuación se incluyen ejemplos de filas del conjunto de datos.

```
name, inspection_result, inspection_closed_business, violation_type, violation_points
100 LB CLAM, Unsatisfactory, FALSE, BLUE, 5
100 PERCENT NUTRICION, Unsatisfactory, FALSE, BLUE, 5
7-ELEVEN #2361-39423A, Complete, FALSE, , 0
```

Para preparar el script PySpark de muestra para EMR

1. Copie el siguiente código de muestra en un nuevo archivo en el editor que prefiera.

```
import argparse

from pyspark.sql import SparkSession

def calculate_red_violations(data_source, output_uri):
    """
    Processes sample food establishment inspection data and queries the data to
 find the top 10 establishments
    with the most Red violations from 2006 to 2020.

    :param data_source: The URI of your food establishment data CSV, such as 's3://
DOC-EXAMPLE-BUCKET/food-establishment-data.csv'.
    :param output_uri: The URI where output is written, such as 's3://DOC-EXAMPLE-
BUCKET/restaurant_violation_results'.
    """
```

```
    with SparkSession.builder.appName("Calculate Red Health
 Violations").getOrCreate() as spark:
        # Load the restaurant violation CSV data
        if data_source is not None:
            restaurants_df = spark.read.option("header", "true").csv(data_source)

        # Create an in-memory DataFrame to query
        restaurants_df.createOrReplaceTempView("restaurant_violations")

        # Create a DataFrame of the top 10 restaurants with the most Red violations
        top_red_violation_restaurants = spark.sql("""SELECT name, count(*) AS
 total_red_violations
          FROM restaurant_violations
          WHERE violation_type = 'RED'
          GROUP BY name
          ORDER BY total_red_violations DESC LIMIT 10""")

        # Write the results to the specified output URI
        top_red_violation_restaurants.write.option("header",
 "true").mode("overwrite").csv(output_uri)

if __name__ == "__main__":
    parser = argparse.ArgumentParser()
    parser.add_argument(
        '--data_source', help="The URI for you CSV restaurant data, like an S3
 bucket location.")
    parser.add_argument(
        '--output_uri', help="The URI where output is saved, like an S3 bucket
 location.")
    args = parser.parse_args()

    calculate_red_violations(args.data_source, args.output_uri)
```

2. Guarde el archivo como `health_violations.py`.
3. Suba `health_violations.py` a Amazon S3 en el bucket que creó para este tutorial. Para obtener instrucciones, consulte Cargar un objeto en un bucket en la Guía de introducción a Amazon Simple Storage Service.

Para preparar los datos de entrada de muestra para EMR

1. Descargue el archivo zip, food_establishment_data.zip.

2. Descomprima y guarde `food_establishment_data.zip` como `food_establishment_data.csv` en su máquina.
3. Suba el archivo CSV en el bucket de S3 que creó para este tutorial. Para obtener instrucciones, consulte Cargar un objeto en un bucket en la Guía de introducción a Amazon Simple Storage Service.

Para obtener más información sobre cómo configurar los datos para EMR, consulte Preparar datos de entrada.

Lanzar un clúster de Amazon EMR

Tras preparar la ubicación de almacenamiento y la aplicación, puede lanzar un clúster de muestra de Amazon EMR. En este paso, lanza un clúster de Apache Spark con la versión de Amazon EMR más reciente.

New console

Para lanzar un clúster con Spark instalado con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En la página Crear clúster, anote los valores predeterminados para la versión, el tipo de instancia, el número de instancias y los permisos. Estos campos se rellenan automáticamente con valores que funcionan para clústeres de uso general.
4. En el campo Nombre del clúster, escriba un nombre único que le ayude a identificarlo, como *`Mi primer clúster`*.
5. En Aplicaciones, elija la opción Spark para instalar Spark en el clúster.

> **Note**
>
> Elija las aplicaciones que desee incluir en su clúster de Amazon EMR antes de lanzarlo. No puede agregar ni eliminar aplicaciones de un clúster después del lanzamiento.

6. En Registros de clúster, seleccione la casilla Publicar los registros específicos del clúster en Amazon S3. Sustituya el valor de Ubicación de Amazon S3 por el bucket de Amazon S3 que creó, seguido por **/logs**. Por ejemplo, **s3://DOC-EXAMPLE-BUCKET/logs**. Al agregar **/logs**, se crea una nueva carpeta llamada "logs" en el bucket, donde Amazon EMR puede copiar los archivos de registro del clúster.
7. En Configuración y permisos de seguridad, elija su par de claves de EC2. En la misma sección, seleccione el menú desplegable Rol de servicio para Amazon EMR y elija EMR_DefaultRole. A continuación, seleccione el menú desplegable Rol de IAM para el perfil de instancia y elija EMR_EC2_DefaultRole.
8. Seleccione Crear clúster para lanzar el clúster y abrir la página de detalles del clúster.
9. Consulte el estado del clúster junto a su nombre. El estado cambia de Iniciando a En ejecución a Esperando a medida que Amazon EMR aprovisiona el clúster. Es posible que tenga que elegir el icono de actualización situado a la derecha o actualizar el navegador para ver las actualizaciones de estado.

Cuando el estado cambia a Esperando, el clúster está activo, en ejecución y listo para aceptar trabajo. Para obtener más información sobre cómo leer el resumen del clúster, consulte Ver el estado y los detalles del clúster. Para obtener más información acerca del estado del clúster, consulte Descripción del ciclo de vida del clúster.

Old console

Para lanzar un clúster con Spark instalado con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Crear clúster para abrir el asistente de opciones rápidas.
3. Anote los valores predeterminados de la versión, el tipo de instancia, el número de instancias y los permisos en la página Crear clúster: Opciones rápidas. Estos campos se rellenan automáticamente con valores que funcionan para clústeres de uso general.
4. Ingrese un nombre de clúster que le ayude a identificarlo. Por ejemplo, *Mi primer clúster*.
5. Deje el registro activado, pero sustituya el valor de la carpeta de S3 por el bucket de Amazon S3 que creó, seguido de **/logs**. Por ejemplo, **s3://DOC-EXAMPLE-BUCKET/logs**. Al

agregar **`/logs`**, se crea una nueva carpeta llamada “logs” en su bucket, donde EMR puede copiar los archivos de registro del clúster.

6. Elige la opción Spark en Aplicaciones para instalar Spark en el clúster.

> **Note**
>
> Elija las aplicaciones que desee incluir en su clúster de Amazon EMR antes de lanzarlo. No puede agregar ni eliminar aplicaciones de un clúster después del lanzamiento.

7. Elija su par de claves de EC2 en Seguridad y acceso.
8. Seleccione Crear clúster para lanzar el clúster y abrir la página de estado del clúster.
9. Consulte el estado del clúster junto a su nombre. El estado cambia de Iniciando a En ejecución a Esperando a medida que Amazon EMR aprovisiona el clúster. Es posible que tenga que elegir el icono de actualización situado a la derecha o actualizar el navegador para ver las actualizaciones de estado.

Cuando el estado cambia a Esperando, el clúster está activo, en ejecución y listo para aceptar trabajo. Para obtener más información sobre cómo leer el resumen del clúster, consulte Ver el estado y los detalles del clúster. Para obtener más información acerca del estado del clúster, consulte Descripción del ciclo de vida del clúster.

CLI

Para lanzar un clúster con Spark instalado con la AWS CLI

1. Cree los roles predeterminados de IAM que, a continuación, podrá utilizar para crear el clúster mediante el siguiente comando.

```
aws emr create-default-roles
```

Para obtener más información sobre `create-default-roles`, consulte la Referencia de los comandos de la AWS CLI.

2. Cree un clúster de Spark con el siguiente comando. Introduzca un nombre para el clúster con la opción `--name` y especifique el nombre del par de claves de EC2 con la opción `--ec2-attributes`.

```
aws emr create-cluster \
--name "<My First EMR Cluster>" \
--release-label <emr-5.36.1> \
--applications Name=Spark \
--ec2-attributes KeyName=<myEMRKeyPairName> \
--instance-type m5.xlarge \
--instance-count 3 \
--use-default-roles
```

Anote los demás valores obligatorios para `--instance-type`, `--instance-count`, y `--use-default-roles`. Estos valores se han elegido para clústeres de uso general. Para obtener más información sobre `create-cluster`, consulte la Referencia de los comandos de la AWS CLI.

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

Debería ver un resultado como el siguiente. El resultado muestra el `ClusterId` y el `ClusterArn` del nuevo clúster. Anote su `ClusterId`. Se utiliza `ClusterId` para comprobar el estado del clúster y enviar trabajos.

```
{
    "ClusterId": "myClusterId",
    "ClusterArn": "myClusterArn"
}
```

3. Compruebe el estado del clúster con el siguiente comando.

```
aws emr describe-cluster --cluster-id <myClusterId>
```

Debería ver un resultado como el siguiente con el objeto `Status` del nuevo clúster.

```
{
    "Cluster": {
```

```
        "Id": "myClusterId",
        "Name": "My First EMR Cluster",
        "Status": {
            "State": "STARTING",
            "StateChangeReason": {
                "Message": "Configuring cluster software"
            }
        }
    }
}
```

El valor `State` cambia de `STARTING` a `RUNNING` a `WAITING` cuando Amazon EMR aprovisiona el clúster.

El estado del clúster cambia a **`WAITING`** cuando un clúster está activo, en ejecución y listo para aceptar trabajo. Para obtener más información acerca del estado del clúster, consulte Descripción del ciclo de vida del clúster.

Paso 2: administrar el clúster de Amazon EMR

Enviar un trabajo a Amazon EMR

Después de lanzar un clúster, puede enviar el trabajo al clúster en ejecución para procesar y analizar los datos. El trabajo se envía a un clúster de Amazon EMR como un paso. Un paso es una unidad de trabajo compuesta por una o más acciones. Por ejemplo, puede enviar un paso para calcular valores o para transferir y procesar datos. Puede enviar pasos cuando al crear un clúster o bien a un clúster en ejecución. En este tutorial, envía `health_violations.py` como un paso a su clúster en ejecución. Para obtener más información sobre los pasos, consulte Enviar trabajo a un clúster.

New console

Para enviar una aplicación de Spark como paso con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, seleccione Clústeres y, a continuación, seleccione el clúster al desee enviar trabajo. El estado del clúster debe estar en Esperando.

3. Seleccione la pestaña Pasos y, a continuación Agregar paso.
4. Configure el paso de acuerdo con las directrices siguientes:

   - En Tipo, seleccione Aplicación de Spark. Deberá ver campos adicionales para el modo implementación, la ubicación de la aplicación y las opciones de spark-submit.
   - En Nombre, escriba un nuevo nombre. Si tiene muchos pasos en un clúster, nombrar cada paso le ayudará a hacer un seguimiento de ellos.
   - Para el modo implementación, deje el valor predeterminado Modo de clúster. Para obtener más información sobre los modos de implementación de Spark, consulte Información general sobre el modo de clúster en la documentación de Apache Spark.
   - En Ubicación de la aplicación, ingrese la ubicación del script `health_violations.py` en Amazon S3, por ejemplo, *`s3://DOC-EXAMPLE-BUCKET/health_violations.py`*.
   - Deje vacío el campo Opciones de Spark-submit. Para obtener más información sobre las opciones de `spark-submit`, consulte Launching applications with spark-submit.
   - En el campo Argumentos, ingrese los siguientes argumentos y valores:

     ```
     --data_source s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv
     --output_uri s3://DOC-EXAMPLE-BUCKET/myOutputFolder
     ```

     Sustituya *`s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv`* por el URI del bucket de S3 de los datos de entrada que preparó en Preparar una aplicación con datos de entrada para Amazon EMR.

     Sustituya *`DOC-EXAMPLE-BUCKET`* por el nombre del bucket que creó para este tutorial y sustituya *`myOutputFolder`* por un nombre para la carpeta de salida del clúster.
   - En Acción si se produce un error en el paso, acepte la opción predeterminada Continuar. De esta forma, si se produce un error en el paso, el clúster seguirá ejecutándose.
5. Seleccione Agregar para enviar el paso. El paso debería aparecer en la consola con el estado Pendiente.
6. Supervise el estado del paso. Debería cambiar de Pendiente a En ejecución a Completado. Para actualizar el estado en la consola, elija el icono de actualización situado a la derecha de Filtro. El script tarda aproximadamente un minuto en ejecutarse. Cuando el paso se haya completado correctamente, el estado cambiará a Completado.

Old console

Para enviar una aplicación de Spark como paso con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. En Lista de clústeres, seleccione el nombre del clúster. El estado del clúster debe estar en Esperando.
3. Elija Steps (Pasos) y, a continuación Add step (Añadir paso).
4. Configure el paso de acuerdo con las directrices siguientes:

   - En Step type (Tipo de paso), elija Spark application (Aplicación de Spark). Debería ver campos adicionales para el modo implementación, las opciones de Spark-submit y Ubicación de la aplicación.
   - En Nombre, deje el valor predeterminado o escriba un nombre nuevo. Si tiene muchos pasos en un clúster, nombrar cada paso le ayudará a hacer un seguimiento de ellos.
   - En Modo implementación, deje el valor predeterminado Clúster. Para obtener más información acerca de los modos de implementación de Spark, consulte Información general sobre el modo de clúster en la documentación de Apache Spark.
   - Deje en blanco el campo Opciones de Spark-submit. Para obtener más información sobre las opciones de `spark-submit`, consulte Launching applications with spark-submit.
   - En Ubicación de la aplicación, ingrese la ubicación del script `health_violations.py` en Amazon S3. Por ejemplo, *`s3://DOC-EXAMPLE-BUCKET/health_violations.py`*.
   - En el campo Argumentos, ingrese los siguientes argumentos y valores:

     ```
     --data_source s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv
     --output_uri s3://DOC-EXAMPLE-BUCKET/myOutputFolder
     ```

     Sustituya *`s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv`* por el URI de S3 de los datos de entrada que preparó en Preparar una aplicación con datos de entrada para Amazon EMR.

     Sustituya *`DOC-EXAMPLE-BUCKET`* por el nombre del bucket que creó para este tutorial y *`myOutputFolder`* por el nombre de la carpeta de salida del clúster.

- En Acción sobre el error, acepte la opción predeterminada Continuar para que, si el paso falla, el clúster continúe ejecutándose.

5. Seleccione Agregar para enviar el paso. El paso debería aparecer en la consola con el estado Pendiente.
6. Compruebe si el estado del paso cambia de Pendiente a En ejecución a Completado. Para actualizar el estado en la consola, elija el icono de actualización situado a la derecha del Filtro. El script tarda aproximadamente un minuto en ejecutarse.

Sabrá que el paso ha finalizado correctamente cuando el estado cambie a Completado.

CLI

Para enviar una solicitud a Spark como un paso con la AWS CLI

1. Asegúrese de tener el `ClusterId` del clúster que lanzó en Lanzar un clúster de Amazon EMR. Puede recuperar el ID del clúster con los siguientes comandos.

```
aws emr list-clusters --cluster-states WAITING
```

2. Envíe `health_violations.py` como un paso con el comando `add-steps` y el `ClusterId`.

   - Puede especificar un nombre para el paso al sustituir *"Mi aplicación de Spark"*. En la matriz `Args`, sustituya *s3://DOC-EXAMPLE-BUCKET/health_violations.py* por la ubicación de la aplicación `health_violations.py`.
   - Sustituya *s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv* por la ubicación de S3 del conjunto de datos `food_establishment_data.csv`.
   - Sustituya *s3://DOC-EXAMPLE-BUCKET/MyOutputFolder* por la ruta de S3 del bucket designado y un nombre para la carpeta de salida del clúster.
   - `ActionOnFailure=CONTINUE` significa que el clúster seguirá ejecutándose si se produce un error en el paso.

```
aws emr add-steps \
--cluster-id <myClusterId> \
--steps Type=Spark,Name="<My Spark
 Application>",ActionOnFailure=CONTINUE,Args=[<s3://DOC-EXAMPLE-
BUCKET/health_violations.py>,--data_source,<s3://DOC-EXAMPLE-BUCKET/
```

```
food_establishment_data.csv>,--output_uri,<s3://DOC-EXAMPLE-BUCKET/
MyOutputFolder>]
```

Para obtener más información acerca envío de pasos mediante la CLI, consulte la Referencia de los comandos de la AWS CLI.

Después de enviar el paso, debería ver un resultado como el siguiente con una lista de `StepIds`. Como ha enviado un paso, solo verá un ID en la lista. Copie el ID de su paso. Utilice el identificador del paso para comprobar el estado.

```
{
    "StepIds": [
        "s-1XXXXXXXXXXA"
    ]
}
```

3. Consulte el estado del paso con el comando `describe-step`.

```
aws emr describe-step --cluster-id <myClusterId> --step-id <s-1XXXXXXXXXXA>
```

Debería ver un resultado como el siguiente, con información sobre el paso.

```
{
    "Step": {
        "Id": "s-1XXXXXXXXXXA",
        "Name": "My Spark Application",
        "Config": {
            "Jar": "command-runner.jar",
            "Properties": {},
            "Args": [
                "spark-submit",
                "s3://DOC-EXAMPLE-BUCKET/health_violations.py",
                "--data_source",
                "s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv",
                "--output_uri",
                "s3://DOC-EXAMPLE-BUCKET/myOutputFolder"
            ]
        },
        "ActionOnFailure": "CONTINUE",
        "Status": {
            "State": "COMPLETED"
```

```
        }
    }
}
```

A medida que el paso se ejecuta, el `State` cambia de `PENDING` a `RUNNING` a `COMPLETED`. El paso tarda aproximadamente un minuto en ejecutarse, por lo que es posible que tenga que comprobar el estado varias veces.

Sabrá que el paso se ha completado correctamente cuando el `State` cambie a **`COMPLETED`**.

Para obtener más información sobre el ciclo de vida del paso, consulte Ejecución de pasos para procesar datos.

Ver los resultados

Cuando un paso se ejecute correctamente, podrá ver los resultados de salida en la carpeta de salida de Amazon S3.

Para ver los resultados de **`health_violations.py`**

1. Abra la consola de Amazon S3 en https://console.aws.amazon.com/s3.
2. Elija el nombre del bucket y, a continuación, la carpeta de salida que especificó al enviar el paso. Por ejemplo, *`DOC-EXAMPLE-BUCKET`* y, a continuación, *`myOutputFolder`*.
3. Compruebe que los siguientes elementos aparecen en la carpeta de salida:
   - Un objeto de tamaño pequeño llamado `_SUCCESS`.
   - Un archivo CSV que comienza con el prefijo `part-` que contiene los resultados.
4. Elija el objeto con los resultados y, a continuación, seleccione Descargar para guardar los resultados en el sistema de archivos local.
5. Abra los resultados con el editor que prefiera. En el archivo de salida, se muestra una lista de los diez establecimientos de alimentación con más infracciones rojas. En el archivo de salida, también se muestra el número total de infracciones rojas de cada establecimiento.

   A continuación se muestra un resultado de `health_violations.py`.

```
name, total_red_violations
SUBWAY, 322
```

```
T-MOBILE PARK, 315
WHOLE FOODS MARKET, 299
PCC COMMUNITY MARKETS, 251
TACO TIME, 240
MCDONALD'S, 177
THAI GINGER, 153
SAFEWAY INC #1508, 143
TAQUERIA EL RINCONSITO, 134
HIMITSU TERIYAKI, 128
```

Para obtener más información sobre el resultado del clúster de Amazon EMR, consulte Configurar una ubicación de salida.

(Opcional) Conectarse al clúster de Amazon EMR en ejecución

Cuando utiliza Amazon EMR, es posible que desee conectarse a un clúster en ejecución para leer los archivos de registro, depurar el clúster o utilizar herramientas de la CLI, como el intérprete de comandos de Spark. Amazon EMR le permite conectarse a un clúster mediante el protocolo Secure Shell (SSH). En esta sección, se explica cómo configurar SSH, conectarse al clúster y ver los archivos de registro de Spark. Para obtener más información acerca de la conexión al clúster, consulte Autenticación en nodos de clúster de Amazon EMR.

Autorizar las conexiones de SSH a su clúster

Antes de conectarse al clúster, debe modificar sus grupos de seguridad para autorizar las conexiones SSH entrantes. Los grupos de seguridad de Amazon EC2 funcionan como firewalls virtuales para controlar el tráfico entrante y saliente del clúster. Cuando creó el clúster para este tutorial, Amazon EMR creó los siguientes grupos de seguridad en su nombre:

ElasticMapReduce-master

: El grupo de seguridad administrado por Amazon EMR, asociado al nodo principal. En un clúster de Amazon EMR, el nodo principal es una instancia de Amazon EC2 que administra el clúster.

ElasticMapReduce-slave

: El grupo de seguridad, asociado a los nodos secundarios y de tareas.

New console

Para permitir el acceso SSH a los orígenes de confianza del grupo de seguridad principal con la nueva consola

Para editar los grupos de seguridad, debe tener permiso para administrar los grupos de seguridad de la VPC en la que se encuentra el clúster. Para obtener más información, consulte Cambio de los permisos de un usuario y el Ejemplo de política que permite administrar los grupos de seguridad de EC2 en la Guía del usuario de IAM.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y, a continuación, seleccione el clúster que desee actualizar. Se abrirá la página de detalles del clúster. La pestaña Propiedades de esta página debe estar preseleccionada.
3. En Redes, en la pestaña Propiedades, seleccione la flecha situada junto a Grupos de seguridad de EC2 (firewall) para ampliar esta sección. En Nodo principal, seleccione el enlace del grupo de seguridad. Cuando haya completado los siguientes pasos, si lo desea, puede volver a este paso, elija Nodos principales y de tareas y repita los pasos siguientes para permitir que el cliente SSH acceda a dichos nodos.
4. Se abrirá la consola de EC2. Elija la pestaña Reglas de entrada y, a continuación, elija Editar reglas de entrada.
5. Compruebe si hay una regla de entrada que permita el acceso público con la siguiente configuración. Si existe, seleccione Eliminar para eliminarla.

   - Tipo

     SSH
   - Puerto

     22
   - Fuente

     0.0.0.0/0 personalizado

> **Warning**
>
> Antes de diciembre de 2020, el grupo de seguridad ElasticMapReduce-master tenía una regla preconfigurada para permitir el tráfico entrante en el puerto 22 desde todos los orígenes. Esta regla se creó para simplificar las conexiones SSH iniciales al nodo maestro. Le recomendamos encarecidamente que elimine esta regla de entrada y que restrinja el tráfico a los orígenes de confianza.

6. Desplácese a la parte inferior de la lista de reglas y seleccione Agregar regla.
7. En Type (Tipo), seleccione SSH. Al seleccionar SSH, se ingresa automáticamente TCP en Protocolo y 22 en Rango de puertos.
8. Como origen, seleccione Mi IP para agregar automáticamente su dirección IP como dirección de origen. También puede agregar un rango de direcciones IP de clientes de confianza personalizadas o crear reglas adicionales para otros clientes. Muchos entornos de red asignan direcciones IP de forma dinámica, por lo que es posible que en el futuro necesite actualizar las direcciones IP de los clientes de confianza.
9. Seleccione Save.
10. Si lo desea, elija Nodos principales y de tareas en la lista y repita los pasos anteriores para permitir que el cliente SSH acceda dichos nodos.

Old console

Para conceder a los orígenes de confianza acceso SSH al grupo de seguridad principal con la consola antigua

Para editar los grupos de seguridad, debe tener permiso para administrar los grupos de seguridad de la VPC en la que se encuentra el clúster. Para obtener más información, consulte Cambio de los permisos de un usuario y el Ejemplo de política que permite administrar los grupos de seguridad de EC2 en la Guía del usuario de IAM.

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Clusters (Clústeres). Seleccione el nombre del clúster que desea modificar.
3. Elija el enlace Grupos de seguridad para nodo maestro de Seguridad y acceso.

4. Elija ElasticMapReduce-master en la lista.
5. Elija la pestaña Reglas de entrada y, a continuación, elija Editar reglas de entrada.
6. Compruebe si hay una regla de entrada que permita el acceso público con la siguiente configuración. Si existe, seleccione Eliminar para eliminarla.

   - Tipo

     SSH
   - Puerto

     22
   - Fuente

     0.0.0.0/0 personalizado

   > **Warning**
   >
   > Antes de diciembre de 2020, el grupo de seguridad ElasticMapReduce-master tenía una regla preconfigurada para permitir el tráfico entrante en el puerto 22 desde todos los orígenes. Esta regla se creó para simplificar las conexiones SSH iniciales al nodo principal. Le recomendamos encarecidamente que elimine esta regla de entrada y que restrinja el tráfico a los orígenes de confianza.

7. Desplácese a la parte inferior de la lista de reglas y seleccione Agregar regla.
8. En Type (Tipo), seleccione SSH.

   Al seleccionar SSH, se ingresa automáticamente TCP en Protocolo y 22 en Rango de puertos.
9. Como origen, seleccione Mi IP para agregar automáticamente su dirección IP como dirección de origen. También puede agregar un rango de direcciones IP de clientes de confianza personalizadas o crear reglas adicionales para otros clientes. Muchos entornos de red asignan direcciones IP de forma dinámica, por lo que es posible que en el futuro necesite actualizar las direcciones IP de los clientes de confianza.
10. Seleccione Save.
11. Si lo desea, puede seleccionar ElasticMapReduce-slave en la lista y repetir los pasos anteriores para permitir el acceso de los clientes SSH a los nodos principales y de tarea.

Conectarse al clúster con la AWS CLI

Independientemente del sistema operativo, puede crear una conexión SSH a su clúster mediante la AWS CLI.

Para conectarse a su clúster y ver los archivos de registro mediante la AWS CLI

1. Use el comando siguiente para abrir una conexión SSH a su clúster. Sustituya *<mykeypair.key>* por la ubicación y el nombre del archivo del par de claves. Por ejemplo, `C:\Users\<username>\.ssh\mykeypair.pem`.

   ```
   aws emr ssh --cluster-id <j-2AL4XXXXXX5T9> --key-pair-file <~/mykeypair.key>
   ```

2. Navegue hasta `/mnt/var/log/spark` para acceder a los registros de Spark en el nodo maestro del clúster. A continuación, consulte los archivos de esa ubicación. Para obtener una lista de los archivos de registro adicionales del nodo maestro, consulte Ver archivos de registro en el nodo principal.

   ```
   cd /mnt/var/log/spark
   ls
   ```

Paso 3: eliminar los recursos de Amazon EMR

Terminar su clúster

Ahora que ha enviado trabajo a su clúster y ha visto los resultados de su aplicación de PySpark, puede terminar el clúster. Al terminar un clúster, se detienen todos los cargos de Amazon EMR y las instancias de Amazon EC2 asociados al clúster.

Al terminar un clúster, Amazon EMR conserva los metadatos del clúster durante dos meses sin costo alguno. Los metadatos archivados le ayudan a clonar el clúster para un nuevo trabajo o a revisitar la configuración del clúster como referencia. Los metadatos no incluyen los datos que el clúster escribe en S3 ni los datos almacenados en el HDFS del clúster.

> **Note**
>
> La consola de Amazon EMR no le permite eliminar un clúster de la vista de lista una vez terminado el clúster. Un clúster terminado desaparece de la consola cuando Amazon EMR borra sus metadatos.

New console

Para terminar el clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. Seleccione Clústeres y, a continuación, elija el clúster que desea terminar.
3. En el menú desplegable Acciones, seleccione Terminar clúster.
4. Seleccione Terminar en el cuadro de diálogo. Según la configuración del clúster, la terminación puede tardar de 5 a 10 minutos. Para obtener más información sobre cómo terminar clústeres en Amazon EMR, consulte Terminar un clúster.

Old console

Para terminar el clúster con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Clústeres y, a continuación, seleccione el clúster que desea terminar. Por ejemplo, *Mi primer clúster de EMR*.
3. Seleccione Terminar para abrir el mensaje Terminar clúster.
4. Elija Terminar en el mensaje abierto. Según la configuración del clúster, la terminación puede tardar de 5 a 10 minutos. Para obtener más información sobre la terminación de clústeres de Amazon EMR, consulte Terminar un clúster.

> **Note**
>
> Si ha seguido el tutorial al pie de la letra, la protección de terminación debería estar desactivada. La protección de terminación de clúster evita la terminación accidental.

Si la protección de terminación está activada, verá un mensaje para cambiar esta opción como medida de precaución antes de terminar el clúster. Seleccione Cambiar, y luego Desactivada.

CLI

Para terminar el clúster con la AWS CLI

1. Inicie el proceso de terminación del clúster con el siguiente comando. Sustituya *<myClusterId>* por el ID del clúster de muestra. El comando no devuelve resultados.

```
aws emr terminate-clusters --cluster-ids <myClusterId>
```

2. Para comprobar que el proceso de terminación del clúster esté en curso, compruebe el estado del clúster con el siguiente comando.

```
aws emr describe-cluster --cluster-id <myClusterId>
```

El siguiente ejemplo está en formato JSON. El clúster `Status` debe cambiar de **`TERMINATING`** a **`TERMINATED`**. Según la configuración del clúster, la terminación puede tardar de 5 a 10 minutos. Para obtener más información sobre la terminación de un clúster de Amazon EMR, consulte Terminar un clúster.

```
{
    "Cluster": {
        "Id": "j-xxxxxxxxxxxxx",
        "Name": "My Cluster Name",
        "Status": {
            "State": "TERMINATED",
            "StateChangeReason": {
                "Code": "USER_REQUEST",
                "Message": "Terminated by user request"
            }
        }
    }
}
```

Eliminar recursos de S3

Para evitar cargos adicionales, debe eliminar el bucket de Amazon S3. Al eliminar el bucket, se eliminan todos los recursos de Amazon S3 de este tutorial. El bucket debe contener:

- El script de PySpark
- El conjunto de datos de entrada
- La carpeta de resultados de salida
- La carpeta de archivos de registro

Es posible que tenga que tomar medidas adicionales para eliminar los archivos almacenados si guardó el script o la salida de PySpark en una ubicación diferente.

> **Note**
>
> Debe terminar el clúster antes de eliminar el bucket. De lo contrario, es posible que no se le permita vaciar el bucket.

Para ello, siga las instrucciones de Eliminar un bucket de S3 en la Guía del usuario de Amazon Simple Storage Service.

Pasos siguientes

Ya ha lanzado su primer clúster de Amazon EMR de principio a fin. También ha completado tareas esenciales de EMR, como preparar y enviar aplicaciones de macrodatos, ver los resultados y terminar un clúster.

Utilice los siguientes temas para obtener más información sobre cómo personalizar el flujo de trabajo de Amazon EMR.

Explorar las aplicaciones de macrodatos para Amazon EMR

Descubra y compare las aplicaciones de macrodatos que puede instalar en un clúster en la Guía de publicación de Amazon EMR. La guía de versiones detalla cada versión de EMR e incluye consejos para usar marcos como Spark y Hadoop en Amazon EMR.

Planificar el hardware, las redes y la seguridad de los clústeres

En este tutorial, creó un clúster de EMR sencillo sin configurar las opciones avanzadas. Las opciones avanzadas le permiten especificar los tipos de instancias de Amazon EC2, las redes de clústeres y la seguridad de los clústeres. Para obtener más información sobre la planificación y el lanzamiento de un clúster que cumpla sus requisitos, consulte Planificar y configurar clústeres y Seguridad en Amazon EMR.

Administrar clústeres

Profundice en el trabajo con clústeres en ejecución en Administrar clústeres. Para administrar un clúster, puede conectarse al clúster, depurar los pasos y hacer un seguimiento de las actividades y el estado del clúster. También puede ajustar los recursos del clúster en respuesta a las demandas de carga de trabajo con el escalado administrado por EMR.

Utilizar una interfaz diferente

Además de la consola de Amazon EMR, puede administrar Amazon EMR mediante la AWS Command Line Interface, la API del servicio web o uno de los muchos AWS SDK compatibles. Para obtener más información, consulte Interfaces de administración.

También puede interactuar con las aplicaciones instaladas en los clústeres de Amazon EMR de muchas maneras. Algunas aplicaciones, como Apache Hadoop, publican interfaces web que puede consultar. Para obtener más información, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Consultar el blog técnico de EMR

Para ver ejemplos de tutoriales y un análisis técnico detallado sobre las nuevas características de Amazon EMR, consulte el blog de macrodatos de AWS.

Novedades de la consola

Amazon EMR ha migrado a una nueva experiencia. La nueva consola ofrece una interfaz actualizada que le proporciona una forma intuitiva de administrar su entorno de Amazon EMR y le brinda un cómodo acceso a la documentación, la información del producto y otros recursos. Esta página describe diferencias importantes entre la experiencia de la consola anterior y la nueva AWS Management Console de Amazon EMR.

¿En qué consola estoy?

Para determinar la consola de Amazon EMR que utiliza actualmente, consulte la URL de la página de la consola en su navegador:

- URL de la nueva consola: https://console.aws.amazon.com/emr
- URL de la consola antigua: https://console.aws.amazon.com/elasticmapreduce

> **Note**
>
> Amazon EMR tiene una nueva experiencia de consola. La consola anterior ha quedado obsoleta y ya no está disponible.

La funcionalidad de la consola de Amazon EMR está migrando a la nueva experiencia por fases. En la siguiente tabla, se enumeran los principales componentes de la consola de Amazon EMR y su estado de migración.

Componente de la consola de Amazon EMR	New console	Consola antigua
EMR Studio[1]	✓	✓
Crear y administrar clústeres	✓	✓
Bloqueo de acceso público	✓	✓
Supervisar eventos de Amazon CloudWatch	✓	✓

Componente de la consola de Amazon EMR	New console	Consola antigua
Configuraciones de seguridad	✓	✓
Clústeres virtuales (Amazon EMR en EKS)	✓	✓
Ver y gestionar sus subredes de Amazon Virtual Private Cloud[2]	✓	✓
Cuadernos[3]	✓	✓

[1] EMR Studio utiliza la nueva experiencia de interfaz tanto en las consolas nuevas como en las antiguas.

[2] En la nueva consola, puede ver y administrar las subredes de Amazon VPC en la sección Redes al crear un clúster. En la consola antigua, utilice el enlace de la barra de navegación de la izquierda para acceder a la lista de subredes de Amazon VPC.

[3] Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Uso de la consola antigua

Amazon EMR tiene una nueva experiencia de consola. La consola anterior ha quedado obsoleta y ya no está disponible.

Resumen de las diferencias

Esta sección destaca las diferencias entre las experiencias de la antigua y la nueva consola de Amazon EMR. Las diferencias se dividen en las siguientes categorías:

- Compatibilidad de clústeres entre la consola antigua y la nueva
- Diferencias a la hora de crear clústeres
- Diferencias al ver o editar los detalles del clúster
- Diferencias a la hora de enumerar y buscar clústeres
- Diferencias al trabajar con configuraciones de seguridad

Compatibilidad de clústeres entre la consola antigua y la nueva

En algunos casos, es posible que un clúster que haya creado en la antigua consola de Amazon EMR no sea compatible con la nueva consola. En la siguiente lista se describen los requisitos de compatibilidad de la nueva consola Amazon EMR.

- La nueva consola admite los clústeres creados en Amazon EMR 5.20.1 y versiones posteriores.
- Puede clonar clústeres que utilizan el escalado automático en la nueva consola, pero solo puede crear nuevos clústeres si desea escalarlos manualmente o utilizar el escalado administrado.

Para crear clústeres que no sean compatibles con la nueva consola y trabajar con ellos, puede usar AWS Command Line Interface (AWS CLI), el AWS SDK o la consola antigua.

Diferencias a la hora de crear clústeres

En la siguiente tabla, se destacan las diferencias que puede esperar al crear clústeres con la nueva consola de Amazon EMR en comparación con la antigua consola de Amazon EMR.

Capability	New console	Consola antigua
Terminología: tipos de nodos de clústeres de Amazon EMR	Principal, secundario, tarea	Maestro, secundario, tarea
Versiones compatibles con Amazon EMR[1]	Amazon EMR 5.20.1 y versiones posteriores	Todas las versiones de Amazon EMR
Lanzamiento rápido de un clúster	Utilice el botón Crear clúster situado en el panel Resumen	Utilice la página Crear clúster: Opciones rápidas

Capability	New console	Consola antigua
Configuración de un tiempo de espera de aprovisionamiento de spot	Defina un periodo de tiempo de espera para aprovisionar instancias para cada flota en el clúster.	No puede personalizar el tiempo de espera de aprovisionamiento al crear un clúster.
Roles de servicio y rol de perfil de instancia de Amazon EC2	La nueva consola no crea roles predeterminados; debe crear roles con la consola de IAM o seleccionar un rol de IAM ya creado	Admite la creación de roles por defecto con las políticas de la v1 y la v2, o puede seleccionar un rol de IAM ya creado
Visibilidad del clúster	Desde la consola de Amazon EMR, no puede hacer que un clúster sea visible para todos los usuarios; su política de IAM determina el acceso al clúster	Desde la consola de Amazon EMR, puede hacer que un clúster sea visible para todos los usuarios si utiliza las políticas de creación de roles de la versión 1, que ya están en desuso
Redes: configurar subredes privadas	Debe configurar los puntos de conexión de Amazon S3 y las puertas de enlace NAT desde sus respectivas consolas Amazon S3 y Amazon VPC	Puede configurar los puntos de conexión Amazon S3 y las puertas de enlace NAT directamente desde el flujo de trabajo Crear clúster de la consola antigua
Vista coherente del sistema de archivos de EMR (EMRFS CV)	Con el lanzamiento de Amazon S3 que cuenta con una sólida coherencia de lectura tras escritura, el 1 de diciembre de 2020, ya no necesitará usar EMRFS CV con sus clústeres de EMR	EMRFS CV está activado, pero puede desactivarlo y eliminar la base de datos de Amazon DynamoDB que utiliza; consulte Vista coherente para obtener más información

Capability	New console	Consola antigua
Debugging	Puede depurar los trabajos mediante la interfaz de usuario de la aplicación en la página de detalles del clúster	Puede utilizar una herramienta de depuración (paso 3 en Opciones avanzadas) para depurar los trabajos de los clústeres que se ejecutan en las versiones 4.1.0 a 5.27.0 de Amazon EMR

[1] No puede crear ni editar clústeres con versiones anteriores a Amazon EMR 5.20.1 en la nueva consola, pero los clústeres existentes creados con versiones anteriores a la 5.20.1 seguirán funcionando. Para crear y editar clústeres con versiones de Amazon EMR anteriores a 5.20.1, utilice la API o la CLI, o vuelva a la consola anterior.

Diferencias a la hora de enumerar y buscar clústeres

En la siguiente tabla destacan las diferencias que puede esperar al ver y buscar clústeres en la vista de lista con la nueva consola de Amazon EMR en comparación con la antigua consola de Amazon EMR.

> **Note**
>
> Tanto en la consola antigua como en la nueva, al aplicar un filtro de datos a la lista de clústeres, este consulta toda la base de datos. Sin embargo, al ingresar una cadena de texto en el cuadro de búsqueda, la búsqueda solo se aplica a los resultados que la lista haya cargado en el cliente.

Capability	New console	Consola antigua
Ver detalles del clúster	Puede seleccionar el ID del clúster para ver detalles exhaustivos del clúster, como las opciones de configura	Puede expandir y contraer cada fila del clúster para ver información como los detalles de configuración y acceder a

Capability	New console	Consola antigua
	ción, las interfaces de usuario de las aplicaciones persistentes y los registros.	los enlaces para la supervisión y los registros del clúster.
Búsqueda de clústeres	Utilice un único campo de búsqueda para ingresar consultas de búsqueda de texto y para crear y aplicar filtros de datos como “Estado = Cualquier estado activo”.	Utilice un menú desplegable para afinar el estado de los clústeres (activo, terminado o con errores) y un campo independiente para introducir una consulta de búsqueda de texto.
Búsqueda de clústeres con errores	Para buscar clústeres con errores, aplique el filtro Estado = Terminado con errores.	Para buscar clústeres con errores, aplique el filtro Clústeres erróneos.

Diferencias al ver o editar los detalles del clúster

En la siguiente tabla, se destacan las diferencias que puede esperar al ver o editar los detalles de un clúster existente con la nueva consola de Amazon EMR en comparación con la antigua consola de Amazon EMR.

Capability	New console	Consola antigua
Ver las instancias de sus grupos de instancias y flotas de instancias, junto con las opciones de escalado, aprovisionamiento, redimensionamiento y terminación	Consulte las opciones y los detalles de las instancias en la pestaña Instancias. Consulte las opciones de terminación en la pestaña Propiedades.	Consulte las opciones de configuración y terminación de la instancia en la pestaña Hardware.
	Vea las configuraciones del clúster en la pestaña	Vea las configuraciones del clúster en la pestaña

Capability	New console	Consola antigua
Visualización de las interfaces de usuario, los registros y las configuraciones (interfaz de usuario de Apache Spark, servicio Spark History, interfaz de usuario de Apache Tez, servidor de cronología YARN)	Configuraciones. Inicie una interfaz de usuario de aplicación persistente y activa para ver los registros de una aplicación desde la pestaña Aplicaciones.	Configuraciones. Inicie una interfaz de usuario de aplicación persistente y activa para ver los registros de una aplicación desde la pestaña Interfaces de usuario de aplicaciones. A partir de enero de 2023, el historial de aplicaciones de alto nivel ya no estará disponible.
Exportación de un clúster a la CLI	La opción está disponible en los menús de acciones de la vista de lista y de detalles del clúster, como “Ver el comando para clonar un clúster”	La opción está disponible en los menús de acciones de la vista de lista de clústeres como “Exportación de la AWS CLI”

Diferencias al trabajar con configuraciones de seguridad

En la siguiente tabla, se destacan las diferencias que puede esperar al configurar las opciones de seguridad con la nueva consola de Amazon EMR en comparación con la antigua consola de Amazon EMR.

Capability	New console	Consola antigua
Clonación de configuraciones de seguridad	✓	
Gobernanza federada mediante Trino y Apache Ranger	✓	
	✓	

Capability	New console	Consola antigua
Uso de un rol de tiempo de ejecución para enviar el trabajo a un clúster[1]		
Autorización del acceso a los datos del sistema de archivos EMR (EMRFS)	Puntos de acceso de Amazon S3	Roles de AWS Identity and Access Management (IAM)
Controles de acceso de AWS Lake Formation	Roles de tiempo de ejecución	Federación SAML

[1] Para transferir un rol durante el envío de pasos, el clúster debe usar una configuración de seguridad con una política de permisos de IAM adjunta para que el usuario pueda transferir solo los roles aprobados y sus trabajos puedan acceder a los recursos de Amazon EMR. Para obtener más información, consulte Roles en tiempo de ejecución para los pasos de Amazon EMR.

Amazon EMR Studio

Amazon EMR Studio es un entorno de desarrollo integrado (IDE) basado en web para cuadernos de Jupyter completamente administrados que se ejecutan en clústeres de Amazon EMR. Puede configurar un EMR Studio para que su equipo desarrolle, visualice y depure aplicaciones escritas en R, Python, Scala y PySpark. EMR Studio está integrado con AWS Identity and Access Management (IAM) e IAM Identity Center para que los usuarios puedan iniciar sesión con sus credenciales corporativas.

Puede crear un EMR Studio sin costo alguno. Cuando utiliza EMR Studio, se aplican cargos por almacenamiento de Amazon S3 y por clústeres de Amazon EMR. Para obtener detalles y aspectos destacados del producto, consulte la página de servicio de Amazon EMR Studio.

Características principales de EMR Studio

Amazon EMR Studio ofrece las siguientes características:

- Autentique a los usuarios con AWS Identity and Access Management (IAM) o AWS IAM Identity Center (IAM Identity Center) y su proveedor de identidad empresarial.
- Lance clústeres de Amazon EMR y acceda a ellos bajo demanda para ejecutar trabajos de cuaderno de Jupyter.
- Conéctese a Amazon EMR en los clústeres de EKS para enviar el trabajo como ejecuciones de trabajo.
- Explore y guarde cuadernos de muestra. Para obtener más información sobre los cuadernos de muestra, consulte el repositorio de GitHub con ejemplos de Cuadernos de EMR Studio.
- Analice los datos con Python, PySpark, Spark Scala, Spark R o SparkSQL e instale kernels y bibliotecas personalizados.
- Colabore en tiempo real con otros usuarios del mismo espacio de trabajo. Para obtener más información, consulte Configuración de la colaboración en el espacio de trabajo.
- Utilice el Explorador de SQL de EMR Studio para examinar su catálogo de datos, ejecutar consultas SQL y descargar los resultados antes de trabajar con los datos de un cuaderno.
- Ejecute cuadernos parametrizados como parte de los flujos de trabajo programados con una herramienta de orquestación como Apache Airflow o Amazon Managed Workflows para Apache Airflow. Para obtener más información, consulte Orquestación de trabajos de análisis en Cuadernos de EMR mediante MWAA en el blog de macrodatos de AWS.

- Enlace repositorios de código como GitHub y BitBucket.
- Haga un seguimiento y depure las tareas mediante el servidor de historial de Spark, la interfaz de usuario de Tez o el servidor de cronogramas YARN.

EMR Studio también cumple con los requisitos de la HIPAA y cuenta con la certificación de HITRUST CSF y de SOC 2. Para obtener más información acerca de la conformidad con la HIPAA de los servicios de AWS, consulte https://aws.amazon.com/compliance/hipaa-compliance/. Para obtener más información sobre la conformidad de los servicios con la CSF de HITRUST de los servicios de AWS, consulte https://aws.amazon.com/compliance/hitrust/. Para obtener más información sobre otros programas de conformidad para los servicios de AWS, consulte Servicios de AWS en el ámbito del programa de conformidad.

Historial de características de Amazon EMR Studio

En esta tabla se enumeran las actualizaciones de la capacidad de Escalado administrado de Amazon EMR.

Fecha de lanzamiento de la nueva versión	Capability
26 de octubre de 2023	Se agregó la posibilidad de crear una aplicación de EMR sin servidor con capacidad interactiva.
28 de febrero de 2023	Se agregó compatibilidad con claves de AWS KMS administrada por el cliente para el almacenamiento de registros de aplicaciones para aplicaciones de EMR sin servidor.
23 de febrero de 2023	Se agregó la creación de roles de IAM con un solo clic para el envío de trabajos de EMR sin servidor. Se agregó la búsqueda de ECR para cuando se selecciona una imagen personalizada para las aplicaciones de EMR sin servidor.
27 de enero de 2023	Los cuadernos de ejecución Headless pueden hacer un seguimiento del progreso de la ejecución de cada celda con magia de `%execute_notebook` .

Fecha de lanzamiento de la nueva versión	Capability
23 de enero de 2023	Las aplicaciones persistentes se han optimizado para acelerar los tiempos de lanzamiento.

Cómo funciona Amazon EMR Studio

Un Amazon EMR Studio es un recurso de Amazon EMR que se crea para un equipo de usuarios. Cada estudio es un entorno de desarrollo integrado basado en web autónomo para cuadernos de Jupyter que se ejecutan en clústeres de Amazon EMR. Los usuarios inician sesión en un estudio con sus credenciales corporativas.

Cada estudio de EMR que cree utiliza los siguientes recursos de AWS:

- Una Amazon Virtual Private Cloud (VPC) con subredes: los usuarios ejecutan kernels y aplicaciones de Studio en Amazon EMR y clústeres de Amazon EMR en EKS de la VPC especificada. Un EMR Studio puede conectarse a cualquier clúster de las subredes que especifique al crear el estudio.
- Roles y políticas de permisos de IAM: para administrar los permisos de los usuarios, debe crear políticas de permisos de IAM que se asocian a la identidad de IAM de un usuario o a un rol de usuario. EMR Studio también utiliza un rol de servicio de IAM y grupos de seguridad para interoperar con otros servicios de AWS. Para obtener más información, consulte Control de acceso y Definir grupos de seguridad para controlar el tráfico de red de EMR Studio.
- Grupos de seguridad: EMR Studio usa grupos de seguridad para establecer un canal de red seguro entre el estudio y un clúster de EMR.
- Una ubicación de copia de seguridad de Amazon S3: EMR Studio guarda el trabajo del cuaderno en una ubicación de Amazon S3.

Los siguientes pasos describen cómo crear y administrar un EMR Studio:

1. Cree un estudio en su Cuenta de AWS con la autenticación de IAM o IAM Identity Center. Para obtener instrucciones, consulte Configuración de un Amazon EMR Studio.
2. Asigne usuarios o grupos al estudio. Utilice políticas de permisos para establecer permisos específicos para cada usuario. Para obtener más información, consulte el tema Asignar y administrar usuarios de EMR Studio.

3. Comience a monitorear las acciones de EMR Studio con eventos de AWS CloudTrail. Para obtener más información, consulte Monitorear las acciones de Amazon EMR Studio.
4. Ofrezca más opciones de clústeres a los usuarios de Studio con plantillas de clúster y puntos de conexión administrados de Amazon EMR en EKS.

Autenticación e inicio de sesión de los usuarios

Amazon EMR Studio admite dos modos de autenticación: el modo de autenticación de IAM y el modo de autenticación de IAM Identity Center. El modo de IAM usa AWS Identity and Access Management (IAM), mientras que el modo de IAM Identity Center usa AWS IAM Identity Center. Cuando crea un EMR Studio, debe elegir el modo de autenticación para todos los usuarios de ese estudio.

Modo de autenticación de IAM

Con el modo de autenticación de IAM, puede utilizar la autenticación de IAM o la federación de IAM.

La autenticación de IAM le permite administrar las identidades de IAM, como los usuarios, los grupos y los roles de IAM. Concede acceso a los usuarios a un estudio con políticas de permisos de IAM y el control de acceso basado en atributos (ABAC).

La federación de IAM le permite establecer la confianza entre un proveedor de identidades (IdP) externo y AWS, de modo que puede administrar las identidades de los usuarios a través de su IdP.

Modo de autenticación de IAM Identity Center

El modo de autenticación de IAM Identity Center permite conceder a los usuarios acceso federado a un EMR Studio. Puede usar IAM Identity Center para autenticar usuarios y grupos desde su directorio de IAM Identity Center, su directorio corporativo existente o un IdP externo, como Azure Active Directory (AD). A continuación, debe administrar los usuarios con su proveedor de identidades (IdP).

EMR Studio admite el uso de los siguientes proveedores de identidades para IAM Identity Center:

- AWS Managed Microsoft AD y Active Directory autoadministrado: para obtener más información, consulte Conectarse al directorio de Microsoft AD.
- Proveedores basados en SAML: para ver una lista completa, consulte Proveedores de identidades compatibles.
- El directorio de IAM Identity Center: para obtener más información, consulte Administrar identidades en IAM Identity Center.

Cómo afecta la autenticación al inicio de sesión y a la asignación de usuarios

El modo de autenticación que elija para EMR Studio afecta a la forma en que los usuarios inician sesión en un estudio, a la forma en que asigna un usuario a un estudio y a la forma en que autoriza (otorga permisos a) los usuarios para que realicen acciones como la creación de nuevos clústeres de Amazon EMR.

En la siguiente tabla se resumen los métodos de inicio de sesión de EMR Studio según el modo de autenticación.

Opciones de inicio de sesión de EMR Studio por modo de autenticación

Modo de autenticación	Método de inicio de sesión	Descripción
• IAM (autenticación y federación) • Centro de identidades de IAM	URL de EMR Studio	Los usuarios inician sesión en un estudio mediante la URL de acceso al estudio. Por ejemplo, `https://xxxxxxxxxxxxxxxxxxxx.emrstudio-prod.us-east-1.amazonaws.com` . Los usuarios introducen las credenciales de IAM cuando se utiliza la autenticación de IAM. Cuando utiliza la federación de IAM o IAM Identity Center, EMR Studio redirige a los usuarios a la URL de inicio de sesión del proveedor de identidades para introducir las credenciales. En el contexto de la federación de identidades, esta opción de inicio de sesión se denomina inicio de sesión iniciado por el proveedor de servicios (SP).
• (Federación de) IAM • Centro de identidades de IAM	Portal del proveedor de identidades (IdP)	Los usuarios inician sesión en el portal de su proveedor de identidades, como Azure Portal, e inician la consola de Amazon EMR. Tras lanzar la consola de Amazon EMR, los usuarios seleccionan y abren un estudio de la lista Studios.

Modo de autenticación	Método de inicio de sesión	Descripción
		También puede configurar EMR Studio como una aplicación SAML para que los usuarios puedan iniciar sesión en un estudio específico desde el portal de su proveedor de identidades. Para obtener instrucciones, consulte Para configurar un EMR Studio como una aplicación SAML en el portal de su IdP. En el contexto de la federación de identidades, esta opción de inicio de sesión se denomina inicio de sesión iniciado por el proveedor de identidades (IdP).
• IAM (autenticación)	AWS Management Console	Los usuarios inician sesión en la AWS Management Console con las credenciales de IAM y abren un estudio de la lista Studios en la consola de Amazon EMR.

En la siguiente tabla se describe la asignación y autorización de usuarios en EMR Studio por modo de autenticación.

Asignación y autorización de usuarios de EMR Studio por modo de autenticación

Modo de autenticación	Asignación de usuarios	Autorización de usuarios
IAM (autenticación y federación)	Permita la acción `CreateStudioPresignedUrl` en una política de permisos de IAM asociada a una identidad de IAM (usuario, grupo o rol). En el caso de los usuarios federados, permita la acción `CreateStudioPresig`	Defina políticas de permisos de IAM que permitan determinadas acciones de EMR Studio. En el caso de los usuarios nativos, asocie la política de permisos de IAM a una identidad de IAM (usuario, grupo o rol). En el caso de los usuarios federados, permita las

Modo de autenticación	Asignación de usuarios	Autorización de usuarios
	`nedUrl` en IAM en la política de permisos que configure para el rol de IAM que utilice para la federación. Utilice el control de acceso basado en atributos (ABAC) para especificar el estudio o estudios a los que puede acceder el usuario. Para obtener instrucciones, consulte Asignar un usuario o grupo a un EMR Studio.	acciones de Studio en la política de permisos que configure para el rol de IAM que utilice para la federación. Para obtener más información, consulte Configurar los permisos de usuario de EMR Studio para Amazon EC2 o Amazon EKS.
Centro de identidades de IAM	Para asignar un usuario a un estudio, asigne el usuario a un estudio con una política de sesión específica. Para obtener instrucciones, consulte Asignar un usuario o grupo a un EMR Studio.	Defina políticas de sesión de IAM que permitan determinadas acciones de EMR Studio. Asigne una política de sesión a un usuario al asignar el usuario a un estudio. Para obtener más información, consulte Permisos de usuario para el modo de autenticación de IAM Identity Center.

Control de acceso

En Amazon EMR Studio, debe configurar los permisos de autorización de usuarios con políticas de AWS Identity and Access Management (IAM) basadas en la identidad. Con estas políticas, debe especificar las acciones y los recursos permitidos, así como las condiciones en las que se permiten las acciones.

Permisos de usuario para el modo de autenticación de IAM

Para establecer los permisos de usuario al utilizar la autenticación de IAM en EMR Studio, debe permitir acciones como `elasticmapreduce:RunJobFlow` en una política de permisos de IAM. Puede crear una o más políticas de permisos para utilizarlas. Por ejemplo, puede crear una política básica que no permita a un usuario crear nuevos clústeres de Amazon EMR y otra política que sí permita la creación de clústeres. Para obtener una lista de las acciones de Studio, consulte Permisos de AWS Identity and Access Management para los usuarios de EMR Studio.

Permisos de usuario para el modo de autenticación de IAM Identity Center

Al utilizar la autenticación de IAM Identity Center, debe crear un único rol de usuario de EMR Studio. El rol de usuario es un rol de IAM dedicado que un estudio asume cuando un usuario inicia sesión.

Asocie las políticas de sesión de IAM al rol de usuario de EMR Studio. Una política de sesión es un tipo especial de política de permisos de IAM que limita lo que un usuario federado puede hacer durante una sesión de inicio de sesión en Studio. Las políticas de sesión le permiten establecer permisos específicos para un usuario o grupo sin crear varios roles de usuario para EMR Studio.

Al asignar usuarios y grupos a un estudio, debe asignar una política de sesión a ese usuario o grupo para aplicar permisos específicos. También puede actualizar la política de sesión de un usuario o grupo en cualquier momento. Amazon EMR almacena cada asignación de políticas de sesión que cree.

Para obtener más información sobre las políticas de sesión, consulte el tema Políticas y permisos en la Guía del usuario de AWS Identity and Access Management.

Workspaces

Los espacios de trabajo son los componentes principales de Amazon EMR Studio. Para organizar los cuadernos, los usuarios crean uno o más espacios de trabajo en un estudio. Para obtener más información, consulte Aprenda los conceptos básicos de los espacios de trabajo.

Al igual que los espacios de trabajo de JupyterLab, un espacio de trabajo conserva el estado del trabajo del cuaderno. Sin embargo, la interfaz de usuario del espacio de trabajo amplía la interfaz de código abierto de JupyterLab con herramientas adicionales que le permiten crear y asociar clústeres de EMR, ejecutar trabajos, explorar cuadernos de muestra y vincular repositorios de Git.

En la siguiente lista se incluyen las principales características de los espacios de trabajo de EMR Studio:

- La visibilidad del espacio de trabajo se basa en Studio. Los espacios de trabajo que cree en un estudio no están visibles en otros estudios.
- De forma predeterminada, los espacios de trabajo son compartidos y pueden verlos todos los usuarios de Studio. Sin embargo, solo un usuario puede abrir y trabajar en un espacio de trabajo a la vez. Para trabajar simultáneamente con otros usuarios, puede Configuración de la colaboración en el espacio de trabajo.
- Puede colaborar simultáneamente con otros usuarios en un espacio de trabajo si activa la colaboración en el espacio de trabajo. Para obtener más información, consulte Configuración de la colaboración en el espacio de trabajo.
- Los cuadernos de un espacio de trabajo comparten el mismo clúster de EMR para ejecutar comandos. Puede asociar un espacio de trabajo a un clúster de Amazon EMR que se ejecute en Amazon EC2 o en un clúster virtual y un punto de conexión administrado de Amazon EMR en EKS.
- Los espacios de trabajo pueden cambiar a otra zona de disponibilidad que asocie a las subredes de un estudio. Puede detener y reiniciar un espacio de trabajo para iniciar el proceso de conmutación por error. Al reiniciar un espacio de trabajo, EMR Studio lanza el espacio de trabajo en una zona de disponibilidad diferente de la VPC del estudio cuando el estudio está configurado con acceso a varias zonas de disponibilidad. Si el estudio solo tiene una zona de disponibilidad, EMR Studio intenta lanzar el espacio de trabajo en una subred diferente. Para obtener más información, consulte Resolver problemas de conectividad con el espacio de trabajo.
- Un espacio de trabajo puede conectarse a clústeres de cualquiera de las subredes asociadas al estudio.

Para obtener más información sobre cómo crear y configurar espacios de trabajo de EMR Studio, consulte Aprenda los conceptos básicos de los espacios de trabajo.

Almacenamiento de cuadernos en Amazon EMR Studio

Cuando utiliza un espacio de trabajo, EMR Studio guarda automáticamente las celdas de los archivos del cuaderno con una cadencia normal en la ubicación de Amazon S3 asociada a su estudio. Este proceso de copia de seguridad conserva el trabajo entre sesiones para que pueda volver a él más adelante sin tener que realizar cambios en un repositorio de Git. Para obtener más información, consulte Guardar contenido del espacio de trabajo.

Al eliminar un archivo de cuaderno de un espacio de trabajo, EMR Studio elimina automáticamente la versión de copia de seguridad de Amazon S3. Sin embargo, si elimina un espacio de trabajo sin

eliminar primero los archivos del cuaderno, los archivos del cuaderno permanecerán en Amazon S3 y seguirán generando gastos de almacenamiento. Para obtener más información, consulte Eliminar un espacio de trabajo y archivos de cuaderno.

Consideraciones sobre EMR Studio

Consideraciones

Tenga en cuenta lo siguiente cuando trabaje con EMR Studio:

- EMR Studio está disponible en las siguientes regiones de AWS: Este de EE. UU. (Norte de Virginia, Ohio), Oeste de EE. UU. (Norte de California, Oregón), Asia-Pacífico (Bombay, Seúl, Singapur, Sídney, Tokio), Canadá (Centro), UE (Fráncfort, Irlanda, Londres, París, Estocolmo) y América del Sur (São Paulo).
- Para permitir a los usuarios aprovisionar nuevos clústeres de EMR que se ejecuten en Amazon EC2 para un espacio de trabajo, puede asociar un estudio de EMR a un conjunto de plantillas de clúster. Los administradores pueden definir plantillas de clúster con Service Catalog y elegir si un usuario o un grupo puede acceder a las plantillas de clúster o a ninguna dentro del estudio.
- Cuando defina los permisos de acceso a los archivos de cuaderno almacenados en Amazon S3 o lea secretos de AWS Secrets Manager, utilice el rol de servicio de Amazon EMR. Estos permisos no admiten políticas de sesión.
- Puede crear varios estudios de EMR Studio para controlar el acceso a los clústeres de EMR en diferentes VPC.
- Use la AWS CLI para configurar Amazon EMR en los clústeres de EKS. A continuación, puede utilizar la interfaz de Studio para asociar clústeres a los espacios de trabajo con un punto de conexión administrado para ejecutar trabajos de cuaderno.
- EMR Studio no admite los siguientes comandos mágicos de Python:
  - `%alias`
  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - `%%js`
  - `%%javascript`
  - Modificar `proxy_user` mediante `%configure`

  - Modificar `KERNEL_USERNAME` mediante `%env` o `%set_env`
- Los clústeres de Amazon EMR en EKS no admiten los comandos de SparkMagic para EMR Studio.
- Para escribir instrucciones de Scala de varias líneas en celdas de cuadernos, asegúrese de que todas las líneas, excepto la última, terminen con un punto. En el siguiente ejemplo, se utiliza la sintaxis correcta para las instrucciones de Scala de varias líneas.

```
val df = spark.sql("SELECT * from table_name).
        filter("col1=='value'").
        limit(50)
```

Problemas conocidos

- Asegúrese de desactivar las herramientas de administración de proxy como FoxyProxy o SwitchyOmega en el navegador antes de crear un estudio. Los proxies activos pueden provocar errores al seleccionar Crear estudio y pueden generar un mensaje de error de red.
- Los kernels que se ejecutan en clústeres de Amazon EMR en EKS pueden no iniciarse debido a problemas de tiempo de espera. Si se produce un error o un problema al iniciar el kernel, cierre el archivo del cuaderno, apague el kernel y vuelva a abrir el archivo del cuaderno.
- La operación Reiniciar el kernel no funciona según lo esperado cuando se utiliza un clúster de Amazon EMR en EKS. Tras seleccionar Reiniciar el kernel, actualice el espacio de trabajo para que el reinicio surta efecto.
- Si un espacio de trabajo no está conectado a un clúster, aparece un mensaje de error cuando un usuario de Studio abre un archivo de cuaderno e intenta seleccionar un kernel. Para ignorar este mensaje de error, pulse Aceptar, pero debe asociar el espacio de trabajo a un clúster y seleccionar un kernel para poder ejecutar el código del cuaderno.
- Cuando utiliza Amazon EMR 6.2.0 con una configuración de seguridad para configurar la seguridad del clúster, la interfaz del espacio de trabajo aparece en blanco y no funciona como se esperaba. Le recomendamos que utilice una versión compatible diferente de Amazon EMR si desea configurar el cifrado de datos o la autorización de Amazon S3 para EMRFS en un clúster. EMR Studio funciona con las versiones 5.32.0 (serie Amazon EMR 5.x) o 6.2.0 (serie Amazon EMR 6.x) y posteriores de Amazon EMR.
- Cuando ejecuta la interfaz de usuario de Spark en el clúster desde un archivo de cuaderno, puede ver información sobre un trabajo después de ejecutar el código del cuaderno. Sin embargo, si inicia

el servidor de historial de Spark desde la lista Clústeres de Studio, es posible que el trabajo no aparezca hasta dentro de dos minutos.

- Cuando Depurar Amazon EMR que se ejecuta en trabajos de Amazon EC2, es posible que los enlaces a la interfaz de usuario de Spark en el clúster no funcionen o no aparezcan. Para regenerar los enlaces, cree una nueva celda del cuaderno y ejecute el comando `%%info`.
- Jupyter Enterprise Gateway no limpia los kernels inactivos del nodo principal de un clúster en las siguientes versiones de Amazon EMR: 5.32.0, 5.33.0, 6.2.0 y 6.3.0. Los kernels inactivos consumen recursos de computación y pueden provocar errores en los clústeres que se ejecutan durante mucho tiempo. Puede configurar la limpieza de kernels inactivos de Jupyter Enterprise Gateway mediante el siguiente script de ejemplo. Puede Conectarse al nodo principal mediante SSH o enviar el script como un paso. Para obtener más información, consulte Ejecutar comandos y scripts en un clúster de Amazon EMR.

```
#!/bin/bash
sudo tee -a /emr/notebook-env/conf/jupyter_enterprise_gateway_config.py << EOF
c.MappingKernelManager.cull_connected = True
c.MappingKernelManager.cull_idle_timeout = 10800
c.MappingKernelManager.cull_interval = 300
EOF
sudo systemctl daemon-reload
sudo systemctl restart jupyter_enterprise_gateway
```

- Cuando utiliza una política de terminación automática con las versiones 5.32.0, 5.33.0, 6.2.0 o 6.3.0 de Amazon EMR, Amazon EMR marca un clúster como inactivo y puede terminarlo automáticamente incluso si tiene un kernel de Python3 activo. Esto se debe a que al ejecutar un kernel de Python3 no se envía ningún trabajo de Spark al clúster. Para utilizar la terminación automática con un kernel de Python3, le recomendamos que utilice la versión 6.4.0 o posterior de Amazon EMR. Para obtener más información sobre la terminación automática, consulte Uso de una política de terminación automática.
- Cuando usa `%%display` para mostrar un DataFrame de Spark en una tabla, es posible que las tablas muy anchas queden truncadas. Puede hacer clic con el botón derecho en la salida y seleccionar Crear nueva vista para la salida para obtener una vista desplazable de la salida.
- Al iniciar un kernel basado en Spark, como PySpark, Spark o SparkR, se inicia una sesión de Spark y, al ejecutar una celda de un cuaderno, se ponen en cola los trabajos de Spark de esa sesión. Cuando interrumpes una celda en ejecución, el trabajo de Spark continúa ejecutándose. Para detener el trabajo de Spark, debe usar la interfaz de usuario de Spark en el clúster. Para

obtener instrucciones sobre cómo conectarse a la interfaz de usuario de Spark, consulte Depurar aplicaciones y trabajos con EMR Studio.

Limitaciones de características

Amazon EMR Studio no admite las siguientes características de Amazon EMR:

- Asociar y ejecutar trabajos en clústeres de EMR con una configuración de seguridad que especifique la autenticación de Kerberos
- Clústeres con varios nodos principales
- Clústeres que utilizan instancias de Amazon EC2 basadas en AWS Graviton2 para las versiones 6.x de Amazon EMR anteriores a la 6.9.0 y las versiones 5.x inferiores a la 5.36.1

Límites de servicio para EMR Studio

En la siguiente tabla se muestran los límites de servicio de EMR Studio.

Elemento	Límite
Estudios de EMR Studio	Máximo de 100 por cuenta de AWS
Subredes	Máximo de 5 asociados a cada EMR Studio
Grupos de IAM Identity Center	Máximo de 5 asignados a cada EMR Studio
Usuarios de IAM Identity Center	Máximo de 100 asignados a cada EMR Studio

Prácticas recomendadas sobre VPC y subredes

Utilice las siguientes prácticas recomendadas para configurar una Amazon Virtual Private Cloud (Amazon VPC) con subredes para EMR Studio:

- Puede especificar un máximo de cinco subredes en la VPC para asociarlas al estudio. Recomendamos que proporcione varias subredes en diferentes zonas de disponibilidad para respaldar la disponibilidad del espacio de trabajo y permitir que los usuarios de Studio accedan a los clústeres de las distintas zonas de disponibilidad. Para obtener más información sobre cómo

trabajar con las VPC, las subredes y las zonas de disponibilidad, consulta VPC y subredes en la Guía del usuario de Amazon Virtual Private Cloud.

- Las subredes que especifique deben poder comunicarse entre sí.
- Para permitir que los usuarios vinculen un espacio de trabajo a repositorios de Git alojados públicamente, debe especificar solo las subredes privadas que tengan acceso a internet a través de la traducción de direcciones de red (NAT). Para obtener más información sobre la configuración de una subred privada para Amazon EMR, consulte Subredes privadas.
- Cuando utiliza Amazon EMR en EKS con EMR Studio, debe haber al menos una subred en común entre su estudio y el clúster de Amazon EKS que utiliza para registrar un clúster virtual. De lo contrario, el punto de conexión administrado no aparecerá como opción en los espacios de trabajo de Studio. Puede crear un clúster de Amazon EKS y asociarlo a una subred que pertenezca al estudio, o bien crear un estudio y especificar las subredes del clúster de EKS.
- Si planea usar Amazon EMR en EKS con EMR Studio, seleccione la misma VPC como los nodos de trabajo del clúster de Amazon EKS.

Requisitos de clúster para Amazon EMR Studio

Clústeres de Amazon EMR que se ejecutan en Amazon EC2

Todos los clústeres de Amazon EMR que se ejecuten en Amazon EC2 y que cree para un espacio de trabajo de EMR Studio deben cumplir los siguientes requisitos. Los clústeres que cree mediante la interfaz de EMR Studio cumplen automáticamente estos requisitos.

- El clúster debe usar las versiones 5.32.0 (serie Amazon EMR 5.x) o 6.2.0 (serie Amazon EMR 6.x) o posteriores de Amazon EMR. Puede crear un clúster mediante la consola de Amazon EMR, AWS Command Line Interface o el SDK y, a continuación, asociarlo a un espacio de trabajo de EMR Studio. Los usuarios del estudio también pueden aprovisionar y asociar clústeres al crear o trabajar en un espacio de trabajo de Amazon EMR. Para obtener más información, consulte Asociar computación a un espacio de trabajo de EMR Studio.
- El clúster debe estar dentro de una Amazon Virtual Private Cloud. No se admite la plataforma EC2-Classic.
- El clúster debe tener Spark, Livy y Jupyter Enterprise Gateway instalados. Si planea usar el clúster para el Explorador de SQL, debe instalar Presto y Spark.
- Para usar el Explorador de SQL, el clúster debe usar la versión 5.34.0 o posterior o la versión 6.4.0 o posterior de Amazon EMR y tener instalado Presto. Si quiere especificar el Catálogo de datos

de AWS Glue como el metaalmacén de Hive para Presto, debe configurarlo en el clúster. Para obtener más información, consulte Uso de Presto con el Catálogo de datos de AWS Glue.

- El clúster debe estar en una subred privada con traducción de direcciones de red (NAT) para usar repositorios de Git alojados públicamente con EMR Studio.

Recomendamos las siguientes configuraciones de clúster cuando trabaje con EMR Studio.

- Configure el modo de despliegue de las sesiones de Spark en el modo de clúster. El modo de clúster coloca los procesos maestros de la aplicación en los nodos básicos y no en el nodo principal de un clúster. De este modo, se alivian las posibles presiones de memoria del nodo principal. Para obtener más información, consulte Cluster Mode Overview en la documentación de Apache Spark.
- Cambie el tiempo de espera de Livy del valor predeterminado de una hora a seis horas, como en el siguiente ejemplo de configuración.

```
{
    "classification":"livy-conf",
        "Properties":{
            "livy.server.session.timeout":"6h",
            "livy.spark.deploy-mode":"cluster"
        }
}
```

- Cree diversas flotas de instancias con hasta 30 instancias y seleccione varios tipos de instancias en su flota de instancias de spot. Por ejemplo, puede especificar los siguientes tipos de instancias optimizadas para memoria para las cargas de trabajo de Spark: r5.2x, r5.4x, r5.8x, r5.12x, r5.16x, r4.2x, r4.4x, r4.8x, r4.12, etc. Para obtener más información, consulte Configurar flotas de instancias.
- Utilice la estrategia de asignación de capacidad optimizada para las instancias de spot para ayudar a Amazon EMR a seleccionar instancias de forma eficaz en función de la información sobre la capacidad en tiempo real de Amazon EC2. Para obtener más información, consulte Estrategia de asignación para flotas de instancias.
- Habilite el escalamiento administrado en su clúster. Establezca el parámetro de número máximo de nodos principales en la capacidad persistente mínima que planea utilizar y configure el escalamiento en función de una flota de tareas bien diversificada que se ejecute en instancias de spot para ahorrar costos. Para obtener más información, consulte Uso del escalado administrado en Amazon EMR.

También le instamos a que mantenga activado el Bloqueo de acceso público de Amazon EMR y que restrinja el tráfico SSH entrante a orígenes de confianza. El acceso entrante a un clúster permite a los usuarios ejecutar cuadernos en el clúster. Para obtener más información, consulte Uso de Bloquear el acceso público de Amazon EMR y Control del tráfico de red con grupos de seguridad.

Clústeres de Amazon EMR en EKS

Además de los clústeres de EMR que se ejecutan en Amazon EC2, puede configurar y administrar clústeres de Amazon EMR en EKS para EMR Studio mediante la AWS CLI. Configure los clústeres de Amazon EMR en EKS siguiendo las pautas que se indican a continuación:

- Cree un punto de conexión HTTPS administrado para el clúster de Amazon EMR en EKS. Los usuarios asocian un espacio de trabajo a un punto de conexión administrado. El clúster de Amazon Elastic Kubernetes Service (EKS) que utilice para registrar un clúster virtual debe tener una subred privada para admitir los puntos de conexión administrados.
- Utilice un clúster de Amazon EKS con al menos una subred privada y una traducción de direcciones de red (NAT) cuando desee utilizar repositorios de Git alojados públicamente.
- Evite utilizar las AMI de Amazon Linux de Arm optimizadas para Amazon EKS, que no son compatibles con los puntos de conexión administrados de Amazon EMR en EKS.
- Evite utilizar clústeres de Amazon EKS exclusivos de AWS Fargate, ya que no son compatibles.

Configuración de Amazon EMR Studio

Esta sección es para los administradores de EMR Studio. En ella se explica cómo configurar un EMR Studio para su equipo y se proporcionan instrucciones para tareas como la asignación de usuarios y grupos, la configuración de plantillas de clústeres y la optimización de Apache Spark para EMR Studio.

Temas

- Permisos de administrador para crear y administrar un EMR Studio
- Configuración de un Amazon EMR Studio
- Administrar un Amazon EMR Studio
- Definir grupos de seguridad para controlar el tráfico de red de EMR Studio
- Crear plantillas de AWS CloudFormation para Amazon EMR Studio
- Establecer el acceso y los permisos para los repositorios basados en Git

- Optimizar los trabajos de Spark en EMR Studio

Permisos de administrador para crear y administrar un EMR Studio

Los permisos de IAM descritos en esta página le permiten crear y administrar un EMR Studio. Para obtener información detallada sobre cada permiso necesario, consulte la Permisos necesarios para administrar un EMR Studio.

Permisos necesarios para administrar un EMR Studio

En la siguiente tabla se enumeran las operaciones relacionadas con la creación y administración de un EMR Studio. En la tabla también se muestran los permisos necesarios para cada operación.

> **Note**
>
> Solo necesita las acciones `SessionMapping` de IAM Identity Center y Studio cuando utiliza el modo de autenticación del IAM Identity Center.

Permisos para crear y administrar un EMR Studio

Operación	Permisos
Crear un estudio	`"elasticmapreduce:CreateStudio", "sso:CreateManagedApplicationInstance", "iam:PassRole"`
Describir un estudio	`"elasticmapreduce:DescribeStudio", "sso:GetManagedApplicationInstance"`
Enumerar los estudios	`"elasticmapreduce:ListStudios"`
Eliminar un estudio	`"elasticmapreduce:DeleteStudio", "sso:DeleteManagedApplicationInstance"`

Operación	Permisos
Additional permissions required when you use IAM Identity Center mode	
Asignar usuarios o grupos a un estudio	`"elasticmapreduce:CreateStudioSessionMapping", "sso:GetProfile", "sso:ListDirectoryAssociations", "sso:ListProfiles", "sso:AssociateProfile" "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup"`
Recuperar los detalles de las tareas de Studio para un usuario o grupo específico	`"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:GetManagedApplicationInstance", "elasticmapreduce:GetStudioSessionMapping"`
Enumerar todos los usuarios y grupos asignados a un estudio	`"elasticmapreduce:ListStudioSessionMappings"`
Actualizar la política de sesión asociada a un usuario o grupo asignado a un estudio	`"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:GetManagedApplicationInstance", "elasticmapreduce:UpdateStudioSessionMapping"`

Operación	Permisos
Eliminar un usuario o grupo de un estudio	`"elasticmapreduce:DeleteStudioSessionMapping", "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:ListDirectoryAssociations", "sso:GetProfile", "sso:GetManagedApplicationInstance" "sso:ListProfiles", "sso:DisassociateProfile"`

Para crear una política con permisos de administrador para EMR Studio

1. Siga las instrucciones de Creación de políticas de IAM para crear una política mediante uno de los siguientes ejemplos. Los permisos que necesita dependen del modo de autenticación de EMR Studio.

   Introduzca sus propios valores para los siguientes elementos:

   - Sustituya *<your-resource-ARN>* para especificar el nombre de recurso de Amazon (ARN) del objeto u objetos que cubre la instrucción en sus casos de uso.
   - Sustituya *<region>* por el código de la Región de AWS donde planea crear el estudio.
   - Sustituya *<aws-account_id>* por el ID de la cuenta de AWS del estudio.
   - Sustituya *<EMRStudio-Service-Role>* y *<EMRStudio-User-Role>* por los nombres de su rol de servicio de EMR Studio y de su rol de usuario de EMR Studio.

   Example Política de ejemplo: permisos de administrador cuando utiliza el modo de autenticación de IAM

   ```
   {
       "Version": "2012-10-17",
       "Statement": [
           {
               "Effect": "Allow",
   ```

```
            "Resource": "arn:aws:elasticmapreduce:<region>:<aws-account-id>:studio/
*",
            "Action": [
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": "<your-resource-ARN>",
            "Action": [
                "elasticmapreduce:ListStudios"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": [
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-Service-Role>"
            ],
            "Action": "iam:PassRole"
        }
    ]
}
```

Example Política de ejemplo: permisos de administrador cuando utiliza el modo de autenticación de IAM Identity Center

> **Note**
>
> Las API de IAM Identity Center e IAM Identity Center Directory no admiten la especificación de un ARN en el elemento de recurso de una declaración de la política de IAM. Para permitir el acceso a IAM Identity Center e IAM Identity Center Directory, los siguientes permisos especifican todos los recursos, “Resource”:“*”, para las acciones de IAM Identity Center. Para obtener información, consulte Acciones, recursos y claves de condición de IAM Identity Center Directory.

```
{
    "Version": "2012-10-17",
    "Statement": [
```

```
        {
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:<region>:<aws-account-id>:studio/
*",
            "Action": [
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio",
                "elasticmapreduce:CreateStudioSessionMapping",
                "elasticmapreduce:GetStudioSessionMapping",
                "elasticmapreduce:UpdateStudioSessionMapping",
                "elasticmapreduce:DeleteStudioSessionMapping"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": "<your-resource-ARN>",
            "Action": [
                "elasticmapreduce:ListStudios",
                "elasticmapreduce:ListStudioSessionMappings"
            ]
        },
        {
            "Effect": "Allow",
            "Resource": [
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-Service-Role>",
                "arn:aws:iam::<aws-account-id>:role/<EMRStudio-User-Role>"
            ],
            "Action": "iam:PassRole"
        },
        {
            "Effect": "Allow",
            "Resource": "*",
            "Action": [
                "sso:CreateManagedApplicationInstance",
                "sso:GetManagedApplicationInstance",
                "sso:DeleteManagedApplicationInstance",
                "sso:AssociateProfile",
                "sso:DisassociateProfile",
                "sso:GetProfile",
                "sso:ListDirectoryAssociations",
                "sso:ListProfiles",
                "sso-directory:SearchUsers",
                "sso-directory:SearchGroups",
```

```
                "sso-directory:DescribeUser",
                "sso-directory:DescribeGroup"
            ]
        }
    ]
}
```

2. Asocie la política a su identidad de IAM (usuario, rol o grupo). Para ver cómo hacerlo, consulte Agregar y eliminar permisos de identidad de IAM.

Configuración de un Amazon EMR Studio

Complete los siguientes pasos para configurar un Amazon EMR Studio.

Antes de comenzar

> **Note**
>
> Si planea usar EMR Studio con Amazon EMR en EKS, le recomendamos que primero configure Amazon EMR en EKS para EMR Studio antes de configurar un estudio.

Antes de configurar un EMR Studio, asegúrese de tener los siguientes elementos:

- Una Cuenta de AWS. Para obtener instrucciones, consulte Configuración de Amazon EMR.
- Permisos para crear y administrar un EMR Studio. Para obtener más información, consulte the section called “Permisos de administrador para crear un EMR Studio”.
- Un bucket de Amazon S3 en el que EMR Studio puede hacer copias de seguridad de los espacios de trabajo y los archivos de cuadernos de su estudio. Para obtener instrucciones, consulte Crear un bucket en la Guía del usuario de Amazon Simple Storage Service (S3).
- Si desea conectarse a un clúster de Amazon EMR en EC2 o Amazon EMR en EKS, o bien, utilizar repositorios de Git, necesitará una Amazon Virtual Private Cloud (VPC) para el estudio y un máximo de cinco subredes. No necesita una VPC para usar EMR Studio con EMR sin servidor. Para obtener consejos sobre cómo configurar las redes, consulte Prácticas recomendadas sobre VPC y subredes.

Para configurar un EMR Studio

1. Elija un modo de autenticación para Amazon EMR Studio
2. Cree los siguientes recursos de Studio.
   - Crear un rol de servicio de EMR Studio
   - Configurar los permisos de usuario de EMR Studio para Amazon EC2 o Amazon EKS
   - (Opcional) Definir grupos de seguridad para controlar el tráfico de red de EMR Studio.
3. Crear un EMR Studio
4. Asignar un usuario o grupo a un EMR Studio

Una vez que haya completado los pasos de configuración, podrá Utilizar un Amazon EMR Studio.

Elija un modo de autenticación para Amazon EMR Studio

EMR Studio admite dos modos de autenticación: el modo de autenticación de IAM y el modo de autenticación de IAM Identity Center. El modo de IAM usa AWS Identity and Access Management (IAM), mientras que el modo de IAM Identity Center usa AWS IAM Identity Center. Cuando crea un EMR Studio, debe elegir el modo de autenticación para todos los usuarios de ese estudio. Para obtener más información sobre los distintos modos de autenticación, consulte Autenticación e inicio de sesión de los usuarios.

Utilice la siguiente tabla para elegir un modo de autenticación para EMR Studio.

En caso de que...	Recomendamos...
Ya esté familiarizado con la autenticación o la federación de IAM o la haya configurado anteriormente	Modo de autenticación de IAM, que ofrece los siguientes beneficios: • Proporciona una configuración rápida para EMR Studio si ya administra identidades como usuarios y grupos en IAM. • Funciona con proveedores de identidad es que son compatibles con OpenID Connect (OIDC) o Security Assertion Markup Language 2.0 (SAML 2.0).

En caso de que...	Recomendamos...
	• Admite el uso de varios proveedores de identidad con la misma Cuenta de AWS. • Disponible en un amplio número de Regiones de AWS. • Cumple con SOC 2.
No tiene experiencia en AWS o Amazon EMR	Modo de autenticación de IAM Identity Center, que ofrece las siguientes características: • Permite una fácil asignación de usuarios y grupos en recursos de AWS. • Funciona con los proveedores de identidades Microsoft Active Directory y SAML 2.0. • Facilita la configuración de la federación de varias cuentas para que no tenga que configurar la federación por separado para cada Cuenta de AWS de la organización.

Configuración del modo de autenticación de IAM para Amazon EMR Studio

Con el modo de autenticación de IAM, puede utilizar la autenticación de IAM o la federación de IAM. La autenticación de IAM le permite administrar las identidades de IAM, como los usuarios, los grupos y los roles de IAM. Concede acceso a los usuarios a un estudio con políticas de permisos de IAM y el control de acceso basado en atributos (ABAC). La federación de IAM le permite establecer la confianza entre un proveedor de identidades (IdP) externo y AWS, de modo que puede administrar las identidades de los usuarios a través de su IdP.

> **Note**
>
> Si ya utiliza IAM para controlar el acceso a los recursos de AWS, o si ya ha configurado su proveedor de identidades (IdP) para IAM, consulte Permisos de usuario para el modo de autenticación de IAM para establecer los permisos de usuario cuando utilice el modo de autenticación de IAM para EMR Studio.

Uso de la federación de IAM para Amazon EMR Studio

Para utilizar la federación de IAM para EMR Studio, debe crear una relación de confianza entre su Cuenta de AWS y su proveedor de identidades (IdP) y permitir que los usuarios federados accedan a AWS Management Console. Los pasos que debe seguir para crear esta relación de confianza varían según el estándar de federación de su IdP.

En general, debe completar las siguientes tareas para configurar la federación con un IdP externo. Para obtener instrucciones completas, consulte Habilitar usuarios federados de SAML 2.0 para acceder a la AWS Management Console y Habilitar el acceso de un agente de identidades personalizado a la AWS Management Console en la Guía del usuario de AWS Identity and Access Management.

1. Recopile información de su IdP. Por lo general, esto implica generar un documento de metadatos para validar las solicitudes de autenticación SAML de su IdP.
2. Cree una entidad de IAM del proveedor de identidades para almacenar información sobre su IdP. Para obtener instrucciones, consulte Creación de proveedores de identidades de IAM.
3. Cree uno o varios roles de IAM para su IdP. EMR Studio asigna un rol a un usuario federado al iniciar sesión. El rol permite al proveedor de identidades solicitar credenciales de seguridad temporales para obtener acceso a AWS. Para obtener instrucciones, consulte Creación de un rol para un proveedor de identidades de terceros (federación). Las políticas de permisos que se asignan al rol determinan lo que los usuarios federados pueden hacer en AWS y en un EMR Studio. Para obtener más información, consulte Permisos de usuario para el modo de autenticación de IAM.
4. (Para los proveedores de SAML) Para completar la confianza de SAML, configure su IdP con información acerca de AWS y los roles que desea que asuman los usuarios federados. Este proceso de configuración crea una relación de confianza entre su IdP y AWS. Para obtener más información, consulte Configuración de una relación de confianza para usuario autenticado y agregación de notificaciones en el proveedor de identidades SAML 2.0.

Para configurar un EMR Studio como una aplicación SAML en el portal de su IdP

Puede configurar un EMR Studio concreto como aplicación SAML mediante un enlace directo al estudio. Esto permite que los usuarios inicien sesión en su portal de IdP y lancen un estudio específico en lugar de tener que navegar por la consola de Amazon EMR.

- Utilice el siguiente formato para configurar un enlace directo a su EMR Studio como URL de destino tras la verificación de la aserción de SAML.

```
https://console.aws.amazon.com/emr/home?region=<aws-region>#studio/<your-studio-id>/start
```

Configuración del modo de autenticación de IAM Identity Center para Amazon EMR Studio

Para preparar AWS IAM Identity Center para EMR Studio, debe configurar su origen de identidades y aprovisionar usuarios y grupos. El aprovisionamiento es el proceso de poner la información de usuarios y grupos a disposición de IAM Identity Center y de las aplicaciones que utilizan IAM Identity Center. Para obtener más información, consulte Aprovisionamiento de usuarios y grupos.

EMR Studio admite el uso de los siguientes proveedores de identidades para IAM Identity Center:

- AWS Managed Microsoft AD y Active Directory autoadministrado: para obtener más información, consulte Conectarse al directorio de Microsoft AD.
- Proveedores basados en SAML: para ver una lista completa, consulte Proveedores de identidades compatibles.
- El directorio de IAM Identity Center: para obtener más información, consulte Administrar identidades en IAM Identity Center.

Para configurar IAM Identity Center para EMR Studio

1. Para configurar IAM Identity Center para EMR Studio, necesita lo siguiente:

   - Una cuenta de administración en su organización de AWS si utiliza varias cuentas en su organización.

     > **Note**
     >
     > Solo debe usar su cuenta de administración para habilitar IAM Identity Center y aprovisionar usuarios y grupos. Tras configurar IAM Identity Center, utilice una cuenta de miembro para crear un EMR Studio y asignar usuarios y grupos. Para obtener más información sobre la terminología de AWS, consulte Terminología y conceptos de AWS Organizations.

   - Si activó IAM Identity Center antes del 25 de noviembre de 2019, es posible que deba habilitar las aplicaciones que utilizan IAM Identity Center para las cuentas de su organización de AWS.

Para obtener más información, consulte Habilitar las aplicaciones integradas en IAM Identity Center en las cuentas de AWS.

- Asegúrese de que los requisitos previos figuren en la página Requisitos previos de IAM Identity Center.

2. Siga las instrucciones de Habilitar IAM Identity Center para habilitar IAM Identity Center en la Región de AWS donde desee crear el EMR Studio.
3. Conecte IAM Identity Center con su proveedor de identidades y aprovisione los usuarios y grupos que desee asignar al estudio.

Si usa…	Haga lo siguiente...
Un directorio de Microsoft AD	1. Siga las instrucciones de Conectarse al directorio de Microsoft AD para conectar su directorio de Active Directory autoadministrado o su directorio de AWS Managed Microsoft AD mediante AWS Directory Service. 2. Para aprovisionar usuarios y grupos para IAM Identity Center, puede sincronizar los datos de identidades de su AD de origen con IAM Identity Center. Puede sincronizar las identidades de su AD de origen de muchas maneras. Una forma es asignar usuarios o grupos de AD a una cuenta de AWS de su organización. Para obtener instrucciones, consulte Inicio de sesión único. La sincronización puede tardar hasta dos horas. Una vez que complete este paso, los usuarios y grupos sincronizados aparecerán en su almacén de identidades.

Si usa…	Haga lo siguiente...
	> **Note** > Los usuarios y los grupos no aparecen en su almacén de identidades hasta que sincronice la información de usuarios y grupos o utilice el aprovisionamiento de usuarios justo a tiempo (JIT). Para obtener más información, consulte Aprovisionamiento cuando los usuarios provienen de Active Directory. 3. (Opcional) Tras sincronizar los usuarios y grupos de AD, puede eliminar su acceso a su cuenta de AWS que configuró en el paso anterior. Para obtener instrucciones, consulte Eliminar el acceso de los usuarios.
Un proveedor de identidades externo	Siga las instrucciones de Conectarse a su proveedor de identidades externo.
El directorio de IAM Identity Center	Al crear usuarios y grupos en IAM Identity Center, el aprovisionamiento es automático. Para obtener más información, consulte Administrar identidades en el IAM Identity Center.

Ahora puede asignar usuarios y grupos de su almacén de identidades a un EMR Studio. Para obtener instrucciones, consulte Asignar un usuario o grupo a un EMR Studio.

Crear un rol de servicio de EMR Studio

Acerca del rol de servicio de EMR Studio

Cada EMR Studio utiliza un rol de IAM con permisos que permiten que el estudio interactúe con otros servicios de AWS. Este rol de servicio debe incluir permisos que permitan que EMR Studio establezca un canal de red seguro entre los espacios de trabajo y los clústeres, almacene los archivos de cuadernos en Amazon S3 Control y acceda a AWS Secrets Manager al vincular un espacio de trabajo a un repositorio de Git.

Utilice el rol de servicio de Studio (en lugar de las políticas de sesión) para definir todos los permisos de acceso a Amazon S3 para almacenar los archivos de cuadernos y para definir los permisos de acceso de AWS Secrets Manager.

Cómo crear un rol de servicio para EMR Studio en Amazon EC2 o Amazon EKS

1. Siga las instrucciones de Creación de un rol para delegar permisos a un servicio de AWS para crear el rol de servicio mediante la siguiente política de confianza.

> **Important**
>
> La siguiente política de confianza incluye las claves de condición globales `aws:SourceArn` y `aws:SourceAccount` para limitar los permisos que concede a EMR Studio a determinados recursos de su cuenta. Si lo hace, podrá protegerse contra el problema del suplente confuso.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "elasticmapreduce.amazonaws.com"
      },
      "Action": "sts:AssumeRole",
      "Condition": {
        "StringEquals": {
          "aws:SourceAccount": "<account-id>"
        },
```

```
        "ArnLike": {
          "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
        }
      }
    }
  ]
}
```

2. Quite los permisos de rol predeterminados. A continuación, incluya los permisos del siguiente ejemplo de política de permisos de IAM. Como opción, puede crear una política personalizada que usa Permisos del rol de servicio de EMR Studio.

   Cuando proceda, cambie `"Resource":"*"` en la siguiente política para especificar el nombre de recurso de Amazon (ARN) del recurso o recursos que la declaración cubre en sus casos de uso.

   > **Important**
   >
   > - El acceso a la API `ModifyNetworkInterfaceAttribute` debe permanecer tal cual en la siguiente política debido a las limitaciones técnicas del control de acceso basado en etiquetas de Amazon EC2 y a la forma en que EMR Studio usa `ModifyNetworkInterfaceAttribute`.
   > - Las siguientes declaraciones deben permanecer sin cambios para que EMR Studio pueda trabajar con el rol de servicio: `AllowAddingEMRTagsDuringDefaultSecurityGroupCreation` y `AllowAddingTagsDuringEC2ENICreation`.
   > - Para usar la política de ejemplo, debe etiquetar los siguientes recursos con la clave "**`for-use-with-amazon-emr-managed-policies`**" y el valor "**`true`**".
   >   - Su Amazon Virtual Private Cloud (VPC) para EMR Studio.
   >   - Cada subred que desee utilizar con el estudio.
   >   - Cualquier grupo de seguridad de EMR Studio personalizado. Debe etiquetar los grupos de seguridad que haya creado durante el período de vista previa de EMR Studio si desea seguir utilizándolos.
   >   - Secretos guardados en AWS Secrets Manager que los usuarios de Studio utilizan para vincular los repositorios de Git a un espacio de trabajo.

Puede aplicar etiquetas a los recursos mediante la pestaña Etiquetas de la pantalla de recursos correspondiente de la AWS Management Console.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowEMRReadOnlyActions",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:ListInstances",
        "elasticmapreduce:DescribeCluster",
        "elasticmapreduce:ListSteps"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowEC2ENIActionsWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterfacePermission",
        "ec2:DeleteNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:network-interface/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowEC2ENIAttributeAction",
      "Effect": "Allow",
      "Action": [
        "ec2:ModifyNetworkInterfaceAttribute"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:instance/*",
        "arn:aws:ec2:*:*:network-interface/*",
```

```
        "arn:aws:ec2:*:*:security-group/*"
      ]
    },
    {
      "Sid": "AllowEC2SecurityGroupActionsWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:AuthorizeSecurityGroupEgress",
        "ec2:AuthorizeSecurityGroupIngress",
        "ec2:RevokeSecurityGroupEgress",
        "ec2:RevokeSecurityGroupIngress",
        "ec2:DeleteNetworkInterfacePermission"
      ],
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowDefaultEC2SecurityGroupsCreationWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSecurityGroup"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:security-group/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowDefaultEC2SecurityGroupsCreationInVPCWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateSecurityGroup"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:vpc/*"
      ],
```

```
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowAddingEMRTagsDuringDefaultSecurityGroupCreation",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:*:*:security-group/*",
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true",
          "ec2:CreateAction": "CreateSecurityGroup"
        }
      }
    },
    {
      "Sid": "AllowEC2ENICreationWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:network-interface/*"
      ],
      "Condition": {
        "StringEquals": {
          "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowEC2ENICreationInSubnetAndSecurityGroupWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:subnet/*",
        "arn:aws:ec2:*:*:security-group/*"
```

```
      ],
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
      }
    },
    {
      "Sid": "AllowAddingTagsDuringEC2ENICreation",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:*:*:network-interface/*",
      "Condition": {
        "StringEquals": {
          "ec2:CreateAction": "CreateNetworkInterface"
        }
      }
    },
    {
      "Sid": "AllowEC2ReadOnlyActions",
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeSecurityGroups",
        "ec2:DescribeNetworkInterfaces",
        "ec2:DescribeTags",
        "ec2:DescribeInstances",
        "ec2:DescribeSubnets",
        "ec2:DescribeVpcs"
      ],
      "Resource": "*"
    },
    {
      "Sid": "AllowSecretsManagerReadOnlyActionsWithEMRTags",
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue"
      ],
      "Resource": "arn:aws:secretsmanager:*:*:secret:*",
      "Condition": {
        "StringEquals": {
          "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
        }
```

```
      }
    },
    {
      "Sid": "AllowWorkspaceCollaboration",
      "Effect": "Allow",
      "Action": [
        "iam:GetUser",
        "iam:GetRole",
        "iam:ListUsers",
        "iam:ListRoles",
        "sso:GetManagedApplicationInstance",
        "sso-directory:SearchUsers"
      ],
      "Resource": "*"
    }
  ]
}
```

3. Conceda a su rol de servicio acceso de lectura y escritura a su ubicación de Amazon S3 para EMR Studio. Utilice el siguiente conjunto mínimo de permisos. Para obtener más información, consulte el ejemplo Amazon S3: permite el acceso de lectura y escritura a objetos en un bucket de S3 mediante programación y en la consola.

```
"s3:PutObject",
"s3:GetObject",
"s3:GetEncryptionConfiguration",
"s3:ListBucket",
"s3:DeleteObject"
```

Si cifra su bucket de Amazon S3, incluya los siguientes permisos para AWS Key Management Service.

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

Rol de servicio mínimo para EMR sin servidor

Si desea ejecutar cargas de trabajo interactivas con EMR sin servidor a través de los cuadernos de EMR Studio, utilice la misma política de confianza que empleó para configurar EMR Studio en la sección anterior: Cómo crear un rol de servicio para EMR Studio en Amazon EC2 o Amazon EKS.

Para su política de IAM, la política mínima viable tiene los siguientes permisos. Actualice *bucket-name* con el nombre del bucket que planea usar al configurar su EMR Studio y espacio de trabajo. EMR Studio usa el bucket como copia de seguridad de los espacios de trabajo y los archivos de cuadernos de su estudio.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ObjectActions",
            "Effect": "Allow",
            "Action": [
               "s3:PutObject",
               "s3:GetObject",
               "s3:DeleteObject"
            ],
            "Resource": ["arn:aws:s3:::bucket-name/*"]
        },
        {
            "Sid": "BucketActions",
            "Effect": "Allow",
            "Action": [
               "s3:ListBucket",
               "s3:GetEncryptionConfiguration"
            ],
            "Resource": ["arn:aws:s3:::bucket-name"]
        }
    ]
}
```

Si pretende usar un bucket de Amazon S3 cifrado, agregue los siguientes permisos a su política:

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
```

```
"kms:DescribeKey"
```

Permisos del rol de servicio de EMR Studio

En la siguiente tabla se enumeran las operaciones que EMR Studio realiza con el rol de servicio, junto con las acciones de IAM necesarias para cada operación.

Operación	Acciones
Establezca un canal de red seguro entre un espacio de trabajo y un clúster de EMR y lleve a cabo las acciones de limpieza necesarias.	`"ec2:CreateNetworkInterface", "ec2:CreateNetworkInterfacePermission", "ec2:DeleteNetworkInterface", "ec2:DeleteNetworkInterfacePermission", "ec2:DescribeNetworkInterfaces", "ec2:ModifyNetworkInterfaceAttribute", "ec2:AuthorizeSecurityGroupEgress", "ec2:AuthorizeSecurityGroupIngress", "ec2:CreateSecurityGroup", "ec2:DescribeSecurityGroups", "ec2:RevokeSecurityGroupEgress", "ec2:DescribeTags", "ec2:DescribeInstances", "ec2:DescribeSubnets", "ec2:DescribeVpcs", "elasticmapreduce:ListInstances", "elasticmapreduce:DescribeCluster", "elasticmapreduce:ListSteps"`
Use las credenciales de Git almacenadas en AWS Secrets Manager para vincular los repositorios de Git a un espacio de trabajo.	`"secretsmanager:GetSecretValue"`
Aplique etiquetas de AWS a la interfaz de red y a los grupos de seguridad predeterminados que EMR Studio crea al configurar el canal de red seguro. Para obtener más información, consulte Tagging	`"ec2:CreateTags"`

Operación	Acciones
AWS resources (Etiquetado de los recursos de).	

Acceda a los archivos y metadatos de los cuadernos en Amazon S3 o cárguelos.	`"s3:PutObject",`
	`"s3:GetObject",`
	`"s3:GetEncryptionConfiguration",`
	`"s3:ListBucket",`
	`"s3:DeleteObject"`
	
	Si usa un bucket de Amazon S3 cifrado, incluya los siguientes permisos.
	
	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Habilite y configure la colaboración en el espacio de trabajo.	`"iam:GetUser",`
	`"iam:GetRole",`
	`"iam:ListUsers",`
	`"iam:ListRoles",`
	`"sso:GetManagedApplicationInstance",`
	`"sso-directory:SearchUsers"`

Configurar los permisos de usuario de EMR Studio para Amazon EC2 o Amazon EKS

Debe configurar las políticas de permisos de usuario para Amazon EMR Studio para poder establecer permisos detallados de usuarios y grupos. Para obtener información sobre cómo funcionan los permisos de usuario en EMR Studio, consulte Control de acceso en Cómo funciona Amazon EMR Studio.

> **Note**
>
> Los permisos que se describen en esta sección no aplican el control de acceso a los datos. Para administrar el acceso a los conjuntos de datos de entrada, debe configurar los permisos

para los clústeres que usa Studio. Para obtener más información, consulte Seguridad en Amazon EMR.

Creación de un rol de usuario de EMR Studio para el modo de autenticación de IAM Identity Center

Debe crear un rol de usuario de EMR Studio cuando usa el modo de autenticación de IAM Identity Center.

Para crear un rol de usuario para EMR Studio

1. Siga las instrucciones que se indican en Creación de un rol para delegar permisos a un servicio de AWS en la Guía del usuario de AWS Identity and Access Management para crear un rol de usuario.

   Utilice la siguiente política de relaciones de confianza al crear el rol.

   ```
   {
     "Version": "2008-10-17",
     "Statement": [
       {
         "Effect": "Allow",
         "Principal": {
           "Service": "elasticmapreduce.amazonaws.com"
         },
         "Action": "sts:AssumeRole"
       }
     ]
   }
   ```

2. Quite los permisos de rol y las políticas predeterminados.
3. Asocie sus políticas de sesión de EMR Studio al rol de usuario antes de asignar usuarios y grupos a un estudio. Para obtener instrucciones sobre cómo crear políticas de sesión, consulte Creación de políticas de permisos para los usuarios de EMR Studio.

Creación de políticas de permisos para los usuarios de EMR Studio

Complete los siguientes pasos para crear políticas de permisos para EMR Studio.

> **Note**
>
> Para configurar los permisos de acceso a Amazon S3 para almacenar los archivos de cuadernos y los permisos de acceso a AWS Secrets Manager para leer secretos al vincular espacios de trabajo a los repositorios de Git, utilice el rol de servicio de EMR Studio.

1. Cree una o varias políticas de permisos de IAM que especifiquen qué acciones puede realizar un usuario en su estudio. Por ejemplo, puede crear tres políticas independientes para los usuarios básicos, intermedios y avanzados de Studio con las políticas de ejemplo de esta página.

   En la tabla Permisos de AWS Identity and Access Management para los usuarios de EMR Studio se desglosan todas las operaciones de Studio que puede realizar un usuario y se enumeran las acciones de IAM mínimas necesarias para realizar esa operación. Para obtener instrucciones, Creación de políticas de IAM.

   La política de permisos debe incluir las siguientes instrucciones.

   ```
   {
               "Sid": "AllowAddingTagsOnSecretsWithEMRStudioPrefix",
               "Effect": "Allow",
               "Action": "secretsmanager:TagResource",
               "Resource": "arn:aws:secretsmanager:*:*:secret:emr-studio-*"
   },
   {
               "Sid": "AllowPassingServiceRoleForWorkspaceCreation",
               "Action": "iam:PassRole",
               "Resource": [
                   "arn:aws:iam::*:role/<your-emr-studio-service-role>"
               ],
               "Effect": "Allow"
   }
   ```

   Definir la propiedad de la colaboración en el espacio de trabajo

   La colaboración en el espacio de trabajo permite que varios usuarios trabajen simultáneamente en el mismo espacio de trabajo y se puede configurar con el panel Colaboración de la interfaz de usuario del espacio de trabajo. Para ver y utilizar el panel Colaboración, el usuario debe tener los siguientes permisos. Cualquier usuario con estos permisos puede ver y usar el panel Colaboración.

```
"elasticmapreduce:UpdateEditor",
"elasticmapreduce:PutWorkspaceAccess",
"elasticmapreduce:DeleteWorkspaceAccess",
"elasticmapreduce:ListWorkspaceAccessIdentities"
```

Para restringir el acceso al panel Colaboración, puede utilizar el control de acceso basado en etiquetas. Cuando un usuario crea un espacio de trabajo, EMR Studio aplica una etiqueta predeterminada con una clave de `creatorUserId` cuyo valor es el ID del usuario que crea el espacio de trabajo.

> **Note**
>
> EMR Studio no agregó la etiqueta `creatorUserId` a los espacios de trabajo que se crearon antes del 16 de noviembre de 2021. Para restringir quién puede configurar la colaboración, le recomendamos que agregue la etiqueta `creatorUserId` manualmente a su espacio de trabajo y, a continuación, utilice el control de acceso basado en etiquetas en sus políticas de permisos de usuario.

La siguiente instrucción de ejemplo permite a un usuario configurar la colaboración para cualquier espacio de trabajo con la clave de etiqueta `creatorUserId` cuyo valor coincida con el ID del usuario (indicado por la variable de política `aws:userId`). En otras palabras, la instrucción permite al usuario configurar la colaboración para los espacios de trabajo que cree. Para obtener más información sobre variables de políticas, consulte Elementos de las políticas de IAM: variables y etiquetas.

```
    {
        "Sid": "UserRolePermissionsForCollaboration",
        "Action": [
            "elasticmapreduce:UpdateEditor",
            "elasticmapreduce:PutWorkspaceAccess",
            "elasticmapreduce:DeleteWorkspaceAccess",
            "elasticmapreduce:ListWorkspaceAccessIdentities"
        ],
        "Resource": "*",
        "Effect": "Allow",
        "Condition": {
            "StringEquals": {
```

```
                "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
            }
        }
    }
```

2. Asocie la política de permisos a su identidad de IAM.

   En la siguiente tabla se resume la identidad de IAM a la que se asocia una política de permisos, según el modo de autenticación de EMR Studio. Para obtener instrucciones acerca de cómo asociar una política, consulte Agregar y quitar permisos de identidad de IAM.

Si usa…	Asociar la política a…
Autenticación de IAM	Sus identidades de IAM (usuarios, grupos de usuarios o roles). Por ejemplo, puede asociar una política de permisos a un usuario en su Cuenta de AWS.
Federación de IAM con un proveedor de identidades (IdP) externo	El rol o roles de IAM que cree para su IdP externo. Por ejemplo, una federación de IAM para SAML 2.0. EMR Studio usa los permisos que asocia a sus roles de IAM para los usuarios con acceso federado a un estudio.
Centro de identidades de IAM	Su rol de usuario de Amazon EMR Studio.

Ejemplo de políticas de usuario

La siguiente política de usuario básico permite la mayoría de las acciones de EMR Studio, pero no permite que un usuario cree nuevos clústeres de Amazon EMR.

Política básica

> **Important**
>
> La política de ejemplo no incluye el permiso `CreateStudioPresignedUrl`, que debe conceder a un usuario al utilizar el modo de autenticación de IAM. Para obtener más información, consulte Asignar un usuario o grupo a un EMR Studio.

La política de ejemplo incluye elementos `Condition` para aplicar el control de acceso basado en etiquetas (TBAC) de modo que pueda utilizar la política con el rol de servicio de ejemplo para EMR Studio. Para obtener más información, consulte Crear un rol de servicio de EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowDefaultEC2SecurityGroupsCreationInVPCWithEMRTags",
         "Effect":"Allow",
         "Action":[
            "ec2:CreateSecurityGroup"
         ],
         "Resource":[
            "arn:aws:ec2:*:*:vpc/*"
         ],
         "Condition":{
            "StringEquals":{
               "aws:ResourceTag/for-use-with-amazon-emr-managed-policies":"true"
            }
         }
      },
      {
         "Sid":"AllowAddingEMRTagsDuringDefaultSecurityGroupCreation",
         "Effect":"Allow",
         "Action":[
            "ec2:CreateTags"
         ],
         "Resource":"arn:aws:ec2:*:*:security-group/*",
         "Condition":{
            "StringEquals":{
               "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true",
               "ec2:CreateAction":"CreateSecurityGroup"
```

```
            }
         }
      },
      {
         "Sid":"AllowSecretManagerListSecrets",
         "Action":[
            "secretsmanager:ListSecrets"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
         "Effect":"Allow",
         "Action":"secretsmanager:CreateSecret",
         "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
         "Condition":{
            "StringEquals":{
               "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
            }
         }
      },
      {
         "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
         "Effect":"Allow",
         "Action":"secretsmanager:TagResource",
         "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
      },
      {
         "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
         "Action":"iam:PassRole",
         "Resource":[
            "arn:aws:iam::*:role/<your-emr-studio-service-role>"
         ],
         "Effect":"Allow"
      },
      {
         "Sid":"AllowS3ListAndLocationPermissions",
         "Action":[
            "s3:ListAllMyBuckets",
            "s3:ListBucket",
            "s3:GetBucketLocation"
         ],
         "Resource":"arn:aws:s3:::*",
```

```
            "Effect":"Allow"
        },
        {
            "Sid":"AllowS3ReadOnlyAccessToLogs",
            "Action":[
                "s3:GetObject"
            ],
            "Resource":[
                "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
            ],
            "Effect":"Allow"
        },
        {
            "Sid":"AllowConfigurationForWorkspaceCollaboration",
            "Action":[
                "elasticmapreduce:UpdateEditor",
                "elasticmapreduce:PutWorkspaceAccess",
                "elasticmapreduce:DeleteWorkspaceAccess",
                "elasticmapreduce:ListWorkspaceAccessIdentities"
            ],
            "Resource":"*",
            "Effect":"Allow",
            "Condition":{
                "StringEquals":{
                    "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
                }
            }
        },
        {
            "Sid":"DescribeNetwork",
            "Effect":"Allow",
            "Action":[
                "ec2:DescribeVpcs",
                "ec2:DescribeSubnets",
                "ec2:DescribeSecurityGroups"
            ],
            "Resource":"*"
        },
        {
            "Sid":"ListIAMRoles",
            "Effect":"Allow",
            "Action":[
                "iam:ListRoles"
            ],
```

```
            "Resource":"*"
        }
    ]
}
```

La siguiente política de usuario intermedio permite la mayoría de las acciones de EMR Studio y permite al usuario crear nuevos clústeres de Amazon EMR mediante una plantilla de clústeres.

Política intermedia

> **Important**
>
> La política de ejemplo no incluye el permiso `CreateStudioPresignedUrl`, que debe conceder a un usuario al utilizar el modo de autenticación de IAM. Para obtener más información, consulte Asignar un usuario o grupo a un EMR Studio.

La política de ejemplo incluye elementos `Condition` para aplicar el control de acceso basado en etiquetas (TBAC) de modo que pueda utilizar la política con el rol de servicio de ejemplo para EMR Studio. Para obtener más información, consulte Crear un rol de servicio de EMR Studio.

```
{
    "Version":"2012-10-17",
    "Statement":[
        {
            "Sid":"AllowEMRBasicActions",
            "Action":[
                "elasticmapreduce:CreateEditor",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:OpenEditorInConsole",
                "elasticmapreduce:AttachEditor",
                "elasticmapreduce:DetachEditor",
                "elasticmapreduce:CreateRepository",
                "elasticmapreduce:DescribeRepository",
                "elasticmapreduce:DeleteRepository",
                "elasticmapreduce:ListRepositories",
                "elasticmapreduce:LinkRepository",
                "elasticmapreduce:UnlinkRepository",
```

```
            "elasticmapreduce:DescribeCluster",
            "elasticmapreduce:ListInstanceGroups",
            "elasticmapreduce:ListBootstrapActions",
            "elasticmapreduce:ListClusters",
            "elasticmapreduce:ListSteps",
            "elasticmapreduce:CreatePersistentAppUI",
            "elasticmapreduce:DescribePersistentAppUI",
            "elasticmapreduce:GetPersistentAppUIPresignedURL",
            "elasticmapreduce:GetOnClusterAppUIPresignedURL"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowEMRContainersBasicActions",
         "Action":[
            "emr-containers:DescribeVirtualCluster",
            "emr-containers:ListVirtualClusters",
            "emr-containers:DescribeManagedEndpoint",
            "emr-containers:ListManagedEndpoints",
            "emr-containers:DescribeJobRun",
            "emr-containers:ListJobRuns"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
          "Sid": "AllowRetrievingManagedEndpointCredentials",
          "Effect": "Allow",
          "Action": [
              "emr-containers:GetManagedEndpointSessionCredentials"
          ],
          "Resource": [
              "arn:aws:emr-containers:<region>:<account-id>:/virtualclusters/<virtual-
cluster-id>/endpoints/<managed-endpoint-id>"
          ],
          "Condition": {
              "StringEquals": {
                  "emr-containers:ExecutionRoleArn": [
                      "arn:aws:iam::<account-id>:role/<emr-on-eks-execution-role>"
                  ]
              }
          }
      },
```

```
    {
       "Sid":"AllowSecretManagerListSecrets",
       "Action":[
          "secretsmanager:ListSecrets"
       ],
       "Resource":"*",
       "Effect":"Allow"
    },
    {
       "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
       "Effect":"Allow",
       "Action":"secretsmanager:CreateSecret",
       "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
       "Condition":{
          "StringEquals":{
             "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
          }
       }
    },
    {
       "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
       "Effect":"Allow",
       "Action":"secretsmanager:TagResource",
       "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
    },
    {
       "Sid":"AllowClusterTemplateRelatedIntermediateActions",
       "Action":[
          "servicecatalog:DescribeProduct",
          "servicecatalog:DescribeProductView",
          "servicecatalog:DescribeProvisioningParameters",
          "servicecatalog:ProvisionProduct",
          "servicecatalog:SearchProducts",
          "servicecatalog:UpdateProvisionedProduct",
          "servicecatalog:ListProvisioningArtifacts",
          "servicecatalog:ListLaunchPaths",
          "servicecatalog:DescribeRecord",
          "cloudformation:DescribeStackResources"
       ],
       "Resource":"*",
       "Effect":"Allow"
    },
    {
       "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
```

```
            "Action":"iam:PassRole",
            "Resource":[
               "arn:aws:iam::*:role/<your-emr-studio-service-role>"
            ],
            "Effect":"Allow"
         },
         {
            "Sid":"AllowS3ListAndLocationPermissions",
            "Action":[
               "s3:ListAllMyBuckets",
               "s3:ListBucket",
               "s3:GetBucketLocation"
            ],
            "Resource":"arn:aws:s3:::*",
            "Effect":"Allow"
         },
         {
            "Sid":"AllowS3ReadOnlyAccessToLogs",
            "Action":[
               "s3:GetObject"
            ],
            "Resource":[
               "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
            ],
            "Effect":"Allow"
         },
         {
            "Sid":"AllowConfigurationForWorkspaceCollaboration",
            "Action":[
               "elasticmapreduce:UpdateEditor",
               "elasticmapreduce:PutWorkspaceAccess",
               "elasticmapreduce:DeleteWorkspaceAccess",
               "elasticmapreduce:ListWorkspaceAccessIdentities"
            ],
            "Resource":"*",
            "Effect":"Allow",
            "Condition":{
               "StringEquals":{
                  "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
               }
            }
         },
         {
            "Sid":"DescribeNetwork",
```

```
        "Effect":"Allow",
        "Action":[
           "ec2:DescribeVpcs",
           "ec2:DescribeSubnets",
           "ec2:DescribeSecurityGroups"
        ],
        "Resource":"*"
     },
     {
        "Sid":"ListIAMRoles",
        "Effect":"Allow",
        "Action":[
           "iam:ListRoles"
        ],
        "Resource":"*"
     },
     {
        "Sid": "AllowServerlessActions",
        "Action": [
           "emr-serverless:CreateApplication",
           "emr-serverless:UpdateApplication",
           "emr-serverless:DeleteApplication",
           "emr-serverless:ListApplications",
           "emr-serverless:GetApplication",
           "emr-serverless:StartApplication",
           "emr-serverless:StopApplication",
           "emr-serverless:StartJobRun",
           "emr-serverless:CancelJobRun",
           "emr-serverless:ListJobRuns",
           "emr-serverless:GetJobRun",
           "emr-serverless:GetDashboardForJobRun",
           "emr-serverless:AccessInteractiveEndpoints"
        ],
        "Resource": "*",
        "Effect": "Allow"
      },
      {
         "Sid": "AllowPassingRuntimeRoleForRunningServerlessJob",
         "Action": "iam:PassRole",
         "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
         "Effect": "Allow"
      }
  ]
```

```
}
```

La siguiente política de usuario avanzada admite todas las acciones de EMR Studio y permite al usuario crear nuevos clústeres de Amazon EMR mediante una plantilla de clústeres o proporcionando una configuración de clúster.

Política avanzada

> **Important**
>
> La política de ejemplo no incluye el permiso `CreateStudioPresignedUrl`, que debe conceder a un usuario al utilizar el modo de autenticación de IAM. Para obtener más información, consulte Asignar un usuario o grupo a un EMR Studio.

La política de ejemplo incluye elementos `Condition` para aplicar el control de acceso basado en etiquetas (TBAC) de modo que pueda utilizar la política con el rol de servicio de ejemplo para EMR Studio. Para obtener más información, consulte Crear un rol de servicio de EMR Studio.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"AllowEMRBasicActions",
         "Action":[
            "elasticmapreduce:CreateEditor",
            "elasticmapreduce:DescribeEditor",
            "elasticmapreduce:ListEditors",
            "elasticmapreduce:StartEditor",
            "elasticmapreduce:StopEditor",
            "elasticmapreduce:DeleteEditor",
            "elasticmapreduce:OpenEditorInConsole",
            "elasticmapreduce:AttachEditor",
            "elasticmapreduce:DetachEditor",
            "elasticmapreduce:CreateRepository",
            "elasticmapreduce:DescribeRepository",
            "elasticmapreduce:DeleteRepository",
            "elasticmapreduce:ListRepositories",
            "elasticmapreduce:LinkRepository",
            "elasticmapreduce:UnlinkRepository",
            "elasticmapreduce:DescribeCluster",
            "elasticmapreduce:ListInstanceGroups",
```

```
            "elasticmapreduce:ListBootstrapActions",
            "elasticmapreduce:ListClusters",
            "elasticmapreduce:ListSteps",
            "elasticmapreduce:CreatePersistentAppUI",
            "elasticmapreduce:DescribePersistentAppUI",
            "elasticmapreduce:GetPersistentAppUIPresignedURL",
            "elasticmapreduce:GetOnClusterAppUIPresignedURL"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
         "Sid":"AllowEMRContainersBasicActions",
         "Action":[
            "emr-containers:DescribeVirtualCluster",
            "emr-containers:ListVirtualClusters",
            "emr-containers:DescribeManagedEndpoint",
            "emr-containers:ListManagedEndpoints",
            "emr-containers:DescribeJobRun",
            "emr-containers:ListJobRuns"
         ],
         "Resource":"*",
         "Effect":"Allow"
      },
      {
          "Sid": "AllowRetrievingManagedEndpointCredentials",
          "Effect": "Allow",
          "Action": [
              "emr-containers:GetManagedEndpointSessionCredentials"
          ],
          "Resource": [
              "arn:aws:emr-containers:<region>:<account-id>:/virtualclusters/<virtual-
cluster-id>/endpoints/<managed-endpoint-id>"
          ],
          "Condition": {
              "StringEquals": {
                  "emr-containers:ExecutionRoleArn": [
                      "arn:aws:iam::<account-id>:role/<emr-on-eks-execution-role>"
                  ]
              }
          }
      },
      {
         "Sid":"AllowSecretManagerListSecrets",
```

```
            "Action":[
               "secretsmanager:ListSecrets"
            ],
            "Resource":"*",
            "Effect":"Allow"
         },
         {
            "Sid":"AllowSecretCreationWithEMRTagsAndEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:CreateSecret",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*",
            "Condition":{
               "StringEquals":{
                  "aws:RequestTag/for-use-with-amazon-emr-managed-policies":"true"
               }
            }
         },
         {
            "Sid":"AllowAddingTagsOnSecretsWithEMRStudioPrefix",
            "Effect":"Allow",
            "Action":"secretsmanager:TagResource",
            "Resource":"arn:aws:secretsmanager:*:*:secret:emr-studio-*"
         },
         {
            "Sid":"AllowClusterTemplateRelatedIntermediateActions",
            "Action":[
               "servicecatalog:DescribeProduct",
               "servicecatalog:DescribeProductView",
               "servicecatalog:DescribeProvisioningParameters",
               "servicecatalog:ProvisionProduct",
               "servicecatalog:SearchProducts",
               "servicecatalog:UpdateProvisionedProduct",
               "servicecatalog:ListProvisioningArtifacts",
               "servicecatalog:ListLaunchPaths",
               "servicecatalog:DescribeRecord",
               "cloudformation:DescribeStackResources"
            ],
            "Resource":"*",
            "Effect":"Allow"
         },
         {
            "Sid":"AllowEMRCreateClusterAdvancedActions",
            "Action":[
               "elasticmapreduce:RunJobFlow"
```

```
        ],
        "Resource":"*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowPassingServiceRoleForWorkspaceCreation",
        "Action":"iam:PassRole",
        "Resource":[
            "arn:aws:iam::*:role/<your-emr-studio-service-role>",
            "arn:aws:iam::*:role/EMR_DefaultRole_V2",
            "arn:aws:iam::*:role/EMR_EC2_DefaultRole"
        ],
        "Effect":"Allow"
    },
    {
        "Sid":"AllowS3ListAndLocationPermissions",
        "Action":[
            "s3:ListAllMyBuckets",
            "s3:ListBucket",
            "s3:GetBucketLocation"
        ],
        "Resource":"arn:aws:s3:::*",
        "Effect":"Allow"
    },
    {
        "Sid":"AllowS3ReadOnlyAccessToLogs",
        "Action":[
            "s3:GetObject"
        ],
        "Resource":[
            "arn:aws:s3:::aws-logs-<aws-account-id>-<region>/elasticmapreduce/*"
        ],
        "Effect":"Allow"
    },
    {
        "Sid":"AllowConfigurationForWorkspaceCollaboration",
        "Action":[
            "elasticmapreduce:UpdateEditor",
            "elasticmapreduce:PutWorkspaceAccess",
            "elasticmapreduce:DeleteWorkspaceAccess",
            "elasticmapreduce:ListWorkspaceAccessIdentities"
        ],
        "Resource":"*",
        "Effect":"Allow",
```

```
         "Condition":{
            "StringEquals":{
               "elasticmapreduce:ResourceTag/creatorUserId":"${aws:userId}"
            }
         }
      },
      {
         "Sid" : "SageMakerDataWranglerForEMRStudio",
         "Effect" : "Allow",
         "Action" : [
            "sagemaker:CreatePresignedDomainUrl",
            "sagemaker:DescribeDomain",
            "sagemaker:ListDomains",
            "sagemaker:ListUserProfiles"
         ],
         "Resource":"*"
      },
      {
         "Sid":"DescribeNetwork",
         "Effect":"Allow",
         "Action":[
            "ec2:DescribeVpcs",
            "ec2:DescribeSubnets",
            "ec2:DescribeSecurityGroups"
         ],
         "Resource":"*"
      },
      {
         "Sid":"ListIAMRoles",
         "Effect":"Allow",
         "Action":[
            "iam:ListRoles"
         ],
         "Resource":"*"
      },
      {
         "Sid": "AllowServerlessActions",
         "Action": [
            "emr-serverless:CreateApplication",
            "emr-serverless:UpdateApplication",
            "emr-serverless:DeleteApplication",
            "emr-serverless:ListApplications",
            "emr-serverless:GetApplication",
            "emr-serverless:StartApplication",
```

```
            "emr-serverless:StopApplication",
            "emr-serverless:StartJobRun",
            "emr-serverless:CancelJobRun",
            "emr-serverless:ListJobRuns",
            "emr-serverless:GetJobRun",
            "emr-serverless:GetDashboardForJobRun",
            "emr-serverless:AccessInteractiveEndpoints"
        ],
        "Resource": "*",
        "Effect": "Allow"
    },
    {
        "Sid": "AllowPassingRuntimeRoleForRunningServerlessJob",
        "Action": "iam:PassRole",
        "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
        "Effect": "Allow"
    }
  ]
}
```

La siguiente política de usuario contiene los permisos de usuario básicos que se necesitan para utilizar una aplicación interactiva de EMR sin servidor con espacios de trabajo de EMR Studio.

Política interactiva de EMR sin servidor

En este ejemplo de política que tiene permisos de usuario para las aplicaciones interactivas de EMR sin servidor con EMR Studio, sustituya los marcadores de posición de *serverless-runtime-role* y *emr-studio-service-role* por su rol de servicio de EMR Studio y rol de tiempo de ejecución de EMR sin servidor correctos.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowServerlessActions",
            "Action": [
                "emr-serverless:CreateApplication",
                "emr-serverless:UpdateApplication",
                "emr-serverless:DeleteApplication",
                "emr-serverless:ListApplications",
                "emr-serverless:GetApplication",
                "emr-serverless:StartApplication",
                "emr-serverless:StopApplication",
```

```
                "emr-serverless:StartJobRun",
                "emr-serverless:CancelJobRun",
                "emr-serverless:ListJobRuns",
                "emr-serverless:GetJobRun",
                "emr-serverless:GetDashboardForJobRun",
                "emr-serverless:AccessInteractiveEndpoints"
            ],
            "Resource": "*",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowEMRBasicActions",
            "Action": [
                "elasticmapreduce:CreateEditor",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:UpdateStudio",
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:OpenEditorInConsole",
                "elasticmapreduce:AttachEditor",
                "elasticmapreduce:DetachEditor",
                "elasticmapreduce:CreateStudio",
                "elasticmapreduce:DescribeStudio",
                "elasticmapreduce:DeleteStudio",
                "elasticmapreduce:ListStudios",
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "*",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowPassingRuntimeRoleForRunningEMRServerlessJob",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/serverless-runtime-role",
            "Effect": "Allow"
        },
        {
            "Sid": "AllowPassingServiceRoleForWorkspaceCreation",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/emr-studio-service-role",
            "Effect": "Allow"
        },
```

```
        {
            "Sid": "AllowS3ListAndGetPermissions",
            "Action": [
                "s3:ListAllMyBuckets",
                "s3:ListBucket",
                "s3:GetBucketLocation",
                "s3:GetObject"
            ],
            "Resource": "arn:aws:s3:::*",
            "Effect": "Allow"
        },
        {
            "Sid":"DescribeNetwork",
            "Effect":"Allow",
            "Action":[
                "ec2:DescribeVpcs",
                "ec2:DescribeSubnets",
                "ec2:DescribeSecurityGroups"
            ],
            "Resource":"*"
        },
        {
            "Sid": "ListIAMRoles",
            "Effect": "Allow",
            "Action": [
                "iam:ListRoles"
            ],
            "Resource": "*"
        }
    ]
}
```

Permisos de AWS Identity and Access Management para los usuarios de EMR Studio

En la tabla siguiente se indican todas las operaciones de Amazon EMR Studio que puede realizar un usuario y se enumeran las acciones de IAM mínimas necesarias para realizar esa operación. Puede permitir estas acciones en sus políticas de permisos de IAM (al usar la autenticación de IAM) o en sus políticas de sesión (al usar la autenticación de IAM Identity Center) para EMR Studio.

En la tabla también se muestran las operaciones permitidas en cada una de las políticas de permisos de ejemplo para EMR Studio. Para obtener más información acerca de las políticas de permisos de ejemplo, consulte Creación de políticas de permisos para los usuarios de EMR Studio.

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
Crear y eliminar espacios de trabajo	Sí	Sí	Sí	`"elasticmapreduce:CreateEditor", "elasticmapreduce:DescribeEditor", "elasticmapreduce:ListEditors", "elasticmapreduce:DeleteEditor"`
Consulte el panel Colaboración, habilite la colaboración en el espacio de trabajo y agregue a colaboradores. Para obtener más información, consulte Establecer la propiedad de la colaboración en el espacio de trabajo.	Sí	Sí	Sí	`"elasticmapreduce:UpdateEditor", "elasticmapreduce:PutWorkspaceAccess", "elasticmapreduce:DeleteWorkspaceAccess", "elasticmapreduce:ListWorkspaceAccessIdentities"`
Consulte una lista de buckets de almacenamiento de Amazon S3 Control en la misma cuenta que Studio al crear un nuevo clúster de EMR y acceda a los registros del contenedor al usar una interfaz de usuario web para depurar aplicaciones	Sí	Sí	Sí	`"s3:ListAllMyBuckets", "s3:ListBucket", "s3:GetBucketLocation", "s3:GetObject"`
Acceder a los espacios de trabajo	Sí	Sí	Sí	`"elasticmapreduce:DescribeEditor",`

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
				`"elasticmapreduce:ListEditors", "elasticmapreduce:StartEditor", "elasticmapreduce:StopEditor", "elasticmapreduce:OpenEditorInConsole"`
Asociar o separar los clústeres de Amazon EMR existentes asociados al espacio de trabajo	Sí	Sí	Sí	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster", "elasticmapreduce:ListInstanceGroups", "elasticmapreduce:ListBootstrapActions"`
Asociar o separar clústeres de Amazon EMR en EKS	Sí	Sí	Sí	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListManagedEndpoints", "emr-containers:DescribeManagedEndpoint", "emr-containers:GetManagedEndpointSessionCredentials"`

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
Adjuntar o desconectar las aplicaciones de EMR sin servidor asociadas al espacio de trabajo	No	Sí	Sí	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-serverless:GetApplication", "emr-serverless:StartApplication", "emr-serverless:ListApplications", "emr-serverless:GetDashboardForJobRun", "emr-serverless:AccessInteractiveEndpoints", "iam:PassRole"` El permiso `PassRole` es necesario para pasar el rol de tiempo de ejecución del trabajo de EMR sin servidor. Para obtener más información, consulte Roles de tiempo de ejecución de trabajo en la Guía del usuario de Amazon EMR sin servidor.

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
Depurar Amazon EMR en trabajos de EC2 con interfaces de usuario de aplicaciones persistentes	Sí	Sí	Sí	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "elasticmapreduce:ListClusters", "elasticmapreduce:ListSteps", "elasticmapreduce:DescribeCluster", "s3:ListBucket", "s3:GetObject"`
Depurar Amazon EMR en trabajos de EC2 con interfaces de usuario de aplicaciones en clústeres	Sí	Sí	Sí	`"elasticmapreduce:GetOnClusterAppUIPresignedURL"`

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
Depurar Amazon EMR en ejecuciones de trabajos de EKS mediante el servidor de historial de Spark	Sí	Sí	Sí	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListJobRuns", "emr-containers:DescribeJobRun", "s3:ListBucket", "s3:GetObject"`
Crear y eliminar repositorios de Git	Sí	Sí	Sí	`"elasticmapreduce:CreateRepository", "elasticmapreduce:DeleteRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository", "secretsmanager:CreateSecret", "secretsmanager:ListSecrets", "secretsmanager:TagResource"`

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
Vincular y desvincular repositorios de Git	Sí	Sí	Sí	`"elasticmapreduce:LinkRepository", "elasticmapreduce:UnlinkRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository"`
Crear nuevos clústeres a partir de plantillas de clústeres predefinidas	No	Sí	Sí	`"servicecatalog:SearchProducts", "servicecatalog:DescribeProduct", "servicecatalog:DescribeProductView", "servicecatalog:DescribeProvisioningParameters", "servicecatalog:ProvisionProduct", "servicecatalog:UpdateProvisionedProduct", "servicecatalog:ListProvisioningArtifacts", "servicecatalog:DescribeRecord", "servicecatalog:ListLaunchPaths", "cloudformation:DescribeStackResources", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
Crear nuevos clústeres proporcionando una configuración de clúster	No	No	Sí	`"elasticmapreduce:RunJobFlow", "iam:PassRole", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`
Asignar un usuario a un estudio al usar el modo de autenticación de IAM. Para obtener más información, consulte Asignar un usuario o grupo a un EMR Studio.	No	No	No	`"elasticmapreduce:CreateStudioPresignedUrl"`

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas

Describir los objetos de la red.	Sí	Sí	Sí	<pre>{
				    "Version": "2012-10-17",
				    "Statement": [
				        {
				            "Sid": "DescribeNetwork",
				            "Effect": "Allow",
				            "Action": [
				                "ec2:DescribeVpcs",
				                "ec2:DescribeSubnets",
				                "ec2:DescribeSecurityGroups"
				            ],
				            "Resource": "*"
				        }
				    ]
				}</pre>

Acción	Básica	Intermedia	Advanced (Avanzado)	Acciones asociadas
Enumerar los roles de IAM.	Sí	Sí	Sí	`{ "Version": "2012-10-17", "Statement": [ { "Sid": "ListIAMRoles", "Effect": "Allow", "Action": [ "iam:ListRoles" ], "Resource": "*" } ] }`
Conéctese a EMR Studio desde Amazon SageMaker Studio y utilice la interfaz visual de Data Wrangler.	No	No	Sí	`"sagemaker:CreatePresignedDomainUrl", "sagemaker:DescribeDomain", "sagemaker:ListDomains", "sagemaker:ListUserProfiles"`

Crear un EMR Studio

Puede crear un EMR Studio para su equipo mediante la consola de Amazon EMR o la AWS CLI. La creación de una instancia de Studio forma parte de la configuración de Amazon EMR Studio.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

Requisitos previos

Antes de crear un estudio, asegúrese de haber completado las tareas anteriores de Configuración de un Amazon EMR Studio.

Para crear un estudio con la AWS CLI, debe tener instalada la última versión. Para obtener más información, consulte Instalación o actualización de la última versión del AWS CLI.

> **Important**
>
> Desactive las herramientas de administración de proxy como FoxyProxy o SwitchyOmega en el navegador antes de crear un estudio. Los proxies activos pueden provocar un mensaje de error de red al seleccionar Crear estudio.

New console

Para crear un EMR Studio con la nueva consola

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/emr.
2. En EMR Studio, en el panel de navegación de la izquierda, elija Comenzar. También puede crear un estudio nuevo desde la página Studios.
3. Seleccione Crear un estudio para abrir la página Crear un estudio.
4. Ingrese un Nombre de Studio y, opcionalmente, una Descripción.
5. Si utiliza la autenticación de IAM para el estudio, puede elegir Agregar nueva etiqueta para agregar una o más etiquetas clave-valor de su elección al estudio. Las etiquetas se utilizan para dar acceso al estudio a determinados usuarios. Para obtener más información, consulte Asignar un usuario o grupo a un EMR Studio.

También puede agregar etiquetas para ayudarle a administrar, identificar, organizar y filtrar los estudios. Para obtener más información, consulte Tagging AWS resources (Etiquetado de los recursos de).

6. En Redes, seleccione una Amazon Virtual Private Cloud (VPC) para el estudio en la lista desplegable.
7. En Subredes, seleccione un máximo de cinco subredes en tu VPC para asociarlas al estudio. Tiene la opción de agregar más subredes después de crear el estudio.
8. En Grupos de seguridad, seleccione los grupos de seguridad predeterminados o grupos de seguridad personalizados. Para obtener más información, consulte Definir grupos de seguridad para controlar el tráfico de red de EMR Studio.

Si selecciona...	Haga lo siguiente...
Los grupos de seguridad predeterminados de EMR Studio	Para habilitar la vinculación de repositorios basada en Git para el estudio, seleccion e Habilitar clústeres o puntos de enlace y repositorio Git. De lo contrario, seleccione Habilitar clústeres/puntos de conexión.
Grupos de seguridad personalizados para su estudio	• En Grupo de seguridad de los clústeres/ puntos de conexión, seleccione el grupo de seguridad del motor que configuró en la lista desplegable. Su estudio usa este grupo de seguridad para permitir el acceso entrante desde los espacios de trabajo asociados. • En Grupo de seguridad del espacio de trabajo, seleccione el grupo de seguridad del espacio de trabajo que configuró en la lista desplegable. Su estudio usa este grupo de seguridad con los espacios de trabajo para proporcionar acceso saliente a los clústeres de Amazon EMR asociados y a los repositorios de Git alojados públicamente.

9. En Autenticación, seleccione un modo de autenticación para el estudio y proporcione la información según la siguiente tabla. Para obtener más información sobre la autenticación en EMR Studio, consulte Elija un modo de autenticación para Amazon EMR Studio.

Si usa…	Haga lo siguiente...
Autenticación o federación de IAM	Seleccione un método de inicio de sesión para el estudio. Si desea que los usuarios federados inicien sesión con la URL de Studio y las credenciales del proveedor de identidad es (IdP), seleccione su IdP en la lista desplegable e introduzca la URL de inicio de sesión de su proveedor de identidades (IdP) y el nombre del parámetro RelayStat e. Para obtener una lista de las direcciones URL de autenticación del IdP y los nombres de RelayState, consulte Parámetros de RelayState y URL de autenticación del proveedor de identidades. A continuación, seleccione su rol de servicio de EMR Studio en la lista desplegable. Para obtener más informaci ón, consulte Crear un rol de servicio de EMR Studio.
Autenticación del Centro de identidades de IAM	Seleccione su rol de servicio y rol de usuario de EMR Studio. Para obtener más información, consulte Crear un rol de servicio de EMR Studio y Creación de un rol de usuario de EMR Studio para el modo de autenticación de IAM Identity Center.

10. En Almacenamiento del espacio de trabajo, seleccione Explorar S3 para seleccionar su bucket de Amazon S3 para hacer copias de seguridad de los espacios de trabajo y los archivos de cuadernos.

    > **Note**
    >
    > Su rol de servicio de EMR Studio debe tener acceso de lectura y escritura al bucket que seleccione.

11. Seleccione Crear un estudio para terminar y vaya a la página Studios. El nuevo estudio aparecerá en la lista con detalles como el nombre del estudio, la fecha de creación y la URL de acceso al estudio.

Después de crear el estudio, siga las instrucciones que aparecen en Asignar un usuario o grupo a un EMR Studio.

Old console

Para crear un EMR Studio con la antigua consola

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/home.
2. Seleccione EMR Studio en el menú de navegación de la izquierda.
3. Seleccione Crear un estudio para abrir la página Crear un estudio.
4. Ingrese un Nombre de Studio y, opcionalmente, una Descripción.
5. Si utiliza la autenticación de IAM para el estudio, puede elegir Agregar nueva etiqueta para agregar una o más etiquetas clave-valor de su elección al estudio. Las etiquetas se utilizan para dar acceso al estudio a determinados usuarios. Para obtener más información, consulte Asignar un usuario o grupo a un EMR Studio.

   También puede agregar etiquetas para ayudarle a administrar, identificar, organizar y filtrar los estudios. Para obtener más información, consulte Tagging AWS resources (Etiquetado de los recursos de).
6. En Redes, seleccione una Amazon Virtual Private Cloud (VPC) para el estudio en la lista desplegable.
7. En Subredes, seleccione un máximo de cinco subredes en tu VPC para asociarlas al estudio. Tiene la opción de agregar más subredes después de crear el estudio.

8. En Grupos de seguridad, seleccione los grupos de seguridad predeterminados o grupos de seguridad personalizados. Para obtener más información, consulte Definir grupos de seguridad para controlar el tráfico de red de EMR Studio.

Si selecciona...	Haga lo siguiente...
Los grupos de seguridad predeterminados de EMR Studio	Para habilitar la vinculación de repositorios basada en Git para el estudio, seleccion e Habilitar clústeres o puntos de enlace y repositorio Git. De lo contrario, seleccione Habilitar clústeres/puntos de conexión.
Grupos de seguridad personalizados para su estudio	• En Grupo de seguridad de los clústeres/ puntos de conexión, seleccione el grupo de seguridad del motor que configuró en la lista desplegable. Su estudio usa este grupo de seguridad para permitir el acceso entrante desde los espacios de trabajo asociados. • En Grupo de seguridad del espacio de trabajo, seleccione el grupo de seguridad del espacio de trabajo que configuró en la lista desplegable. Su estudio usa este grupo de seguridad con los espacios de trabajo para proporcionar acceso saliente a los clústeres de Amazon EMR asociados y a los repositorios de Git alojados públicamente.

9. En Autenticación, seleccione un modo de autenticación para el estudio y proporcione la información según la siguiente tabla. Para obtener más información sobre la autenticación en EMR Studio, consulte Elija un modo de autenticación para Amazon EMR Studio.

Si usa…	Haga lo siguiente...
Autenticación o federación de IAM	Seleccione un método de inicio de sesión para el estudio.

Si usa…	Haga lo siguiente...
	Si desea que los usuarios federados inicien sesión con la URL de Studio y las credenciales del proveedor de identidad es (IdP), seleccione su IdP en la lista desplegable e introduzca la URL de inicio de sesión de su proveedor de identidades (IdP) y el nombre del parámetro RelayStat e. Para obtener una lista de las direcciones URL de autenticación del IdP y los nombres de RelayState, consulte Parámetros de RelayState y URL de autenticación del proveedor de identidades. A continuación, seleccione su rol de servicio de EMR Studio en la lista desplegable. Para obtener más informaci ón, consulte Crear un rol de servicio de EMR Studio.
Autenticación del Centro de identidades de IAM	Seleccione su rol de servicio y rol de usuario de EMR Studio. Para obtener más información, consulte Crear un rol de servicio de EMR Studio y Creación de un rol de usuario de EMR Studio para el modo de autenticación de IAM Identity Center.

10. En Almacenamiento del espacio de trabajo, seleccione Explorar S3 para seleccionar su bucket de Amazon S3 para hacer copias de seguridad de los espacios de trabajo y los archivos de cuadernos.

> **Note**
>
> Su rol de servicio de EMR Studio debe tener acceso de lectura y escritura al bucket que seleccione.

11. Seleccione Crear un estudio para terminar y vaya a la página Studios. El nuevo estudio aparecerá en la lista con detalles como el nombre del estudio, la fecha de creación y la URL de acceso al estudio.

Después de crear el estudio, siga las instrucciones que aparecen en Asignar un usuario o grupo a un EMR Studio.

CLI

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

Example Comando de la CLI para crear un EMR Studio con el modo de autenticación de IAM

El siguiente comando de ejemplo de la AWS CLI crea un EMR Studio con el modo de autenticación de IAM. Cuando utiliza la autenticación o federación de IAM del estudio, no tiene que especificar un `--user-role`.

Para permitir que los usuarios federados inicien sesión con la URL de Studio y las credenciales de su proveedor de identidades (IdP), especifique su `--idp-auth-url` y `--idp-relay-state-parameter-name`. Para obtener una lista de las direcciones URL de autenticación del IdP y los nombres de RelayState, consulte Parámetros de RelayState y URL de autenticación del proveedor de identidades.

```
aws emr create-studio \
--name <example-studio-name> \
--auth-mode IAM \
--vpc-id <example-vpc-id> \
--subnet-ids <subnet-id-1> <subnet-id-2>... <subnet-id-5>  \
```

```
--service-role <example-studio-service-role-name> \
--workspace-security-group-id <example-workspace-sg-id> \
--engine-security-group-id <example-engine-sg-id> \
--default-s3-location <example-s3-location> \
--idp-auth-url <https://EXAMPLE/login/> \
--idp-relay-state-parameter-name <example-RelayState>
```

Example Comando de la CLI para crear un EMR Studio con el modo de autenticación de IAM Identity Center

El siguiente comando de ejemplo de la AWS CLI crea un EMR Studio que usa el modo de autenticación de IAM Identity Center. Al utilizar la autenticación de IAM Identity Center, debe especificar un `--user-role`.

Para obtener más información acerca del modo de autenticación de IAM Identity Center, consulte Configuración del modo de autenticación de IAM Identity Center para Amazon EMR Studio.

```
aws emr create-studio \
--name <example-studio-name> \
--auth-mode SSO \
--vpc-id <example-vpc-id> \
--subnet-ids <subnet-id-1> <subnet-id-2>... <subnet-id-5>  \
--service-role <example-studio-service-role-name> \
--user-role <example-studio-user-role-name> \
--workspace-security-group-id <example-workspace-sg-id> \
--engine-security-group-id <example-engine-sg-id> \
--default-s3-location <example-s3-location>
```

Example Salida de la CLI para **`aws emr create-studio`**

A continuación, se muestra un ejemplo de la salida que aparece después de crear un estudio.

```
{
    StudioId: "es-123XXXXXXXXX",
    Url: "https://es-123XXXXXXXXX.emrstudio-prod.us-east-1.amazonaws.com"
}
```

Para obtener más información sobre el comando `create-studio`, consulte Referencia del comando AWS CLI.

Parámetros de RelayState y URL de autenticación del proveedor de identidades

Si utiliza la federación de IAM y quiere que los usuarios inicien sesión con la URL de Studio y las credenciales de su proveedor de identidades (IdP), puede especificar la URL de inicio de sesión de su proveedor de identidades (IdP) y el nombre del parámetro RelayState cuando Crear un EMR Studio.

En la siguiente tabla se muestra la URL de autenticación estándar y el nombre del parámetro RelayState de algunos proveedores de identidades populares.

Proveedor de identidades	Parámetro	URL de autenticación
Auth0	`RelayState`	`https://<sub_domain> .auth0.com/samlp/<app_id>`
Cuentas de Google	`RelayState`	`https://accounts.google.com/o/saml2/initsso?idpid= <idp_id>&spid=<sp_id>&forceauthn=false`
Microsoft Azure	`RelayState`	`https://myapps.microsoft.com/signin/ <app_name> /<app_id>?tenantId= <tenant_id>`
Okta	`RelayState`	`https://<sub_domain> .okta.com/app/<app_name> /<app_id>/sso/saml`
PingFederate	`TargetResource`	`https://<host>/idp/<idp_id>/startSSO.ping?PartnerSpId= <sp_id>`
PingOne	`TargetResource`	`https://sso.connect.pingidentity.com/sso/sp/initsso?saasid= <app_id>&idpid=<idp_id>`

Asignar y administrar usuarios de EMR Studio

Después de crear un EMR Studio, puede asignarle usuarios y grupos. El método que utilice para asignar, actualizar y eliminar usuarios depende del modo de autenticación de Studio.

- Cuando utiliza el modo de autenticación de IAM, debe configurar la asignación de usuarios y los permisos de EMR Studio en IAM o con IAM y su proveedor de identidades.
- Con el modo de autenticación de IAM Identity Center, debe utilizar la consola de administración de Amazon EMR o la AWS CLI para administrar a los usuarios.

Para obtener más información sobre la autenticación en Amazon EMR Studio, consulte Elija un modo de autenticación para Amazon EMR Studio.

Asignar un usuario o grupo a un EMR Studio

IAM

Cuando utilice Configuración del modo de autenticación de IAM para Amazon EMR Studio, debe permitir la acción `CreateStudioPresignedUrl` en la política de permisos de IAM de un usuario y restringirlo a un estudio concreto. Puede incluir `CreateStudioPresignedUrl` en su Permisos de usuario para el modo de autenticación de IAM o utilizar una política independiente.

Para restringir a un usuario a un estudio (o conjunto de estudios), puede usar el control de acceso basado en atributos (ABAC) o especificar el nombre de recurso de Amazon (ARN) de un estudio en el elemento `Resource` de la política de permisos.

Example Asignar un usuario a un estudio mediante el ARN del estudio

El siguiente ejemplo de política proporciona a un usuario acceso a un EMR Studio concreto al permitir la acción `CreateStudioPresignedUrl` y especificar el nombre de recurso de Amazon (ARN) del estudio en el elemento `Resource`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-
id>:studio/<studio-id>"
        }
    ]
```

```
}
```

Example Asignar un usuario a un estudio con ABAC para la autenticación de IAM

Hay varias formas de configurar el control de acceso basado en atributos (ABAC) en un estudio. Por ejemplo, puede asociar una o más etiquetas a un EMR Studio y, a continuación, crear una política de IAM que restrinja la acción `CreateStudioPresignedUrl` a un estudio concreto o a un conjunto de estudios con esas etiquetas.

Puede agregar etiquetas durante o después de la creación del estudio. Para agregar etiquetas a un estudio existente, puede utilizar el comando AWS CLI `emr add-tags`. El siguiente ejemplo agrega una etiqueta con el par clave-valor `Team = Data Analytics` a un EMR Studio.

```
aws emr add-tags --resource-id <example-studio-id> --tags Team="Data Analytics"
```

El siguiente ejemplo de política de permisos permite la acción `CreateStudioPresignedUrl` para los estudios de EMR Studio con el par clave-valor de etiqueta `Team = DataAnalytics`. Para obtener más información sobre el uso de etiquetas para el control de acceso, consulte Controlar el acceso a y para los usuarios y roles utilizando etiquetas o Controlar el acceso a los recursos de AWS utilizando etiquetas.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-id>:studio/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/Team": "Data Analytics"
                }
            }
        }
    ]
}
```

Example Asignar un usuario a un estudio mediante la clave de condición global aws:SourceIdentity

Al utilizar la federación de IAM, puede utilizar la clave de condición global `aws:SourceIdentity` en una política de permisos para que los usuarios accedan a Studio cuando asuman su rol de IAM para la federación.

Primero debe configurar su proveedor de identidades (IdP) para que devuelva una cadena de identificación, como una dirección de correo electrónico o un nombre de usuario, cuando un usuario se autentique y asuma su rol de IAM para la federación. IAM establece la clave de condición global `aws:SourceIdentity` en la cadena de identificación devuelta por el IdP.

Para obtener más información, consulte la entrada del blog Cómo relacionar la actividad de los roles de IAM con la identidad corporativa en el blog de seguridad de AWS y la entrada aws:SourceIdentity en la referencia de claves de condición globales.

El siguiente ejemplo de política permite la acción `CreateStudioPresignedUrl` y proporciona a los usuarios una `aws:SourceIdentity` que coincide con el acceso al EMR Studio de *<example-source-identity>* especificado por *<example-studio-arn>*.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "elasticmapreduce:CreateStudioPresignedUrl",
      "Resource": "<example-studio-arn>",
      "Condition": {
        "StringLike": {
          "aws:SourceIdentity": "<example-source-identity>"
        }
      }
    }
  ]
}
```

IAM Identity Center

Al asignar un usuario o un grupo a un EMR Studio, debe especificar una política de sesión que define permisos detallados, como la capacidad de crear un nuevo clúster de EMR, para ese

usuario o grupo. Amazon EMR almacena estas asignaciones de políticas de sesión. Puede actualizar la política de sesión de un usuario o grupo después de la asignación.

> **Note**
>
> Los permisos finales de un usuario o grupo son una intersección de los permisos definidos en su rol de usuario de EMR Studio y los permisos definidos en la política de sesión para ese usuario o grupo. Si un usuario pertenece a más de un grupo asignado al estudio, EMR Studio utiliza una unión de permisos para ese usuario.

Para asignar usuarios o grupos a un EMR Studio mediante la consola de Amazon EMR

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione EMR Studio en el menú de navegación de la izquierda.
3. Elija el nombre de su estudio en la lista Studios o seleccione el estudio y elija Ver detalles para abrir la página de detalles del estudio.
4. Elija la pestaña Agregar usuarios para ver la tabla de búsqueda Usuarios y Grupos.
5. Seleccione la pestaña Usuarios o la pestaña Grupos e introduzca un término de búsqueda en la barra de búsqueda para buscar un usuario o grupo.
6. Seleccione uno o más usuarios o grupos de la lista de resultados de la búsqueda. Puede cambiar de una pestaña a otra de Usuarios y Grupos.
7. Tras seleccionar los usuarios y grupos para agregarlos al estudio, seleccione Agregar. Debería ver los usuarios y grupos en la lista Usuarios del estudio. La lista puede tardar unos segundos en actualizarse.
8. Siga las instrucciones que se indican en Actualizar los permisos de un usuario o grupo asignado a un estudio para ajustar los permisos de Studio de un usuario o un grupo.

Para asignar un usuario o grupo a un EMR Studio con la AWS CLI

Inserte sus propios valores para los siguientes argumentos de `create-studio-session-mapping`. Para obtener más información sobre el comando `create-studio-session-mapping`, consulte la Referencia de comandos de la AWS CLI.

- **`--studio-id`**: el ID del estudio al que quiere asignar el usuario o grupo. Para obtener instrucciones sobre cómo recuperar un ID de Studio, consulte Ver detalles del estudio.
- **`--identity-name`**: el nombre del usuario o grupo del Almacén de identidades. Para obtener más información, consulte UserName para los usuarios y DisplayName para los grupos en la Referencia de la API del Almacén de identidades.
- **`--identity-type`**: utilice USER o GROUP para especificar el tipo de identidad.
- **`--session-policy-arn`**: el Nombre de recurso de Amazon (ARN) de la política de sesión que desee asociar al usuario o grupo. Por ejemplo, **`arn:aws:iam::`*`<aws-account-id>`*`:policy/`*`EMRStudio_Advanced_User_Policy`**. Para obtener más información, consulte Creación de políticas de permisos para los usuarios de EMR Studio.

```
aws emr create-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-name <example-identity-name> \
 --identity-type <USER-or-GROUP> \
 --session-policy-arn <example-session-policy-arn>
```

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

Utilice el comando `get-studio-session-mapping` para comprobar la nueva asignación. Sustituya *<example-identity-name>* por el nombre de IAM Identity Center del usuario o grupo que actualizó.

```
aws emr get-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <user-or-group-name> \
```

Actualizar los permisos de un usuario o grupo asignado a un estudio

IAM

Para actualizar los permisos de un usuario o grupo al utilizar el modo de autenticación de IAM, debe utilizar IAM para cambiar las políticas de permisos de IAM asociadas a sus identidades de IAM (usuarios, grupos o roles).

Para obtener más información, consulte Permisos de usuario para el modo de autenticación de IAM.

IAM Identity Center

Para actualizar los permisos de EMR Studio de un usuario o grupo mediante la consola

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione EMR Studio en el menú de navegación de la izquierda.
3. Elija el nombre de su estudio en la lista Studios o seleccione el estudio y elija Ver detalles para abrir la página de detalles del estudio.
4. En la lista Usuarios del estudio de la página de detalles del estudio, busque el usuario o grupo que desee actualizar. Puede buscar por nombre o tipo de identidad.
5. Seleccione el usuario o grupo que desee actualizar y seleccione Asignar política para abrir el cuadro de diálogo Política de sesión.
6. Seleccione una política para aplicarla al usuario o grupo que eligió en el paso 5 y seleccione Aplicar política. La lista Usuarios del estudio mostrará el nombre de la política en la columna Política de sesión del usuario o grupo que haya actualizado.

Para actualizar los permisos de EMR Studio de un usuario o grupo mediante la AWS CLI

Inserte sus propios valores para los siguientes argumentos de `update-studio-session-mappings`. Para obtener más información sobre el comando `update-studio-session-mappings`, consulte la Referencia de comandos de la AWS CLI.

```
aws emr update-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-name <name-of-user-or-group-to-update> \
 --session-policy-arn <new-session-policy-arn-to-apply> \
```

```
 --identity-type <USER-or-GROUP> \
```

Utilice el comando `get-studio-session-mapping` para comprobar la nueva asignación de la política de sesión. Sustituya *<example-identity-name>* por el nombre de IAM Identity Center del usuario o grupo que actualizó.

```
aws emr get-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <user-or-group-name> \
```

Eliminar un usuario o grupo de un estudio

IAM

Para eliminar un usuario o un grupo de un EMR Studio al utilizar el modo de autenticación de IAM, debe revocar el acceso del usuario al estudio reconfigurando la política de permisos de IAM del usuario.

En el siguiente ejemplo de política, suponga que tiene un EMR Studio con el par clave-valor de etiqueta `Team = Quality Assurance`. Según la política, el usuario puede acceder a los estudios etiquetados con la clave `Team` cuyo valor sea igual a `Data Analytics` o `Quality Assurance`. Para eliminar al usuario del estudio etiquetado con `Team = Quality Assurance`, elimine `Quality Assurance` de la lista de valores de etiqueta.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowCreateStudioPresignedUrl",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:CreateStudioPresignedUrl"
            ],
            "Resource": "arn:aws:elasticmapreduce:<region>:<account-id>:studio/*",
            "Condition": {
                "StringEquals": {
                    "emr:ResourceTag/Team": [
                        "Data Analytics",
                        "Quality Assurance"
```

```
                    ]
                }
            }
        }
    ]
}
```

IAM Identity Center

Para eliminar un usuario o un grupo de un EMR Studio mediante la consola

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione EMR Studio en el menú de navegación de la izquierda.
3. Elija el nombre de su estudio en la lista Studios o seleccione el estudio y elija Ver detalles para abrir la página de detalles del estudio.
4. En la lista Usuarios del estudio, en la página de detalles del estudio, busque el usuario o grupo que desee eliminar del estudio. Puede buscar por nombre o tipo de identidad.
5. Seleccione el usuario o grupo que desea eliminar, seleccione Eliminar y confirme la eliminación. El usuario o grupo que ha eliminado desaparece de la lista Usuarios del estudio.

Para eliminar un usuario o un grupo de un EMR Studio mediante la AWS CLI

Inserte sus propios valores para los siguientes argumentos de `delete-studio-session-mapping`. Para obtener más información sobre el comando `delete-studio-session-mapping`, consulte la Referencia de comandos de la AWS CLI.

```
aws emr delete-studio-session-mapping \
 --studio-id <example-studio-id> \
 --identity-type <USER-or-GROUP> \
 --identity-name <name-of-user-or-group-to-delete> \
```

Administrar un Amazon EMR Studio

En esta sección se incluyen instrucciones que le ayudarán a monitorear, actualizar o eliminar un recurso de EMR Studio. Para obtener información sobre la asignación de usuarios o la actualización de los permisos de los usuarios, consulte Asignar y administrar usuarios de EMR Studio.

Ver detalles del estudio

New console

Para ver detalles sobre un EMR Studio con la nueva consola

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/emr.
2. En la barra de navegación de la izquierda, en EMR Studio, seleccione Studios.
3. Seleccione el estudio en la lista Studios para abrir la página de detalles del estudio. En la página de detalles del estudio se incluye información sobre los ajustes del estudio, como la descripción, la VPC y las subredes del estudio.

Old console

Para ver detalles sobre un EMR Studio con la antigua consola

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/home.
2. Seleccione EMR Studio en el menú de navegación de la izquierda.
3. Seleccione el estudio en la lista Studios para abrir la página de detalles del estudio. En la página de detalles del estudio se incluye información sobre los ajustes del estudio, como la descripción, la VPC y las subredes del estudio.

CLI

Para recuperar los detalles de un EMR Studio por ID de estudio mediante la AWS CLI

Utilice el siguiente comando `describe-studio` de la AWS CLI para obtener información detallada sobre un EMR Studio en particular. Para obtener más información, consulte Referencia de comandos de la AWS CLI.

```
aws emr describe-studio \
 --studio-id <id-of-studio-to-describe> \
```

Para recuperar una lista de los estudios de EMR Studio mediante la AWS CLI

Use el siguiente comando `list-studios` de la AWS CLI. Para obtener más información, consulte Referencia de comandos de la AWS CLI.

```
aws emr list-studios
```

A continuación, se muestra un ejemplo del valor devuelto para el comando `list-studios` en formato JSON.

```
{
    "Studios": [
        {
            "AuthMode": "IAM",
            "VpcId": "vpc-b21XXXXX",
            "Name": "example-studio-name",
            "Url": "https://es-7HWP74SNGDXXXXXXXXXXXXXX.emrstudio-prod.us-
east-1.amazonaws.com",
            "CreationTime": 1605672582.781,
            "StudioId": "es-7HWP74SNGDXXXXXXXXXXXXXX",
            "Description": "example studio description"
        }
    ]
}
```

Monitorear las acciones de Amazon EMR Studio

Ver la actividad de EMR Studio y la API

EMR Studio se integra con AWS CloudTrail, un servicio que proporciona un registro de las acciones que realiza un usuario, un rol de IAM u otro servicio de AWS en EMR Studio. CloudTrail captura las llamadas a la API de EMR Studio como eventos. Puede ver los eventos con la consola de CloudTrail en https://console.aws.amazon.com/cloudtrail/.

Los eventos de EMR Studio proporcionan información como qué estudio o usuario de IAM realiza una solicitud y qué tipo de solicitud es.

> **Note**
>
> Las acciones realizadas en el clúster, como la ejecución de tareas de cuaderno, no emiten AWS CloudTrail.

También puede crear un registro para la entrega continua de eventos de CloudTrail en EMR Studio a un bucket de Amazon S3. Para obtener más información, consulte la Guía del usuario de AWS CloudTrail.

Ejemplo de evento de CloudTrail: un usuario llama a la API DescribeStudio

El siguiente es un ejemplo de evento de AWS CloudTrail que se crea cuando un usuario, `admin`, llama a la API DescribeStudio. CloudTrail registra el nombre de usuario como `admin`.

> **Note**
>
> Para proteger los detalles del estudio, el evento de la API de EMR Studio para DescribeStudio excluye un valor para `responseElements`.


```
{
   "eventVersion":"1.08",
   "userIdentity":{
      "type":"IAMUser",
      "principalId":"AIDXXXXXXXXXXXXXXXXX",
      "arn":"arn:aws:iam::653XXXXXXXXX:user/admin",
      "accountId":"653XXXXXXXXX",
      "accessKeyId":"AKIAIOSFODNN7EXAMPLE",
      "userName":"admin"
   },
   "eventTime":"2021-01-07T19:13:58Z",
   "eventSource":"elasticmapreduce.amazonaws.com",
   "eventName":"DescribeStudio",
   "awsRegion":"us-east-1",
   "sourceIPAddress":"72.XX.XXX.XX",
   "userAgent":"aws-cli/1.18.188 Python/3.8.5 Darwin/18.7.0 botocore/1.19.28",
   "requestParameters":{
      "studioId":"es-9O5XXXXXXXXXXXXXXXXXXXX"
   },
   "responseElements":null,
   "requestID":"0fxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "eventID":"b0xxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "readOnly":true,
   "eventType":"AwsApiCall",
   "managementEvent":true,
   "eventCategory":"Management",
   "recipientAccountId":"653XXXXXXXXX"
```

```
}
```

Consultar la actividad de los trabajos y los usuarios de Spark

Para ver la actividad de trabajos de Spark realizada por los usuarios de Amazon EMR Studio, puede configurar la suplantación de usuarios en un clúster. Con la suplantación de usuarios, cada trabajo de Spark que se envía desde un espacio de trabajo se asocia al usuario de Studio que ejecutó el código.

Cuando la suplantación de usuarios está habilitada, Amazon EMR crea un directorio de usuarios de HDFS en el nodo principal del clúster para cada usuario que ejecuta código en el espacio de trabajo. Por ejemplo, si el usuario `studio-user-1@example.com` ejecuta código, puede conectarse al nodo principal y comprobar si `hadoop fs -ls /user` tiene un directorio para `studio-user-1@example.com`.

Para configurar la suplantación de usuarios de Spark, establezca las siguientes propiedades en las siguientes clasificaciones de configuración:

- `core-site`
- `livy-conf`

```
[
    {
        "Classification": "core-site",
        "Properties": {
          "hadoop.proxyuser.livy.groups": "*",
          "hadoop.proxyuser.livy.hosts": "*"
        }
    },
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.impersonation.enabled": "true"
        }
    }
]
```

Para consultar las páginas del servidor de historial, consulte Depurar aplicaciones y trabajos con EMR Studio. También puede conectarse al nodo principal del clúster mediante SSH para ver las

interfaces web de la aplicación. Para obtener más información, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Actualizar un Amazon EMR Studio

Tras crear un EMR Studio, puede actualizar los siguientes atributos mediante la AWS CLI:

- Nombre
- Descripción
- Ubicación de S3 predeterminada
- Subredes

Para actualizar un EMR Studio mediante la AWS CLI

Utilice el comando `update-studio` de la AWS CLI para actualizar un EMR Studio. Para obtener más información, consulte Referencia de comandos de la AWS CLI.

> **Note**
>
> Un estudio se puede utilizar con un máximo de 5 subredes. Estas subredes deben pertenecer a la misma VPC que el estudio. La lista de identificadores de subred que envíe al comando `update-studio` puede incluir nuevos identificadores de subred, pero también debe incluir todos los identificadores de subred que ya haya asociado al estudio. No se pueden eliminar subredes de un estudio.

```
aws emr update-studio \
 --studio-id <example-studio-id-to-update> \
 --name <example-new-studio-name> \
 --subnet-ids <old-subnet-id-1 old-subnet-id-2 old-subnet-id-3 new-subnet-id> \
```

Para verificar los cambios, use el comando `describe-studio` de la AWS CLI y especifique su ID de Studio. Para obtener más información, consulte Referencia de comandos de la AWS CLI.

```
aws emr describe-studio \
 --studio-id <id-of-updated-studio> \
```

Eliminar un Amazon EMR Studio y un espacio de trabajo

Al eliminar un estudio, EMR Studio elimina todas las asignaciones de usuarios y grupos de IAM Identity Center asociadas al estudio.

> **Note**
>
> Al eliminar un estudio, Amazon EMR no elimina los espacios de trabajo asociados a ese estudio. Debe eliminar los espacios de trabajo de su estudio por separado.

Eliminar espacios de trabajo

Console

Dado que cada espacio de trabajo de EMR Studio es una instancia de cuaderno de EMR, puede utilizar la consola de administración de Amazon EMR para eliminar los espacios de trabajo. Puede eliminar espacios de trabajo mediante la consola de Amazon EMR antes o después de eliminar su estudio.

Para eliminar un espacio de trabajo mediante la consola de Amazon EMR

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Cuadernos.
3. Seleccione los espacios de trabajo que desea eliminar.
4. Seleccione Eliminar y, a continuación, Eliminar de nuevo para confirmar la acción.
5. Siga las instrucciones para eliminar objetos en la Guía del usuario de la consola de Amazon Simple Storage Service para eliminar los archivos de cuadernos asociados al espacio de trabajo eliminado de Amazon S3.

EMR Studio UI

From the Workspace UI

Eliminar un espacio de trabajo y sus archivos de copia de seguridad asociados de EMR Studio

1. Inicie sesión en su EMR Studio con la URL de acceso de su estudio y elija Espacios de trabajo en el menú de navegación de la izquierda.
2. Busque su espacio de trabajo en la lista y seleccione la casilla de verificación que aparece junto a su nombre. Puede seleccionar varios espacios de trabajo para eliminarlos al mismo tiempo.
3. Seleccione Eliminar en la esquina superior derecha de la lista de Espacios de trabajo y confirme que desea eliminar los espacios de trabajo seleccionados. Elija Delete para confirmar.
4. Si desea eliminar los archivos de cuadernos que estaban asociados al espacio de trabajo eliminado de Amazon S3, siga las instrucciones para eliminar objetos en la Guía del usuario de la consola de Amazon Simple Storage Service. Si usted no creó el estudio, consulte a su administrador del estudio para determinar la ubicación de la copia de seguridad de Amazon S3 para el espacio de trabajo eliminado.

From the Workspaces list

Eliminar un espacio de trabajo y sus archivos de copia de seguridad asociados de la lista de espacios de trabajo

1. Navegue hasta la lista Espacios de trabajo en la consola.
2. Seleccione el espacio de trabajo que desee eliminar de la lista y, a continuación, seleccione Acciones.
3. Elija Eliminar (Delete).
4. Si desea eliminar los archivos de cuadernos que estaban asociados al espacio de trabajo eliminado de Amazon S3, siga las instrucciones para eliminar objetos en la Guía del usuario de la consola de Amazon Simple Storage Service. Si usted no creó el estudio, consulte a su administrador del estudio para determinar la ubicación de la copia de seguridad de Amazon S3 para el espacio de trabajo eliminado.

Eliminar un EMR Studio

New console

Para eliminar un EMR Studio con la nueva consola

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/emr.
2. En la barra de navegación de la izquierda, en EMR Studio, seleccione Studios.
3. Seleccione el estudio en la lista Studios con el botón situado a la izquierda del nombre del estudio. Elija Eliminar (Delete).

Old console

Para eliminar un EMR Studio con la antigua consola

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/home.
2. Seleccione EMR Studio en el menú de navegación de la izquierda.
3. Seleccione el estudio en la lista Studios y elija Eliminar.

CLI

Para eliminar un EMR Studio con la AWS CLI

Utilice el comando `delete-studio` de la AWS CLI para eliminar un EMR Studio. Para obtener más información, consulte Referencia de comandos de la AWS CLI.

```
aws emr delete-studio --studio-id <id-of-studio-to-delete>
```

Definir grupos de seguridad para controlar el tráfico de red de EMR Studio

Acerca de los grupos de seguridad de EMR Studio

Amazon EMR Studio utiliza dos grupos de seguridad para controlar el tráfico de red entre los espacios de trabajo del estudio y un clúster de Amazon EMR asociado que se ejecuta en Amazon EC2:

- Un grupo de seguridad del motor que utiliza el puerto 18888 para comunicarse con un clúster de Amazon EMR asociado que se ejecuta en Amazon EC2.

- Un grupo de seguridad de espacio de trabajo asociado a los espacios de trabajo de un estudio. Este grupo de seguridad incluye una regla HTTPS saliente para permitir que el espacio de trabajo enrute el tráfico a internet y debe permitir el tráfico saliente a internet en el puerto 443 para poder vincular los repositorios de Git a un espacio de trabajo.

EMR Studio usa estos grupos de seguridad además de cualquier grupo de seguridad asociado a un clúster de EMR conectado a un espacio de trabajo.

Debe crear estos grupos de seguridad cuando utilice la AWS CLI para crear un estudio.

> **Note**
>
> Puede personalizar los grupos de seguridad de EMR Studio con reglas adaptadas a su entorno, pero debe incluir las reglas que se indican en esta página. El grupo de seguridad del espacio de trabajo no puede permitir ningún tráfico entrante y el grupo de seguridad del motor debe permitir el tráfico entrante del grupo de seguridad del espacio de trabajo.

Usar los grupos de seguridad predeterminados de EMR Studio

Al utilizar la consola de Amazon EMR, puede elegir los siguientes grupos de seguridad predeterminados. EMR Studio crea los grupos de seguridad predeterminados en su nombre, los cuales contienen las reglas de entrada y salida mínimas necesarias para los espacios de trabajo de un EMR Studio.

- `DefaultEngineSecurityGroup`
- `DefaultWorkspaceSecurityGroupGit` o `DefaultWorkspaceSecurityGroupWithoutGit`

Requisitos previos

Para crear los grupos de seguridad de EMR Studio, necesita una Amazon Virtual Private Cloud (VPC) para el estudio. Debe elegir esta VPC al crear los grupos de seguridad. Debe ser la misma VPC que especifique al crear el estudio. Si planea usar Amazon EMR en EKS con EMR Studio, seleccione la VPC para los nodos de trabajo del clúster de Amazon EKS.

Instrucciones

Siga las instrucciones de Creación de un grupo de seguridad de la Guía del usuario de Amazon EC2 para instancias de Linux para crear un grupo de seguridad del motor y un grupo de seguridad del

espacio de trabajo en la VPC. Los grupos de seguridad deben incluir las reglas que se resumen en las siguientes tablas.

Al crear grupos de seguridad para EMR Studio, anote los ID de ambos. Al crear un estudio, debe especificar cada grupo de seguridad por ID.

Grupo de seguridad del motor

EMR Studio usa el puerto 18888 para comunicarse con un clúster asociado.

Reglas de entrada

Tipo	Protocolo	Port	Destino	Descripción
TCP	TCP	18888	Su grupo de seguridad del espacio de trabajo de EMR Studio.	Permita el tráfico desde cualquier recurso del grupo de seguridad del espacio de trabajo para EMR Studio.

Grupo de seguridad del espacio de trabajo

Este grupo de seguridad está asociado a los espacios de trabajo de un EMR Studio.

Reglas de salida

Tipo	Protocolo	Port	Destino	Descripción
TCP	TCP	18888	Su grupo de seguridad del motor de EMR Studio.	Permita el tráfico a cualquier recurso del grupo de seguridad del motor para EMR Studio.
HTTPS	TCP	443	0.0.0.0/0	Permita que el tráfico de internet vincule los repositorios de Git alojados públicamente con los espacios de trabajo.

Crear plantillas de AWS CloudFormation para Amazon EMR Studio

Acerca de las plantillas de clúster de EMR Studio

Puede crear plantillas de AWS CloudFormation para ayudar a los usuarios de EMR Studio a lanzar nuevos clústeres de Amazon EMR en un espacio de trabajo. Las plantillas de CloudFormation son archivos de texto con formato de tipo JSON o YAML. En una plantilla, debe describir una pila de recursos de AWS e indicar a CloudFormation cómo aprovisionar esos recursos por usted. En el caso de EMR Studio, puede crear una o más plantillas que describan un clúster de Amazon EMR.

Puede organizar sus plantillas en AWS Service Catalog. AWS Service Catalog le permite crear y administrar los servicios de TI más desplegados, denominados productos en AWS. Debe recopilar sus plantillas como productos en una cartera que comparta con los usuarios de EMR Studio. Tras crear las plantillas de clúster, los usuarios de Studio pueden lanzar un nuevo clúster para un espacio de trabajo con una de sus plantillas. Los usuarios deben tener permiso para crear nuevos clústeres a partir de plantillas. Puede configurar los permisos de usuario en las políticas de permisos de EMR Studio.

Para obtener más información sobre las plantillas de CloudFormation, consulte Plantillas en la Guía del usuario de AWS CloudFormation. Para obtener más información acerca de AWS Service Catalog, consulte ¿Qué es AWS Service Catalog.

En el siguiente video se muestra cómo configurar plantillas de clúster en AWS Service Catalog para EMR Studio. También puede obtener más información en la entrada del blog Crear un entorno de autoservicio para cada línea de negocio mediante Amazon EMR y Service Catalog.

Parámetros de plantilla opcionales

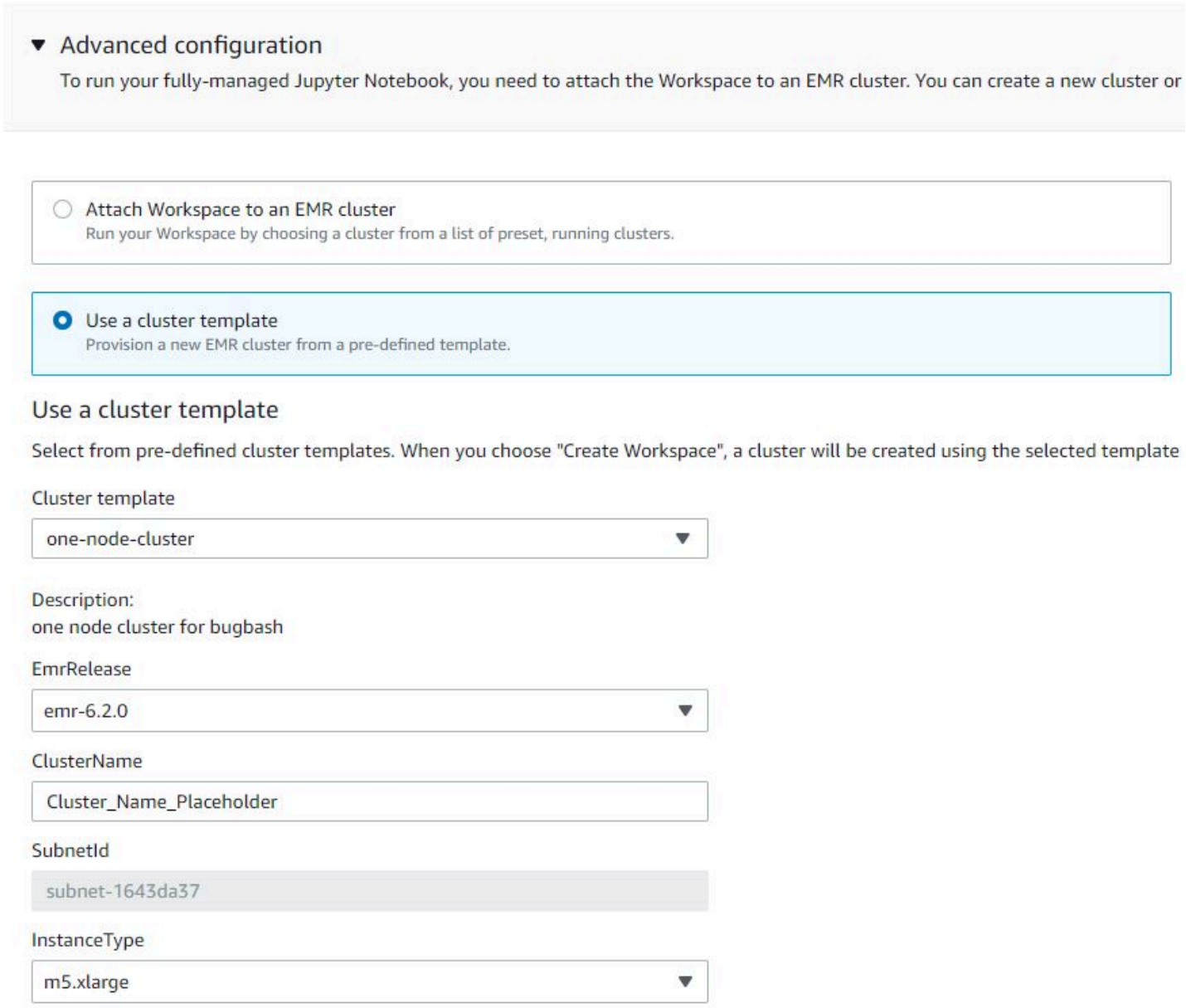
Puede incluir opciones adicionales en la sección `Parameters` de la plantilla. Los parámetros permiten a los usuarios de Studio introducir o seleccionar valores personalizados para un clúster. Por ejemplo, puede agregar un parámetro que permita a los usuarios seleccionar una versión concreta de Amazon EMR. Para obtener más información, consulte Parámetros en la Guía del usuario de AWS CloudFormation.

En la siguiente sección `Parameters` de ejemplo se definen parámetros de entrada adicionales, como `ClusterName`, la versión de `EmrRelease` y `ClusterInstanceType`.

```
Parameters:
  ClusterName:
```

```
    Type: "String"
    Default: "Cluster_Name_Placeholder"
  EmrRelease:
    Type: "String"
    Default: "emr-6.2.0"
    AllowedValues:
    - "emr-6.2.0"
    - "emr-5.32.0"
  ClusterInstanceType:
    Type: "String"
    Default: "m5.xlarge"
    AllowedValues:
    - "m5.xlarge"
    - "m5.2xlarge"
```

Al agregar parámetros, los usuarios de Studio ven opciones de formulario adicionales tras seleccionar una plantilla de clúster. En la siguiente imagen se muestran opciones de formato adicionales para la versión de EmrRelease, ClusterName e InstanceType.



Requisitos previos

Antes de crear una plantilla de clúster, asegúrese de tener permisos de IAM para acceder a la vista de la consola de administrador de Service Catalog. También necesita los permisos de IAM necesarios para realizar las tareas administrativas de Service Catalog. Para obtener más información, consulte Conceder permisos a los administradores de Service Catalog.

Instrucciones

Para crear plantillas de clústeres de EMR mediante Service Catalog

1. Cree una o más plantillas de CloudFormation. Puede elegir el lugar donde almacena sus plantillas. Como las plantillas son archivos de texto con formato, puede cargarlas en Amazon S3 o guardarlas en su sistema de archivos local. Para obtener más información sobre las plantillas de CloudFormation, consulte Plantillas en la Guía del usuario de AWS CloudFormation.

   Use las siguientes reglas para asignar un nombre a sus plantillas o compare sus nombres con el patrón `[a-zA-Z0-9][a-zA-Z0-9._-]*`.

   - El nombre de las plantillas debe comenzar por una letra o un número.
   - Los nombres de plantillas solo pueden constar de letras, números, puntos (.), guiones bajos (_) y guiones (-).

   Cada plantilla de clúster que cree debe incluir las siguientes opciones:

   Parámetros de entrada

   - ClusterName: nombre del clúster para ayudar a los usuarios a identificarlo una vez aprovisionado.

   Salida

   - `ClusterId`: el ID del clúster de EMR recién aprovisionado.

   A continuación, se muestra un ejemplo de plantilla de AWS CloudFormation en formato YAML para un clúster con dos nodos. En la plantilla de ejemplo se incluyen las opciones de plantilla necesarias y se definen parámetros de entrada adicionales para `EmrRelease` y `ClusterInstanceType`.

```
awsTemplateFormatVersion: 2010-09-09

Parameters:
  ClusterName:
    Type: "String"
    Default: "Example_Two_Node_Cluster"
  EmrRelease:
```

```
    Type: "String"
    Default: "emr-6.2.0"
    AllowedValues:
    - "emr-6.2.0"
    - "emr-5.32.0"
  ClusterInstanceType:
    Type: "String"
    Default: "m5.xlarge"
    AllowedValues:
    - "m5.xlarge"
    - "m5.2xlarge"

Resources:
  EmrCluster:
    Type: AWS::EMR::Cluster
    Properties:
      Applications:
      - Name: Spark
      - Name: Livy
      - Name: JupyterEnterpriseGateway
      - Name: Hive
      EbsRootVolumeSize: '10'
      Name: !Ref ClusterName
      JobFlowRole: EMR_EC2_DefaultRole
      ServiceRole: EMR_DefaultRole_V2
      ReleaseLabel: !Ref EmrRelease
      VisibleToAllUsers: true
      LogUri:
        Fn::Sub: 's3://aws-logs-${AWS::AccountId}-${AWS::Region}/elasticmapreduce/'
      Instances:
        TerminationProtected: false
        Ec2SubnetId: 'subnet-ab12345c'
        MasterInstanceGroup:
          InstanceCount: 1
          InstanceType: !Ref ClusterInstanceType
        CoreInstanceGroup:
          InstanceCount: 1
          InstanceType: !Ref ClusterInstanceType
          Market: ON_DEMAND
          Name: Core

Outputs:
  ClusterId:
    Value:
```

```
      Ref: EmrCluster
    Description: The ID of the  EMR cluster
```

2. Cree una cartera para sus plantillas de clúster en la misma cuenta de AWS que su estudio.

   a. Abra la consola de AWS Service Catalog en https://console.aws.amazon.com/servicecatalog/.

   b. En el menú de navegación izquierdo, seleccione Carteras.

   c. En la página Crear cartera, introduzca la información solicitada.

   d. Seleccione Crear. AWS Service Catalog crea la cartera y muestra sus detalles.

3. Utilice los siguientes pasos para agregar plantillas de clústeres como productos de AWS Service Catalog.

   a. Navegue a la página Productos en Administración en la consola de administración de AWS Service Catalog.

   b. Seleccione Cargar un producto nuevo.

   c. Introduzca el nombre del producto y el propietario.

   d. Especifique el archivo de plantilla en Detalles de la versión.

   e. Seleccione Revisar para revisar la configuración del producto y, a continuación, seleccione Crear producto.

4. Complete los siguientes pasos para agregar sus productos a su cartera.

   a. Navegue a la página Productos en la consola de administración de AWS Service Catalog.

   b. Seleccione su producto, seleccione Acciones y, a continuación, seleccione Agregar producto a la cartera.

   c. Seleccione su cartera y, a continuación, seleccione Agregar producto a la cartera.

5. Cree una restricción de lanzamiento para sus productos. Una restricción de lanzamiento es un rol de IAM que especifica los permisos de los usuarios para lanzar un producto. Puede personalizar sus restricciones de lanzamiento, pero debe permitir permisos para usar CloudFormation, Amazon EMR y AWS Service Catalog. Para obtener más información e instrucciones, consulte Restricciones de lanzamiento de Service Catalog.

6. Aplique su restricción de lanzamiento a cada producto de su cartera. Debe aplicar la restricción de lanzamiento a cada producto de forma individual.

   a. Seleccione su cartera en la página Carteras de la consola de administración de AWS Service Catalog.

b. Elija la pestaña Constraints (Restricciones) y elija Create constraint (Crear restricción).

c. Seleccione su producto y, en Tipo de restricción, Lanzamiento. Elija Continue (Continuar).

d. Seleccione su rol de restricción de lanzamiento en la sección Restricción de lanzamiento y, a continuación, seleccione Crear.

7. Concede acceso a su cartera.

a. Seleccione su cartera en la página Carteras de la consola de administración de AWS Service Catalog.

b. Amplíe la pestaña Grupos, roles y usuarios y seleccione Agregar grupos, roles y usuarios.

c. Busque su rol de IAM de EMR Studio en la pestaña Roles, seleccione su rol y elija Agregar acceso.

Si usa…	Conceder acceso a…
Autenticación de IAM	Sus usuarios nativos
Federación de IAM	Su rol de IAM en la federación
Federación de IAM Identity Center	Su rol de usuario de EMR Studio

Establecer el acceso y los permisos para los repositorios basados en Git

EMR Studio admite los siguientes servicios basados en Git:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

Para permitir que los usuarios de EMR Studio asocien un repositorio de Git a un espacio de trabajo, configure los siguientes requisitos de acceso y permisos. También puede configurar los repositorios basados en Git que aloje en una red privada siguiendo las instrucciones que se indican en Configurar un repositorio de Git alojado de forma privada para EMR Studio.

Acceso a internet en clúster

Tanto los clústeres de Amazon EMR que se ejecutan en Amazon EC2 como los de Amazon EMR en EKS asociados a espacios de trabajo de Studio deben estar en una subred privada que utilice una puerta de enlace de traducción de direcciones de red (NAT) o deben poder acceder a internet a través de una puerta de enlace privada virtual. Para obtener más información, consulte Opciones de Amazon VPC.

Los grupos de seguridad que utilice con EMR Studio también deben incluir una regla de salida que permita que los espacios de trabajo enruten el tráfico a internet desde un clúster de EMR asociado. Para obtener más información, consulte Definir grupos de seguridad para controlar el tráfico de red de EMR Studio.

> **Important**
>
> Si la interfaz de red se encuentra en una subred pública, no podrá comunicarse con internet a través de una puerta de enlace de Internet (IGW).

Permisos de AWS Secrets Manager

Para permitir que los usuarios de EMR Studio accedan a los repositorios de Git con secretos almacenados en AWS Secrets Manager, agregue una política de permisos al rol de servicio de EMR Studio que permite la operación `secretsmanager:GetSecretValue`.

Para obtener información sobre cómo vincular los repositorios basados en Git a los espacios de trabajo, consulte Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio.

Configurar un repositorio de Git alojado de forma privada para EMR Studio

Siga las instrucciones siguientes para configurar los repositorios alojados de forma privada para Amazon EMR Studio. Proporcione un archivo de configuración con información sobre sus servidores DNS y Git. EMR Studio utiliza esta información para configurar los espacios de trabajo que pueden enrutar el tráfico a sus repositorios autoadministrados.

> **Note**
>
> Si configura `DnsServerIpV4`, EMR Studio utilizará su servidor DNS para resolver tanto su `GitServerDnsName` como su punto de conexión de Amazon EMR, como

`elasticmapreduce.us-east-1.amazonaws.com`. Para configurar un punto de conexión para Amazon EMR, conéctese a su punto de conexión a través de la VPC que utilice con su estudio. De este modo, se garantiza que el punto de conexión de Amazon EMR se especifique en una IP privada. Para obtener más información, consulte Conexión a Amazon EMR mediante un punto de conexión de VPC de tipo interfaz.

Requisitos previos

Antes de configurar un repositorio de Git alojado de forma privada para EMR Studio, necesita una ubicación de almacenamiento de Amazon S3 en la que EMR Studio pueda realizar copias de seguridad de los espacios de trabajo y los archivos de cuadernos del estudio. Utilice el mismo bucket de S3 que especifique al crear un estudio.

Para configurar un repositorio de Git alojado de forma privada para EMR Studio

1. Cree un archivo de configuración mediante la siguiente plantilla. Incluya los siguientes valores para cada servidor de Git que desee especificar en la configuración:

   - **`DnsServerIpV4`**: la dirección IPv4 de su servidor de DNS. Si proporciona valores para `DnsServerIpV4` y `GitServerIpV4List`, el valor de `DnsServerIpV4` tiene prioridad y EMR Studio utiliza `DnsServerIpV4` para resolver el `GitServerDnsName`.

     > **Note**
     >
     > Para usar repositorios de Git alojados de forma privada, su servidor DNS debe permitir el acceso entrante desde EMR Studio. Le instamos a que proteja su servidor DNS contra otros accesos no autorizados.

   - **`GitServerDnsName`**: el nombre de DNS del servidor de Git. Por ejemplo `"git.example.com"`.
   - **`GitServerIpV4List`**: una lista de direcciones IPv4 que pertenecen a sus servidores de Git.

```
[
    {
        "Type": "PrivatelyHostedGitConfig",
        "Value": [
            {
```

```
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<enterprise.git.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            },
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<git.example.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            }
        ]
    }
]
```

2. Guarde su archivo de configuración como `configuration.json`.
3. Cargue el archivo de configuración en su ubicación de almacenamiento de Amazon S3, en una carpeta llamada `life-cycle-configuration`. Por ejemplo, si su ubicación de S3 predeterminada es `s3://DOC-EXAMPLE-BUCKET/studios`, el archivo de configuración estaría en `s3://DOC-EXAMPLE-BUCKET/studios/`**`life-cycle-configuration/configuration.json`**.

> **Important**
>
> Le instamos a que restrinja el acceso a su carpeta `life-cycle-configuration` a los administradores de Studio y a su rol de servicio de EMR Studio, y a que proteja `configuration.json` contra el acceso no autorizado. Para obtener instrucciones, consulte Controlar el acceso a un bucket con políticas de usuario o Prácticas recomendadas de seguridad para Amazon S3.

Para ver las instrucciones de carga, consulte Creación de una carpeta y Carga de objetos en la Guía del usuario de Amazon Simple Storage Service. Para aplicar la configuración a un espacio de trabajo existente, ciérrelo y reinícielo después de cargar el archivo de configuración en Amazon S3.

Optimizar los trabajos de Spark en EMR Studio

Al ejecutar un trabajo de Spark con EMR Studio, hay algunos pasos que puede seguir para asegurarse de que optimiza los recursos de su clúster de Amazon EMR.

Prolongar la sesión de Livy

Si utiliza Apache Livy junto con Spark en su clúster de Amazon EMR, le recomendamos que aumente el tiempo de espera de la sesión de Livy de la siguiente manera:

- Al crear un clúster de Amazon EMR, defina esta clasificación de configuración en el campo Introducir configuración.

```
[
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.server.session.timeout": "8h"
        }
    }
]
```

- En el caso de un clúster de EMR que ya se esté ejecutando, conéctese a su clúster mediante `ssh` y establezca la clasificación de configuración `livy-conf` en `/etc/livy/conf/livy.conf`.

```
[
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.server.session.timeout": "8h"
        }
    }
]
```

  Es posible que necesite reiniciar Livy después de cambiar la configuración.

- Si no quiere que se agote el tiempo de espera de su sesión de Livy, configure la propiedad `livy.server.session.timeout-check` como `false` en `/etc/livy/conf/livy.conf`.

Ejecutar Spark en el modo de clúster

En el modo de clúster, el controlador de Spark se ejecuta en un nodo de núcleo en lugar de en el nodo principal, lo que mejora la utilización de los recursos en el nodo principal.

Para ejecutar su aplicación de Spark en el modo de clúster en lugar del modo cliente predeterminado, elija el modo Clúster al configurar el Modo de implementación mientras configura el paso de Spark en el nuevo clúster de Amazon EMR. Para obtener más información, consulte Cluster mode overview en la documentación de Apache Spark.

Aumentar la memoria del controlador de Spark

Para aumentar la memoria del controlador de Spark, configure su sesión de Spark con el comando mágico `%%configure` de su cuaderno de EMR, como en el siguiente ejemplo.

```
%%configure -f
{"driverMemory": "6000M"}
```

Utilizar un Amazon EMR Studio

Esta sección contiene temas que lo ayudan a configurar un Amazon EMR Studio e interactuar con él.

En el siguiente video, se incluye información práctica, como la creación de un nuevo espacio de trabajo y el lanzamiento de un nuevo clúster de Amazon EMR con una plantilla de clúster. El video también muestra un cuaderno de muestra.

En esta sección se incluyen los temas siguientes, que le ayudarán a trabajar en un EMR Studio:

- Aprenda los conceptos básicos de los espacios de trabajo
- Configuración de la colaboración en el espacio de trabajo
- Ejecutar un espacio de trabajo de EMR Studio con un rol de tiempo de ejecución
- Ejecutar los cuadernos del espacio de trabajo mediante programación
- Examinar los datos con SQL Explorer
- Asociar computación a un espacio de trabajo de EMR Studio
- Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio
- Depurar aplicaciones y trabajos con EMR Studio
- Instalar kernels y bibliotecas en un espacio de trabajo de EMR Studio

- Mejorar los kernels con comandos magic
- Usar cuadernos multilingües con kernels de Spark

Aprenda los conceptos básicos de los espacios de trabajo

Cuando utiliza un EMR Studio, puede crear y configurar diferentes espacios de trabajo para organizar y ejecutar cuadernos. En esta sección se describe la creación de espacios de trabajo y el trabajo con ellos. Para obtener información general conceptual, consulte Workspaces en la página Cómo funciona Amazon EMR Studio.

En esta sección se cubren los temas siguientes, que le ayudarán a usar espacios de trabajo de EMR Studio:

- Crear un espacio de trabajo de EMR Studio
- Lanzar un espacio de trabajo
- Explicación de la interfaz de usuario del espacio de trabajo
- Explorar ejemplos de cuadernos
- Guardar contenido del espacio de trabajo
- Eliminar un espacio de trabajo y archivos de cuaderno
- Explicación del estado del espacio de trabajo
- Resolver problemas de conectividad con el espacio de trabajo

Crear un espacio de trabajo de EMR Studio

Puede crear espacios de trabajo de EMR Studio para ejecutar código de cuadernos mediante la interfaz de EMR Studio.

Para crear un espacio de trabajo en un EMR Studio

1. Inicie sesión en su EMR Studio.
2. Seleccione Crear un espacio de trabajo.
3. Ingrese un Nombre de espacio de trabajo y una Description. Asignar un nombre a un espacio de trabajo le ayuda a identificarlo en la página Espacios de trabajo.
4. Si quiere trabajar con otros usuarios de Studio en este espacio de trabajo en tiempo real, active la colaboración en el espacio de trabajo. Puede configurar los colaboradores después de lanzar el espacio de trabajo.

5. Si quiere asociar un clúster a un espacio de trabajo, expanda la sección Configuración avanzada. Si lo prefiere, puede asociar un clúster más adelante. Para obtener más información, consulte Asociar computación a un espacio de trabajo de EMR Studio.

   > **Note**
   >
   > Para aprovisionar un clúster nuevo, necesita los permisos de acceso de su administrador.

   Elija una de las opciones de clúster para el espacio de trabajo y asocie el clúster. Para obtener más información sobre el aprovisionamiento de un clúster al crear un espacio de trabajo, consulte Crear y asociar un nuevo clúster de EMR a un espacio de trabajo de EMR Studio.
6. Seleccione Crear un espacio de trabajo en la parte inferior derecha de la página.

Tras crear un espacio de trabajo, EMR Studio abrirá la página Espacios de trabajo. En la parte superior de la página, verá un banner verde indicando que la operación se ha realizado correctamente y encontrará el espacio de trabajo recién creado en la lista.

De forma predeterminada, los espacios de trabajo son compartidos y pueden verlos todos los usuarios de Studio. Sin embargo, solo un usuario puede abrir y trabajar en un espacio de trabajo a la vez. Para trabajar simultáneamente con otros usuarios, puede Configuración de la colaboración en el espacio de trabajo.

Lanzar un espacio de trabajo

Para empezar a trabajar con los archivos de cuaderno, abra un espacio de trabajo para acceder al editor del cuaderno. La página Espacios de trabajo de un estudio muestra todos los espacios de trabajo a los que tiene acceso, con detalles como el nombre, el estado, la hora de creación y la última modificación.

> **Note**
>
> Si tenía cuadernos de EMR en la antigua consola de Amazon EMR, los encontrará en la nueva consola como espacios de trabajo de EMR Studio. Los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales para acceder o crear espacios de trabajo. Si ha creado recientemente un cuaderno en la consola anterior, puede que tenga que actualizar la lista de espacios de trabajo para poder verlo en la nueva consola. Para obtener

más información sobre la transición, consulte Cuadernos de Amazon EMR está disponible como Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola.

Para iniciar un espacio de trabajo para editar y ejecutar cuadernos

1. En la página Espacios de trabajo de su estudio, busque el espacio de trabajo. Puede filtrar la lista por palabra clave o por valor de columna.
2. Elija el nombre del espacio de trabajo para iniciarlo en una nueva pestaña del navegador. Si está inactivo, es posible que el espacio de trabajo tarde unos minutos en abrirse. Como alternativa, seleccione la fila del espacio de trabajo y, a continuación, seleccione Iniciar espacio de trabajo. Puede elegir entre las siguientes opciones:

   - Inicio rápido: inicie rápidamente su espacio de trabajo con las opciones predeterminadas. Seleccione Inicio rápido si quiere asociar clústeres al espacio de trabajo de JupyterLab.
   - Inicio con opciones: inicie su espacio de trabajo con opciones personalizadas. Puede elegir lanzarlo en Jupyter o JupyterLab, asociar su espacio de trabajo a un clúster de EMR y seleccionar sus grupos de seguridad.

> **Note**
>
> Solo un usuario puede abrir y trabajar en un espacio de trabajo a la vez. Si selecciona un espacio de trabajo que ya está en uso, EMR Studio mostrará una notificación cuando intente abrirlo. La columna Usuario de la página Espacios de trabajo muestra el usuario que trabaja en el espacio de trabajo.

Explicación de la interfaz de usuario del espacio de trabajo

La interfaz de usuario del espacio de trabajo de EMR Studio se basa en la interfaz de JupyterLab con pestañas con iconos en la barra lateral izquierda. Al hacer una pausa sobre un icono, aparece información sobre herramientas que muestra el nombre de la pestaña. Seleccione las pestañas de la barra lateral izquierda para acceder a los siguientes paneles.

- Explorador de archivos: muestra los archivos y directorios del espacio de trabajo, así como los archivos y directorios de los repositorios de Git enlazados.

- Ejecución de kernels y terminales: muestra todos los kernels y terminales que se ejecutan en el espacio de trabajo. Para obtener más información, consulte Managing Kernels and Terminals en la documentación oficial de JupyterLab.
- Git: proporciona una interfaz gráfica de usuario para ejecutar comandos en los repositorios de Git asociados al espacio de trabajo. Este panel es una extensión de JupyterLab llamada jupyterlab-git. Para obtener más información, consulte jupyterlab-git.
- Clústeres de EMR: le permite asociar un clúster al espacio de trabajo o separarlo a fin de ejecutar el código del cuaderno. El panel de configuración del clúster de EMR también ofrece opciones de configuración avanzadas para ayudarle a crear y asociar un clúster nuevo al espacio de trabajo. Para obtener más información, consulte Crear y asociar un nuevo clúster de EMR a un espacio de trabajo de EMR Studio.
- Repositorio de Git de Amazon EMR: le ayuda a vincular el espacio de trabajo con hasta tres repositorios de Git. Para obtener información más detallada e instrucciones, consulte Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio.
- Ejemplos de cuadernos: proporciona una lista de ejemplos de cuadernos que puede guardar en el espacio de trabajo. También puede acceder a los ejemplos seleccionando Ejemplos de cuadernos en la página de inicio del espacio de trabajo.
- Comandos: ofrece una forma de buscar y ejecutar comandos de JupyterLab mediante el teclado. Para obtener más información, consulte la página Command Palette en la documentación de JupyterLab.
- Herramientas de cuaderno: le permiten seleccionar y configurar opciones como el tipo de diapositiva de la celda y los metadatos. La opción Herramientas del cuaderno aparece en la barra lateral izquierda después de abrir un archivo de cuaderno.
- Pestañas abiertas: muestra los documentos y las actividades abiertos en el área de trabajo principal para que pueda ir directamente a una pestaña abierta. Para obtener más información, consulte la página Tabs and Single-Document Mode en la documentación de JupyterLab.
- Colaboración: le permite activar o desactivar la colaboración en el espacio de trabajo y administrar a los colaboradores. Para ver el panel Colaboración, debe tener los permisos necesarios. Para obtener más información, consulte Establecer la propiedad de la colaboración en el espacio de trabajo.

Explorar ejemplos de cuadernos

Cada espacio de trabajo de EMR Studio incluye un conjunto de ejemplos de cuadernos que puede utilizar para explorar las características de EMR Studio. Para editar o ejecutar un ejemplo de cuaderno, puede guardarlo en el espacio de trabajo.

Para guardar un ejemplo de cuaderno en un espacio de trabajo

1. En la barra lateral izquierda, seleccione la pestaña Ejemplos de cuadernos para abrir el panel Ejemplos de cuadernos. También puede acceder a los ejemplos seleccionando Ejemplos de cuadernos en la página de inicio del espacio de trabajo.
2. Elija un ejemplo de cuaderno para ver su vista previa en el área de trabajo principal. El ejemplo es de solo lectura.
3. Para guardar el ejemplo del cuaderno en el espacio de trabajo, seleccione Guardar en el espacio de trabajo. EMR Studio guarda el ejemplo en su directorio de inicio. Tras guardar un ejemplo de cuaderno en el espacio de trabajo, puede cambiar su nombre, editarlo y ejecutarlo.

Para obtener más información sobre los cuadernos de ejemplo, consulte el repositorio de GitHub con ejemplos de cuadernos de EMR Studio.

Guardar contenido del espacio de trabajo

Cuando trabaja en el editor del cuaderno de un espacio de trabajo, EMR Studio guarda el contenido de las celdas del cuaderno y la salida para usted en la ubicación de Amazon S3 asociada al estudio. Este proceso de copia de seguridad preserva el trabajo entre sesiones.

También puede guardar un cuaderno pulsando CTRL+S en la pestaña del cuaderno abierto o utilizando una de las opciones de guardado de la sección Archivo.

Otra forma de hacer copias de seguridad de los archivos de cuaderno en un espacio de trabajo consiste en asociar el espacio de trabajo a un repositorio basado en Git y sincronizar los cambios con el repositorio remoto. De este modo, también podrá guardar y compartir cuadernos con los miembros del equipo que usen un espacio de trabajo o un estudio diferente. Para obtener instrucciones, consulte Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio.

Eliminar un espacio de trabajo y archivos de cuaderno

Al eliminar un archivo de cuaderno de un espacio de trabajo de EMR Studio, se elimina el archivo del Explorador de archivos y EMR Studio elimina su copia de seguridad en Amazon S3. No tiene que

tomar ninguna otra medida para evitar los cargos de almacenamiento cuando elimina un archivo de un espacio de trabajo.

Al eliminar un espacio de trabajo completo, los archivos de cuaderno y las carpetas permanecerán en la ubicación de almacenamiento de Amazon S3. Los archivos siguen acumulando gastos de almacenamiento. Para evitar gastos de almacenamiento, elimine de Amazon S3 todos los archivos y carpetas de los que se haya hecho una copia de seguridad y que estén asociados a su espacio de trabajo eliminado.

Para eliminar un archivo de cuaderno de un espacio de trabajo de EMR Studio

1. Seleccione el panel del Explorador de archivos en la barra lateral izquierda del espacio de trabajo.
2. Seleccione el archivo o la carpeta que desea eliminar. Haga clic con el botón derecho en la selección y seleccione Eliminar. El archivo desaparecerá de la lista. EMR Studio elimina automáticamente el archivo o la carpeta de Amazon S3.

From the Workspace UI

Eliminar un espacio de trabajo y sus archivos de copia de seguridad asociados de EMR Studio

1. Inicie sesión en su EMR Studio con la URL de acceso de su estudio y elija Espacios de trabajo en el menú de navegación de la izquierda.
2. Busque su espacio de trabajo en la lista y seleccione la casilla de verificación que aparece junto a su nombre. Puede seleccionar varios espacios de trabajo para eliminarlos al mismo tiempo.
3. Seleccione Eliminar en la esquina superior derecha de la lista de Espacios de trabajo y confirme que desea eliminar los espacios de trabajo seleccionados. Elija Delete para confirmar.
4. Si desea eliminar los archivos de cuadernos que estaban asociados al espacio de trabajo eliminado de Amazon S3, siga las instrucciones para eliminar objetos en la Guía del usuario de la consola de Amazon Simple Storage Service. Si usted no creó el estudio, consulte a su administrador del estudio para determinar la ubicación de la copia de seguridad de Amazon S3 para el espacio de trabajo eliminado.

From the Workspaces list

Eliminar un espacio de trabajo y sus archivos de copia de seguridad asociados de la lista de espacios de trabajo

1. Navegue hasta la lista Espacios de trabajo en la consola.
2. Seleccione el espacio de trabajo que desee eliminar de la lista y, a continuación, seleccione Acciones.
3. Elija Eliminar (Delete).
4. Si desea eliminar los archivos de cuadernos que estaban asociados al espacio de trabajo eliminado de Amazon S3, siga las instrucciones para eliminar objetos en la Guía del usuario de la consola de Amazon Simple Storage Service. Si usted no creó el estudio, consulte a su administrador del estudio para determinar la ubicación de la copia de seguridad de Amazon S3 para el espacio de trabajo eliminado.

Explicación del estado del espacio de trabajo

Tras crear un espacio de trabajo de EMR Studio, aparece como una fila en la lista Espacios de trabajo de su estudio con su nombre, estado, hora de creación y fecha de la última modificación. En la tabla siguiente se describen los estados de los espacios de trabajo.

Estado	Descripción
Starting	El espacio de trabajo se está preparando, pero aún no está listo para usarse. No puede abrir un espacio de trabajo cuando se encuentra en el estado Iniciando.
Ready	Puede abrir el espacio de trabajo para usar el editor de cuadernos, pero debe asociar el espacio de trabajo a un clúster de EMR antes de poder ejecutar el código del cuaderno.
Asociando	El espacio de trabajo se está asociando a un clúster.
Attached (Asociado)	El espacio de trabajo se ha asociado a un clúster de EMR y está listo para que escriba

Estado	Descripción
	y ejecute el código del cuaderno. Si el estado de un espacio de trabajo no es Asociado, debe asociarlo a un clúster para poder ejecutar el código del cuaderno.
Inactivo	El espacio de trabajo se ha detenido. Para reactivar un espacio de trabajo inactivo, selecciónelo en la lista de espacios de trabajo. El estado cambia de Inactivo a Empezando y a Preparado al seleccionar el espacio de trabajo.
Stopping	El espacio de trabajo se está apagando y pasará a estar Inactivo. Al detener un espacio de trabajo, se cancelan los kernels de cuadernos correspondientes. EMR Studio detiene los cuadernos que han estado inactivos durante mucho tiempo.
Eliminando	Al eliminar un espacio de trabajo, EMR Studio lo marca para su eliminación e inicia el proceso de eliminación. Una vez finalizado el proceso de eliminación, el espacio de trabajo desaparec e de la lista. Al eliminar un espacio de trabajo, los archivos de cuaderno permanecerán en la ubicación de almacenamiento de Amazon S3.

Resolver problemas de conectividad con el espacio de trabajo

Para resolver los problemas de conectividad de un espacio de trabajo, puede detenerlo y reiniciarlo. Al reiniciar un espacio de trabajo, EMR Studio lanza el espacio de trabajo en una zona de disponibilidad diferente o en una subred diferente que esté asociada a su estudio.

Para detener y reiniciar un espacio de trabajo de EMR Studio

1. Cierre el espacio de trabajo en su navegador.
2. Navegue hasta la lista Espacios de trabajo en la consola.

3. En la lista, seleccione su espacio de trabajo y, a continuación, seleccione Acciones.
4. Seleccione Detener y espera a que el estado del espacio de trabajo cambie de Deteniendo a Inactivo.
5. Vuelva a seleccionar Acciones y, a continuación, seleccione Iniciar para reiniciar el espacio de trabajo.
6. Espere a que el estado del espacio de trabajo cambie de Empezando a Listo y, a continuación, elija el nombre del espacio de trabajo para volver a abrirlo en una nueva pestaña del navegador.

Configuración de la colaboración en el espacio de trabajo

La colaboración en el espacio de trabajo le permite escribir y ejecutar código de cuadernos de forma simultánea con otros miembros del equipo. Cuando trabaje en el mismo archivo de cuaderno, verá los cambios a medida que los hagan los colaboradores. Puede habilitar la colaboración al crear un espacio de trabajo o activar y desactivar la colaboración en un espacio de trabajo existente.

> **Note**
>
> La colaboración en el espacio de trabajo de EMR Studio no es compatible con las aplicaciones interactivas de EMR sin servidor.

Requisitos previos

Antes de configurar la colaboración en un espacio de trabajo, asegúrese de completar las siguientes tareas:

- Asegúrese de que el administrador de EMR Studio le haya otorgado los permisos necesarios. Por ejemplo, la siguiente instrucción permite a un usuario configurar la colaboración para cualquier espacio de trabajo con la clave de etiqueta `creatorUserId` cuyo valor coincida con el ID del usuario (indicado por la variable de política `aws:userId`).

```
{
    "Sid": "UserRolePermissionsForCollaboration",
    "Action": [
        "elasticmapreduce:UpdateEditor",
        "elasticmapreduce:PutWorkspaceAccess",
        "elasticmapreduce:DeleteWorkspaceAccess",
        "elasticmapreduce:ListWorkspaceAccessIdentities"
```

```
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Condition": {
        "StringEquals": {
            "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
        }
    }
}
```

- Asegúrese de que el rol de servicio asociado a su EMR Studio tenga los permisos necesarios para habilitar y configurar la colaboración en el espacio de trabajo, como se muestra en la siguiente declaración de ejemplo.

```
{
    "Sid": "AllowWorkspaceCollaboration",
    "Effect": "Allow",
    "Action": [
        "iam:GetUser",
        "iam:GetRole",
        "iam:ListUsers",
        "iam:ListRoles",
        "sso:GetManagedApplicationInstance",
        "sso-directory:SearchUsers"
    ],
    "Resource": "*"
}
```

  Para obtener más información, consulte Crear un rol de servicio de EMR Studio.

Para habilitar la colaboración en el espacio de trabajo y agregar colaboradores

1. En su espacio de trabajo, seleccione el icono de Colaboración en la pantalla de inicio o en la parte inferior del panel izquierdo.

   > **Note**
   >
   > No verá el panel Colaboración a menos que el administrador de Studio le haya dado permiso para configurar la colaboración en el espacio de trabajo. Para obtener más

información, consulte Establecer la propiedad de la colaboración en el espacio de trabajo.

2. Asegúrese de que la opción Permitir la colaboración en el espacio de trabajo esté activada. Cuando habilita la colaboración, solo usted y los colaboradores que agregue podrán ver el espacio de trabajo en la lista de la página Espacios de trabajo de Studio.
3. Introduzca el nombre de un colaborador. En el espacio de trabajo se puede tener un máximo de cinco colaboradores, incluido usted. Un colaborador puede ser cualquier usuario con acceso a su EMR Studio. Si no introduce un colaborador, el espacio de trabajo es un espacio de trabajo privado al que solo usted puede acceder.

   En la siguiente tabla se especifican los valores de colaborador aplicables que se deben introducir en función del tipo de identidad del propietario.

   > **Note**
   >
   > Un propietario solo puede invitar a colaboradores con el mismo tipo de identidad. Por ejemplo, un usuario solo puede agregar a otros usuarios, y un usuario de IAM Identity Center solo puede agregar a otros usuarios de IAM Identity Center.

Modo de autenticación	Valor que se debe introducir en Nombre del colaborador
Autenticación de IAM	un nombre de usuario. Este es el nombre que ve un usuario al iniciar sesión en la AWS Management Console.
Federación de IAM	El nombre de un rol de IAM y un nombre de sesión opcional. Para agregar todos los usuarios federados que asumen el mismo rol de IAM, especifiq ue el nombre de un rol de IAM para la federación. Para agregar un único usuario como colaborador, especifique un rol y un nombre

Modo de autenticación	Valor que se debe introducir en Nombre del colaborador
	de sesión. Por ejemplo, `MyRoleNam e:MySessionName` .
SSO	Un nombre de usuario de IAM Identity Center, como `user@example.com.` .

4. Elija Add (Agregar). El colaborador ahora puede ver el espacio de trabajo en su página Espacios de trabajo de EMR Studio e iniciar el espacio de trabajo para usarlo en tiempo real con usted.

> **Note**
>
> Si inhabilita la colaboración en el espacio de trabajo, el espacio de trabajo vuelve a su estado compartido y todos los usuarios de Studio lo pueden ver. En el estado compartido, solo un usuario de Studio puede abrir un espacio de trabajo y trabajar en él a la vez.

Ejecutar un espacio de trabajo de EMR Studio con un rol de tiempo de ejecución

> **Note**
>
> La funcionalidad del rol de tiempo de ejecución que se describe en esta página solo se aplica a Amazon EMR en ejecución en Amazon EC2 y no hace referencia a la funcionalidad del rol de tiempo de ejecución en las aplicaciones interactivas de EMR sin servidor. Para obtener más información sobre cómo utilizar los roles de tiempo de ejecución en EMR sin servidor, consulte Rol de tiempo de ejecución de tareas en la Guía del usuario de Amazon EMR sin servidor.

Un rol en tiempo de ejecución es un rol de AWS Identity and Access Management (IAM) que puede especificar al enviar un trabajo o una consulta a un clúster de Amazon EMR. El trabajo o la consulta que envíe a su clúster de EMR utiliza el rol de tiempo de ejecución para acceder a los recursos de AWS, como los objetos de Amazon S3.

Al asociar un espacio de trabajo de EMR Studio a un clúster de EMR que usa Amazon EMR 6.11 o una versión posterior, puede seleccionar un rol de tiempo de ejecución para el trabajo o la consulta que envíe para usarlo cuando acceda a los recursos de AWS. Sin embargo, si el clúster de EMR no admite roles de tiempo de ejecución, el clúster de EMR no asumirá el rol cuando acceda a los recursos de AWS.

Antes de poder utilizar un rol de tiempo de ejecución con un espacio de trabajo de Amazon EMR Studio, un administrador debe configurar los permisos de usuario para que el usuario de Studio pueda llamar a la API `elasticmapreduce:GetClusterSessionCredentials` en el rol de tiempo de ejecución. A continuación, lance un nuevo clúster con un rol de tiempo de ejecución que pueda usar con su espacio de trabajo de Amazon EMR Studio.

En esta página

- Configurar los permisos de usuario para el rol de tiempo de ejecución
- Lanzar un clúster nuevo con un rol de tiempo de ejecución
- Utilizar el clúster de EMR con un rol de tiempo de ejecución en los espacios de trabajo
- Consideraciones

Configurar los permisos de usuario para el rol de tiempo de ejecución

Configure los permisos de usuario para que el usuario de Studio pueda llamar a la API `elasticmapreduce:GetClusterSessionCredentials` en el rol de tiempo de ejecución que quiera usar. También debe configurar the section called “Permisos de usuario del estudio (EC2, EKS)” para que el usuario pueda empezar a usar Studio.

> **Warning**
>
> Para conceder este permiso, cree una condición basada en la clave de contexto `elasticmapreduce:ExecutionRoleArn` cuando conceda a la persona que llama acceso para llamar a las API de `GetClusterSessionCredentials`. En los siguientes ejemplos, se muestra cómo hacerlo.

```
{
      "Sid": "AllowSpecificExecRoleArn",
      "Effect": "Allow",
```

```
      "Action": [
          "elasticmapreduce:GetClusterSessionCredentials"
      ],
      "Resource": "*",
      "Condition": {
          "StringEquals": {
              "elasticmapreduce:ExecutionRoleArn": [
                  "arn:aws:iam::111122223333:role/test-emr-demo1",
                  "arn:aws:iam::111122223333:role/test-emr-demo2"
              ]
          }
      }
  }
```

En el siguiente ejemplo, se muestra cómo permitir que una entidad principal de IAM utilice un rol de IAM denominado `test-emr-demo3` como rol de tiempo de ejecución. Además, el titular de la política solo podrá acceder a los clústeres de Amazon EMR con el ID de clúster `j-123456789`.

```
{
    "Sid":"AllowSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:GetClusterSessionCredentials"
    ],
    "Resource": [
          "arn:aws:elasticmapreduce:<region>:111122223333:cluster/j-123456789"
     ],
    "Condition":{
        "StringEquals":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::111122223333:role/test-emr-demo3"
            ]
        }
    }
}
```

El siguiente ejemplo permite que una entidad principal de IAM utilice cualquier rol de IAM cuyo nombre comience por la cadena `test-emr-demo4` como rol de tiempo de ejecución. Además, el titular de la política solo podrá acceder a los clústeres de Amazon EMR etiquetados con el par clave-valor `tagKey: tagValue`.

```
{
```

```
    "Sid":"AllowSpecificExecRoleArn",
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:GetClusterSessionCredentials"
    ],
    "Resource": "*",
    "Condition":{
        "StringEquals":{
             "elasticmapreduce:ResourceTag/tagKey": "tagValue"
        },
        "StringLike":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::111122223333:role/test-emr-demo4*"
            ]
        }
    }
}
```

Lanzar un clúster nuevo con un rol de tiempo de ejecución

Ahora que ya tiene los permisos necesarios, lance un nuevo clúster con un rol de tiempo de ejecución que pueda usar con su espacio de trabajo de Amazon EMR Studio.

Si ya ha lanzado un clúster nuevo con un rol de tiempo de ejecución, puede pasar directamente a la sección the section called “Utilizar el clúster con el espacio de trabajo”.

1. En primer lugar, complete los requisitos previos de la sección Roles en tiempo de ejecución para los pasos de Amazon EMR.
2. A continuación, lance un clúster con la siguiente configuración para usar los roles de tiempo de ejecución con los espacios de trabajo de Amazon EMR Studio. Para obtener instrucciones sobre cómo lanzar el clúster, consulte Especificación de una configuración de seguridad para un clúster.

   - Seleccione la etiqueta de versión emr-6.11.0 o posterior.
   - Seleccione Spark, Livy y Jupyter Enterprise Gateway como aplicaciones de clúster.
   - Use la configuración de seguridad que creó en el paso anterior.
   - Si lo desea, puede habilitar Lake Formation para el clúster de EMR. Para obtener más información, consulte Habilitación de Lake Formation con Amazon EMR.

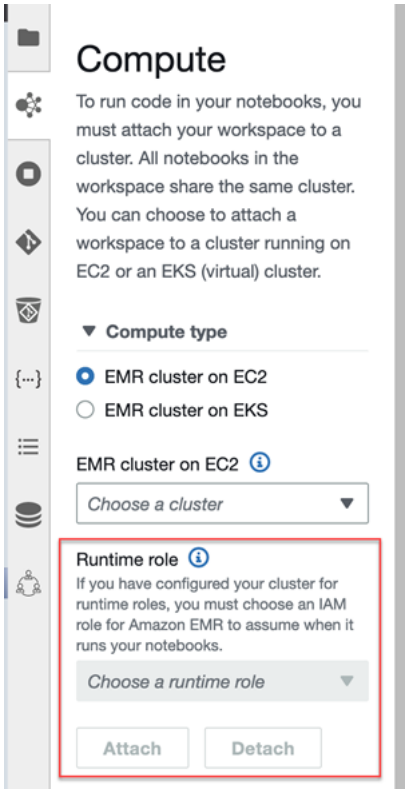
Después de lanzar el clúster, estará listo para usar el clúster habilitado para roles de tiempo de ejecución con un espacio de trabajo de EMR Studio.

> **Note**
>
> El valor de ExecutionRoleArn no es compatible actualmente con la operación de la API StartNotebookExecution cuando el valor de `ExecutionEngineConfig.Type` es `EMR`.

Utilizar el clúster de EMR con un rol de tiempo de ejecución en los espacios de trabajo

Después de configurar y lanzar el clúster, puede usar el clúster habilitado para roles de tiempo de ejecución con su espacio de trabajo de EMR Studio.

1. Cree un nuevo espacio de trabajo o lance uno existente. Para obtener más información, consulte Crear un espacio de trabajo de EMR Studio.
2. Seleccione la pestaña Clústeres de EMR en la barra lateral izquierda de su espacio de trabajo abierto, expanda la sección Tipo de computación y seleccione su clúster en el menú Clúster de EMR en EC2 y el rol de tiempo de ejecución en el menú Rol de tiempo de ejecución.



3. Seleccione Asociar para asociar el clúster con el rol de tiempo de ejecución a su espacio de trabajo.

Consideraciones

Tenga en cuenta las siguientes consideraciones cuando utilice un clúster habilitado para roles de tiempo de ejecución con su espacio de trabajo de Amazon EMR Studio:

- Solo puede seleccionar un rol de tiempo de ejecución al asociar un espacio de trabajo de EMR Studio a un clúster de EMR que utilice la versión 6.11 de Amazon EMR o superior.
- La funcionalidad del rol de tiempo de ejecución que se describe en esta página solo es compatible con Amazon EMR en ejecución en Amazon EC2 y no es compatible con las aplicaciones interactivas de EMR sin servidor. Para obtener más información sobre los roles de tiempo de ejecución de EMR sin servidor, consulte Roles de tiempo de ejecución de tareas en la Guía del usuario de Amazon EMR sin servidor.
- Si bien necesita configurar permisos adicionales antes de poder especificar un rol de tiempo de ejecución al enviar un trabajo a un clúster, no necesita permisos adicionales para acceder a los archivos generados por un espacio de trabajo de EMR Studio. Los permisos de estos archivos son los mismos que los de los archivos generados a partir de los clústeres sin roles de tiempo de ejecución.
- No puede usar el Explorador de SQL en un espacio de trabajo de EMR Studio con un clúster que tenga un rol de tiempo de ejecución. Amazon EMR deshabilita el Explorador de SQL en la interfaz de usuario cuando un espacio de trabajo está conectado a un clúster de EMR habilitado para los roles de tiempo de ejecución.
- No puede usar el modo de colaboración en un espacio de trabajo de EMR Studio con un clúster que tenga un rol de tiempo de ejecución. Amazon EMR deshabilita las capacidades de colaboración en el espacio de trabajo cuando un espacio de trabajo está conectado a un clúster de EMR con roles en tiempo de ejecución. Solo podrá acceder al espacio de trabajo el usuario que lo asoció.
- Es posible que aparezca la advertencia “¡Puede que la página no sea segura!” en la interfaz de usuario de Spark para un clúster habilitado para los roles de tiempo de ejecución. En tal caso, omita la alerta para seguir viendo la interfaz de usuario de Spark.

Ejecutar los cuadernos del espacio de trabajo mediante programación

> **Note**
>
> Las aplicaciones interactivas de Amazon EMR sin servidor no admiten la ejecución programática de cuadernos.

Puede ejecutar los cuadernos de sus espacios de trabajo de Amazon EMR Studio mediante programación con un script o en la AWS CLI. Para obtener información sobre cómo ejecutar su cuaderno mediante programación, consulte Ejemplos de comandos para ejecutar Cuadernos de EMR mediante programación.

Examinar los datos con SQL Explorer

> **Note**
>
> El Explorador de SQL para EMR Studio no es compatible con las aplicaciones interactivas de Amazon EMR sin servidor.

En este tema, se proporciona información que lo ayudará a comenzar a trabajar con el Explorador de SQL de Amazon EMR Studio. El Explorador de SQL es una herramienta de una sola página en su espacio de trabajo que le ayuda a comprender los orígenes de datos del catálogo de datos del clúster de EMR. Puede usar el Explorador de SQL para examinar sus datos, ejecutar consultas SQL para recuperar datos y descargar los resultados de las consultas.

El Explorador de SQL es compatible con Presto. Antes de usar el Explorador de SQL, asegúrese de que su clúster use la versión 5.34.0 o posterior o la versión 6.4.0 o posterior de Amazon EMR y tenga instalado Presto. El Explorador de SQL de Amazon EMR Studio no admite los clústeres de Presto que haya configurado con cifrado en tránsito. Esto se debe a que Presto se ejecuta en modo TLS en estos clústeres.

Explorar el catálogo de datos del clúster

El Explorador de SQL proporciona una interfaz de explorador del catálogo que puede utilizar para explorar y comprender cómo se organizan los datos. Por ejemplo, puede usar el explorador del catálogo de datos para comprobar los nombres de las tablas y columnas antes de escribir una consulta SQL.

Para examinar el catálogo de datos

1. Abra el Explorador de SQL en el espacio de trabajo.
2. Asegúrese de que su espacio de trabajo esté asociado a un clúster de EMR que se ejecute en EC2 y que utilice la versión 6.4.0 o posterior de Amazon EMR con Presto instalado. Puede elegir un clúster existente o crear uno nuevo. Para obtener más información, consulte Asociar computación a un espacio de trabajo de EMR Studio.
3. Seleccione una base de datos de la lista desplegable para explorarla.
4. Expanda una tabla de la base de datos para ver los nombres de las columnas de la tabla. También puede introducir una palabra clave en la barra de búsqueda para filtrar los resultados de la tabla.

Ejecutar una consulta SQL para recuperar datos

Para recuperar datos con una consulta SQL y descargar los resultados

1. Abra el Explorador de SQL en el espacio de trabajo.
2. Asegúrese de que su espacio de trabajo esté asociado a un clúster de EMR que se ejecute en EC2 con Presto y Spark instalados. Puede elegir un clúster existente o crear uno nuevo. Para obtener más información, consulte Asociar computación a un espacio de trabajo de EMR Studio.
3. Seleccione Abrir editor para abrir una nueva pestaña de editor en su espacio de trabajo.
4. Redacte la consulta SQL en la pestaña del editor.
5. Elija Run (Ejecutar).
6. Consulte los resultados de la consulta en la vista previa de los resultados. El Explorador de SQL muestra los primeros 100 resultados de forma predeterminada. Puede elegir un número diferente de resultados para mostrarlos (hasta 1000) mediante el menú desplegable Vista previa de los primeros 100 resultados de la consulta.
7. Para descargar los resultados en formato CSV, seleccione Descargar resultados. Puede descargar hasta 1000 filas de resultados.

Asociar computación a un espacio de trabajo de EMR Studio

Amazon EMR Studio ejecuta comandos de cuadernos mediante un kernel en un clúster de EMR. Para poder seleccionar un kernel, debe asociar el espacio de trabajo a un clúster que utilice instancias de Amazon EC2, a un clúster de Amazon EMR en EKS o a una aplicación de EMR sin

servidor. EMR Studio le permite asociar espacios de trabajo a clústeres nuevos o existentes y le brinda la flexibilidad de cambiar los clústeres sin cerrar el espacio de trabajo.

En esta sección se cubren los temas siguientes, que le ayudarán a trabajar con clústeres de EMR Studio y aprovisionarlos:

- Asociar un clúster de Amazon EC2 a un espacio de trabajo de EMR Studio
- Asociar un clúster de Amazon EMR en EKS a un espacio de trabajo de EMR Studio
- Asociar una aplicación de Amazon EMR sin servidor a un espacio de trabajo de EMR Studio
- Crear y asociar un nuevo clúster de EMR a un espacio de trabajo de EMR Studio
- Separar la computación de un espacio de trabajo de EMR Studio

Asociar un clúster de Amazon EC2 a un espacio de trabajo de EMR Studio

Puede asociar un clúster de EMR que se ejecute en Amazon EC2 a un espacio de trabajo al crear el espacio de trabajo o bien asociar un clúster a un espacio de trabajo existente. Si desea crear y asociar un clúster nuevo, consulte Crear y asociar un nuevo clúster de EMR a un espacio de trabajo de EMR Studio.

On create

Asociar un clúster de computación de Amazon EMR al crear un espacio de trabajo

1. En el cuadro de diálogo Crear un espacio de trabajo, asegúrese de haber seleccionado una subred para el nuevo espacio de trabajo. Amplíe la sección Configuración avanzada.
2. En el cuadro de diálogo Crear un espacio de trabajo, asegúrese de haber seleccionado una subred para el nuevo espacio de trabajo. Amplíe la sección Configuración avanzada.
3. Seleccione Asociar espacio de trabajo a un clúster de EMR.
4. En la lista desplegable de clústeres de EMR, seleccione un clúster de EMR existente para asociarlo al espacio de trabajo.

Tras asociar un clúster, termine de crear el espacio de trabajo. Cuando abra el nuevo espacio de trabajo por primera vez y seleccione el panel Clústeres de EMR, debería ver el clúster seleccionado asociado.

On launch

Asociar un clúster de computación de Amazon EMR al iniciar el espacio de trabajo

1. Navegue hasta la lista de espacios de trabajo y seleccione la fila del espacio de trabajo que desee iniciar. A continuación, seleccione Iniciar espacio de trabajo > Iniciar con opciones.
2. Seleccione un clúster de EMR para asociarlo al espacio de trabajo.

Tras asociar un clúster, termine de crear el espacio de trabajo. Cuando abra el nuevo espacio de trabajo por primera vez y seleccione el panel Clústeres de EMR, debería ver el clúster seleccionado asociado.

In JupyterLab

Asocie un espacio de trabajo a un clúster de computación de Amazon EMR en JupyterLab

1. Seleccione su espacio de trabajo y, a continuación, seleccione Iniciar espacio de trabajo > Inicio rápido.
2. En JupyterLab, abra la pestaña Clúster en la barra lateral izquierda.
3. Seleccione el menú desplegable Clúster de EMR en EC2 o seleccione un clúster de Amazon EMR en EKS.
4. Seleccione Asociar para asociar el clúster al espacio de trabajo.

Tras asociar el clúster, termine de crear el espacio de trabajo. Cuando abra el nuevo espacio de trabajo por primera vez y seleccione el panel Clústeres de EMR, debería ver el clúster seleccionado asociado.

In the Workspace UI

Asocie un espacio de trabajo a un clúster de computación de Amazon EMR desde la interfaz de usuario del espacio de trabajo

1. En el espacio de trabajo que desee asociar a un clúster, elija el icono Clústeres de EMR en la barra lateral izquierda para abrir el panel Clúster.
2. En Tipo de clúster, expanda el menú desplegable y seleccione Clúster de EMR en EC2.
3. Seleccione un clúster de la lista desplegable. Es posible que primero tenga que separar un clúster existente para habilitar la lista desplegable de selección del clúster.

4. Elija Attach (Adjuntar). Cuando el clúster esté asociado, debería aparecer un mensaje de confirmación.

Asociar un clúster de Amazon EMR en EKS a un espacio de trabajo de EMR Studio

Además de utilizar clústeres de Amazon EMR que se ejecutan en Amazon EC2, puede asociar un espacio de trabajo a un clúster de Amazon EMR en EKS para ejecutar código de cuadernos. Para obtener más información sobre Amazon EMR en EKS, consulte Qué es Amazon EMR en EKS.

Para poder conectar un espacio de trabajo a un clúster de Amazon EMR en EKS, el administrador de Studio debe concederle permisos de acceso.

On create

Para asociar un clúster de Amazon EMR en EKS al crear un espacio de trabajo

1. En el cuadro de diálogo Crear un espacio de trabajo, amplíe la sección Configuración avanzada.
2. Seleccione Asociar espacio de trabajo a un clúster de Amazon EMR en EKS.
3. En Clúster de Amazon EMR en EKS, elija un clúster de la lista desplegable.
4. En Seleccione un punto de conexión, seleccione un punto de conexión administrado para asociarlo al espacio de trabajo. Un punto de conexión administrado es una puerta de enlace que permite a EMR Studio comunicarse con el clúster elegido.
5. Seleccione Crear un espacio de trabajo para finalizar el proceso de creación del espacio de trabajo y asociar el clúster seleccionado.

Tras asociar un clúster, puede terminar el proceso de creación del espacio de trabajo. Cuando abra el nuevo espacio de trabajo por primera vez y seleccione el panel Clústeres de EMR, debería ver que el clúster seleccionado está asociado.

In the Workspace UI

Para asociar un espacio de trabajo a un clúster de Amazon EMR en EKS desde la interfaz de usuario del espacio de trabajo

1. En el espacio de trabajo que desee asociar a un clúster, elija el icono Clústeres de EMR en la barra lateral izquierda para abrir el panel Clúster.
2. Expanda el menú desplegable Tipo de clúster y seleccione Clústeres de EMR en EKS.

3. En Clúster de EMR en EKS, seleccione un clúster de la lista desplegable.
4. En Punto de conexión, seleccione un punto de conexión administrado para asociarlo al espacio de trabajo. Un punto de conexión administrado es una puerta de enlace que permite a EMR Studio comunicarse con el clúster elegido.
5. Elija Attach (Adjuntar). Cuando el clúster esté asociado, debería aparecer un mensaje de confirmación.

Asociar una aplicación de Amazon EMR sin servidor a un espacio de trabajo de EMR Studio

Puede asociar un espacio de trabajo a una aplicación de EMR sin servidor para ejecutar cargas de trabajo interactivas. Para obtener más información, consulte Uso de cuadernos para ejecutar cargas de trabajo interactivas con EMR sin servidor a través de EMR Studio.

Example Asociar un espacio de trabajo a una aplicación de EMR sin servidor en JupyterLab

Para que pueda conectar un espacio de trabajo a una aplicación de EMR sin servidor, el administrador de su cuenta debe concederle los permisos de acceso, tal y como se describe en Required permissions for interactive workloads.

1. Diríjase a EMR Studio, seleccione su espacio de trabajo y, a continuación, seleccione Iniciar espacio de trabajo > Inicio rápido.
2. En JupyterLab, abra la pestaña Clúster en la barra lateral izquierda.
3. Seleccione EMR sin servidor como opción de computación y, a continuación, seleccione una aplicación de EMR sin servidor y un rol de tiempo de ejecución.
4. Seleccione Asociar para asociar el clúster al espacio de trabajo.

Ahora, cuando abra este espacio de trabajo, debería ver la aplicación seleccionada que asoció.

Crear y asociar un nuevo clúster de EMR a un espacio de trabajo de EMR Studio

Los usuarios avanzados de EMR Studio pueden aprovisionar nuevos clústeres de EMR que se ejecuten en Amazon EC2 para usarlos con un espacio de trabajo. El nuevo clúster tiene todas las aplicaciones de macrodatos necesarias para EMR Studio instaladas de forma predeterminada.

Para crear clústeres, el administrador de Studio primero debe concederle el permiso mediante una política de sesión. Para obtener más información, consulte Creación de políticas de permisos para los usuarios de EMR Studio.

Puede crear un clúster nuevo en el cuadro de diálogo Crear un espacio de trabajo o desde el panel Clúster de la interfaz de usuario del espacio de trabajo. En cualquier caso, tiene dos opciones para crear un clúster:

1. Crear un clúster de EMR: cree un clúster de EMR eligiendo el tipo y el recuento de instancias de Amazon EC2.
2. Utilizar una plantilla de clúster: aprovisione un clúster seleccionando una plantilla de clúster predefinida. Esta opción aparece si tiene permiso para usar plantillas de clúster.

Para crear un clúster de EMR proporcionando una configuración de clúster

1. Elija un punto de inicio.

Para…	Haga lo siguiente...
Cree el clúster al crear un espacio de trabajo con el cuadro de diálogo Crear un espacio de trabajo.	Amplíe la sección Configuración avanzada en el cuadro de diálogo Crear un espacio de trabajo y seleccione Crear un clúster de EMR.
Cree el clúster desde el panel de clústeres de EMR en la interfaz de usuario del espacio de trabajo después de haber creado un espacio de trabajo.	Elija la pestaña Clústeres de EMR en la barra lateral izquierda de un espacio de trabajo abierto, expanda la sección Configura ción avanzada y elija Crear clúster.

2. Ingrese un nombre de clúster. Asignar un nombre al clúster le ayudará a encontrarlo más adelante en la lista de clústeres de EMR Studio.
3. En Versión de Amazon EMR, seleccione una versión de Amazon EMR para el clúster.
4. En Instancia, seleccione el tipo y el número de instancias de Amazon EC2 del clúster. Para obtener más información sobre la selección de los tipos de instancia, consulte Configuración de instancias de Amazon EC2. Se utilizará una instancia como nodo principal.

5. Seleccione una subred en la que EMR Studio pueda lanzar el nuevo clúster. El administrador de Studio aprueba previamente cada opción de subred. Su espacio de trabajo debería poder conectarse a un clúster de cualquier subred de la lista.
6. Seleccione un URI de S3 para el almacenamiento de registros.
7. Seleccione Crear clúster de EMR para aprovisionarlo. Si usa el cuadro de diálogo Crear un espacio de trabajo, seleccione Crear un espacio de trabajo para crear el espacio de trabajo y aprovisionar el clúster. Una vez que EMR Studio aprovisione el nuevo clúster, lo asocia al espacio de trabajo.

Para crear un clúster mediante una plantilla de clúster

1. Elija un punto de inicio.

Para…	Haga lo siguiente...
Cree el clúster al crear un espacio de trabajo con el cuadro de diálogo Crear un espacio de trabajo.	Expanda la sección Configuración avanzada en el cuadro de diálogo Crear un espacio de trabajo y seleccione Usar una plantilla de clúster.
Cree el clúster desde el panel de clústeres de EMR en la interfaz de usuario del espacio de trabajo.	Seleccione la pestaña Clústeres de EMR en la barra lateral izquierda de un espacio de trabajo abierto, expanda la sección Configuración avanzada y seleccione Plantilla de clúster.

2. En la lista desplegable, seleccione una plantilla de clúster. Cada plantilla de clúster disponible incluye una breve descripción para ayudarle a realizar una selección.
3. La plantilla de clúster que elija puede tener parámetros adicionales, como la versión de lanzamiento de Amazon EMR o el nombre del clúster. Puede elegir o insertar valores, o bien utilizar los valores predeterminados que haya seleccionado el administrador.
4. Seleccione una subred en la que EMR Studio pueda lanzar el nuevo clúster. El administrador de Studio aprueba previamente cada opción de subred. Su espacio de trabajo debería poder conectarse a un clúster de cualquier subred.
5. Seleccione Usar plantilla de clúster para aprovisionar el clúster y asociarlo al espacio de trabajo. EMR Studio tardará unos minutos en crear el clúster. Si usa el cuadro de diálogo Crear un

espacio de trabajo, seleccione Crear un espacio de trabajo para crear el espacio de trabajo y aprovisionar el clúster. Una vez que EMR Studio aprovisione el nuevo clúster, lo asocia al espacio de trabajo.

Separar la computación de un espacio de trabajo de EMR Studio

Para intercambiar el clúster asociado a un espacio de trabajo, puede separar un clúster de la interfaz de usuario del espacio de trabajo.

Para separar un clúster de un espacio de trabajo

1. En el espacio de trabajo que desee separar de un clúster, elija el icono Clústeres de EMR en la barra lateral izquierda para abrir el panel Clúster.
2. En Seleccionar clúster, seleccione Separar y espere a que EMR Studio separe el clúster. Cuando se separe el clúster, verá un mensaje indicándole que la operación se ha realizado correctamente.

Para separar una aplicación de EMR sin servidor de un espacio de trabajo de EMR Studio

Para intercambiar la computación asociada a un espacio de trabajo, puede separar la aplicación de la interfaz de usuario del espacio de trabajo.

1. En el espacio de trabajo que desee separar de un clúster, elija el icono Computación de Amazon EMR en la barra lateral izquierda para abrir el panel Computación.
2. En Seleccionar computación, seleccione Separar y espere a que EMR Studio separe la aplicación. Cuando se separe la aplicación, verá un mensaje indicándole que la operación se ha realizado correctamente.

Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio

Acerca de los repositorios de Git para EMR Studio

Puede asociar un máximo de tres repositorios de Git a un espacio de trabajo de EMR Studio. De forma predeterminada, cada espacio de trabajo le permite elegir en una lista de repositorios de Git que están asociados a la misma cuenta de AWS que el estudio. También puede crear un nuevo repositorio de Git como recurso para un espacio de trabajo.

Puede ejecutar comandos de Git como los siguientes mediante un comando de terminal mientras está conectado al nodo principal de un clúster.

```
!git pull origin <branch-name>
```

También puede utilizar la extensión jupyterlab-git. Ábralo desde la barra lateral izquierda seleccionando el icono de Git. Para obtener información sobre la extensión jupyterlab-git de JupyterLab, consulte jupyterlab-git.

Requisitos previos

- Para asociar un repositorio de Git a un espacio de trabajo, el estudio debe estar configurado para permitir la vinculación de repositorios de Git. El administrador de Studio debe tomar medidas para Establecer el acceso y los permisos para los repositorios basados en Git.
- Si usa un repositorio de CodeCommit, debe usar credenciales de Git y HTTPS. No se admiten las claves SSH ni HTTPS con el ayudante de credenciales de AWS Command Line Interface. CodeCommit tampoco admite los tokens de acceso personal (PAT). Para obtener más información, consulte Uso de IAM con CodeCommit en la Guía del usuario de IAM y Configuración para usuarios de HTTPS con credenciales de Git en la Guía del usuario de AWS CodeCommit.

Instrucciones

Para vincular un repositorio de Git asociado a un espacio de trabajo

1. Abra el espacio de trabajo que quiera vincular a un repositorio desde la lista Espacios de trabajo del estudio.
2. En la barra lateral izquierda, seleccione el icono del repositorio de Git de Amazon EMR para abrir el panel de herramientas del repositorio de Git.
3. En Repositorios de Git, expanda la lista desplegable y seleccione un máximo de tres repositorios para vincularlos al espacio de trabajo. EMR Studio registra su selección y comienza a vincular cada repositorio.

Es posible que el proceso de vinculación tarde algún tiempo en completarse. Puede ver el estado de cada repositorio que haya seleccionado en el panel de herramientas del repositorio de Git. Después de que EMR Studio vincule un repositorio a un espacio de trabajo, debería ver los archivos que pertenecen a ese repositorio en el panel del Explorador de archivos.

Para agregar un nuevo repositorio de Git a un espacio de trabajo como recurso

1. Abra el espacio de trabajo que quiera vincular a un repositorio desde la lista de espacios de trabajo del estudio.
2. En la barra lateral izquierda, seleccione el icono del repositorio de Git de Amazon EMR para abrir el panel de herramientas del repositorio de Git.
3. Seleccione Agregar nuevo repositorio de Git.
4. En Nombre del repositorio, escriba un nombre descriptivo para el repositorio de EMR Studio. Los nombres pueden contener caracteres alfanuméricos, guiones o guiones bajos.
5. En Git repository URL (URL del repositorio de Git), escriba la URL del repositorio. Cuando utiliza un repositorio de CodeCommit, esta es la URL que se copia cuando selecciona Clonar URL y, a continuación, Clonar HTTPS. Por ejemplo, `https://git-codecommit.us-west-2.amazonaws.com/v1/repos/[MyCodeCommitRepoName]`.
6. En Ramificación, introduzca el nombre de una ramificación existente que quiera seleccionar.
7. En Credenciales de Git, seleccione una opción de acuerdo con las siguientes pautas. EMR Studio accede a sus credenciales de Git mediante los secretos almacenados en Secrets Manager.

> **Note**
>
> Si utiliza un repositorio de GitHub, le recomendamos que utilice un token de acceso personal (PAT) para autenticarse. A partir del 13 de agosto de 2021, GitHub requerirá la autenticación basada en tokens y ya no aceptará contraseñas al autenticar las operaciones de Git. Para obtener más información, consulta la publicación sobre los requisitos de autenticación con token para las operaciones de Git en The GitHub Blog.

Opción	Descripción
Crear un nuevo secreto	Seleccione esta opción para asociar las credenciales de Git existentes a un nuevo secreto que se creará en AWS Secrets Manager para usted. Realice una de las siguientes acciones basadas en las

Opción	Descripción
	credenciales de Git que utilice para el repositorio. Si usa un nombre de usuario y una contraseña de Git para acceder al repositorio, seleccione Nombre de usuario y contraseña, escriba el nombre del secreto que va a usar en Secrets Manager y, a continuación, escriba el nombre de usuario y la contraseña que asociar al secreto. - O BIEN - Si utiliza un token de acceso personal para acceder al repositorio, seleccione Token de acceso personal (PAT), escriba el nombre del secreto que usará en Secrets Manager y, a continuación, escriba su token de acceso personal. Para obtener más información, consulte Creating a personal access token for the command line for GitHub yPersonal access tokens for Bitbucket. Los repositorios de CodeCommit no admiten esta opción.
Usar un repositorio público sin credenciales	Elija esta opción para acceder a un repositorio público.

Opción	Descripción
Usar un secreto de AWS existente	Elija esta opción si ya ha guardado sus credenciales como un secreto en Secrets Manager y, a continuación, seleccione el nombre del secreto en la lista. Si selecciona un secreto asociado con un nombre de usuario y contraseña de Git, el secreto debe estar en el formato `{"gitUsername": " MyUserName ", "gitPassword": " MyPassword "}`.

8. Seleccione Agregar repositorio para crear el nuevo repositorio. Después de que EMR Studio cree el nuevo repositorio, verá un mensaje de confirmación. El nuevo repositorio aparece en la lista desplegable de Repositorios de Git.
9. Para vincular el nuevo repositorio a su espacio de trabajo, selecciónelo en la lista desplegable de Repositorios de Git.

Es posible que el proceso de vinculación tarde algún tiempo en completarse. Después de que EMR Studio vincule el nuevo repositorio al espacio de trabajo, debería aparecer una nueva carpeta con el mismo nombre que el repositorio en el panel del Explorador de archivos.

Para abrir un repositorio vinculado diferente, navegue hasta su carpeta en el Explorador de archivos.

Depurar aplicaciones y trabajos con EMR Studio

Con Amazon EMR Studio, puede lanzar interfaces de aplicaciones de datos para analizar las aplicaciones y las ejecuciones de trabajos en el navegador.

También puede lanzar las interfaces de usuario persistentes y fuera del clúster para Amazon EMR que se ejecutan en clústeres de EC2 desde la consola de Amazon EMR. Para obtener más información, consulte Ver interfaces de usuario de aplicaciones persistentes.

> **Note**
>
> En función de la configuración del navegador, es posible que necesite habilitar las ventanas emergentes para que se abra la interfaz de usuario de una aplicación.

Para obtener información sobre la configuración y el uso de las interfaces de la aplicación, consulte The YARN Timeline Server, Monitoring and Instrumentation o Tez UI Overview.

Depurar Amazon EMR que se ejecuta en trabajos de Amazon EC2

Workspace UI

Iniciar una interfaz de usuario en el clúster a partir de un archivo de cuaderno

Si utiliza las versiones 5.33.0 y posteriores de Amazon EMR, puede iniciar la interfaz de usuario web de Spark (la interfaz de usuario de Spark o el servidor de historial de Spark) desde un cuaderno en el espacio de trabajo.

Las interfaces de usuario integradas en el clúster funcionan con los kernels de PySpark, Spark o SparkR. El tamaño máximo de archivo visible para los registros de eventos o contenedores de Spark es de 10 MB. Si sus archivos de registro superan los 10 MB, le recomendamos que utilice el servidor de historial de Spark persistente en lugar de la interfaz de usuario de Spark integrada en el clúster para depurar los trabajos.

> **Important**
>
> Para que EMR Studio pueda lanzar interfaces de usuario de aplicaciones en un clúster desde un espacio de trabajo, un clúster debe poder comunicarse con Amazon API Gateway. Debe configurar el clúster de EMR para permitir el tráfico de red saliente a Amazon API Gateway y asegurarse de que se pueda acceder a Amazon API Gateway desde el clúster.
> La interfaz de usuario de Spark accede a los registros del contenedor resolviendo los nombres de host. Si utiliza un nombre de dominio personalizado, debe asegurarse de que Amazon DNS o el servidor DNS que especifique puedan resolver los nombres de host de los nodos de su clúster. Para ello, defina las opciones del protocolo de configuración dinámica de host (DHCP) para la Amazon Virtual Private Cloud (VPC) asociada al clúster. Para obtener más información sobre las opciones de DHCP, consulte los conjuntos de opciones de DHCP en la Guía del usuario de Amazon Virtual Private Cloud.

1. En su EMR Studio, abra el espacio de trabajo que desee usar y asegúrese de que esté conectado a un clúster de Amazon EMR que se ejecute en EC2. Para obtener instrucciones, consulte Asociar computación a un espacio de trabajo de EMR Studio.

2. Abra un archivo de cuaderno y use el kernel de PySpark, Spark o SparkR. Para seleccionar un kernel, seleccione su nombre en la parte superior derecha de la barra de herramientas del cuaderno para abrir el cuadro de diálogo Seleccionar kernel. El nombre aparece como No hay ningún kernel si no se ha seleccionado ningún kernel.
3. Ejecute el código de su cuaderno. El fragmento siguiente aparece como salida en el cuaderno al iniciar el contexto de Spark. Puede que tarde unos segundos en aparecer. Si ha iniciado el contexto de Spark, puede ejecutar el comando `%%info` para acceder a un enlace que le llevará a la interfaz de usuario de Spark en cualquier momento.

> **Note**
>
> Si los enlaces de la interfaz de usuario de Spark no funcionan o no aparecen después de unos segundos, cree una nueva celda de cuaderno y ejecute el comando `%%info` para regenerar los enlaces.

4. Para iniciar la interfaz de usuario de Spark, seleccione Vincular en la interfaz de usuario de Spark. Si la aplicación de Spark se está ejecutando, la interfaz de usuario de Spark se abrirá en una pestaña nueva. Si la aplicación se ha completado, en su lugar se abrirá el servidor de historial de Spark.

   Tras iniciar la interfaz de usuario de Spark, puede modificar la URL en el navegador para abrir el YARN ResourceManager o el servidor YARN Timeline. Agregue una de las siguientes rutas después de `amazonaws.com`.

Interfaz de usuario web	Ruta	Ejemplo de URL modificada
YARN ResourceManager	/rm	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*rm*
Yarn Timeline Server	/yts	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*yts*
Servidor de historial de Spark	/shs	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*shs*

Studio UI

Iniciar el servidor YARN Timeline, el servidor de historial de Spark o la interfaz de usuario de Tez persistentes desde la interfaz de usuario de EMR Studio

1. En su EMR Studio, seleccione Amazon EMR en EC2 en la parte izquierda de la página para abrir la lista de clústeres de Amazon EMR en EC2.
2. Filtre la lista de clústeres por nombre, estado o identificador introduciendo valores en el cuadro de búsqueda. También puede buscar por intervalo de tiempo de creación.
3. Seleccione un clúster y, a continuación, seleccione Iniciar las interfaces de usuario de la aplicación para seleccionar una interfaz de usuario de la aplicación. La interfaz de usuario de la aplicación se abre en una nueva pestaña del navegador y puede que tarde en cargarse.

Depurar un EMR Studio en ejecución en EMR sin servidor

Al igual que Amazon EMR que se ejecuta en Amazon EC2, puede utilizar la interfaz de usuario del espacio de trabajo para analizar sus aplicaciones de EMR sin servidor. En la interfaz de usuario del espacio de trabajo, si utiliza las versiones 6.14.0 y posteriores de Amazon EMR, puede iniciar la interfaz de usuario web de Spark (la interfaz de usuario de Spark o el servidor de historial de Spark) desde un cuaderno en el espacio de trabajo. Para su comodidad, también ofrecemos un enlace para acceder rápidamente a los registros de controladores de Spark.

Depurar Amazon EMR en ejecuciones de trabajos de EKS con el servidor de historial de Spark

Cuando envía una ejecución de trabajo a un clúster de Amazon EMR en EKS, puede acceder a los registros de esa ejecución de trabajos mediante el servidor de historial de Spark. El servidor de historial de Spark proporciona herramientas para monitorear las aplicaciones de Spark, como una lista de las etapas y tareas del programador, un resumen del tamaño de los RDD y el uso de memoria, e información sobre el entorno. Puede iniciar el servidor de historial de Spark para las ejecuciones de trabajos de Amazon EMR en EKS de las siguientes maneras:

- Cuando envíe una ejecución de trabajo con EMR Studio con un punto de conexión administrado de Amazon EMR en EKS, puede lanzar el servidor de historial de Spark desde un archivo de cuaderno de su espacio de trabajo.
- Cuando envíe una ejecución de trabajo con la AWS CLI o el AWS SDK de Amazon EMR en EKS, puede lanzar el servidor de historial de Spark desde la interfaz de usuario de EMR Studio.

Para obtener información sobre cómo utilizar el servidor de historial de Spark, consulte Supervisión e instrumentación en la documentación de Apache Spark. Para obtener más información sobre las ejecuciones de trabajos, consulte Conceptos y componentes en la Guía de desarrollo de Amazon EMR en EKS.

Para iniciar el servidor de historial de Spark desde un archivo de cuaderno en su espacio de trabajo de EMR Studio

1. Abra un espacio de trabajo que esté conectado a un clúster de Amazon EMR en EKS.
2. Seleccione y abra el archivo de su cuaderno en el espacio de trabajo.
3. Elija la interfaz de usuario de Spark en la parte superior del archivo del cuaderno para abrir el servidor de historial de Spark persistente en una pestaña nueva.

Para iniciar el servidor de historial de Spark desde la interfaz de usuario de EMR Studio

> **Note**
>
> La lista Trabajos de la interfaz de usuario de EMR Studio muestra solo las ejecuciones de trabajos que envíe mediante el uso de la AWS CLI o el AWS SDK para Amazon EMR en EKS.

1. En su EMR Studio, seleccione Amazon EMR en EKS en la parte izquierda de la página.
2. Busque el clúster virtual de Amazon EMR en EKS que utilizó para enviar la ejecución de su trabajo. Puede filtrar la lista de clústeres por estado o identificador introduciendo valores en el cuadro de búsqueda.
3. Seleccione el clúster para abrir su página de detalles. En la página de detalles se muestra información sobre el clúster, como el identificador, el espacio de nombres y el estado. En la página también se muestra una lista de todas las ejecuciones de trabajos enviadas a ese clúster.
4. En la página de detalles del clúster, seleccione una ejecución de trabajos para depurarla.
5. En la parte superior derecha de la lista Trabajos, seleccione Iniciar servidor de historial de Spark para abrir la interfaz de la aplicación en una nueva pestaña del navegador.

Instalar kernels y bibliotecas en un espacio de trabajo de EMR Studio

Cada espacio de trabajo de Amazon EMR Studio viene con un conjunto de bibliotecas y kernels preinstalados.

Kernels y bibliotecas de los clústeres que se ejecutan en Amazon EC2

También puede personalizar el entorno de EMR Studio de las siguientes maneras cuando utilice clústeres de EMR que se ejecuten en Amazon EC2:

- Instale los kernels de cuadernos de Jupyter y las bibliotecas de Python en el nodo principal de un clúster: al instalar bibliotecas con esta opción, todos los espacios de trabajo asociados al mismo clúster comparten esas bibliotecas. Puede instalar kernels o bibliotecas desde una celda del cuaderno o mientras está conectado mediante SSH al nodo principal de un clúster.
- Usar bibliotecas para cuadernos: cuando los usuarios del espacio de trabajo instalan y usan bibliotecas desde la celda de un cuaderno, esas bibliotecas solo están disponibles para ese cuaderno. Esta opción permite que diferentes cuadernos que utilizan el mismo clúster funcionen sin tener que preocuparse por los conflictos en las versiones de las bibliotecas.

Los espacios de trabajo de EMR Studio tienen la misma arquitectura subyacente que Cuadernos de EMR. Puede instalar y utilizar los kernels de cuadernos de Jupyter y las bibliotecas de Python con EMR Studio del mismo modo que lo haría con Cuadernos de EMR. Para obtener instrucciones, consulte Instalación y uso de kernels y bibliotecas.

Kernels y bibliotecas de los clústeres de Amazon EMR en EKS

Los clústeres de Amazon EMR en EKS incluyen los kernels de PySpark y Python 3.7 con un conjunto de bibliotecas preinstaladas. Amazon EMR en EKS no admite la instalación de bibliotecas o clústeres adicionales.

Cada clúster de Amazon EMR en EKS viene con las siguientes bibliotecas de Python y PySpark instaladas:

- Python – boto3, cffi, future, ggplot, jupyter, kubernetes, matplotlib, numpy, pandas, plotly, pycryptodomex, py4j, requests, scikit-learn, scipy, seaborn
- PySpark – ggplot, jupyter, matplotlib, numpy, pandas, plotly, pycryptodomex, py4j, requests, scikit-learn, scipy, seaborn

Kernels y bibliotecas en aplicaciones de EMR sin servidor

Cada aplicación de EMR sin servidor viene con las siguientes bibliotecas de Python y PySpark instaladas:

- Python – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn
- PySpark – ggplot, matplotlib,numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn

Mejorar los kernels con comandos magic

Información general

EMR Studio y Cuadernos de EMR admiten comandos magic. Los comandos Magic o magics son mejoras proporcionadas por el kernel de IPython para ayudar a ejecutar y analizar datos. IPython es un entorno de intérprete de comandos interactivo creado con Python.

Amazon EMR también es compatible con Sparkmagic, un paquete que proporciona kernels relacionados con Spark (kernels de PySpark, SparkR y Scala) con comandos magic específicos y que utiliza Livy en el clúster para enviar los trabajos de Spark.

Puede usar comandos magic siempre que tenga un kernel de Python en su cuaderno de EMR. Del mismo modo, cualquier kernel relacionado con Spark admite los comandos de Sparkmagic.

Los comandos Magic, también llamados magics, vienen en dos variedades:

- magics de línea: estos comandos magic se indican con un único `%` como prefijo y funcionan en una sola línea de código.
- magics de celda: estos comandos magic se indican con un doble `%%` como prefijo y funcionan en varias líneas de código.

Para ver todos los comandos magic disponibles, consulte Enumerar los comandos magic y Sparkmagic.

Condiciones y limitaciones

- EMR sin servidor no admite `%%sh` para ejecutar `spark-submit`. No es compatible con magic de Cuadernos de EMR.
- Los clústeres de Amazon EMR en EKS no admiten los comandos de Sparkmagic para EMR Studio. Esto se debe a que los kernels de Spark que se utilizan con los puntos de conexión administrados están integrados en Kubernetes, y Sparkmagic y Livy no los admiten. Puede establecer la configuración de Spark directamente en el objeto SparkContext como solución alternativa, como se muestra en el siguiente ejemplo.

```
spark.conf.set("spark.driver.maxResultSize", '6g')
```

- Las siguientes acciones y comandos magic están prohibidos por AWS:
  - `%alias`
  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - Modificar `proxy_user` con `%configure`
  - Modificar `KERNEL_USERNAME` con `%env` o `%set_env`

Enumerar los comandos magic y Sparkmagic

Utilice los siguientes comandos para enumerar los comandos magic disponibles:

- `%lsmagic` muestra una lista de todas las funciones magic disponibles actualmente.
- `%%help` muestra las funciones magic relacionadas con Spark que están disponibles actualmente en el paquete de Sparkmagic.

Usar `%%configure` para configurar Spark

Uno de los comandos más útiles de Sparkmagic es el comando `%%configure`, que configura los parámetros de creación de sesiones. En cuanto a los ajustes de `conf`, puede establecer cualquier configuración de Spark que se mencione en la documentación de configuración de Apache Spark.

Example Agregar un archivo JAR externo a los Cuadernos de EMR desde el repositorio de Maven o Amazon S3

Puede usar el siguiente enfoque para agregar una dependencia de un archivo JAR externo a cualquier kernel relacionado con Spark compatible con Sparkmagic.

```
%%configure -f
{"conf": {
    "spark.jars.packages": "com.jsuereth:scala-arm_2.11:2.0,ml.combust.bundle:bundle-
ml_2.11:0.13.0,com.databricks:dbutils-api_2.11:0.0.3",
    "spark.jars": "s3://DOC-EXAMPLE-BUCKET/my-jar.jar"
    }
}
```

Example : Configurar Hudi

Puede utilizar el editor de cuadernos para configurar el cuaderno de EMR de modo que utilice Hudi.

```
%%configure
{ "conf": {
     "spark.jars": "hdfs://apps/hudi/lib/hudi-spark-bundle.jar,hdfs:///apps/hudi/lib/
spark-spark-avro.jar",
     "spark.serializer": "org.apache.spark.serializer.KryoSerializer",
     "spark.sql.hive.convertMetastoreParquet":"false"
     }
}
```

Usar `%%sh` para ejecutar `spark-submit`

La magic `%%sh` ejecuta comandos del intérprete de comandos en un subproceso en una instancia del clúster asociado. Normalmente, utilizaría uno de los kernels relacionados con Spark para ejecutar las aplicaciones de Spark en su clúster asociado. Sin embargo, si quiere usar un kernel de Python para enviar una aplicación de Spark, puede usar la siguiente magic y reemplazar el nombre del bucket por el nombre de su bucket en minúsculas.

```
%%sh
spark-submit --master yarn --deploy-mode cluster s3://DOC-EXAMPLE-BUCKET/test.py
```

En este ejemplo, el clúster tiene que acceder a la ubicación de `s3://DOC-EXAMPLE-BUCKET/test.py` o se producirá un error en el comando.

Con la magic `%%sh`, puede usar cualquier comando de Linux. Si desea ejecutar algún comando de Spark o YARN, use una de las siguientes opciones para crear un usuario de Hadoop de `emr-notebook` y concederle permisos para ejecutar los comandos:

- Para crear un usuario nuevo de forma explícita, ejecute los siguientes comandos.

  ```
  hadoop fs -mkdir /user/emr-notebook
  hadoop fs -chown emr-notebook /user/emr-notebook
  ```

- Puede activar la suplantación de usuarios en Livy, lo que crea el usuario automáticamente. Para obtener más información, consulte Habilitar la suplantación de usuario para supervisar la actividad del usuario y del trabajo de Spark.

Usar `%%display` para visualizar los marcos de datos de Spark

Puede usar magic `%%display` para ver un marco de datos de Spark. Para usar esta magic, ejecute el siguiente comando.

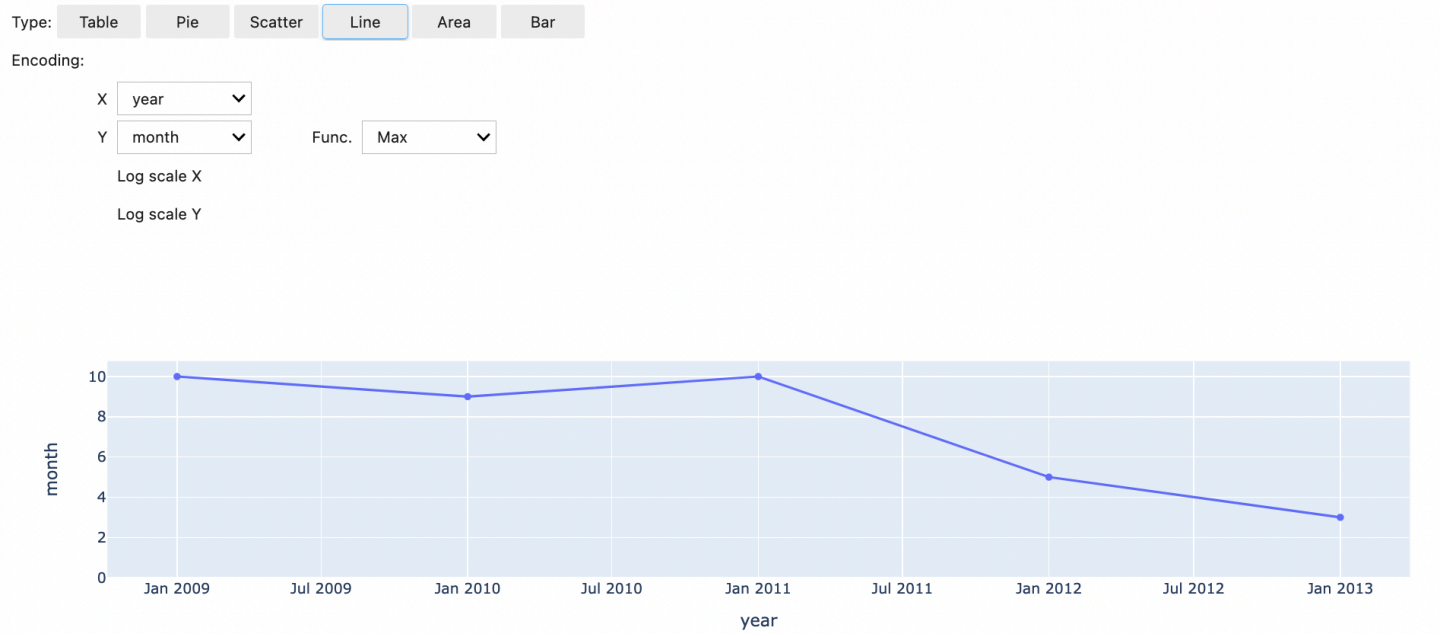
```
%%display df
```

Elija ver los resultados en formato de tabla, como se muestra en la siguiente imagen.

Type: Table Pie Scatter Line Area Bar

	year	month	total_passengers	total_trips
2012-01-01	3	26866837	16146923
2011-01-01	3	26091246	16066350
2013-01-01	3	26965079	15749228
2011-01-01	10	26287953	15707756
2009-01-01	10	26202049	15604551
2012-01-01	5	26278817	15567525
2011-01-01	5	25508952	15554868
2010-01-01	9	25533166	15540209
2010-01-01	5	26002858	15481351
2012-01-01	4	25900645	15477914

También puede optar por visualizar sus datos con cinco tipos de gráficos. Las opciones son gráficos circulares, de dispersión, de líneas, de áreas y de barras.



Usar la magic de Cuadernos de EMR

Amazon EMR proporciona la siguiente magic de Cuadernos de EMR que puede utilizar con los kernels basados en Python3 y Spark:

- `%mount_workspace_dir`: monta el directorio del espacio de trabajo en el clúster para que pueda importar y ejecutar código desde otros archivos de su espacio de trabajo

> **Note**
>
> Con `%mount_workspace_dir`, solo el kernel de Python 3 puede acceder a sus sistemas de archivos locales. Los ejecutores de Spark no tendrán acceso al directorio montado con este kernel.

- `%umount_workspace_dir`: desmonta el directorio del espacio de trabajo del clúster
- `%generate_s3_download_url`: genera un enlace de descarga temporal en la salida del cuaderno para un objeto de Amazon S3

Requisitos previos

Antes de instalar la magic de Cuadernos de EMR, complete las siguientes tareas:

- Asegúrese de que Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2) tenga acceso de lectura a Amazon S3. `EMR_EC2_DefaultRole` con la política administrada `AmazonElasticMapReduceforEC2Role` cumple con este requisito. Si usa un rol o una política personalizados, asegúrese de que tenga los permisos de S3 necesarios.

> **Note**
>
> La magic de Cuadernos de EMR se ejecuta en un clúster como usuario del cuaderno y utiliza el perfil de instancia EC2 para interactuar con Amazon S3. Al montar un directorio del espacio de trabajo en un clúster de EMR, todos los espacios de trabajo y cuadernos de EMR con permiso para asociarse a ese clúster pueden acceder al directorio montado. De forma predeterminada, los directorios se montan como de solo lectura. Si bien `s3fs-fuse` y `goofys` permiten los montajes de lectura y escritura, le recomendamos especialmente que no modifique los parámetros de montaje para montar directorios en modo de lectura y escritura. Si permite el acceso de escritura, todos los cambios que se realicen en el directorio se escribirán en el bucket de S3. Para evitar que se eliminen o

sobrescriban accidentalmente, puede habilitar el control de versiones en su bucket de S3. Para obtener más información, consulte Uso del control de versiones en buckets de S3.

- Ejecute uno de los siguientes scripts en su clúster para instalar las dependencias de la magic de Cuadernos de EMR. Para ejecutar un script, puede Usar acciones de arranque personalizadas o seguir las instrucciones de Ejecutar comandos y scripts en un clúster de Amazon EMR cuando ya tenga un clúster en ejecución.

  Puede elegir qué dependencia desea instalar. Tanto s3fs-fuse como goofys son herramientas FUSE (sistema de archivos en el espacio de usuario) que permiten montar un bucket de Amazon S3 como un sistema de archivos local en un clúster. La herramienta `s3fs` proporciona una experiencia similar a POSIX. La herramienta `goofys` es una buena opción si prefiere el rendimiento en lugar de un sistema de archivos compatible con POSIX.

```
#!/bin/sh

# Install the s3fs dependency for EMR Notebooks magics sudo amazon-linux-extras
 install epel -y
sudo yum install s3fs-fuse -y
```

  O BIEN

```
#!/bin/sh

# Install the goofys dependency for EMR Notebooks magics sudo wget https://
github.com/kahing/goofys/releases/latest/download/goofys -P /usr/bin/
sudo chmod ugo+x /usr/bin/goofys
```

Instalar la magic de Cuadernos de EMR

> **Note**
>
> Con las versiones de 6.0 a 6.9.0 y de 5.0 a 5.36.0 de Amazon EMR, solo las versiones 0.2.0 y superiores del paquete `emr-notebooks-magics` admiten la magic `%mount_workspace_dir`.

Siga los pasos que se indican a continuación para instalar la magic de Cuadernos de EMR.

1. En su cuaderno, ejecute los siguientes comandos para instalar el paquete emr-notebooks-magics.

```
%pip install boto3 --upgrade
%pip install botocore --upgrade
%pip install emr-notebooks-magics --upgrade
```

2. Reinicie el kernel para cargar la magic de Cuadernos de EMR.
3. Verifique la instalación con el siguiente comando, que debería mostrar el texto de ayuda de la salida de `%mount_workspace_dir`.

```
%mount_workspace_dir?
```

Montar un directorio del espacio de trabajo con **`%mount_workspace_dir`**

La magic `%mount_workspace_dir` le permite montar el directorio del espacio de trabajo en su clúster de EMR para que pueda importar y ejecutar otros archivos, módulos o paquetes almacenados en su directorio.

En el siguiente ejemplo, se monta todo el directorio del espacio de trabajo en un clúster y se especifica el argumento opcional *<--fuse-type>* para usar goofys para montar el directorio.

```
%mount_workspace_dir . <--fuse-type goofys>
```

Para comprobar que el directorio del espacio de trabajo esté montado, utilice el siguiente ejemplo para mostrar el directorio de trabajo actual con el comando `ls`. La salida debe mostrar todos los archivos del espacio de trabajo.

```
%%sh
ls
```

Cuando termine de realizar cambios en su espacio de trabajo, puede desmontar el directorio del espacio de trabajo con el siguiente comando:

> **Note**
>
> El directorio del espacio de trabajo permanece montado en el clúster incluso cuando el espacio de trabajo está detenido o separado. Debe desmontar el directorio del espacio de trabajo de forma explícita.

```
%umount_workspace_dir
```

Descargar un objeto de Amazon S3 con **`%generate_s3_download_url`**

El comando `generate_s3_download_url`, crea una URL prefirmada para un objeto almacenado en Amazon S3. Puede usar la URL prefirmada para descargar el objeto a su máquina local. Por ejemplo, puede ejecutar `generate_s3_download_url` para descargar el resultado de una consulta SQL que su código escribe en Amazon S3.

De forma predeterminada, la URL prefirmada es válida durante 60 minutos. Puede cambiar el tiempo de caducidad especificando un número de segundos para la marca `--expires-in`. Por ejemplo, `--expires-in 1800` crea una URL que es válida durante 30 minutos.

El siguiente ejemplo genera un enlace de descarga para un objeto especificando la ruta completa de Amazon S3: *`s3://EXAMPLE-DOC-BUCKET/path/to/my/object`*.

```
%generate_s3_download_url s3://EXAMPLE-DOC-BUCKET/path/to/my/object
```

Para obtener más información sobre el uso de `generate_s3_download_url`, ejecute el siguiente comando para mostrar el texto de ayuda.

```
%generate_s3_download_url?
```

Ejecutar un cuaderno en el modo sin encabezado con **`%execute_notebook`**

Con la magic `%execute_notebook`, puede ejecutar otro cuaderno en el modo sin encabezado y ver el resultado de cada celda que haya ejecutado. Esta magic requiere permisos adicionales para el rol de instancia que comparten Amazon EMR y Amazon EC2. Para obtener más información sobre cómo conceder permisos adicionales, ejecute el comando `%execute_notebook?`.

Durante un trabajo de larga duración, es posible que el sistema entre en modo de suspensión debido a la inactividad o que pierda temporalmente la conectividad a internet. Esto podría interrumpir la

conexión entre su navegador y el servidor de Jupyter. En este caso, podría perder la salida de las celdas que ha ejecutado y enviado desde el servidor de Jupyter.

Si ejecuta el cuaderno en el modo sin encabezado con la magic `%execute_notebook`, Cuadernos de EMR captura el resultado de las celdas que se han ejecutado, incluso si la red local sufre una interrupción. Cuadernos de EMR guarda la salida de forma incremental en un nuevo cuaderno con el mismo nombre que el cuaderno que ha utilizado. A continuación, Cuadernos de EMR coloca el cuaderno en una nueva carpeta dentro del espacio de trabajo. Las ejecuciones sin encabezado se realizan en el mismo clúster y utilizan el rol de servicio `EMR_Notebook_DefaultRole`, pero los argumentos adicionales pueden alterar los valores predeterminados.

Para ejecutar un cuaderno en el modo sin encabezado, utilice el siguiente comando:

```
%execute_notebook <relative-file-path>
```

Para especificar un ID de clúster y un rol de servicio para una ejecución sin encabezado, use el siguiente comando:

```
%execute_notebook <notebook_name>.ipynb --cluster-id <emr-cluster-id> --service-role
 <emr-notebook-service-role>
```

Cuando Amazon EMR y Amazon EC2 comparten un rol de instancia, el rol requiere los siguientes permisos adicionales:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:StartNotebookExecution",
                "elasticmapreduce:DescribeNotebookExecution",
                "ec2:DescribeInstances"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
```

```
            ],
            "Resource": "arn:aws:iam::<AccoundId>:role/EMR_Notebooks_DefaultRole"
        }
    ]
}
```

> **Note**
>
> Para usar magic `%execute_notebook`, instale la versión 0.2.3 o una superior del paquete `emr-notebooks-magics`.

Usar cuadernos multilingües con kernels de Spark

Cada kernel de cuadernos de Jupyter tiene un lenguaje predeterminado. Por ejemplo, el lenguaje predeterminado del kernel de Spark es Scala, y el lenguaje predeterminado de los kernels de PySpark es Python. Con Amazon EMR 6.4.0 y las versiones posteriores, EMR Studio admite los cuadernos de varios lenguajes. Esto significa que cada kernel de EMR Studio puede admitir los siguientes lenguajes además del lenguaje predeterminado: Python, Spark, R y Spark SQL.

Para activar esta característica, especifique uno de los siguientes comandos magic al principio de cualquier celda.

Lenguaje	Comando
Python	`%%pyspark`
Scala	`%%scalaspark`
R	`%%rspark` No se admite para cargas de trabajo interactivas con EMR sin servidor.
Spark SQL	`%%sql`

Cuando se invocan, estos comandos ejecutan toda la celda dentro de la misma sesión de Spark utilizando el intérprete del lenguaje correspondiente.

La magic de celda `%%pyspark` permite a los usuarios escribir código de PySpark en todos los kernels de Spark.

```
%%pyspark
a = 1
```

La magic de celda `%%sql` permite a los usuarios ejecutar código Spark-SQL en todos los kernels de Spark.

```
%%sql
SHOW TABLES
```

La magic de celda `%%rspark` permite a los usuarios ejecutar código SparkR en todos los kernels de Spark.

```
%%rspark
a <- 1
```

La magic de celda `%%scalaspark` permite a los usuarios ejecutar código Spark Scala en todos los kernels de Spark.

```
%%scalaspark
val a = 1
```

Compartir datos entre intérpretes de lenguajes con tablas temporales

También puede compartir datos entre intérpretes de lenguajes mediante tablas temporales. En el siguiente ejemplo, se usa `%%pyspark` en una celda para crear una tabla temporal en Python y se usa `%%scalaspark` en la celda siguiente para leer los datos de esa tabla en Scala.

```
%%pyspark
df=spark.sql("SELECT * from nyc_top_trips_report LIMIT 20")
# create a temporary table called nyc_top_trips_report_view in python
df.createOrReplaceTempView("nyc_top_trips_report_view")
```

```
%%scalaspark
// read the temp table in scala
val df=spark.sql("SELECT * from nyc_top_trips_report_view")
```

```
df.show(5)
```

Información general de Cuadernos de Amazon EMR

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Puede usar Cuadernos de Amazon EMR junto con clústeres de Amazon EMR que ejecuten Apache Spark para crear y abrir las interfaces de Cuaderno de Jupyter y JupyterLab dentro de la consola de Amazon EMR. Un cuaderno de EMR es un cuaderno “sin servidor” que puede usar para ejecutar consultas y código. A diferencia de un cuaderno tradicional, el contenido de un cuaderno de EMR (ecuaciones, consultas, modelos, código y texto narrativo dentro de las celdas del cuaderno) se ejecutan en un cliente. Los comandos se ejecutan utilizando un kernel en el clúster de EMR. El contenido del cuaderno también se guarda en Amazon S3 separado de los datos del clúster para mayor durabilidad y flexibilidad en la reutilización.

Puede iniciar un clúster, asociar un cuaderno de EMR para análisis y, a continuación, terminar el clúster. También puede cerrar un bloc de notas asociado a un clúster en ejecución y cambiar a otro. Varios usuarios pueden asociar cuadernos al mismo clúster de forma simultánea y compartir entre sí archivos de cuaderno en Amazon S3. Estas características le permiten ejecutar clústeres bajo demanda para ahorrar costes y reducir el tiempo dedicado a reconfigurar blocs de notas para diferentes clústeres y conjuntos de datos.

También puede ejecutar un cuaderno de EMR mediante programación con la API de Amazon EMR, sin necesidad de interactuar con la consola de Amazon EMR (“ejecución Headless”). Debe incluir una celda en el cuaderno de EMR que tenga una etiqueta de parámetros. Esa celda permite que un script pase nuevos valores de entrada al cuaderno. Los cuadernos parametrizados se pueden reutilizar con diferentes conjuntos de valores de entrada. No es necesario hacer copias del mismo cuaderno para editarlo y ejecutarlo con nuevos valores de entrada. Amazon EMR crea y guarda el cuaderno de salida en S3 para cada ejecución del cuaderno parametrizado. Para ver muestras de

códigos de la API de cuadernos de EMR, consulte Ejemplos de comandos para ejecutar Cuadernos de EMR mediante programación.

> **Important**
>
> La capacidad de Cuadernos de EMR admite clústeres que utilizan la versión 5.18.0 de Amazon EMR y versiones posteriores. Le recomendamos que utilice Cuadernos de EMR con clústeres que utilicen la última versión de Amazon EMR, o al menos las versiones 5.30.0, 5.32.0 o 6.2.0. Con estas versiones, los kernels de Jupyter se ejecutan en el clúster asociado, en lugar de hacerlo en una instancia de Jupyter. Este cambio mejora el rendimiento y mejora su capacidad para personalizar kernels y bibliotecas. Para obtener más información, consulte Diferencias en capacidades por versión de clúster.

Se aplicarán los cargos correspondientes para el almacenamiento de Amazon S3 y los clústeres de Amazon EMR.

Cuadernos de Amazon EMR está disponible como Espacios de trabajo de Amazon EMR Studio en la nueva consola

Hacer la transición de Cuadernos de EMR a Espacios de trabajo

En la nueva consola de Amazon EMR, hemos fusionado Cuadernos de EMR con Espacios de trabajo de Amazon EMR Studio en una sola experiencia. Cuando utiliza un EMR Studio, puede crear y configurar diferentes espacios de trabajo para organizar y ejecutar cuadernos. Si tenía cuadernos de Amazon EMR en la consola anterior, estos están disponibles como espacios de trabajo de EMR Studio en la nueva consola.

Amazon EMR creó estos nuevos espacios de trabajo de EMR Studio. El número de Studios que hemos creado corresponde al número de VPC distintas que utiliza en Cuadernos de EMR. Por ejemplo, si se conecta a clústeres de EMR en dos VPC diferentes de Cuadernos de EMR, crearemos dos nuevos EMR Studios. Sus cuadernos se distribuyen entre los nuevos Studios.

> **Important**
>
> Desactivamos la opción para crear nuevos cuadernos en la antigua consola de Amazon EMR. En su lugar, utilice Crear espacio de trabajo en la nueva consola de Amazon EMR.

Para obtener más información sobre Espacios de trabajo de Amazon EMR Studio, consulte Aprenda los conceptos básicos de los espacios de trabajo. Para obtener información general conceptual sobre EMR Studio, consulte Workspaces en la página Cómo funciona Amazon EMR Studio.

¿Qué necesita hacer?

Si bien puede seguir utilizando sus cuadernos actuales en la consola anterior, le recomendamos que utilice Espacios de trabajo de Amazon EMR Studio en la nueva consola. Debe configurar permisos de rol adicionales para activar las capacidades en EMR Studio que no están disponibles en Cuadernos de EMR.

> **Note**
>
> Como mínimo, para ver los cuadernos de EMR existentes como espacios de trabajo de EMR Studio y para crear nuevos espacios de trabajo, los usuarios deben tener los permisos `elasticmapreduce:ListStudios` y `elasticmapreduce:CreateStudioPresignedUrl` en sus roles. Para acceder a todas las características de EMR Studio, consulte Habilitación de las características de EMR Studio para los usuarios de Cuadernos de EMR para ver la lista completa de permisos adicionales que necesitarán los usuarios de Cuadernos de EMR.

Capacidades mejoradas en EMR Studio más allá de los cuadernos de EMR

Con Amazon EMR Studio, puede configurar y utilizar las siguientes capacidades que no están disponibles en Cuadernos de EMR:

- Explorar clústeres de EMR y adjuntar a ellos desde JupyterLab
- Explorar clústeres virtuales de Cuadernos de EMR y adjuntar a ellos desde JupyterLab
- Conectarse a repositorios de Git desde JupyterLab
- Colaborar con otros miembros de su equipo para escribir y ejecutar código de cuaderno
- Examinar los datos con SQL Explorer
- Aprovisionar clústeres de EMR con Service Catalog

Para obtener una lista completa de las capacidades de Amazon EMR Studio, consulte Características principales de EMR Studio.

Habilitación de las características de EMR Studio para los usuarios de Cuadernos de EMR

Los nuevos EMR Studios que crearemos como parte de esta fusión utilizan el rol de IAM `EMR_Notebooks_DefaultRole` existente como rol de servicio de EMR Studio.

Los usuarios que lleven a cabo la transición a EMR Studio desde Cuadernos de EMR y deseen utilizar las capacidades adicionales de EMR Studio necesitarán varios permisos de rol nuevos. Agregue los siguientes permisos a los roles de los usuarios de Cuadernos de EMR que planean usar EMR Studio.

> **Note**
>
> Como mínimo, para ver los cuadernos de EMR existentes como espacios de trabajo de EMR Studio y para crear nuevos espacios de trabajo, los usuarios deben tener los permisos `elasticmapreduce:ListStudios` y `elasticmapreduce:CreateStudioPresignedUrl` en sus roles. Para utilizar todas las características de EMR Studio, agregue todos los permisos que se indican a continuación. Los usuarios administradores también necesitan permiso para crear y administrar un EMR Studio. Para obtener más información, consulte Permisos de administrador para crear y administrar un EMR Studio.

```
"elasticmapreduce:DescribeStudio",
"elasticmapreduce:ListStudios",
"elasticmapreduce:CreateStudioPresignedUrl",
"elasticmapreduce:UpdateEditor",
"elasticmapreduce:PutWorkspaceAccess",
"elasticmapreduce:DeleteWorkspaceAccess",
"elasticmapreduce:ListWorkspaceAccessIdentities",
"emr-containers:ListVirtualClusters",
"emr-containers:DescribeVirtualCluster",
"emr-containers:ListManagedEndpoints",
"emr-containers:DescribeManagedEndpoint",
"emr-containers:CreateAccessTokenForManagedEndpoint",
"emr-containers:ListJobRuns",
"emr-containers:DescribeJobRun",
"servicecatalog:SearchProducts",
"servicecatalog:DescribeProduct",
"servicecatalog:DescribeProductView",
```

```
"servicecatalog:DescribeProvisioningParameters",
"servicecatalog:ProvisionProduct",
"servicecatalog:UpdateProvisionedProduct",
"servicecatalog:ListProvisioningArtifacts",
"servicecatalog:DescribeRecord",
"servicecatalog:ListLaunchPaths",
"cloudformation:DescribeStackResources"
```

Los siguientes permisos también son necesarios para usar las capacidades de colaboración de EMR Studio, pero no eran necesarios con Cuadernos de EMR.

```
"sso-directory:SearchUsers",
"iam:GetUser",
"iam:GetRole",
"iam:ListUsers",
"iam:ListRoles",
"sso:GetManagedApplicationInstance"
```

Consideraciones al usar Cuadernos de EMR

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Tenga en cuenta los siguientes requisitos al crear clústeres y desarrollar soluciones mediante los cuadernos de EMR.

Requisitos del clúster

- Habilite el acceso público por bloques de Amazon EMR: el acceso entrante a un clúster permite a los usuarios del clúster ejecutar kernels de cuadernos. Asegúrese de que solo los usuarios autorizados puedan acceder al clúster. Recomendamos encarecidamente que deje habilitado el

acceso público de bloqueo y que limite el tráfico SSH entrante solo a fuentes de confianza. Para obtener más información, consulte Uso de Bloquear el acceso público de Amazon EMR y Control del tráfico de red con grupos de seguridad.

- Utilice un clúster compatible: un clúster conectado a un cuaderno debe cumplir los siguientes requisitos:
  - Solo se admiten clústeres creados mediante Amazon EMR. Puede crear un clúster de forma independiente en Amazon EMR y, a continuación, asociar un cuaderno de EMR, o puede crear un clúster compatible al crear un cuaderno de EMR.
  - Solo se admiten los clústeres creados con la versión 5.18.0 de Amazon EMR o versiones posteriores. Consulte the section called “Diferencias en capacidades por versión de clúster”.
  - Los clústeres creados con instancias de Amazon EC2 con procesadores AMD EPYC, por ejemplo, los tipos de instancias m5a.* y r5a.*, no son compatibles.
  - Cuadernos de EMR solo funciona con clústeres creados con `VisibleToAllUsers` establecidos en `true`. `VisibleToAllUsers` es `true` de forma predeterminada.
  - El clúster debe lanzarse dentro de una EC2-VPC. Se admiten subredes públicas y privadas. No se admite la plataforma EC2-Classic.
  - El clúster se debe lanzar con Hadoop, Spark y Livy instalados. Se pueden instalar otras aplicaciones, pero Cuadernos de EMR actualmente solo admite clústeres de Spark.

    > **Important**
    >
    > Para las versiones 5.32.0 y posteriores de Amazon EMR, o 6.2.0 y posteriores, su clúster también debe ejecutar la aplicación Jupyter Enterprise Gateway para poder funcionar con Cuadernos de EMR.

  - No se admiten los clústeres que utilizan la autenticación de Kerberos.
  - Los clústeres integrados con AWS Lake Formation admiten solo la instalación de bibliotecas adaptadas al bloc de notas. En el clúster no se admite la instalación de kernels ni bibliotecas.
  - No se admiten clústeres con varios nodos principales.
  - No se admiten los clústeres que utilizan instancias de Amazon EC2 basadas en AWS Graviton2.

Diferencias en capacidades por versión de clúster

Le recomendamos que utilice Cuadernos de EMR con clústeres creados con las versiones 5.30.0, 5.32.0 o posteriores, o 6.2.0 o posteriores de Amazon EMR. Con estas versiones, Cuadernos de

EMR ejecuta los kernels en el clúster Amazon EMR adjunto. Los kernels y las bibliotecas se pueden instalar directamente en el nodo principal del clúster. El uso de Cuadernos de EMR con estas versiones de clúster tiene las siguientes ventajas:

- Rendimiento mejorado: los kernels de los cuadernos se ejecutan en clústeres con los tipos de instancias de EC2 que usted seleccione. Las versiones anteriores ejecutan kernels en una instancia especializada a la que no se puede cambiar el tamaño, acceder o personalizar.
- Posibilidad de agregar y personalizar kernels: puede conectarse al clúster para instalar los paquetes del kernel mediante `conda` y `pip`. Además, se admite la instalación de `pip` mediante comandos de terminal dentro de celdas de bloc de notas. En versiones anteriores, solo estaban disponibles los kernels preinstalados (Python, PySpark, Spark y SparkR). Para obtener más información, consulte Instalación de kernels y bibliotecas de Python en un nodo principal del clúster.
- Capacidad para instalar bibliotecas de Python: puede instalar bibliotecas de Python en el nodo principal del clúster mediante `conda` y `pip`. Recomendamos utilizar `conda`. Con versiones anteriores, solo se admiten bibliotecas adaptadas al bloc de notas para PySpark.

Funciones de Cuadernos de EMR admitidas por la versión del clúster

Versión de lanzamiento del clúster	Bibliotecas adaptadas al bloc de notas para PySpark	Instalación del kernel en el clúster	Instalación de la biblioteca de Python en el nodo principal
Antes de 5.18.0	Cuadernos de EMR no es compatible		
5.18.0–5.25.0	No	No	No
5.26.0–5.29.0	Sí	No	No
5.30.0	Sí	Sí	Sí
6.0.0	No	No	No
5.32.0 y versiones posteriores, y 6.2.0 y versiones posteriores	Sí	Sí	Sí

Límites para cuadernos asociados de forma simultánea

Cuando cree un clúster que admita cuadernos, tenga en cuenta el tipo de instancia de EC2 del nodo principal del clúster. Las restricciones de memoria de esta instancia EC2 determinan el número de blocs de notas que pueden estar listos de forma simultánea para ejecutar código y consultas en el clúster.

Tipo de instancia de EC2 del nodo principal	Número de Cuadernos de EMR
*.medium	2
*.large	4
*.xlarge	8
*.2xlarge	16
*.4xlarge	24
*.8xlarge	24
*.16xlarge	24

Versiones de cuaderno de Jupyter y Python

Cuadernos de EMR ejecuta la versión 6.0.2 de Cuaderno de Jupyter y Python 3.6.5 independientemente de la versión de Amazon EMR del clúster asociado.

Consideraciones en torno a la seguridad

Uso de ubicaciones de S3 cifradas

Si especifica una ubicación cifrada en Amazon S3 para almacenar archivos de cuaderno, debe configurar Rol de servicio para Cuadernos de Amazon EMR como usuario clave. El rol de servicio predeterminado es `EMR_Notebooks_DefaultRole`. Si utiliza una clave de AWS KMS para el cifrado, consulte Uso políticas de claves en AWS KMS en la Guía para desarrolladores de AWS Key Management Service y el artículo de soporte técnico para agregar usuarios de claves.

Creación de un bloc de notas

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Puede crear un cuaderno de EMR con la antigua consola de Amazon EMR. No se admite la creación de cuadernos mediante la AWS CLI o la API de Amazon EMR.

Para crear un bloc de notas de EMR

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/.
2. Elija Notebooks (Blocs de notas), Create notebook (Crear bloc de notas).
3. Introduzca el Notebook name (Nombre del bloc de notas) y una Notebook description (Descripción del bloc de notas) opcional.
4. Si tiene un clúster activo al que desea asociar el cuaderno, deje el valor predeterminado Elegir un clúster existente seleccionado, haga clic en Elegir, seleccione un clúster en la lista y, a continuación, haga clic en Elegir clúster. Para obtener información sobre los requisitos de clúster para Cuadernos de EMR, consulte Consideraciones al usar Cuadernos de EMR.

   —o bien—

   Elija Create a cluster (Crear un clúster), introduzca un Cluster name (Nombre de clúster) y elija opciones según las siguientes directrices. El clúster se crea en la VPC predeterminada para la cuenta mediante instancias bajo demanda.

Opción	Descripción
Cluster name (Nombre del clúster)	Nombre descriptivo utilizado para identificar el clúster.

Opción	Descripción
Release (Versión)	No se puede modificar. El valor predeterminado es la versión de Amazon EMR más reciente (5.36.1).
Aplicaciones	No se puede modificar. Enumera las aplicaciones que están instaladas en el clúster.
Instance (Instancia)	Introduzca el número de instancias y seleccione el tipo de instancia EC2. Una instancia se utiliza para el nodo principal. El resto se utiliza para nodos principales. El tipo de instancia determina el número de blocs de notas que se pueden asociar simultáneamente al clúster. Para obtener más información, consulte Límites para cuadernos asociados de forma simultánea.
Rol de EMR	Deje el valor predeterminado o elija el vínculo para el que desea especificar un rol de servicio personalizado de Amazon EMR. Para obtener más información, consulte Rol de servicio para Amazon EMR (rol de EMR).
Perfil de instancia EC2	Deje el valor predeterminado o elija el enlace para especificar un rol de servicio personalizado para las instancias EC2. Para obtener más información, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2).
EC2 key pair (Par de claves de EC2)	Elija un par de claves EC2 para poder conectarse a instancias de clúster. Para obtener más información, consulte Conectarse al nodo principal mediante SSH.

Opción	Descripción
Terminación automática	La terminación automática es compatible con las versiones 5.30.0 y 6.1.0 y posteriores de Amazon EMR. Seleccione la casilla de verificación para habilitar la terminación automática y, a continuación, especifique la cantidad de tiempo de inactividad tras el cual el clúster debe apagarse automáticamente. Para obtener más información, consulte Uso de una política de terminación automática.

5. En Security groups (Grupos de seguridad), elija Use default security groups (Usar grupos de seguridad predeterminados). También puede elegir Choose security groups (Elegir grupos de seguridad) y seleccionar grupos de seguridad personalizados disponibles en la VPC del clúster. Seleccione uno para la instancia principal y otro para la instancia de cliente de cuaderno. Para obtener más información, consulte the section called “Grupos de seguridad para Cuadernos de Amazon EMR”.
6. En Rol de servicio de AWS, deje el valor predeterminado o elija un rol personalizado en la lista. La instancia de cliente para el bloc de notas utiliza este rol. Para obtener más información, consulte Rol de servicio para Cuadernos de Amazon EMR.
7. En Ubicación del cuaderno, elija la ubicación de Amazon S3 donde se guarda el archivo del cuaderno o especifique su propia ubicación. Si el bucket y la carpeta no existen, Amazon EMR los creará.

   Amazon EMR crea una carpeta cuyo nombre es el valor indicado en ID del cuaderno y guarda el cuaderno en un archivo denominado *`NotebookName`*`.ipynb`. Por ejemplo, si especifica la ubicación de Amazon S3 `s3://MyBucket/MyNotebooks` para un cuaderno denominado `MyFirstEMRManagedNotebook`, el archivo del cuaderno se guardará en `s3://MyBucket/MyNotebooks/`*`NotebookID`*`/MyFirstEMRManagedNotebook.ipynb`.

   Si especifica una ubicación cifrada en Amazon S3, debe configurar Rol de servicio para Cuadernos de Amazon EMR como usuario de claves. El rol de servicio predeterminado es `EMR_Notebooks_DefaultRole`. Si utiliza una clave de AWS KMS para el cifrado, consulte

Uso políticas de claves en AWS KMS en la Guía para desarrolladores de AWS Key Management Service y el artículo de soporte técnico para agregar usuarios de claves.

8. Opcionalmente, si ha agregado un repositorio basado en Git a Amazon EMR que desea asociar con este cuaderno, elija Repositorio de Git, haga clic en Elegir repositorio y, a continuación, seleccione un repositorio de la lista. Para obtener más información, consulte Asociación de repositorios basados en Git con Cuadernos de EMR.
9. También puede elegir Tags (Etiquetas) y, a continuación, añadir etiquetas de clave-valor adicionales para el bloc de notas.

> **Important**
>
> Se aplicará una etiqueta predeterminada con la cadena Key (Clave) establecida en `creatorUserID` y el valor definido como su ID de usuario de IAM para poder obtener acceso. Recomendamos que no cambie ni elimine esta etiqueta, ya que se puede utilizar para controlar el acceso. Para obtener más información, consulte Uso de etiquetas de clúster y cuaderno con políticas de IAM para el control de acceso.

10. Elija Create Notebook (Crear bloc de notas).

Uso de los Cuadernos de EMR

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Después de crear un cuaderno de EMR, el cuaderno tarda unos momentos en iniciarse. El campo Status (Estado) de la lista Notebooks (Blocs de notas) muestra Starting (Comenzando). Puede abrir un bloc de notas cuando su estado sea Ready (Listo). Es posible que el bloc de notas tarde algo más de tiempo en estar Ready (Listo) si lo ha creado al mismo tiempo que un clúster.

> **Tip**
>
> Actualice el navegador o elija el icono de actualización situado encima de la lista de blocs de notas para actualizar el estado del bloc de notas.

Descripción de los estados del cuaderno

Un cuaderno de EMR puede tener los siguientes valores en el campo Estado de la lista Cuadernos.

Estado	Significado
Ready	Puede abrir el bloc de notas con el editor de blocs de notas. Mientras el bloc de notas tenga el estado Ready (Listo), puede detenerlo o eliminarlo. Para cambiar de clúster, primero debe detener el bloc de notas. Si un bloc de notas que se encuentra en el estado Ready (Listo) está inactivo durante un periodo prolongado de tiempo, se detiene automátic amente.
Starting	El bloc de notas se está creando y asociando al clúster. Mientras se inicia un bloc de notas, no se puede abrir el editor de blocs de notas, detenerlo, eliminarlo ni cambiar de clúster.
Pendiente	El bloc de notas se ha creado y está esperando a que finalice la integración con el clúster. El clúster puede estar todavía aprovisionando recursos o respondiendo a otras solicitudes. Puede abrir el editor de blocs de notas con el bloc de notas en modo local. El código que se basa en los procesos del clúster no se ejecuta y genera un error.
Stopping (Deteniéndose)	El bloc de notas se está cerrando o el clúster al que está asociado se está terminando.

Estado	Significado
	Mientras se detiene un bloc de notas, no se puede abrir el editor de blocs de notas, detenerlo, eliminarlo ni cambiar de clúster.
Stopped (Detenido)	El bloc de notas se ha cerrado. Puede iniciar el bloc de notas en el mismo clúster, siempre que este siga ejecutándose. Puede cambiar de clúster y eliminar el clúster.
Deleting (Eliminándose)	El clúster se está eliminando de la lista de clústeres disponibles. El archivo del cuaderno, *NotebookName* `.ipynb` , permanece en Amazon S3 y continúa acumulando los cargos de almacenamiento correspondientes.

Uso del editor de cuadernos

Una ventaja de usar un cuaderno de EMR es que puede lanzar el cuaderno en Jupyter o JupyterLab directamente desde la consola.

Con Cuadernos de EMR, el editor de cuadernos al que se obtiene acceso desde la consola de Amazon EMR es el conocido editor de cuadernos de Jupyter de código abierto o JupyterLab. Dado que el editor de cuadernos se lanza en la consola de Amazon EMR, resulta más eficaz configurar el acceso que en el caso de un cuaderno alojado en un clúster de Amazon EMR. No es necesario configurar el cliente del usuario para poder disponer de acceso web a través de SSH, reglas de grupos de seguridad y configuraciones de proxy. Si el usuario tiene permisos suficientes, solo tiene que abrir el editor de cuadernos en la consola de Amazon EMR.

Solo un usuario puede abrir un cuaderno de EMR a la vez desde Amazon EMR. Si otro usuario intenta abrir un cuaderno de EMR que ya está abierto, se produce un error.

> **Important**
>
> Amazon EMR crea una URL prefirmada única para cada sesión del editor de cuadernos, que es válida solamente durante un breve periodo. Le recomendamos que no comparten la URL del editor de bloc de notas. De hacerlo se crea un riesgo de seguridad ya que los

destinatarios de la URL adoptan sus permisos para editar el bloc de notas y ejecutar el código del bloc de notas durante la vida útil de la URL. Si otras personas necesitan acceso a un cuaderno, proporcione permisos a su usuario mediante políticas de permisos y asegúrese de que el rol de servicio de Cuadernos de EMR tenga acceso a la ubicación de Amazon S3. Para obtener más información, consulte the section called “Seguridad” y Rol de servicio para Cuadernos de Amazon EMR.

Para abrir el editor de cuadernos con un cuaderno de EMR

1. Seleccione un bloc de notas cuyo Status (Estado) sea Ready (Listo) o Pending (Pendiente) en la lista Notebooks (Blocs de notas).
2. Elija Open in JupyterLab (Abrir en JupyterLab) o Open in Jupyter (Abrir en Jupyter).

   El editor de blocs de notas de JupyterLab o Jupyter se abre en una pestaña nueva del navegador.
3. En el menú Kernel, elija Change kernel (Cambiar kernel) y, a continuación, seleccione el kernel correspondiente al lenguaje de programación que utilice.

   Ahora está preparado para escribir y ejecutar código desde el editor de blocs de notas.

Cómo guardar el contenido de un cuaderno

Cuando se trabaja en el editor de cuadernos, el contenido de las celdas del cuaderno y la salida se guardan automáticamente de forma periódica en el archivo del cuaderno en Amazon S3. Un bloc de notas que no ha sufrido cambios desde la última vez que se ha editado una celda muestra la cadena (autosaved) (guardado automático) junto al nombre del bloc de notas en el editor. Si los cambios no se han guardado todavía, aparece la cadena unsaved changes (cambios sin guardar).

Puede guardar un bloc de notas manualmente. En el menú Archivo, seleccione Guardar y punto de control o pulse CTRL+S. De este modo, se crea un archivo denominado *NotebookName*`.ipynb` en una carpeta de puntos de control dentro de la carpeta del cuaderno en Amazon S3. Por ejemplo, `s3://`*MyBucket*`/`*MyNotebookFolder*`/`*NotebookID*`/checkpoints/`*NotebookName*`.ipynb`. En esta ubicación, solo se guarda el archivo de punto de comprobación más reciente.

Cambio de clústeres

Puede cambiar el clúster al que está asociado un cuaderno de EMR sin cambiar el contenido del propio cuaderno. Solo se pueden cambiar de clúster los blocs de notas que tengan el estado Stopped (Detenido).

Para cambiar el clúster de un cuaderno de EMR

1. Si el bloc de notas que desea cambiar se está ejecutando, selecciónelo en la lista Notebooks (Blocs de notas) y elija Stop (Detener).
2. Cuando el estado del bloc de notas sea Stopped (Detenido), selecciónelo en la lista Notebooks (Blocs de notas) y, a continuación, elija View details (Ver detalles).
3. Elija Change cluster (Cambiar clúster).
4. Si tiene un clúster activo que ejecuta Hadoop, Spark y Livy al que desea asociar el bloc de notas, deje el valor predeterminado y seleccione un clúster en la lista. En la lista solo aparecerán los clústeres que cumplan los requisitos.

   —o—

   Elija Create a cluster (Crear clúster) y, a continuación, elija las opciones del clúster. Para obtener más información, consulte Requisitos del clúster.
5. Elija una opción para Security groups (Grupos de seguridad) y, a continuación, elija Change cluster and start notebook (Cambiar clúster e iniciar bloc de notas).

Eliminación de cuadernos y archivos de cuadernos

Cuando se elimina un cuaderno de EMR mediante la consola de Amazon EMR, se elimina el cuaderno de la lista de cuadernos disponibles. Sin embargo, los archivos de cuadernos permanecen en Amazon S3 y continúan acumulando cargos de almacenamiento.

Para eliminar un bloc de notas y sus archivos asociados

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/.
2. Elija Notebooks (Blocs de notas), seleccione el bloc de notas en la lista y, a continuación, elija View details (Ver detalles).
3. Elija el icono de carpeta que se encuentra junto a Notebook location (Ubicación del bloc de notas) y copie la URL, que tiene el patrón `s3://MyNotebookLocationPath/NotebookID/`.

4. Elija Eliminar (Delete).

   El bloc de notas se elimina de la lista y los detalles de este ya no se pueden consultar.

5. Siga las instrucciones de ¿Cómo elimino carpetas de un bucket de S3? en la Guía del usuario de Amazon Simple Storage Service. Vaya al bucket y la carpeta del paso 3.

   —o—

   Si ha instalado la AWS CLI, abra un símbolo del sistema y escriba el comando que aparece al final de este párrafo. Sustituya la ubicación de Amazon S3 por la ubicación que ha copiado anteriormente. Asegúrese de que la AWS CLI se ha configurado con las claves de acceso de un usuario que tenga permisos para eliminar la ubicación de Amazon S3. Para obtener más información, consulte Configuración de AWS CLI en la Guía del usuario de AWS Command Line Interface.

   ```
   aws s3 rm s3://MyNotebookLocationPath/NotebookID
   ```

Uso compartido de archivos de cuadernos

Cada cuaderno de EMR se guarda en Amazon S3 como un archivo denominado *NotebookName*.ipynb. Siempre que un archivo de cuaderno sea compatible con la misma versión de Cuaderno de Jupyter en la que está basado Cuadernos de EMR, puede abrir el cuaderno como un cuaderno de EMR.

La forma más fácil de abrir un archivo de bloc de notas de otro usuario es guardar el archivo*.ipynb de otro usuario en el sistema de archivos local y, a continuación, utilizar la función de carga en los editores Jupyter y JupyterLab.

Puede emplear este proceso para utilizar blocs de notas de EMR compartidos por otros usuarios, blocs de notas compartidos en la comunidad de Jupyter o para restaurar un bloc de notas que se ha eliminado de la consola mientras aún se conserva el archivo de bloc de notas.

Para utilizar un archivo de cuaderno diferente como base de un cuaderno de EMR

1. Antes de continuar, cierre el editor de cuadernos para los cuadernos con los que va a trabajar y, a continuación, detenga el cuaderno si se trata de un cuaderno de EMR.

2. Cree un cuaderno de EMR y asígnele un nombre. El nombre que escriba para el bloc de notas será el nombre del archivo que necesita reemplazar. El nombre de archivo nuevo debe coincidir exactamente con el nombre de este archivo.
3. Anote la ubicación en Amazon S3 que ha elegido para el cuaderno. El archivo que va a sustituir está en una carpeta con una ruta y un nombre de archivo que tienen el siguiente patrón: `s3://MyNotebookLocation/NotebookID/MyNotebookName.ipynb`.
4. Detenga el bloc de notas.
5. Sustituya el antiguo archivo de cuaderno en la ubicación de Amazon S3 por el nuevo y utilice exactamente el mismo nombre.

   El siguiente comando de la AWS CLI para Amazon S3 sustituye un archivo guardado en un equipo local denominado `SharedNotebook.ipynb` por un cuaderno de EMR llamado MyNotebook, un ID de `e-12A3BCDEFJHIJKLMNO45PQRST`, y que se ha creado con `MyBucket/MyNotebooksFolder` especificado en Amazon S3. Para obtener más información sobre el uso de la consola de Amazon S3 para copiar y reemplazar archivos, consulte Carga, descarga y administración de objetos en la Guía del usuario de Amazon Simple Storage Service.

   ```
   aws s3 cp SharedNotebook.ipynb s3://MyBucket/
   MyNotebooksFolder/-12A3BCDEFJHIJKLMNO45PQRST/MyNotebook.ipynb
   ```

Ejemplos de comandos para ejecutar Cuadernos de EMR mediante programación

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Información general

Puede ejecutar cuadernos de EMR con las API de ejecución desde un script o desde la línea de comandos. Al iniciar, detener, enumerar y describir las ejecuciones de cuadernos de EMR fuera de la consola de AWS, puede controlar un cuaderno de EMR mediante programación. Puede pasar diferentes valores de parámetros a un cuaderno con una celda de cuaderno parametrizada. Esto elimina la necesidad de crear una copia del cuaderno para cada nuevo conjunto de valores de parámetros. Para obtener más información, consulte Acciones de la API de Amazon EMR.

Puede programar o agrupar ejecuciones de cuadernos EMR con Eventos de Amazon CloudWatch y AWS Lambda. Para obtener más información, consulte Uso de AWS Lambda con Eventos de Amazon CloudWatch.

Permisos de rol para la ejecución programática

Para utilizar la ejecución programática con Cuadernos de EMR, debe configurar los permisos de usuario con las siguientes políticas:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowExecutionActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:StartNotebookExecution",
                "elasticmapreduce:DescribeNotebookExecution",
                "elasticmapreduce:ListNotebookExecutions"
            ],
            "Resource": "*"
        },
        {
            "Sid": "AllowPassingServiceRole",
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "arn:aws:iam::account-id:role/EMR_Notebooks_DefaultRole"
        }
    ]
}
```

Al ejecutar mediante programación Cuadernos de EMR en un clúster de Cuadernos de EMR, debe agregar los siguientes permisos adicionales:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "AllowRetrievingManagedEndpointCredentials",
            "Effect": "Allow",
            "Action": [
                "emr-containers:GetManagedEndpointSessionCredentials"
            ],
            "Resource": [
                "arn:aws:emr-containers:region:account-id:/virtualclusters/virtual-
cluster-id/endpoints/managed-endpoint-id"
            ],
            "Condition": {
                "StringEquals": {
                    "emr-containers:ExecutionRoleArn": [
                        "arn:aws:iam::account-id:role/emr-on-eks-execution-role"
                    ]
                }
            }
        },
        {
            "Sid": "AllowDescribingManagedEndpoint",
            "Effect": "Allow",
            "Action": [
                "emr-containers:DescribeManagedEndpoint"
            ],
            "Resource": [
                "arn:aws:emr-containers:region:account-id:/virtualclusters/virtual-
cluster-id/endpoints/managed-endpoint-id"
            ]
        }
    ]
}
```

Limitaciones de la ejecución programática

- Se admite un máximo de 100 ejecuciones simultáneas por cuenta y Región de AWS.
- Una ejecución termina si dura más de 30 días.

- Las aplicaciones interactivas de Amazon EMR sin servidor no admiten la ejecución programática de cuadernos.

Ejemplos de ejecución programática de cuadernos de EMR

En las siguientes secciones se proporcionan varios ejemplos de ejecución programática de cuadernos de EMR con la AWS CLI, el SDK de Boto3 (Python) y Ruby:

- Ejemplos de comandos de la CLI de ejecución de cuadernos
- Ejemplos de Python para la ejecución de cuadernos
- Ejemplos de Ruby de ejecución de cuadernos

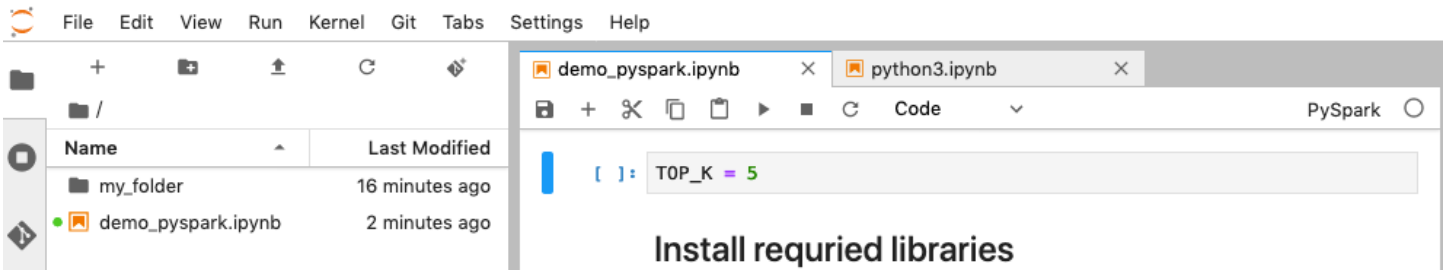
También puede ejecutar cuadernos parametrizados como parte de los flujos de trabajo programados con una herramienta de orquestación como Apache Airflow o Amazon Managed Workflows para Apache Airflow (MWAA). Para obtener más información, consulte Orquestación de trabajos de análisis en Cuadernos de EMR mediante MWAA en el blog de macrodatos de AWS.

Ejemplos de comandos de la CLI de ejecución de cuadernos

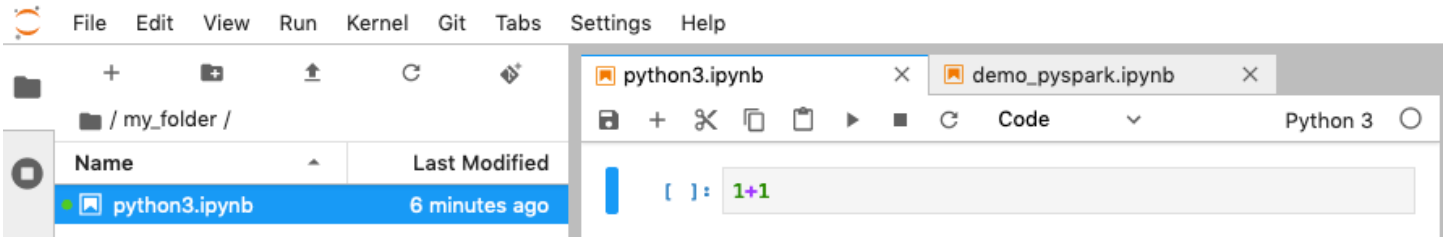
> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

El siguiente ejemplo utiliza el cuaderno de demostración de la consola de Cuadernos de EMR. Para localizar el cuaderno, utilice la ruta del archivo relativa al directorio principal. En este ejemplo, hay dos archivos de cuadernos que puede ejecutar: `demo_pyspark.ipynb` y `my_folder/python3.ipynb`.

La ruta relativa del archivo `demo_pyspark.ipynb` es `demo_pyspark.ipynb`, como se muestra a continuación.



La ruta relativa de `python3.ipynb` es `my_folder/python3.ipynb`, como se muestra a continuación.



Para obtener información sobre las acciones `NotebookExecution` de la API de Amazon EMR, consulte Acciones de la API de Amazon EMR.

Ejecutar un cuaderno

Puede utilizar la AWS CLI para ejecutar su cuaderno con la acción `start-notebook-execution`, tal como se muestra en los siguientes ejemplos.

Example – Ejecución de un cuaderno de EMR en un espacio de trabajo de EMR Studio con un clúster de Amazon EMR (que se ejecuta en Amazon EC2)

```
aws emr --region us-east-1 \
start-notebook-execution \
--editor-id e-ABCDEFG123456 \
--notebook-params '{"input_param":"my-value", "good_superhero":["superman", "batman"]}'
 \
--relative-path test.ipynb \
--notebook-execution-name my-execution \
--execution-engine '{"Id" : "j-1234ABCD123"}' \
--service-role EMR_Notebooks_DefaultRole

{
    "NotebookExecutionId": "ex-ABCDEFGHIJ1234ABCD"
}
```

Example – Ejecución de un cuaderno de EMR en un espacio de trabajo de EMR Studio con un clúster de Cuadernos de EMR

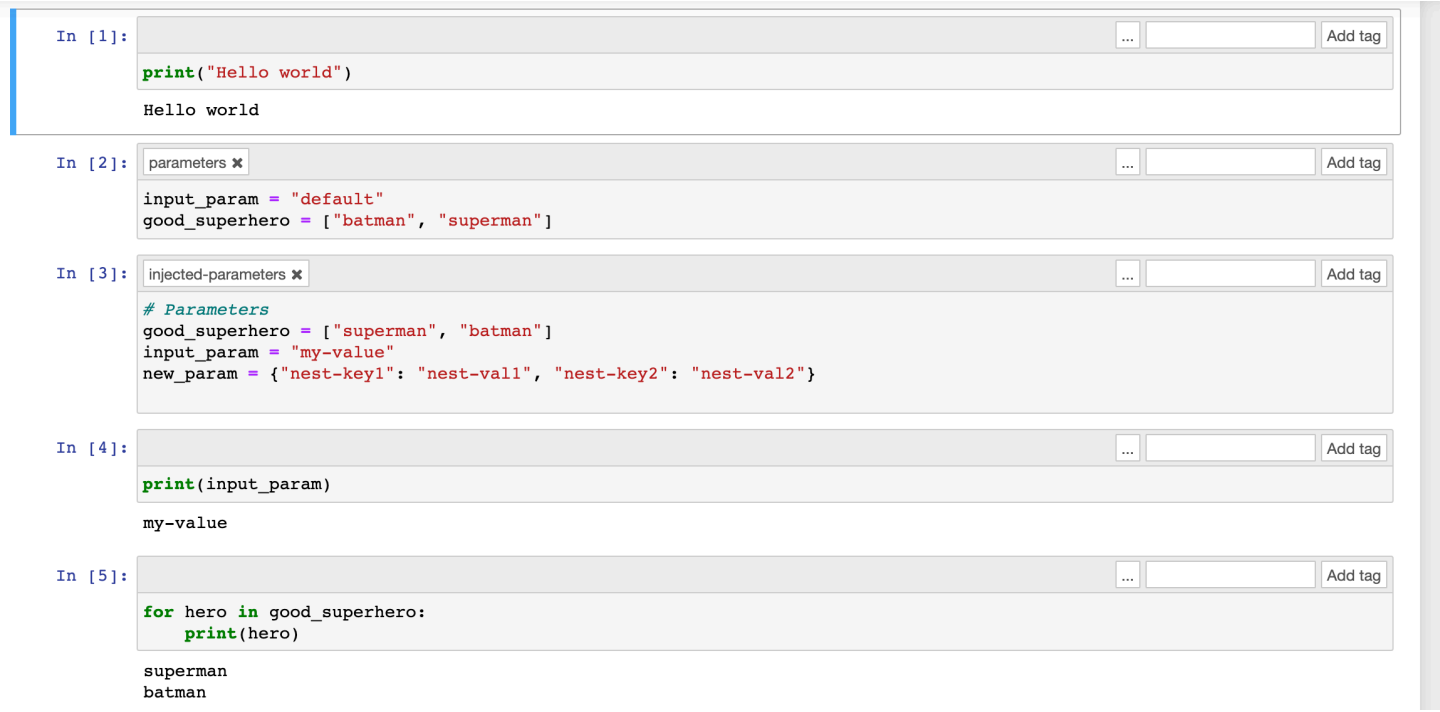
```
aws emr start-notebook-execution \
    --region us-east-1 \
    --service-role EMR_Notebooks_DefaultRole \
    --environment-variables '{"KERNEL_EXTRA_SPARK_OPTS": "--conf
 spark.executor.instances=1", "KERNEL_LAUNCH_TIMEOUT": "350"}' \
    --output-notebook-format HTML \
    --execution-engine Id=arn:aws:emr-containers:us-west-2:account-id:/
virtualclusters/ABCDEFG/
endpoints/ABCDEF,Type=EMR_ON_EKS,ExecutionRoleArn=arn:aws:iam::account-
id:role/execution-role \
    --editor-id e-ABCDEFG \
    --relative-path EMRonEKS-spark_python.ipynb
```

Example – Ejecución de un cuaderno de EMR al especificar su ubicación en Amazon S3

```
aws emr start-notebook-execution \
    --region us-east-1 \
    --notebook-execution-name my-execution-on-emr-on-eks-cluster \
    --service-role EMR_Notebooks_DefaultRole \
    --environment-variables '{"KERNEL_EXTRA_SPARK_OPTS": "--conf
 spark.executor.instances=1", "KERNEL_LAUNCH_TIMEOUT": "350"}' \
    --output-notebook-format HTML \
    --execution-engine Id=arn:aws:emr-containers:us-west-2:account-id:/
virtualclusters/ABCDEF/
endpoints/ABCDEF,Type=EMR_ON_EKS,ExecutionRoleArn=arn:aws:iam::account-
id:role/execution-role \
    --notebook-s3-location '{"Bucket": "your-s3-bucket","Key": "s3-prefix-to-notebook-
location/EMRonEKS-spark_python.ipynb"}' \
    --output-notebook-s3-location '{"Bucket": "your-s3-bucket","Key": "s3-prefix-for-
storing-output-notebook"}'
```

Salida de bloc de notas

Este es el resultado de un ejemplo de cuaderno. La celda 3 muestra los valores de los parámetros recién inyectados.



Describir un cuaderno

Puede utilizar la acción `describe-notebook-execution` para acceder a la información sobre la ejecución de un cuaderno específico.

```
aws emr --region us-east-1 \
describe-notebook-execution --notebook-execution-id ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE

{
    "NotebookExecution": {
        "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
        "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
        "ExecutionEngine": {
            "Id": "j-2QMOV6JAX1TS2",
            "Type": "EMR",
            "MasterInstanceSecurityGroupId": "sg-05ce12e58cd4f715e"
        },
        "NotebookExecutionName": "my-execution",
        "NotebookParams": "{\"input_param\":\"my-value\", \"good_superhero\":
[\"superman\", \"batman\"]}",
        "Status": "FINISHED",
        "StartTime": 1593490857.009,
        "Arn": "arn:aws:elasticmapreduce:us-east-1:123456789012:notebook-execution/ex-
IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
```

```
        "LastStateChangeReason": "Execution is finished for cluster j-2QMOV6JAX1TS2.",
        "NotebookInstanceSecurityGroupId": "sg-0683b0a39966d4a6a",
        "Tags": []
    }
}
```

Detener un cuaderno

Si su cuaderno está ejecutando una ejecución que desea detener, puede hacerlo con el comando `stop-notebook-execution`.

```
# stop a running execution
aws emr --region us-east-1 \
stop-notebook-execution --notebook-execution-id ex-IZWZX78UVPAATC8LHJR129B1RBN4T


# describe it
aws emr --region us-east-1 \
describe-notebook-execution --notebook-execution-id ex-IZWZX78UVPAATC8LHJR129B1RBN4T

{
    "NotebookExecution": {
        "NotebookExecutionId": "ex-IZWZX78UVPAATC8LHJR129B1RBN4T",
        "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
        "ExecutionEngine": {
            "Id": "j-2QMOV6JAX1TS2",
            "Type": "EMR"
        },
        "NotebookExecutionName": "my-execution",
        "NotebookParams": "{\"input_param\":\"my-value\", \"good_superhero\":
[\"superman\", \"batman\"]}",
        "Status": "STOPPED",
        "StartTime": 1593490876.241,
        "Arn": "arn:aws:elasticmapreduce:us-east-1:123456789012:editor-execution/ex-
IZWZX78UVPAATC8LHJR129B1RBN4T",
        "LastStateChangeReason": "Execution is stopped for cluster j-2QMOV6JAX1TS2.
 Internal error",
        "Tags": []
    }
}
```

Enumerar las ejecuciones de un cuaderno por hora de inicio

Puede pasar un parámetro `--from` a `list-notebook-executions` para enumerar las ejecuciones de su cuaderno por hora de inicio.

```
# filter by start time
aws emr --region us-east-1 \
list-notebook-executions --from 1593400000.000

{
    "NotebookExecutions": [
        {
            "NotebookExecutionId": "ex-IZWZX78UVPAATC8LHJR129B1RBN4T",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "STOPPED",
            "StartTime": 1593490876.241
        },
        {
            "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "RUNNING",
            "StartTime": 1593490857.009
        },
        {
            "NotebookExecutionId": "ex-IZWZYRS0M14L5V95WZ9OQ399SKMNW",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "STOPPED",
            "StartTime": 1593490292.995
        },
        {
            "NotebookExecutionId": "ex-IZX009ZK83IVY5E33VH8MDMELVK8K",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
            "StartTime": 1593489834.765
        },
        {
            "NotebookExecutionId": "ex-IZWZXOZF88JWDF9J09GJ91R57VI0N",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
```

```
            "Status": "FAILED",
            "StartTime": 1593488934.688
        }
    ]
}
```

Enumerar las ejecuciones de un cuaderno por hora de inicio y estado

El comando `list-notebook-executions` también puede utilizar un parámetro `--status` para filtrar los resultados.

```
# filter by start time and status
aws emr --region us-east-1 \
list-notebook-executions --from 1593400000.000 --status FINISHED
{
    "NotebookExecutions": [
        {
            "NotebookExecutionId": "ex-IZWZZVR9DKQ9WQ7VZWXJZR29UGHTE",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
            "StartTime": 1593490857.009
        },
        {
            "NotebookExecutionId": "ex-IZX009ZK83IVY5E33VH8MDMELVK8K",
            "EditorId": "e-BKTM2DIHXBEDRU44ANWRKIU8N",
            "NotebookExecutionName": "my-execution",
            "Status": "FINISHED",
            "StartTime": 1593489834.765
        }
    ]
}
```

Ejemplos de Python para la ejecución de cuadernos

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a

ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

El siguiente ejemplo de código es un archivo de SDK para Python (Boto3) denominado `demo.py`, que muestra las API de ejecución del cuaderno.

Para obtener información sobre las acciones `NotebookExecution` de la API de Amazon EMR, consulte Acciones de la API de Amazon EMR.

```
import boto3,time

emr = boto3.client(
    'emr',
    region_name='us-west-1'
)

start_resp = emr.start_notebook_execution(
    EditorId='e-40AC8ZO6EGGCPJ4DLO48KGGGI',
    RelativePath='boto3_demo.ipynb',
    ExecutionEngine={'Id':'j-1HYZS6JQKV11Q'},
    ServiceRole='EMR_Notebooks_DefaultRole'
)

execution_id = start_resp["NotebookExecutionId"]
print(execution_id)
print("\n")

describe_response = emr.describe_notebook_execution(NotebookExecutionId=execution_id)

print(describe_response)
print("\n")

list_response = emr.list_notebook_executions()
print("Existing notebook executions:\n")
for execution in list_response['NotebookExecutions']:
    print(execution)
    print("\n")

print("Sleeping for 5 sec...")
time.sleep(5)
```

```
print("Stop execution " + execution_id)
emr.stop_notebook_execution(NotebookExecutionId=execution_id)
describe_response = emr.describe_notebook_execution(NotebookExecutionId=execution_id)
print(describe_response)
print("\n")
```

Este es el resultado de ejecutar `demo.py`.

```
ex-IZX56YJDW1D29Q1PHR32WABU2SAPK

{'NotebookExecution': {'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK',
 'EditorId': 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'ExecutionEngine': {'Id':
 'j-1HYZS6JQKV11Q', 'Type': 'EMR'}, 'NotebookExecutionName': '', 'Status': 'STARTING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal()),
 'Arn': 'arn:aws:elasticmapreduce:us-west-1:123456789012:notebook-execution/ex-
IZX56YJDW1D29Q1PHR32WABU2SAPK', 'LastStateChangeReason': 'Execution is starting
 for cluster j-1HYZS6JQKV11Q.', 'Tags': []}, 'ResponseMetadata': {'RequestId':
 '70f12c5f-1dda-45b7-adf6-964987d373b7', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-
amzn-requestid': '70f12c5f-1dda-45b7-adf6-964987d373b7', 'content-type': 'application/
x-amz-json-1.1', 'content-length': '448', 'date': 'Wed, 19 Aug 2020 00:49:22 GMT'},
 'RetryAttempts': 0}}

Existing notebook executions:

{'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'STARTING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5ABS5PR1E5AHMFYEMX3JJIORRB', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'RUNNING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 48, 36, 373000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5GLVXIU1HNI8BWVW057F6MF4VE', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 45, 14, 646000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 46, 26, 543000, tzinfo=tzlocal())}
```

```
{'NotebookExecutionId': 'ex-IZX5CV8YDUO8JAIWMXN2VH32RUIT1', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 43, 5, 807000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 44, 31, 632000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX5AS0PPW55CEDEURZ9NSOWSUJZ6', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 42, 29, 265000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 43, 48, 320000, tzinfo=tzlocal())}


{'NotebookExecutionId': 'ex-IZX57YF5Q53BKWLR4I5QZ14HJ7DRS', 'EditorId':
 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'NotebookExecutionName': '', 'Status': 'FINISHED',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 38, 37, 81000, tzinfo=tzlocal()),
 'EndTime': datetime.datetime(2020, 8, 19, 0, 40, 39, 646000, tzinfo=tzlocal())}

Sleeping for 5 sec...
Stop execution ex-IZX56YJDW1D29Q1PHR32WABU2SAPK
{'NotebookExecution': {'NotebookExecutionId': 'ex-IZX56YJDW1D29Q1PHR32WABU2SAPK',
 'EditorId': 'e-40AC8ZO6EGGCPJ4DLO48KGGGI', 'ExecutionEngine': {'Id':
 'j-1HYZS6JQKV11Q', 'Type': 'EMR'}, 'NotebookExecutionName': '', 'Status': 'STOPPING',
 'StartTime': datetime.datetime(2020, 8, 19, 0, 49, 19, 418000, tzinfo=tzlocal()),
 'Arn': 'arn:aws:elasticmapreduce:us-west-1:123456789012:notebook-execution/ex-
IZX56YJDW1D29Q1PHR32WABU2SAPK', 'LastStateChangeReason': 'Execution is being stopped
 for cluster j-1HYZS6JQKV11Q.', 'Tags': []}, 'ResponseMetadata': {'RequestId':
 '2a77ef73-c1c6-467c-a1d1-7204ab2f6a53', 'HTTPStatusCode': 200, 'HTTPHeaders': {'x-
amzn-requestid': '2a77ef73-c1c6-467c-a1d1-7204ab2f6a53', 'content-type': 'application/
x-amz-json-1.1', 'content-length': '453', 'date': 'Wed, 19 Aug 2020 00:49:30 GMT'},
 'RetryAttempts': 0}}
```

Ejemplos de Ruby de ejecución de cuadernos

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para

obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Los siguientes son ejemplos de código de Ruby que muestran el uso de la API de ejecución de cuadernos.

```
# prepare an Amazon EMR client

emr = Aws::EMR::Client.new(
  region: 'us-east-1',
  access_key_id: 'AKIA...JKPKA',
  secret_access_key: 'rLMeu...vU0OLrAC1',
)
```

Iniciar la ejecución del cuaderno y obtener el identificador de ejecución

En este ejemplo, el editor de Amazon S3 y el cuaderno de EMR son `s3://mybucket/notebooks/e-EA8VGAA429FEQTC8HC9ZHWISK/test.ipynb`.

Para obtener información sobre las acciones `NotebookExecution` de la API de Amazon EMR, consulte Acciones de la API de Amazon EMR.

```
start_response = emr.start_notebook_execution({
    editor_id: "e-EA8VGAA429FEQTC8HC9ZHWISK",
    relative_path: "test.ipynb",

    execution_engine: {id: "j-3U82I95AMALGE"},

    service_role: "EMR_Notebooks_DefaultRole",
})


notebook_execution_id = start_resp.notebook_execution_id
```

Descripción de la ejecución del cuaderno e impresión de los detalles

```
describe_resp = emr.describe_notebook_execution({
    notebook_execution_id: notebook_execution_id
})
puts describe_resp.notebook_execution
```

El resultado de los comandos anteriores será el siguiente.

```
{
:notebook_execution_id=>"ex-IZX3VTVZWVWPP27KUB90BZ7V9IEDG",
:editor_id=>"e-EA8VGAA429FEQTC8HC9ZHWISK",
:execution_engine=>{:id=>"j-3U82I95AMALGE", :type=>"EMR", :master_instance_security_group_id=>n
:notebook_execution_name=>"",
:notebook_params=>nil,
:status=>"STARTING",
:start_time=>2020-07-23 15:07:07 -0700,
:end_time=>nil,
:arn=>"arn:aws:elasticmapreduce:us-east-1:123456789012:notebook-execution/ex-
IZX3VTVZWVWPP27KUB90BZ7V9IEDG",
:output_notebook_uri=>nil,
:last_state_change_reason=>"Execution is starting for cluster
 j-3U82I95AMALGE.", :notebook_instance_security_group_id=>nil,
:tags=>[]
}
```

Filtros para cuadernos

```
"EditorId": "e-XXXX",           [Optional]
"From" : "1593400000.000",    [Optional]
"To" :
```

Detener la ejecución del cuaderno

```
stop_resp = emr.stop_notebook_execution({
    notebook_execution_id: notebook_execution_id
})
```

Habilitar la suplantación de usuario para supervisar la actividad del usuario y del trabajo de Spark

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no

podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Cuadernos de EMR le permite configurar la suplantación de usuarios en un clúster de Spark. Esta característica le ayuda a realizar un seguimiento de la actividad de los trabajos iniciados desde el propio editor de blocs de notas. Además, Cuadernos de EMR dispone de un widget de cuaderno de Jupyter para ver los detalles de los trabajos de Spark junto con la salida de las consultas en el editor de cuadernos. El widget está disponible de forma predeterminada y no requiere ninguna configuración especial. Sin embargo, para ver los servidores del historial, el cliente debe configurarse para ver las interfaces web de Amazon EMR que se alojan en el nodo principal.

Configuración de la suplantación de usuarios de Spark

De forma predeterminada, los trabajos de Spark que los usuarios envían mediante el editor de blocs de notas parecen proceder de una identidad de usuario de `livy` indefinida. Es posible configurar la suplantación de usuarios para dicho clúster para que estos trabajos se asocien a la identidad del usuario que ejecutó el código. Se crean directorios de usuario de HDFS en el nodo principal para cada identidad de usuario que ejecuta código en el cuaderno. Por ejemplo, si el usuario `NbUser1` ejecuta código desde el editor de cuadernos, puede conectarse con el nodo principal y comprobar que `hadoop fs -ls /user` muestra el directorio `/user/user_NbUser1`.

Para habilitar esta característica, establezca las propiedades de las clasificaciones de configuración `core-site` y `livy-conf`. Esta característica no está disponible de forma predeterminada si se indica a Amazon EMR que cree un clúster junto con un cuaderno. Para obtener más información sobre cómo usar clasificaciones de configuración para personalizar aplicaciones, consulte Configuración de aplicaciones en la Guía de publicación de Amazon EMR.

Utilice los siguientes valores y clasificaciones de configuración para habilitar la suplantación de usuarios en Cuadernos de EMR:

```
[
    {
        "Classification": "core-site",
        "Properties": {
```

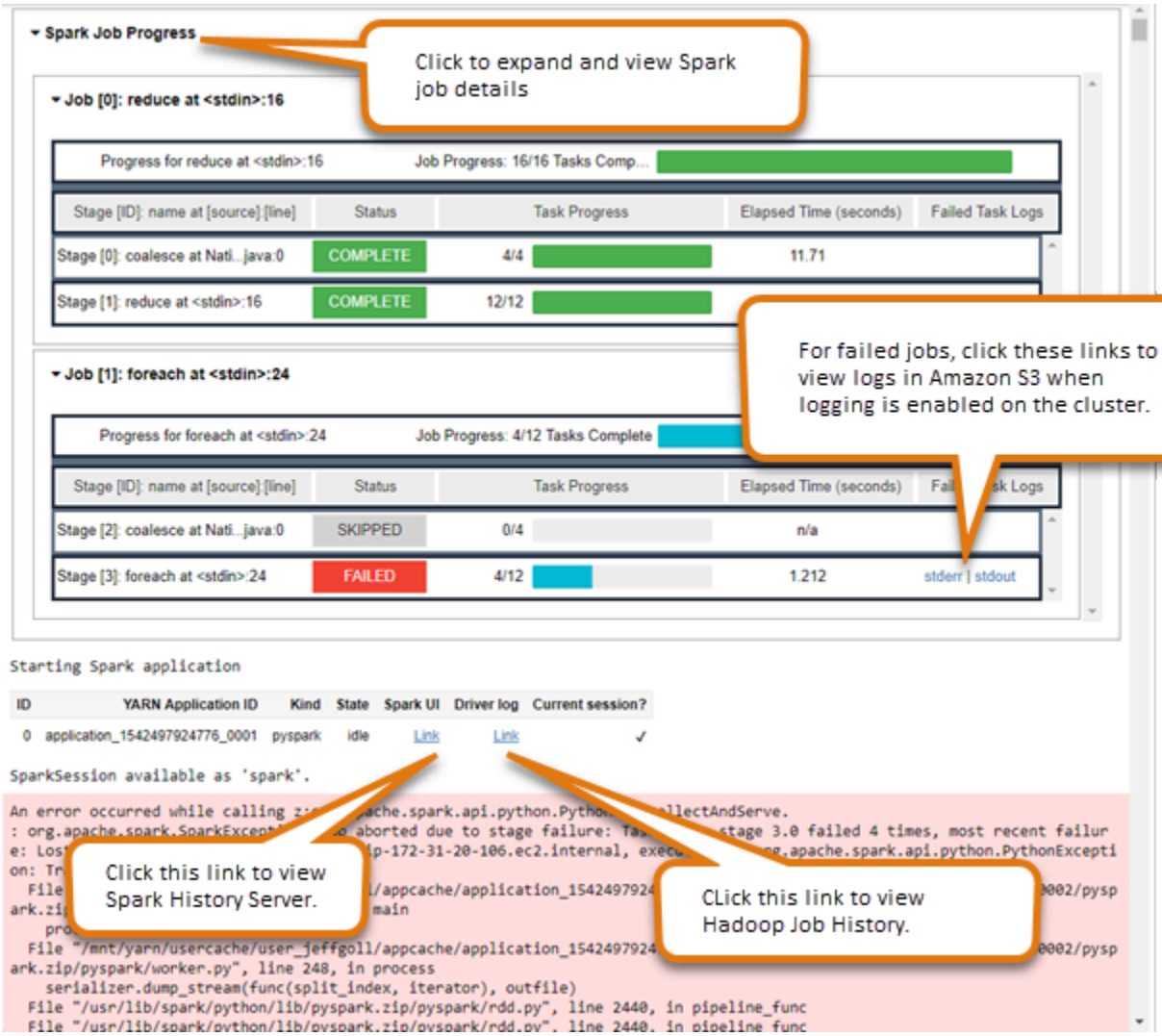
```
          "hadoop.proxyuser.livy.groups": "*",
          "hadoop.proxyuser.livy.hosts": "*"
        }
    },
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.impersonation.enabled": "true"
        }
    }
]
```

Uso del widget de supervisión de trabajos de Spark

Cuando se ejecuta código en el editor de blocs de notas que ejecuta a su vez trabajos de Spark en el clúster de EMR, la salida incluye un widget de Jupyter Notebook para la monitorización de trabajos de Spark. El widget proporciona detalles del trabajo y enlaces útiles a la página del servidor de historial de Spark y a la página del historial de trabajos de Hadoop, junto con los enlaces correspondientes a registros de trabajos en Amazon S3 para los trabajos con errores.

Para ver las páginas del servidor de historial en el nodo principal del clúster, debe configurar un cliente SSH y un proxy de forma adecuada. Para obtener más información, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR. Para ver los registros en Amazon S3, el registro de clúster debe estar habilitado, que es la opción predeterminada para los clústeres nuevos. Para obtener más información, consulte Ver los archivos de registro archivados en Amazon S3.

A continuación, se muestra un ejemplo de supervisión de trabajos de Spark.



Control de acceso y seguridad de los cuadernos de EMR

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Dispone de varias características que le pueden ayudar a personalizar el nivel de seguridad de Cuadernos de EMR. Esto ayuda a garantizar que solo los usuarios autorizados tengan acceso a un cuaderno de EMR, puedan trabajar con los cuadernos y utilizar el editor de cuadernos para ejecutar código en el clúster. Estas características operan junto con las características de seguridad disponibles para Amazon EMR y los clústeres de Amazon EMR. Para obtener más información, consulte Seguridad en Amazon EMR.

- Puede utilizar instrucciones de política de AWS Identity and Access Management junto con etiquetas del bloc de notas para limitar el acceso. Para obtener más información, consulte Cómo funciona Amazon EMR con IAM y Ejemplo de instrucciones de políticas basadas en identidades para Cuadernos de Amazon EMR.
- Los grupos de seguridad de Amazon EC2 actúan como firewalls virtuales que controlan el tráfico de red entre la instancia principal del clúster y el editor de cuadernos. Puede utilizar los valores predeterminados o personalizar estos grupos de seguridad. Para obtener más información, consulte Especificación de grupos de seguridad de EC2 para Cuadernos de Amazon EMR.
- Debe especificar un rol de servicio de AWS que determine los permisos que tiene un cuaderno de EMR al interactuar con los demás servicios de AWS. Para obtener más información, consulte Rol de servicio para Cuadernos de Amazon EMR.

Instalación y uso de kernels y bibliotecas

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Cada cuaderno de EMR viene con un conjunto de bibliotecas y kernels preinstalados. Puede instalar bibliotecas y kernels adicionales en un clúster de EMR si el clúster tiene acceso al repositorio donde se encuentran los kernels y bibliotecas. Por ejemplo, para clústeres en subredes privadas, es posible que tenga que configurar la traducción de direcciones de red (NAT) y proporcionar una ruta para

que el clúster acceda al repositorio PyPI público para instalar una biblioteca. Para obtener más información acerca de cómo configurar el acceso externo para diferentes configuraciones de red, consulte Escenarios y ejemplos en la Guía del usuario de Amazon VPC.

Las aplicaciones de EMR sin servidor vienen con las siguientes bibliotecas preinstaladas para Python y PySpark:

- Bibliotecas Python: ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, scipy
- Bibliotecas PySpark: ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, scipy

Instalación de kernels y bibliotecas de Python en un nodo principal del clúster

Con la versión 5.30.0 de Amazon EMR y posteriores, excepto 6.0.0, puede instalar bibliotecas y kernels de Python adicionales en el nodo principal del clúster. Después de la instalación, estos kernels y bibliotecas están disponibles para cualquier usuario que ejecute un cuaderno de EMR asociado al clúster. Las bibliotecas de Python instaladas de esta manera solo están disponibles para procesos que se ejecutan en el nodo principal. Las bibliotecas no están instaladas en nodos principales o de tareas y no están disponibles para los ejecutores que se ejecutan en esos nodos.

> **Note**
>
> Para las versiones 5.30.1, 5.31.0 y 6.1.0 de Amazon EMR, debe tomar medidas adicionales para instalar los kernels y las bibliotecas en el nodo principal de un clúster.
> Para habilitar la característica, haga lo siguiente:
>
> 1. Asegúrese de que la política de permisos adjunta al rol de servicio de Cuadernos de EMR permita la siguiente acción:
>
>    ```
>    elasticmapreduce:ListSteps
>    ```
>
>   Para obtener más información, consulte Rol de servicio para Cuadernos de EMR.
> 2. Utilice la AWS CLI para ejecutar un paso en el clúster que configura Cuadernos de EMR, como se muestra en el siguiente ejemplo. Debe usar el nombre del paso `EMRNotebooksSetup`. Sustituya *`us-east-1`* por la región en la que reside el clúster. Para obtener más información, consulte Adición de pasos a un clúster mediante la 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"]
```

Puede instalar kernels y bibliotecas con `pip` o `conda` en el directorio `/emr/notebook-env/bin` del nodo principal.

Example – Instalación de bibliotecas de Python

Desde el kernel de Python3, ejecute la magia de `%pip` como un comando desde el interior de una celda de un cuaderno para instalar las bibliotecas de Python.

```
%pip install pmdarima
```

Es posible que necesite reiniciar el kernel para usar los paquetes actualizados. También puede usar la magia de Spark de `%%sh` para invocar `pip`.

```
%%sh
/emr/notebook-env/bin/pip install -U matplotlib
/emr/notebook-env/bin/pip install -U pmdarima
```

Si utiliza un kernel de PySpark, puede instalar bibliotecas en el clúster mediante comandos `pip` o utilizar bibliotecas adaptadas a un cuaderno desde un cuaderno de PySpark.

Para ejecutar comandos `pip` en el clúster desde la terminal, conéctese primero al nodo principal mediante SSH, tal como se muestra en los siguientes comandos.

```
sudo pip3 install -U matplotlib
sudo pip3 install -U pmdarima
```

Como alternativa, puede utilizar bibliotecas adaptadas a un cuaderno. En el caso de las bibliotecas adaptadas a un cuaderno, la instalación de la biblioteca se limita al ámbito de la sesión y se hace en todos los ejecutores de Spark. Para obtener más información, consulte Uso de bibliotecas adaptadas a los cuadernos.

Si quiere empaquetar varias bibliotecas de Python en un kernel de PySpark, también puede crear un entorno virtual de Python aislado. Para ver ejemplos, consulte Using Virtualenv.

Para crear un entorno virtual de Python en una sesión, utilice la propiedad `spark.yarn.dist.archives` de Spark del comando mágico `%%configure` en la primera celda de un cuaderno, tal como se muestra en el siguiente ejemplo.

```
%%configure -f
{
   "conf": {
   "spark.yarn.appMasterEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.appMasterEnv.PYSPARK_DRIVER_PYTHON":"./environment/bin/python",
   "spark.yarn.dist.archives":"s3://DOC-EXAMPLE-BUCKET/prefix/
my_pyspark_venv.tar.gz#environment",
   "spark.submit.deployMode":"cluster"
   }
}
```

De manera similar, puede crear un entorno ejecutor de Spark.

```
%%configure -f
{
   "conf": {
   "spark.yarn.appMasterEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.appMasterEnv.PYSPARK_DRIVER_PYTHON":"./environment/bin/python",
   "spark.executorEnv.PYSPARK_PYTHON":"./environment/bin/python",
   "spark.yarn.dist.archives":"s3://DOC-EXAMPLE-BUCKET/prefix/
my_pyspark_venv.tar.gz#environment",
   "spark.submit.deployMode":"cluster"
   }
}
```

También puede utilizar `conda` para instalar bibliotecas de Python. No necesita tener acceso a sudo para poder utilizar `conda`. Debe conectarse al nodo principal mediante SSH y luego ejecutar `conda` desde el terminal. Para obtener más información, consulte Conectarse al nodo principal mediante SSH.

Example – Instalación de kernels

El siguiente ejemplo muestra cómo instalar el kernel de Kotlin mediante un comando de terminal mientras está conectado al nodo principal de un clúster:

```
sudo /emr/notebook-env/bin/conda install kotlin-jupyter-kernel -c jetbrains
```

> **Note**
>
> Estas instrucciones no instalan las dependencias del kernel. Si su kernel tiene dependencias de terceros, es posible que deba seguir pasos de configuración adicionales antes de poder utilizar el kernel con su cuaderno.

Consideraciones y limitaciones de las bibliotecas adaptadas al cuaderno

Tenga en cuenta lo siguiente al utilizar bibliotecas adaptadas al cuaderno:

- Las bibliotecas adaptadas al cuaderno están disponibles para los clústeres que cree con las versiones 5.26.0 y posteriores de Amazon EMR.
- Las bibliotecas adaptadas al bloc de notas están pensadas para usarse solo con el kernel PySpark.
- Cualquier usuario puede instalar bibliotecas adaptadas al bloc de notas adicionales desde una celda de bloc de notas. Estas bibliotecas solo están disponibles para ese usuario de bloc de notas durante una sola sesión de bloc de notas. Si otros usuarios necesitan las mismas bibliotecas o el mismo usuario necesita las mismas bibliotecas en una sesión diferente, la biblioteca debe volver a instalarse.
- Solo puede desinstalar las bibliotecas que se hayan instalado con la API de `install_pypi_package`. No puede desinstalar ninguna biblioteca que esté preinstalada en el clúster.
- Si se instalan las mismas bibliotecas con diferentes versiones en el clúster y como bibliotecas adaptadas al bloc de notas, la versión de la biblioteca con ámbito de bloc de notas anula la versión de la biblioteca del clúster.

Uso de bibliotecas adaptadas al cuaderno

Para instalar bibliotecas, el clúster de Amazon EMR debe tener acceso al repositorio PyPI en el que se encuentran las bibliotecas.

Los siguientes ejemplos muestran comandos simples para enumerar, instalar y desinstalar bibliotecas desde una celda de bloc de notas mediante las API y kernel de PySpark. Para obtener

ejemplos adicionales, consulte Instalación de bibliotecas de Python en un clúster en ejecución con Cuadernos de EMR en el blog de macrodatos de AWS.

Example – Enumeración de bibliotecas actuales

El siguiente comando muestra los paquetes de Python disponibles para la sesión de bloc de notas de Spark actual. Esto muestra las bibliotecas instaladas en el clúster y las bibliotecas adaptadas al bloc de notas.

```
sc.list_packages()
```

Example – Instalación de la biblioteca de Celery

El siguiente comando instala la biblioteca de Celery como una biblioteca adaptada al bloc de notas.

```
sc.install_pypi_package("celery")
```

Después de instalar la biblioteca, el siguiente comando confirma que la biblioteca está disponible en el controlador Spark y los ejecutores.

```
import celery
sc.range(1,10000,1,100).map(lambda x: celery.__version__).collect()
```

Example – Instalación de la biblioteca de Arrow, con la versión y el repositorio especificados

El siguiente comando instala la biblioteca Arrow como una biblioteca adaptada al bloc de notas, con una especificación de la versión de la biblioteca y la URL del repositorio.

```
sc.install_pypi_package("arrow==0.14.0", "https://pypi.org/simple")
```

Example – Desinstalación de una biblioteca

El siguiente comando desinstala la biblioteca Arrow y la elimina de la sesión actual como biblioteca adaptada al bloc de notas.

```
sc.uninstall_package("arrow")
```

Asociación de repositorios basados en Git con Cuadernos de EMR

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Puede asociar repositorios basados en Git con sus cuadernos de Amazon EMR para guardar los cuadernos en un entorno con control de versiones. Puede asociar hasta tres repositorios a un bloc de notas. Se admiten los siguientes servicios basados en Git:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

La asociación de repositorios basados en Git con su bloc de notas tiene las siguientes ventajas.

- Control de versiones: puede registrar los cambios de código en un sistema de control de versiones para que pueda revisar el historial de sus cambios y revertir selectivamente algunos de ellos.
- Colaboración: los compañeros que trabajan en diferentes cuadernos pueden compartir código a través de repositorios remotos basados en Git. Puede clonar o combinar código de repositorios remotos e insertar los cambios en esos repositorios remotos.
- Reutilización de código: muchos cuadernos de Jupyter que muestran técnicas de análisis de datos o machine learning están disponibles en repositorios de Git alojados públicamente como, por ejemplo, en GitHub. Puede asociar sus blocs de notas a un repositorio para reutilizar los blocs de notas de Jupyter incluidos en un repositorio.

Para usar repositorios basados en Git con Cuadernos de EMR, agregue los repositorios como recursos en la consola de Amazon EMR, asocie credenciales a los repositorios que requieran

autenticación y vincule los repositorios con sus cuadernos. Puede ver una lista de repositorios que se almacenan en su cuenta y obtener más información sobre cada repositorio en la consola de Amazon EMR. Puede asociar un repositorio basado en Git existente con un bloc de notas al crearlo.

Temas

- Requisitos previos y consideraciones
- Agregar un repositorio basado en Git a Amazon EMR
- Actualizar o eliminar un repositorio basado en Git
- Vincular o desvincular un repositorio basado en Git
- Crear un nuevo cuaderno con un repositorio de Git asociado
- Uso de repositorios de Git en un cuaderno

Requisitos previos y consideraciones

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Tenga en cuenta lo siguiente al planear la integración de un repositorio basado en Git con Cuadernos de EMR.

AWS CodeCommit

Si usa un repositorio CodeCommit, debe usar credenciales de Git y HTTPS con CodeCommit. No se admiten las claves SSH ni HTTPS con el ayudante de credenciales de la AWS CLI. CodeCommit no admite los tokens de acceso personal (PAT). Para obtener más información, consulte Uso de IAM con CodeCommit: credenciales de Git, claves SSH y claves de acceso de AWS en la Guía del usuario de IAM y Configuración para usuarios de HTTPS con credenciales de Git en la Guía del usuario de AWS CodeCommit.

Consideraciones de acceso y permisos

Antes de asociar un repositorio a su cuaderno, debe asegurarse de que el clúster, el rol de IAM de Cuadernos de EMR y los grupos de seguridad tengan la configuración y los permisos correctos. También puede configurar los repositorios basados en Git que aloje en una red privada siguiendo las instrucciones que se indican en Configurar un repositorio Git alojado de forma privada para Cuadernos de EMR.

- Acceso a Internet del clúster: la interfaz de red que se inicia solo tiene una dirección IP privada. Esto significa que el clúster al que se conecta el bloc de notas debe estar en una subred privada con una gateway de traducción de direcciones de red (NAT) o debe poder obtener acceso a Internet a través de una gateway privada virtual. Para obtener más información, consulte Opciones de Amazon VPC.

  Los grupos de seguridad del bloc de notas deben incluir una regla de salida que permita que el bloc de notas envíe tráfico a Internet desde el clúster. Le recomendamos que cree sus propios grupos de seguridad. Para obtener más información, consulte Especificación de grupos de seguridad de EC2 para Cuadernos de EMR.

  > **Important**
  >
  > Si la interfaz de red se inicia en una subred pública, no podrá comunicarse con Internet a través de una puerta de enlace de Internet (IGW).

- Permisos para AWS Secrets Manager: si utilizas Secrets Manager para almacenar los secretos que utiliza para acceder a un repositorio, el the section called “Rol de Cuadernos de Amazon EMR” debe tener una política de permisos adjunta que permita la acción `secretsmanager:GetSecretValue`.

Configurar un repositorio Git alojado de forma privada para Cuadernos de EMR

Utilice las siguientes instrucciones para configurar los repositorios alojados de forma privada para Cuadernos de EMR. Debe proporcionar un archivo de configuración con información sobre sus servidores de DNS y Git. Amazon EMR utiliza esta información para configurar cuadernos de EMR que puedan dirigir el tráfico a sus repositorios alojados de forma privada.

Requisitos previos

Para obtener acceso a un repositorio de Git alojado de forma privada para Cuadernos de EMR, debe tener lo siguiente:

- Una ubicación de Amazon S3 Control en la que se guardarán los archivos de su cuaderno de EMR.

Para configurar uno o más repositorios de Git alojados de forma privada para Cuadernos de EMR

1. Cree un archivo de configuración con la plantilla proporcionada. Incluya los siguientes valores para cada servidor de Git que desee especificar en la configuración:

   - **`DnsServerIpV4`**: la dirección IPv4 de su servidor de DNS. Si proporciona valores para `DnsServerIpV4` y `GitServerIpV4List`, el valor de `DnsServerIpV4` tiene prioridad y se utilizará para resolver el `GitServerDnsName`.

     > **Note**
     >
     > Para usar repositorios de Git alojados de forma privada, su servidor de DNS debe permitir el acceso entrante desde Cuadernos de EMR. Le recomendamos que proteja su servidor de DNS contra otros accesos no autorizados.

   - **`GitServerDnsName`**: el nombre de DNS del servidor de Git. Por ejemplo `"git.example.com"`.
   - **`GitServerIpV4List`**: una lista de direcciones IPv4 que pertenecen a sus servidores de Git.

```
[
    {
        "Type": "PrivatelyHostedGitConfig",
        "Value": [
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
                "GitServerDnsName": "<enterprise.git.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            },
            {
                "DnsServerIpV4": "<10.24.34.xxx>",
```

```
                "GitServerDnsName": "<git.example.com>",
                "GitServerIpV4List": [
                    "<xxx.xxx.xxx.xxx>",
                    "<xxx.xxx.xxx.xxx>"
                ]
            }
        ]
    }
]
```

2. Guarde su archivo de configuración como `configuration.json`.
3. Cargue el archivo de configuración en la ubicación de almacenamiento designada de Amazon S3, en una carpeta llamada `life-cycle-configuration`. Por ejemplo, si su ubicación de S3 predeterminada es `s3://DOC-EXAMPLE-BUCKET/notebooks`, el archivo de configuración debe estar ubicado en `s3://DOC-EXAMPLE-BUCKET/notebooks/`**`life-cycle-configuration/configuration.json`**.

> **Important**
>
> Le recomendamos que restrinja el acceso a su carpeta `life-cycle-configuration` únicamente a los administradores de Cuadernos de EMR y al rol de servicio de Cuadernos de EMR. También debe protegerse contra el acceso no autorizado de `configuration.json`. Para obtener instrucciones, consulte Controlar el acceso a un bucket con políticas de usuario o Prácticas recomendadas de seguridad para Amazon S3.

Para ver las instrucciones de carga, consulte Creación de una carpeta y Carga de objetos en la Guía del usuario de Amazon Simple Storage Service.

Agregar un repositorio basado en Git a Amazon EMR

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a

ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Consulte las siguientes secciones para obtener información sobre cómo agregar un repositorio basado en Git a un cuaderno de EMR de la consola antigua o a un Espacio de trabajo de EMR Studio de la nueva consola.

New console

Como Cuadernos de EMR es Espacios de trabajo de EMR Studio en la nueva consola, puede seguir las instrucciones de Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio para asociar hasta tres repositorios de Git a su espacio de trabajo.

También puede utilizar la extensión JupyterLab Git. Seleccione el icono de Git en la barra lateral izquierda de su cuaderno de JupyterLab para acceder a la extensión. Para obtener información sobre la extensión, consulte el repositorio de GitHub de jupyterlab-git.

Para asociar un repositorio de Git a un espacio de trabajo, el administrador de Studio debe tomar medidas para configurar Studio y permitir la vinculación de repositorios de Git. Para obtener más información, consulte Establecer el acceso y los permisos para los repositorios basados en Git.

Old console

Para agregar un repositorio basado en Git como recurso en su cuenta de Amazon EMR con la consola antigua

1. Abra la consola de Amazon EMR antigua en https://console.aws.amazon.com/elasticmapreduce.
2. Elija Git repositories (Repositorios de Git) y, a continuación, seleccione Add repository (Añadir repositorio).
3. En Nombre del repositorio, escriba el nombre que desee utilizar para el repositorio en Amazon EMR.

   Los nombres solo pueden contener caracteres alfanuméricos, guiones (-) o guiones bajos (_).
4. En Git repository URL (URL del repositorio de Git), escriba la URL del repositorio. Cuando se utiliza un repositorio de CodeCommit, esta es la URL que se copia cuando elige Clonar

URL y, a continuación, Clonar HTTPS, por ejemplo, `https://git-codecommit.us-west-2.amazonaws.com/v1/repos/MyCodeCommitRepoName`.

5. En Branch (Rama), escriba un nombre de rama.
6. En Git credentials (Credenciales de Git), elija las opciones de acuerdo con las siguientes pautas. Puede usar un nombre de usuario y una contraseña de Git o un token de acceso personal (PAT) para autenticarse en su repositorio. Cuadernos de EMR accede a sus credenciales de Git mediante los secretos almacenados en Secrets Manager.

> **Note**
>
> Si utiliza un repositorio de GitHub, le recomendamos que utilice un token de acceso personal (PAT) para autenticarse. A partir del 13 de agosto de 2021, GitHub ya no aceptará contraseñas al autenticar las operaciones de Git. Para obtener más información, consulta la publicación sobre los requisitos de autenticación con token para las operaciones de Git en The GitHub Blog.

Opción	Descripción
Usar un secreto de AWS existente	Elija esta opción si ya ha guardado sus credenciales como un secreto en Secrets Manager y, a continuación, seleccione el nombre del secreto en la lista. Si selecciona un secreto asociado con un nombre de usuario y contraseña de Git, el secreto debe estar en el formato `{"gitUsername": " MyUserName ", "gitPassword": " MyPassword "}`.
Crear un nuevo secreto	Elija esta opción para asociar las credencia les de Git existentes con un nuevo secreto que cree en Secrets Manager. Realice una de las siguientes acciones basadas en las credenciales de Git que utilice para el repositorio.

Opción	Descripción
	Si usa un nombre de usuario y una contraseña de Git para acceder al repositorio, seleccione Nombre de usuario y contraseña, escriba el nombre del secreto que va a usar en Secrets Manager y, a continuación, escriba el nombre de usuario y la contraseña que asociar al secreto. - O BIEN - Si utiliza un token de acceso personal para acceder al repositorio, seleccione Token de acceso personal (PAT), escriba el nombre del secreto que usará en Secrets Manager y, a continuación, escriba su token de acceso personal. Para obtener más información, consulte Creating a personal access token for the command line for GitHub yPersonal access tokens for Bitbucket. Los repositorios de CodeCommit no admiten esta opción.
Usar un repositorio público sin credenciales	Elija esta opción para acceder a un repositorio público.

7. Elija Add repository (Añadir repositorio).

Actualizar o eliminar un repositorio basado en Git

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a

ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Consulte las siguientes secciones para obtener información sobre cómo eliminar un repositorio basado en Git de un cuaderno de EMR de la consola antigua o de un Espacio de trabajo de EMR Studio de la nueva consola.

New console

Como Cuadernos de EMR es Espacios de trabajo de EMR Studio en la nueva consola, puede consultar Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio para obtener más información sobre cómo trabajar con los repositorios de Git en su espacio de trabajo. Sin embargo, en este momento, no puede eliminar los repositorios de Git de los espacios de trabajo.

Old console

Para actualizar un repositorio basado en Git en la consola antigua

1. En la página Git repositories (Repositorios de Git), elija el repositorio que desea actualizar.
2. En la página del repositorio, elija Edit repository (Editar repositorio).
3. Actualice el valor de Git credentials (Credenciales de Git) en la página del repositorio.

Para eliminar un repositorio de Git en la consola antigua

1. En la página Git repositories (Repositorios de Git), elija el repositorio que desea eliminar.
2. En la página del repositorio, elija todos los blocs de notas que están actualmente vinculados al repositorio. Elija Unlink notebook (Desvincular bloc de notas).
3. En la página del repositorio, elija Delete (Eliminar).

> **Note**
>
> Para eliminar el repositorio local de Git de Amazon EMR, debe desvincular antes todos los cuadernos de este repositorio. Para obtener más información, consulte Vincular o desvincular un repositorio basado en Git. Al eliminar un repositorio de Git no se eliminan

los secretos creados para el repositorio. Puede eliminar el secreto en AWS Secrets Manager.

Vincular o desvincular un repositorio basado en Git

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Siga estos pasos para vincular o desvincular un repositorio basado en Git a un cuaderno de EMR de la consola anterior o a un Espacio de trabajo de EMR Studio de la nueva consola.

New console

Como Cuadernos de EMR es Espacios de trabajo de EMR Studio en la nueva consola, puede consultar Vincular repositorios basados en Git a un espacio de trabajo de EMR Studio para obtener más información sobre cómo trabajar con los repositorios de Git en su espacio de trabajo. Sin embargo, en este momento, no puede eliminar los repositorios de Git de los espacios de trabajo.

Old console

Para vincular un repositorio basado en Git a un bloc de notas de EMR

El repositorio se puede vincular a un bloc de notas cuando este último tenga el estado Ready (Listo).

1. En la lista Notebooks (Blocs de notas) elija el bloc de notas que desea actualizar.
2. En la sección Git repositories (Repositorios de Git), en la página Notebook (Bloc de notas), elija Link new repository (Vincular nuevo repositorio).

3. En la lista de repositorios de la ventana Link Git repository to notebook (Vincular repositorio de Git a bloc de notas), seleccione los repositorios que desee vincular a su bloc de notas y, a continuación, elija Link repository (Vincular repositorio).

O bien

1. En la página Git repositories (Repositorios de Git), elija el repositorio que desea vincular a su bloc de notas.
2. En la lista EMR notebooks (Blocs de notas de EMR), elija Link new notebook (Vincular nuevo bloc de notas) para vincular este repositorio a un bloc de notas existente.

Para desvincular un repositorio de Git de un bloc de notas de EMR

1. En la lista Notebooks (Blocs de notas) elija el bloc de notas que desea actualizar.
2. En la lista Git repositories (Repositorios de Git), seleccione el repositorio que desea desvincular del bloc de notas y, a continuación, elija Unlink repository (Desvincular repositorio).

O bien

1. En la página Git repositories (Repositorios de Git), elija el repositorio que desea actualizar.
2. En la lista EMR notebooks (Blocs de notas de EMR), seleccione el bloc de notas que desea desvincular del repositorio y, a continuación, elija Unlink notebook (Desvincular bloc de notas).

> **Note**
>
> Cuando se vincula un repositorio de Git a un bloc de notas, el repositorio remoto se clona en su bloc de notas local de Jupyter. Cuando se desvincula el repositorio de Git de un cuaderno, solo se desconecta el cuaderno del repositorio remoto, pero no se elimina el repositorio local de Git.

Descripción del estado del repositorio

Un repositorio Git puede tener cualquiera de los siguientes estados en la lista de repositorios. Para obtener más información sobre cómo vincular EMR Notebooks con los repositorios Git, consulte Vincular o desvincular un repositorio basado en Git.

Estado	Significado
Linking (En proceso de vinculación)	El repositorio de Git se está vinculando al bloc de notas. Mientras el repositorio tiene el estado Linking (En proceso de vinculación), no se puede detener el bloc de notas.
Linked (Vinculado)	El repositorio de Git está vinculado al bloc de notas. Mientras el repositorio tenga el estado Linked (Vinculado), estará conectado al repositorio remoto.
Link Failed (Error al vincular)	El repositorio de Git no pudo vincularse al bloc de notas. Puede intentar vincularlo de nuevo.
Unlinking (En proceso de desvinculación)	El repositorio de Git se está desvinculando del bloc de notas. Mientras el repositorio tenga el estado Unlinking (En proceso de desvinculación), no podrá detener el bloc de notas. Cuando se desvincula un repositorio de Git de un bloc de notas solo se desconecta del repositorio remoto: no se elimina ningún código del bloc de notas.
Unlink Failed (Error al desvincular)	El repositorio de Git no se pudo desvincular del bloc de notas. Puede intentar desvincularlo de nuevo.

Crear un nuevo cuaderno con un repositorio de Git asociado

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Para crear un cuaderno y asociarlo con repositorios de Git en la consola de Amazon EMR antigua

1. Siga las instrucciones de Creación de un bloc de notas.
2. En Security group (Grupo de seguridad), elija Use your own security group (Usar su propio grupo de seguridad).

   > **Note**
   >
   > Los grupos de seguridad del bloc de notas deben incluir una regla de salida para permitir que el bloc de notas envíe tráfico a Internet a través del clúster. Le recomendamos que cree sus propios grupos de seguridad. Para obtener más información, consulte Especificación de grupos de seguridad de EC2 para Cuadernos de EMR.

3. En Git repositories (Repositorios de Git), elija el repositorio que desee asociar al bloc de notas.

   1. Elija un repositorio almacenado como recurso en su cuenta y, a continuación, elija Save (Guardar).
   2. Para añadir un nuevo repositorio como recurso a su cuenta, elija add a new repository (añadir un nuevo repositorio). Complete el flujo de trabajo Add repository (Añadir repositorio) en una ventana nueva.

Uso de repositorios de Git en un cuaderno

> **Note**
>
> Cuadernos de EMR está disponible como Espacios de trabajo de EMR Studio en la nueva consola. Puede seguir utilizando sus cuadernos existentes en la consola antigua, pero no podrá crear nuevos cuadernos en la consola antigua. El botón Crear espacio de trabajo de la nueva consola sustituye a esta funcionalidad. Para crear espacios de trabajo o acceder a ellos, los usuarios de Cuadernos de EMR necesitan permisos de rol de IAM adicionales. Para obtener más información, consulte Cuadernos de Amazon EMR es Espacios de trabajo de Amazon EMR Studio en la nueva consola y Novedades de la consola

Puede elegir Open in JupyterLab (Abrir en JupyterLab) o Open in Jupyter (Abrir en Jupyter) al abrir un bloc de notas.

Si elige abrir el bloc de notas en Jupyter, se muestra una lista expandible de archivos y carpetas dentro del bloc de notas. Puede ejecutar manualmente comandos de Git como los siguientes en una celda del bloc de notas.

```
!git pull origin primary
```

Para abrir cualquiera de los repositorios adicionales, desplácese hasta otras carpetas.

Si elige abrir el cuaderno con una interfaz de JupyterLab, puede usar la extensión JupyterLab de Git preinstalada. Para obtener más información acerca de la extensión, consulte jupyterlab-git.

Planificar y configurar clústeres

En esta sección se explican las opciones de configuración y las instrucciones de planificación, configuración y lanzamiento de clústeres con Amazon EMR. Antes de lanzar un clúster, elija opciones del sistema en función de los datos que está procesando y de los requisitos de costo, velocidad, capacidad, disponibilidad, seguridad y capacidad de administración. Las opciones incluyen:

- La región donde ejecutar un clúster, dónde y cómo almacenar datos y cómo extraer resultados. Consulte Configurar la ubicación del clúster y el almacenamiento de datos.
- Si está ejecutando clústeres de Amazon EMR en Outposts o zonas locales. Consulte Clústeres de EMR en AWS Outposts o Clústeres de EMR en Zonas locales de AWS.
- Si un clúster es de ejecución prolongada o transitorio y el software que ejecuta. Consulte Configuración de un clúster para que continúe o termine después de la ejecución de pasos y Configuración de software del clúster.
- Si un clúster tiene un solo nodo principal o tres nodos principales. Consulte Planificación y configuración de nodos principales.
- Las opciones de hardware y red que optimizan costos, el desempeño y la disponibilidad para la aplicación. Consulte Configuración del hardware y las redes de los clústeres.
- Cómo configurar los clústeres para que pueda administrarlos con más facilidad y monitorizar la actividad, el desempeño y la salud. Consulte Configurar el registro y la depuración de un clúster y Etiquetado de clústeres.
- Cómo autenticar y autorizar el acceso a los recursos de clústeres y cómo cifrar los datos. Consulte Seguridad en Amazon EMR.
- Cómo integrarlo con otro software y servicios. Consulte Integración de controladores y aplicaciones de terceros.

Lanzamiento rápido de un clúster

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para lanzar un clúster rápidamente utilizando la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr/clusters.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En la página Crear clúster, ingrese o seleccione valores para los campos proporcionados. En el panel de resumen persistente se muestra una vista en tiempo real de las opciones de clúster seleccionadas actualmente. Seleccione un encabezado en el panel de resumen para ir a la sección correspondiente y realizar los ajustes. Debe completar todas las configuraciones requeridas antes de seleccionar Crear clúster.
4. Seleccione Crear clúster para aceptar la configuración tal como se muestra.
5. Se abrirá la página de detalles del clúster. Consulte el estado del clúster junto a su nombre. El estado debería cambiar de Iniciando a En ejecución y a Esperando durante el proceso de creación del clúster. Es posible que tenga que elegir el icono de actualización situado en la parte superior derecha o actualizar el navegador para recibir las actualizaciones.

   Cuando el estado cambia a Esperando, el clúster está activo, ejecutándose y listo para aceptar pasos y conexiones SSH.

Old console

Utilice la página Crear clúster: opciones rápidas de la antigua consola de Amazon EMR para crear rápidamente un clúster que realice tareas sencillas o con fines de evaluación o pruebas. Opciones rápidas utiliza valores predeterminados para las opciones de configuración, como el software, las redes y la seguridad del clúster.

Para lanzar un clúster con Opciones rápidas desde la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Clústeres y, a continuación, seleccione Crear clúster para abrir la página Opciones rápidas.

3. En la página Crear clúster: opciones rápidas, introduzca o seleccione valores para los campos proporcionados.
4. Seleccione Crear clúster para lanzar el clúster y abrir la página de estado del clúster.
5. En la página de estado del clúster, consulte el estado de este junto a su nombre. El estado debería cambiar de Iniciando a En ejecución y a Esperando durante el proceso de creación del clúster. Es posible que tenga que elegir el icono de actualización situado a la derecha o actualizar el navegador para recibir las actualizaciones.

   Cuando el estado cambia a Esperando, el clúster está activo, ejecutándose y listo para aceptar pasos y conexiones SSH.

Configurar la ubicación del clúster y el almacenamiento de datos

Esta sección describe cómo configurar la región de un clúster, los diferentes sistemas de archivos disponibles cuando utilice Amazon EMR y cómo utilizarlos. También aborda la preparación o carga de datos en Amazon EMR si es necesario, así como la manera de preparar una ubicación de salida para archivos de registro y cualquier archivo de datos de salida que configure.

Temas

- Elegir una región de AWS
- Utilizar almacenamiento y sistemas de archivos
- Preparar datos de entrada
- Configurar una ubicación de salida

Elegir una región de AWS

Amazon Web Services se ejecuta en los servidores de los centros de datos de todo el mundo. Los centros de datos están organizados en regiones geográficas. Al lanzar un clúster de Amazon EMR, debe especificar una región. Puede elegir la región para reducir la latencia, minimizar los costos o satisfacer los requisitos reglamentarios. Para ver la lista de regiones y puntos de conexión que admite Amazon EMR, consulte Regiones y puntos de conexión en la Referencia general de Amazon Web Services.

Para obtener el máximo rendimiento, debe lanzar el clúster en la misma región donde se encuentren sus datos. Por ejemplo, si el bucket de Amazon S3 que almacena sus datos de entrada se encuentra en la región Oeste de EE. UU. (Oregón), debe lanzar el clúster en esa misma región para evitar las

tarifas de transferencia de datos entre regiones. Si utiliza un bucket de Amazon S3 para recibir la salida del clúster, también debería crearlo en la región Oeste de EE. UU. (Oregón).

Si tiene previsto asociar un par de claves de Amazon EC2 con el clúster (necesarias para utilizar SSH para iniciar sesión en el nodo maestro), el par de claves debe crearse en la misma región que el clúster. Del mismo modo, los grupos de seguridad que crea Amazon EMR para administrar el clúster se crean en la misma región que este.

Si se ha registrado para obtener una Cuenta de AWS el 17 de mayo de 2017 o después de esa fecha, la región predeterminada para acceder a un recurso desde la AWS Management Console es Este de EE. UU. (Ohio) (us-east-2); en el caso de las cuentas antiguas, la región predeterminada es Oeste de EE. UU. (Oregón) (us-west-2) o Este de EE. UU. (Norte de Virginia) (us-east-1). Para obtener más información, consulte Regiones y puntos de enlace.

Algunas características de AWS solo están disponibles en algunas regiones. Por ejemplo, las instancias de informática del clúster solo están disponibles en la región Este de EE. UU. (Norte de Virginia) y la región Asia-Pacífico (Sídney) admite solo Hadoop 1.0.3 y versiones posteriores. Cuando elija una región, compruebe que admite las características que desea utilizar.

Para obtener el máximo rendimiento, utilice la misma región para todos los recursos de AWS que se utilizarán con el clúster. La siguiente tabla asigna los nombres de región entre servicios. Para obtener una lista de las regiones de Amazon EMR, consulte Regiones de AWS y puntos de conexión en la Referencia general de Amazon Web Services.

Elegir una región mediante la consola

La región predeterminada aparece del lado izquierdo de la información de la cuenta en la barra de navegación. Para cambiar de región tanto en la consola nueva como en la antigua, seleccione el menú desplegable Región y seleccione una nueva opción.

Especificación de una región con AWS CLI

Para especificar una región predeterminada en la AWS CLI, utilice el comando aws configure o la variable de entorno `AWS_DEFAULT_REGION`. Para más información, consulte Configuración de la región de AWS en la Guía del usuario de AWS Command Line Interface.

Elegir una región mediante un SDK o la API

Para elegir una región con un SDK, configure su aplicación para que utilice el punto de conexión de esa región. Si está creando una aplicación cliente mediante un SDK de AWS, puede cambiar

el punto de conexión del cliente llamando `setEndpoint`, tal y como se muestra en el ejemplo siguiente:

```
client.setEndpoint("elasticmapreduce.us-west-2.amazonaws.com");
```

Después de que la aplicación haya especificado una región definiendo el punto de conexión, puede configurar la zona de disponibilidad de las instancias de EC2 de su clúster. Las zonas de disponibilidad son regiones geográficas diferentes que se han diseñado para estar aisladas de errores que se produzcan en otras zonas de disponibilidad y que proporcionan conectividad de red de baja latencia a otras zonas de disponibilidad de la misma región. Una región consta de una o varias zonas de disponibilidad. Para optimizar el rendimiento y reducir la latencia, todos los recursos deben estar situados en la misma zona de disponibilidad que el clúster que los utiliza.

Utilizar almacenamiento y sistemas de archivos

Amazon EMR y Hadoop proporcionan una variedad de sistemas de archivos que puede utilizar al momento de procesar pasos de clústeres. Puede especificar qué sistema de archivos utilizar mediante el prefijo del URI utilizado para acceder a los datos. Por ejemplo, `s3://DOC-EXAMPLE-BUCKET1/path` hace referencia a un bucket de Amazon S3 mediante EMRFS. En la siguiente tabla se muestran los sistemas de archivos disponibles, con recomendaciones sobre la hora que es mejor utilizar cada una de ellas.

Amazon EMR y Hadoop suelen utilizar dos o más de los siguientes sistemas de archivos al procesar un clúster. HDFS y EMRFS son los dos sistemas de archivos principales que se utilizan con Amazon EMR.

> **Important**
>
> A partir de la versión 5.22.0 de Amazon EMR, este servicio utiliza AWS Signature Version 4 exclusivamente para autenticar las solicitudes a Amazon S3. Las versiones anteriores de Amazon EMR utilizan AWS Signature Version 2 en algunos casos, a menos que las notas de la versión indiquen que se utiliza exclusivamente Signature Version 4. Para más información, consulte Autenticación de solicitudes (AWS Signature Version 4) y Autenticación de solicitudes (AWS Signature Version 2) en la Guía para desarrolladores de Amazon Simple Storage Service.

Sistema de archivos	Prefix	Descripción
HDFS	`hdfs://` (o sin prefijo)	HDFS es un sistema de archivos distribuido, escalable y portátil para Hadoop. Una ventaja de HDFS es el reconocimiento de datos entre los nodos de clúster de Hadoop que administran los clústeres y los nodos de clúster de Hadoop que administran los pasos individuales. Para obtener más información, consulte la documentación de Hadoop. Los nodos principales y los nodos secundarios utilizan HDFS. Una ventaja es que es rápido; una desventaj a es que se trata de almacenamiento efímero que se reclama cuando el clúster finaliza. Es mejor utilizarlo para almacenar en caché los resultados producidos por pasos de flujos de trabajo intermedios.
EMRFS	`s3://`	EMRFS es una implementación del sistema de archivos Hadoop utilizada para lectura y escritura de archivos desde Amazon EMR directamente en Amazon S3. EMRFS ofrece la comodidad de almacenar los datos persistentes en Amazon S3 para su uso con Hadoop además de ofrecer características como el cifrado del lado del servidor de Amazon S3, la consistencia de lectura tras escritura y la consistencia de lista. **Note** Anteriormente, Amazon EMR utilizaba los sistemas de archivos `s3n` y `s3a`. Aunque estos todavía funcionan, se recomienda utilizar el esquema de URI de `s3` para un mejor rendimiento, seguridad y fiabilidad.
sistema de archivos local		El sistema de archivos local se refiere a un disco conectado a nivel local. Cuando se crea un clúster

Sistema de archivos	Prefix	Descripción
		de Hadoop, cada nodo se crea a partir de una instancia EC2 que viene con un bloque preconfig urado de almacenamiento en disco preasociado que se denomina almacén de instancias. Los datos en volúmenes del almacén de instancias se conservan solo durante la vida de su instancia EC2. Los volúmenes de almacén de instancias son ideales para el almacenamiento temporal de datos que cambian constantemente, como los búferes, las cachés, los datos de pruebas y otro contenido temporal. Para más información, consulte Almacenamiento de instancias de Amazon EC2. HDFS utiliza el sistema de archivos local, pero Python también se ejecuta desde el sistema de archivos local y puede optar por almacenar archivos de aplicaciones adicionales en volúmenes de almacenes de instancias.
Sistema de archivos de bloques de Amazon S3 (heredado)	`s3bfs://`	El sistema de archivos de bloque de Amazon S3 es un sistema de almacenamiento de archivos heredado. Recomendamos encarecidamente evitar el uso de este sistema. **Important** Le recomendamos que no use este sistema de archivos, ya que puede activar una condición de carrera que podrían provocar un error del clúster. Sin embargo, es posible que aplicacio nes heredadas lo requieran.

Acceso a sistemas de archivo

Puede especificar qué sistema de archivos utilizar mediante el prefijo del identificador de recursos uniforme (URI) utilizado para acceder a los datos. Los siguientes procedimientos ilustran cómo hacer referencia a diferentes tipos de sistemas de archivos.

Para acceder a una HDFS local

- Especifique el prefijo `hdfs:///` en el URI. Amazon EMR resuelve rutas que no especifican un prefijo en el URI al HDFS local. Por ejemplo, los dos siguientes URI resolverían la misma ubicación en HDFS.

```
hdfs:///path-to-data

/path-to-data
```

Para acceder a una HDFS remota

- Incluya la dirección IP del nodo principal en el URI, tal y como se muestra en los siguientes ejemplos.

```
hdfs://master-ip-address/path-to-data

master-ip-address/path-to-data
```

Para acceder a Amazon S3

- Utilice el prefijo `s3://`.

```
s3://bucket-name/path-to-file-in-bucket
```

Para acceder al sistema de archivos de bloques de Amazon S3

- Se utiliza únicamente por las aplicaciones heredadas que requieren el sistema de archivos de bloques de Amazon S3. Para acceder a datos o almacenar datos con este sistema de archivos, utilice el prefijo `s3bfs://` en el URI.

  El sistema de archivos de bloques de Amazon S3 es un sistema de archivos heredado que se utilizaba para soportar cargas de más de 5 GB en Amazon S3. Con la funcionalidad de carga multiparte que Amazon EMR proporciona a través del SDK de AWS para Java, puede cargar archivos de hasta 5 TB en el sistema de archivos nativo de Amazon S3 y el sistema de archivos de bloques de Amazon S3 está obsoleto.

  > **Warning**
  >
  > Dado que este sistema de archivos heredado puede crear condiciones de carrera que podrían provocar daños en el sistema de archivos, debe evitar este formato y usar EMRFS en su lugar.

```
s3bfs://bucket-name/path-to-file-in-bucket
```

Preparar datos de entrada

La mayoría de los clústeres cargan los datos de entrada y, a continuación, procesan dichos datos. Para cargar datos, tiene que estar en una ubicación a la que el clúster pueda acceder y en un formato que el clúster pueda procesar. El escenario más habitual consiste en cargar los datos de entrada en Amazon S3. Amazon EMR proporciona herramientas para su clúster para importar o leer datos desde Amazon S3.

El formato de entrada predeterminado en Hadoop son archivos de texto, aunque puede personalizar Hadoop y utilizar herramientas para importar los datos almacenados en otros formatos.

Temas

- Tipos de entrada que Amazon EMR puede aceptar
- Cómo obtener datos en Amazon EMR

Tipos de entrada que Amazon EMR puede aceptar

El formato de entrada predeterminado para un clúster son archivos de texto con cada línea separada por un carácter de nueva línea (\n), que es el formato de entrada usado con más frecuencia.

Si los datos de entrada se encuentran en un formato que no sean los archivos de texto predeterminados, puede utilizar la interfaz de Hadoop `InputFormat` para especificar otros tipos de entrada. Puede incluso crear una subclase de la clase `FileInputFormat` para gestionar tipos de datos personalizados. Para obtener más información, consulte http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/InputFormat.html.

Si utiliza Hive, puede utilizar un serializador/deserializador (SerDe) para leer datos desde un formato dado en HDFS. Para obtener más información, consulte https://cwiki.apache.org/confluence/display/Hive/SerDe.

Cómo obtener datos en Amazon EMR

Amazon EMR proporciona varias formas de obtener datos en un clúster. La forma más común consiste en cargar los datos en Amazon S3 y utilizar las características integradas de Amazon EMR para cargar los datos en el clúster. También puede utilizar la característica DistributedCache de Hadoop para transferir archivos desde un sistema de archivos distribuido al sistema de archivos local. La implementación de Hive proporcionada por Amazon EMR (Hive versión 0.7.1.1 y posteriores) incluye la funcionalidad que puede utilizar para importar y exportar datos entre DynamoDB y un clúster de Amazon EMR. Si tiene una gran cantidad de datos on-premises para procesar, puede encontrar útil el servicio AWS Direct Connect.

Temas

- Descargar datos en Amazon S3
- Carga de datos con AWS DataSync
- Importar archivos con caché distribuida
- Cómo procesar archivos comprimidos
- Importación de datos de DynamoDB a Hive
- Conexión a los datos con AWS Direct Connect
- Cargar grandes cantidades de datos con AWS Snowball

Descargar datos en Amazon S3

Para obtener información sobre cómo cargar objetos en Amazon S3, consulte Agregar un objeto a un bucket en la Guía del usuario de Amazon Simple Storage Service. Para más información sobre el uso de Amazon S3 con Hadoop, consulte http://wiki.apache.org/hadoop/AmazonS3.

Temas

- Creación y configuración de un bucket de Amazon S3
- Configurar cargas multiparte para Amazon S3
- Prácticas recomendadas

Creación y configuración de un bucket de Amazon S3

Amazon EMR utiliza AWS SDK for Java con Amazon S3 para almacenar datos de entrada y de salida, y archivos de registro. Amazon S3 denomina estas ubicaciones de almacenamiento buckets. Los buckets tienen ciertas restricciones y limitaciones para cumplir con los requisitos de Amazon S3 y DNS. Para obtener más información, consulte Restricciones y limitaciones de los buckets en la Guía del usuario de Amazon Simple Storage Service.

En esta sección, se muestra cómo utilizar la AWS Management Console de Amazon S3 para crear y después definir permisos para un bucket de Amazon S3. También puede crear y establecer permisos para un bucket de Amazon S3 con la API o la AWS CLI de Amazon S3. También puede utilizar curl junto con una modificación para pasar los parámetros de autenticación correspondientes a Amazon S3.

Consulte los siguientes recursos:

- Para crear un bucket mediante la consola, consulte Crear un bucket en la Guía del usuario de Amazon S3.
- Para crear y trabajar con buckets mediante la AWS CLI, consulte Uso de comandos de S3 de alto nivel con la AWS Command Line Interface en la Guía del usuario de Amazon S3.
- Para crear un bucket mediante un SDK, consulte Ejemplos de creación de un bucket en la Guía del usuario de Amazon Simple Storage Service.
- Para trabajar con buckets mediante curl, consulte Herramienta de autenticación de Amazon S3 para curl.
- Para más información acerca de la especificación de buckets específicos de una región, consulte Acceso a un bucket en la Guía del usuario de Amazon Simple Storage Service.

- Para trabajar con buckets mediante Puntos de acceso de Amazon S3, consulte Uso de un alias de estilo de bucket para su punto de acceso en la Guía del usuario de Amazon S3. Puede utilizar fácilmente Puntos de acceso de Amazon S3 con el alias del punto de acceso de Amazon S3 en lugar del nombre del bucket de Amazon S3. Puede utilizar el alias del punto de acceso Amazon S3 para aplicaciones nuevas y existentes, incluidas Spark, Hive y Presto, entre otras.

> **Note**
>
> Si habilita el registro para un bucket, se habilitan únicamente los registros de acceso al bucket, no los registros del clúster de Amazon EMR.

Durante la creación de un bucket o posteriormente, puede definir los permisos adecuados para tener acceso a él, en función de su aplicación. Lo habitual es que se conceda a sí mismo (el propietario) acceso de lectura y de escritura, así como acceso de lectura a los usuarios autenticados.

Los buckets de Amazon S3 necesarios deben existir para poder crear un clúster. Debe cargar en Amazon S3 todos los scripts o datos necesarios a los que se hace referencia en el clúster. En la siguiente tabla se describen ubicaciones de ejemplo de datos, scripts y archivos de registro.

Configurar cargas multiparte para Amazon S3

Amazon EMR admite la carga multiparte de Amazon S3 a través del SDK de AWS para Java. La carga multiparte permite cargar un solo objeto como un conjunto de partes. Puede cargar estas partes del objeto de forma independiente y en cualquier orden. Si la transmisión de cualquier parte falla, puede retransmitir esta parte sin que las demás partes se vean afectadas. Después de cargar todas las partes del objeto, Amazon S3 las combina y crea el objeto.

Para obtener más información, consulte la sección Información general sobre la carga multiparte en la Guía de del usuario de Amazon Simple Storage Service.

Además, Amazon EMR ofrece propiedades que le permiten controlar de manera más precisa la limpieza de partes de carga multiparte con errores.

En la siguiente tabla, se describen las propiedades de configuración de Amazon EMR para la carga multiparte. Puede configurarlas utilizando la clasificación de configuración `core-site`. Para más información, consulte Configurar aplicaciones en la Guía de versiones de Amazon EMR.

Nombre de parámetro de configuración	Valor predeterminado	Descripción
`fs.s3n.multipart.uploads.enabled`	`true`	Un tipo booleano que indica si se debe habilitar las cargas multiparte. Cuando la vista consistente de EMRFS se ha habilitado, las cargas multiparte se habilitan de forma predeterminada y si se configura este valor en `false`, no se tiene en cuenta.
`fs.s3n.multipart.uploads.split.size`	`134217728`	Especifica el tamaño máximo de una parte, en bytes, para que EMRFS inicie la carga de una parte nueva cuando las cargas multiparte están habilitadas. El valor mínimo es `5242880` (5 MB). Si se especifica un valor menor, se utiliza `5242880`. El valor máximo es `5368709120` (5 GB). Si se especifica un valor mayor, se utiliza `5368709120` . Si se ha desactivado el cifrado del cliente de EMRFS y el confirmador optimizado de Amazon S3 también se ha desactivado, este valor también controla el tamaño máximo que puede alcanzar un archivo de datos para que EMRFS utilice cargas multiparte en lugar de una solicitud `PutObject` para cargar el archivo. Para obtener más información, consulte
`fs.s3n.ssl.enabled`	`true`	Un tipo booleano que indica si se debe utilizar http o https.
`fs.s3.buckets.create.enabled`	`false`	Un tipo booleano que indica si se debe crear un bucket en caso de que no exista. Si se establece en `false`, se

Nombre de parámetro de configuración	Valor predeterminado	Descripción
		provocará una excepción en las operaciones `CreateBucket` .
`fs.s3.multipart.clean.enabled`	`false`	Un tipo booleano que indica si se debe habilitar la limpieza periódica en segundo plano de las cargas multiparte incompletas.
`fs.s3.multipart.clean.age.threshold`	`604800`	Un tipo long que especifica la edad mínima de una carga multiparte, en segundos, antes de que se la incluya en la limpieza. El valor predeterminado es una semana.
`fs.s3.multipart.clean.jitter.max`	`10000`	Un tipo entero que especifica la cantidad máxima de retraso de fluctuación aleatoria, en segundos, que se agrega al retraso fijo de 15 minutos antes de programar la próxima ronda de limpieza.

Desactivar cargas multiparte

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para deshabilitar las cargas multiparte con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.

3. En Configuración de software, introduzca la configuración siguiente: `classification=core-site,properties=[fs.s3n.multipart.uploads.enabled=false]`.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para deshabilitar las cargas multiparte con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Crear clúster e Ir a las opciones avanzadas.
3. En Editar configuración de software, introduzca la configuración siguiente: `classification=core-site,properties=[fs.s3n.multipart.uploads.enabled=false]`
4. Continúe con la creación del clúster.

CLI

Para deshabilitar la carga multiparte mediante la AWS CLI

Este procedimiento explica cómo deshabilitar la carga multiparte utilizando la AWS CLI. Para deshabilitar la carga multiparte, escriba el comando `create-cluster` con el parámetro `--bootstrap-actions`.

1. Cree un archivo `myConfig.json` con el siguiente contenido y guárdelo en el mismo directorio en el que va a ejecutar el comando:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "fs.s3n.multipart.uploads.enabled": "false"
    }
  }
```

```
]
```

2. Escriba el comando siguiente y sustituya *myKey* por el nombre del par de claves de EC2.

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-5.36.1 --applications Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --configurations file://myConfig.json
```

CLI

Para desactivar la carga multiparte mediante la API

- Para obtener información sobre cómo utilizar las cargas multiparte de Amazon S3 mediante programación, consulte Uso del SDK de AWS para Java para cargas multiparte en la Guía del usuario de Amazon Simple Storage Service.

  Para más información sobre el SDK de AWS para Java, consulte SDK de AWS para Java.

Prácticas recomendadas

A continuación se indican las recomendaciones para utilizar buckets de Amazon S3 con clústeres de EMR.

Habilitación del control de versiones

El control de versiones es una configuración recomendada para su bucket de Amazon S3. Al habilitar el control de versiones, se asegura de que incluso si se eliminan o se sobrescriben involuntariamente los datos, se puedan recuperar. Para más información, consulte Uso del control de versiones en la Guía del usuario de Amazon Simple Storage Service.

Limpiar las cargas multiparte con errores

Los componentes del clúster de EMR utilizan cargas multiparte a través del SDK de AWS para Java con las API de Amazon S3 para escribir los archivos de registro y los datos de salida en Amazon S3 de forma predeterminada. Para más información sobre cómo utilizar Amazon EMR para cambiar las propiedades relacionadas con esta configuración, consulte Configurar cargas multiparte para Amazon S3. En ocasiones, la carga de un archivo de gran tamaño puede dar lugar a una carga multiparte de Amazon S3 incompleta. Cuando una carga multiparte no se puede completar de forma satisfactoria, la carga multiparte en curso sigue ocupando su bucket e incurre en costos de almacenamiento. Le recomendamos las siguientes opciones para evitar un almacenamiento de archivos excesivo:

- En el caso de los buckets que utilice con Amazon EMR, utilice una regla de configuración del ciclo de vida en Amazon S3 que elimine las cargas multiparte incompletas tres días después de la fecha de inicio de la carga. Las reglas de configuración del ciclo de vida le permiten controlar la clase de almacenamiento y el ciclo de vida de los objetos. Para más información, consulte Administración del ciclo de vida de los objetos y Anulación de cargas multiparte incompletas con la política de ciclo de vida de buckets.
- Habilite la característica de limpieza de cargas multiparte de Amazon EMR estableciendo `fs.s3.multipart.clean.enabled` en `TRUE` y ajustando otros parámetros de limpieza. Esta característica es útil con volúmenes altos y a gran escala, así como con clústeres que tienen un tiempo de actividad limitado. En este caso, el parámetro `DaysAfterIntitiation` de una regla de configuración de ciclo de vida puede ser demasiado largo, incluso si se establece en su valor mínimo, provocando picos en el almacenamiento de Amazon S3. La limpieza multiparte de Amazon EMR permite un control más preciso. Para obtener más información, consulte Configurar cargas multiparte para Amazon S3.

Administrar los marcadores de versiones

Se recomienda habilitar una regla de configuración de ciclo de vida en Amazon S3 para eliminar los marcadores de eliminación de objetos vencidos de los buckets con control de versiones que utilice con Amazon EMR. Al eliminar un objeto de un bucket con control de versiones, se crea un marcador de eliminación. Si todas las versiones anteriores del objeto caducan posteriormente, queda un marcador de eliminación de objeto caducado en el bucket. Aunque no se aplican cargos por los marcadores de eliminación, si borra los marcadores caducados puede mejorar el rendimiento de las solicitudes de LIST. Para más información, consulte Configuración del ciclo de vida de un bucket con el control de versiones en la Guía del usuario de Amazon Simple Storage Service.

Prácticas recomendadas de rendimiento

En función de las cargas de trabajo, los tipos específicos de uso de clústeres de EMR y las aplicaciones en dichos clústeres pueden dar lugar a un número elevado de solicitudes en su bucket. Para más información, consulte Consideraciones sobre la tasa de solicitudes y el rendimiento en la Guía del usuario de Amazon Simple Storage Service.

Carga de datos con AWS DataSync

AWS DataSync es un servicio de transferencia de datos en línea que simplifica, automatiza y acelera el proceso de transferencia de datos entre el almacenamiento en las instalaciones y los servicios de almacenamiento de AWS o entre servicios de almacenamiento de AWS. DataSync es compatible con una variedad de sistemas de almacenamiento en las instalaciones, como el Sistema de archivos distribuido de Hadoop (HDFS), los servidores de archivos NAS y el almacenamiento de objetos autoadministrado.

La forma más común de tener datos en un clúster es mediante la carga de datos en Amazon S3 y el uso de características integradas de Amazon EMR para cargar los datos en el clúster.

DataSync puede ayudarlo a realizar las siguientes tareas:

- Replicar el HDFS de su clúster de Hadoop en Amazon S3 para garantizar la continuidad empresarial
- Copiar el HDFS a Amazon S3 para rellenar sus lagos de datos
- Transferir datos entre el HDFS de su clúster de Hadoop y Amazon S3 para su análisis y procesamiento

Para cargar datos en su bucket de S3, primero debe implementar uno o más agentes de DataSync en la misma red que su almacenamiento en las instalaciones. Un agente es una máquina virtual (VM) que se utiliza para leer o escribir datos en una ubicación autoadministrada. A continuación, active los agentes en la Cuenta de AWS y la Región de AWS donde se encuentre su bucket de S3.

Una vez activado el agente, debe crear una ubicación de origen para el almacenamiento en las instalaciones, una ubicación de destino para el bucket de S3 y una tarea. Una tarea es un conjunto de dos ubicaciones (origen y destino) y un conjunto de opciones predeterminadas que se utilizan para controlar el comportamiento de la tarea.

Por último, ejecute la tarea de DataSync para transferir los datos del origen al destino.

Para obtener más información, consulte Introducción a AWS DataSync.

Importar archivos con caché distribuida

Temas

- Tipos de archivo admitidos
- Ubicación de los archivos en caché
- Acceso a archivos almacenados en caché desde aplicaciones de streaming
- Acceso a archivos almacenados en caché desde aplicaciones de streaming

DistributedCache es una característica de Hadoop que puede aumentar la eficacia cuando una tarea de asignación o de reducción necesita acceder a datos comunes. Si el clúster depende de aplicaciones existentes o binarias que no se instalaron al crear el clúster, puede utilizar DistributedCache para importar estos archivos. Esta característica permite a un nodo de clúster leer los archivos importados desde su sistema de archivos local, en lugar de recuperar los archivos desde otros nodos del clúster.

Para obtener más información, visite http://hadoop.apache.org/docs/stable/api/org/apache/hadoop/filecache/DistributedCache.html.

Invoca DistributedCache al crear el clúster. Los archivos se almacenan en caché justo antes de empezar el trabajo de Hadoop y los archivos permanecen almacenados en la memoria caché durante el tiempo que dura el trabajo. Puede almacenar en caché los archivos almacenados en cualquier sistema de archivos compatible con Hadoop, por ejemplo, HDFS o Amazon S3. El tamaño predeterminado de la caché de archivos es de 10 GB. Para cambiar el tamaño de la caché, vuelva a configurar el parámetro de Hadoop, `local.cache.size` a través de la acción de arranque. Para obtener más información, consulte Crear acciones de arranque para instalar software adicional.

Tipos de archivo admitidos

DistributedCache permite tanto archivos individuales como conjuntos de archivos. Los archivos individuales se almacenan en caché como de solo lectura. Los archivos ejecutables y binarios tienen establecidos permisos de ejecución.

Los archivos de almacenamiento son uno o más archivos empaquetados a través de una utilidad, por ejemplo `gzip`. DistributedCache transfiere los archivos comprimidos a cada nodo principal y descomprime el archivo como parte del proceso de almacenamiento en caché. DistributedCache es compatible con los siguientes formatos de compresión:

- zip
- tgz
- tar.gz
- tar
- jar

Ubicación de los archivos en caché

DistributedCache únicamente copia los archivos en los nodos básicos. Si no hay nodos básicos en el clúster, DistributedCache copia los archivos en el nodo principal.

DistributedCache asocia los archivos en caché al directorio de trabajo actual del asignador y reductor utilizando symlinks. Un symlink es un alias a una ubicación de archivo, no la ubicación de archivo real. El valor del parámetro, `yarn.nodemanager.local-dirs` en `yarn-site.xml`, especifica la ubicación de los archivos temporales. Amazon EMR establece este parámetro en `/mnt/mapred` o alguna variación basada en el tipo de instancia y la versión de EMR. Por ejemplo, una configuración podría tener `/mnt/mapred` y `/mnt1/mapred` dado que el tipo de instancia tiene dos volúmenes efímeros. Los archivos de caché se encuentran en un subdirectorio de la ubicación de archivos temporales en `/mnt/mapred/taskTracker/archive`.

Si almacena en caché un solo archivo, DistributedCache colocará el archivo en el directorio `archive`. Si almacena en caché un conjunto de archivos, DistributedCache descomprime el archivo, crea un subdirectorio en `/archive` con el mismo nombre que el del conjunto de archivos. Los archivos individuales se encuentran en el nuevo subdirectorio.

Puede utilizar DistributedCache solo cuando se utilice streaming.

Acceso a archivos almacenados en caché desde aplicaciones de streaming

Para acceder a los archivos almacenados en la memoria caché de sus aplicaciones de mapeador o reductor, asegúrese de añadir el directorio de trabajo actual (./) en la ruta de la aplicación y hacer referencia a los archivos en caché aunque estén presentes en el directorio de trabajo actual.

Acceso a archivos almacenados en caché desde aplicaciones de streaming

Puede utilizar la AWS Management Console y la AWS CLI para crear clústeres de caché que utilizan la caché distribuida.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para especificar los archivos de caché distribuida utilizando la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Pasos, seleccione Agregar paso. Se abrirá el cuadro de diálogo Agregar paso. En el campo Argumentos, incluya los archivos y conjuntos de archivos que desea guardar en la memoria caché. El tamaño del archivo (o tamaño total de los archivos en un archivo de almacenamiento) debe ser menor que el tamaño de la memoria caché asignado.

   Si desea agregar un archivo individual en la caché distribuida, especifique `-cacheFile` seguido del nombre y ubicación del archivo, el símbolo numeral (#) y el nombre que desee asignarle al archivo cuando se coloque en la caché local. En el ejemplo siguiente se muestra cómo agregar un archivo individual a la caché distribuida.

   ```
   -cacheFile \
   s3://DOC-EXAMPLE-BUCKET/file-name#cache-file-name
   ```

   Si desea agregar un conjunto de archivos en la caché distribuida, introduzca `-cacheArchive` seguido de la ubicación de los archivos en Amazon S3, el símbolo numeral (#) y a continuación el nombre que desee asignarle a la colección de archivos en la caché local. En el ejemplo siguiente se muestra cómo agregar un archivo de almacenamiento a la caché distribuida.

   ```
   -cacheArchive \
   s3://DOC-EXAMPLE-BUCKET/archive-name#cache-archive-name
   ```

Ingrese los valores adecuados en el resto de los campos del cuadro de diálogo. Las opciones varían según el tipo de paso. Para agregar el paso y salir del cuadro de diálogo, elija Agregar paso.

4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para especificar archivos de caché distribuida utilizando la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Step execution (Ejecución de pasos) como modo de lanzamiento.
4. En la sección Steps (Pasos), en el campo Add step (Añadir paso), elija Streaming program (Programa de streaming) en la lista y haga clic en Configure and add (Configurar y añadir).
5. En el campo Argumentos, incluya los archivos y conjuntos de archivos que desea guardar en la caché y haga clic en Agregar. El tamaño del archivo (o tamaño total de los archivos en un archivo de almacenamiento) debe ser menor que el tamaño de la memoria caché asignado.

   Si desea agregar un archivo individual en la caché distribuida, especifique `-cacheFile` seguido del nombre y ubicación del archivo, el símbolo numeral (#) y el nombre que desee asignarle al archivo cuando se coloque en la caché local. En el ejemplo siguiente se muestra cómo agregar un archivo individual a la caché distribuida.

   ```
   -cacheFile \
   s3://DOC-EXAMPLE-BUCKET/file_name#cache_file_name
   ```

   Si desea agregar un conjunto de archivos en la caché distribuida, introduzca `-cacheArchive` seguido de la ubicación de los archivos en Amazon S3, el símbolo numeral (#) y a continuación el nombre que desee asignarle a la colección de archivos en la caché local. En el ejemplo siguiente se muestra cómo agregar un archivo de almacenamiento a la caché distribuida.

   ```
   -cacheArchive \
   ```

```
s3://DOC-EXAMPLE-BUCKET/archive_name#cache_archive_name
```

6. Continúe con la configuración y el lanzamiento del clúster. El clúster copia los archivos en la ubicación de caché antes de procesar ningún paso de clúster.

CLI

Para especificar archivos de caché distribuida con la AWS CLI

- Para enviar un paso de streaming cuando se crea un clúster, escriba el comando `create-cluster` con el parámetro `--steps`. Para especificar archivos de caché distribuida con la AWS CLI, especifique los argumentos adecuados al enviar un paso de streaming.

  Si desea agregar un archivo individual en la caché distribuida, especifique `-cacheFile` seguido del nombre y ubicación del archivo, el símbolo numeral (#) y el nombre que desee asignarle al archivo cuando se coloque en la caché local.

  Si desea agregar un conjunto de archivos en la caché distribuida, introduzca `-cacheArchive` seguido de la ubicación de los archivos en Amazon S3, el símbolo numeral (#) y a continuación el nombre que desee asignarle a la colección de archivos en la caché local. En el ejemplo siguiente se muestra cómo agregar un archivo de almacenamiento a la caché distribuida.

  Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Example 1

Escriba el siguiente comando para lanzar un clúster y enviar un paso de streaming que utiliza `-cacheFile` para añadir un archivo `sample_dataset_cached.dat`, a la caché.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications 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","s3://my_bucket/my_mapper.py
 s3://my_bucket/my_reducer.py","-mapper","my_mapper.py","-reducer","my_reducer.py","-
input","s3://my_bucket/my_input","-output","s3://my_bucket/my_output", "-
cacheFile","s3://my_bucket/sample_dataset.dat#sample_dataset_cached.dat"]
```

Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos básicos. Todos los nodos utilizarán el tipo de instancia especificado en el comando.

Si no ha creado con anterioridad el rol de servicio de EMR predeterminado y el perfil de instancia de EC2, escriba `aws emr create-default-roles` para crearlos antes de escribir el subcomando `create-cluster`.

Example 2

El siguiente comando muestra la creación de un clúster de streaming y utiliza `-cacheArchive` para añadir un archivo de almacenamiento de archivos a la caché.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications 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","s3://my_bucket/my_mapper.py
 s3://my_bucket/my_reducer.py","-mapper","my_mapper.py","-reducer","my_reducer.py,"-
input","s3://my_bucket/my_input","-output","s3://my_bucket/my_output", "-
cacheArchive","s3://my_bucket/sample_dataset.tgz#sample_dataset_cached"]
```

Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos básicos. Todos los nodos utilizarán el tipo de instancia especificado en el comando.

Si no ha creado con anterioridad el rol de servicio de EMR predeterminado y el perfil de instancia de EC2, escriba `aws emr create-default-roles` para crearlos antes de escribir el subcomando `create-cluster`.

Cómo procesar archivos comprimidos

Hadoop comprueba la extensión de archivo para detectar archivos comprimidos. Los tipos de compresión compatibles con Hadoop son: gzip, bzip2 y LZO. No tiene que tomar ninguna medida adicional para extraer archivos utilizando estos tipos de compresión; Hadoop se encarga por usted.

Para indexar los archivos LZO, puede utilizar la biblioteca hadoop-lzo que se puede descargar desde https://github.com/kevinweil/hadoop-lzo. Tenga en cuenta que, dado que se trata de una biblioteca de terceros, Amazon EMR no ofrece soporte al desarrollador sobre cómo utilizar esta herramienta. Para obtener información sobre su uso, consulte el archivo readme de hadoop-lzo.

Importación de datos de DynamoDB a Hive

La implementación de Hive proporcionada por Amazon EMR incluye funcionalidad que puede utilizar para importar y exportar datos entre DynamoDB y un clúster de Amazon EMR. Esto resulta útil si los datos de entrada se almacenan en DynamoDB. Para más información, consulte Exportación, importación, consulta y unión de tablas en DynamoDB mediante Amazon EMR.

Conexión a los datos con AWS Direct Connect

AWS Direct Connect es un servicio que puede utilizar para establecer una conexión de red privada dedicada a Amazon Web Services desde su centro de datos, oficina o entorno de coubicación. Si dispone de grandes cantidades de datos de entrada, el uso de AWS Direct Connect puede reducir los costos de red, mejorar el consumo de ancho de banda y proporcionar una experiencia de red más coherente que las conexiones basadas en Internet. Para obtener más información, consulte la Guía del usuario de AWS Direct Connect.

Cargar grandes cantidades de datos con AWS Snowball

AWS Snowball es un servicio que puede utilizar para transferir grandes cantidades de datos entre Amazon Simple Storage Service (Amazon S3) y su ubicación de almacenamiento de datos local a velocidades superiores a las de Internet. Snowball admite dos tipos de trabajos: de importación y de exportación. Los trabajos de importación implican una transferencia de datos desde un origen en las instalaciones a un bucket de Amazon S3. Los trabajos de exportación implican una transferencia de datos de un bucket de Amazon S3 a un origen en las instalaciones. Para ambos tipos de trabajos, los dispositivos Snowball protegen y aseguran sus datos, mientras que los transportadores regionales los trasladan entre Amazon S3 y su ubicación de almacenamiento de datos local. Los dispositivos Snowball son físicamente robustos y están protegidos por AWS Key Management Service (AWS KMS). Para más información, consulte Guía para desarrolladores de AWS Snowball Edge.

Configurar una ubicación de salida

El formato de salida más común de un clúster de Amazon EMR son los archivos de texto, ya sea comprimidos o sin comprimir. Normalmente, están escritos en un bucket de Amazon S3. Este bucket se debe crear antes de lanzar el clúster. Puede especificar el bucket de S3 como ubicación de salida al lanzar el clúster.

Para obtener más información, consulte los siguientes temas:

Temas

- Creación y configuración de un bucket de Amazon S3
- ¿Qué formatos puede devolver Amazon EMR?
- Cómo escribir datos en un bucket de Amazon S3 que no le pertenece
- Comprimir la salida de su clúster

Creación y configuración de un bucket de Amazon S3

Amazon EMR (Amazon EMR) utiliza Amazon S3 para almacenar datos de entrada y de salida, y archivos de registro. Amazon S3 denomina estas ubicaciones de almacenamiento buckets. Los buckets tienen ciertas restricciones y limitaciones para cumplir con los requisitos de Amazon S3 y DNS. Para obtener más información, consulte Restricciones y limitaciones de los buckets en la Guía para desarrolladores de Amazon Simple Storage Service.

Para crear un bucket de Amazon S3, siga las instrucciones que se muestran en Creación de un bucket en la Guía del usuario de Amazon Simple Storage Service.

> **Note**
>
> Si habilita el registro en el asistente de Crear un bucket, solo se habilitan los registros de acceso al bucket, no los registros del clúster.

> **Note**
>
> Para más información acerca de cómo especificar buckets específicos de una región, consulte Buckets y regiones en la Guía para desarrolladores de Amazon Simple Storage Service y Puntos de conexión de región disponibles para los SDK de AWS.

Después de crear su bucket, puede definir los permisos correspondientes. Lo habitual es que se conceda (el propietario) acceso de lectura y escritura. Se recomienda encarecidamente que siga las prácticas recomendadas de seguridad para Amazon S3 al configurar su bucket.

Los buckets de Amazon S3 necesarios deben existir para poder crear un clúster. Debe cargar en Amazon S3 todos los scripts o datos necesarios a los que se hace referencia en el clúster. En la siguiente tabla se describen ubicaciones de ejemplo de datos, scripts y archivos de registro.

Información	Ejemplo de ubicación en Amazon S3
script o programa	`s3://DOC-EXAMPLE-BUCKET1 /script/MapperScript.py`
archivos de registro	`s3://DOC-EXAMPLE-BUCKET1 /logs`
datos de entrada	`s3://DOC-EXAMPLE-BUCKET1 /input`
datos de salida	`s3://DOC-EXAMPLE-BUCKET1 /output`

¿Qué formatos puede devolver Amazon EMR?

El formato de salida predeterminado para un clúster es texto con parejas de clave, valor escritas en líneas individuales de los archivos de texto. Este es el formato de salida usado con más frecuencia.

Si los datos de salida se tienen que escribir en un formato que no sean los archivos de texto predeterminados, puede utilizar la interfaz de Hadoop `OutputFormat` para especificar otros tipos de salida. Puede incluso crear una subclase de la clase `FileOutputFormat` para gestionar tipos de datos personalizados. Para obtener más información, consulte http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/OutputFormat.html.

Si está lanzando un clúster de Hive, puede utilizar un serializador/deserializador (SerDe) para generar datos desde HDFS en un formato dado. Para obtener más información, consulte https://cwiki.apache.org/confluence/display/Hive/SerDe.

Cómo escribir datos en un bucket de Amazon S3 que no le pertenece

Cuando escriba un archivo en un bucket de Amazon Simple Storage Service (Amazon S3), de forma predeterminada, usted es el único capaz de leer dicho archivo. La suposición es que escribirá archivos en sus propios buckets y esta configuración predeterminada protege la privacidad de sus archivos.

Sin embargo, si está ejecutando un clúster y desea que la salida escriba en el bucket de Amazon S3 de otro usuario de AWS y desea que otro usuario de AWS pueda leer la salida, debe hacer dos cosas:

- Hacer que el otro usuario de AWS le otorgue permisos de escritura para su bucket de Amazon S3. El clúster que lanza se ejecuta con sus credenciales de AWS, por lo que los clústeres que lance también podrán escribir en el bucket del otro usuario de AWS.
- Establezca los permisos de lectura del otro usuario de AWS en los archivos que usted o el clúster escriben en el bucket de Amazon S3. La forma más sencilla de definir estos permisos de lectura consiste en utilizar listas de control de acceso (ACL) predefinidas, un conjunto de políticas de acceso predefinidas por Amazon S3.

Para obtener información sobre cómo el otro usuario de AWS puede otorgarle permisos para escribir archivos en el bucket de Amazon S3 del otro usuario, consulte Edición de permisos de bucket en la Guía del usuario de Amazon Simple Storage Service.

Para que el clúster utilice las ACL predefinidas cuando escriba archivos en Amazon S3, defina la opción de configuración del clúster `fs.s3.canned.acl` en la ACL predefinida que desee utilizar. En la siguiente tabla se muestran las ACL predefinidas definidas en la actualidad.

ACL predefinidas	Descripción
`AuthenticatedRead`	Especifica que al usuario se le otorga `Permission.FullControl` y al beneficiario del grupo `GroupGrantee.AuthenticatedUsers` se le otorga acceso `Permission.Read`.
`BucketOwnerFullControl`	Especifica que al propietario del bucket se le otorga `Permission.FullControl`. El propietario del bucket no es necesariamente el mismo que el propietario del objeto.
`BucketOwnerRead`	Especifica que al propietario del bucket se le otorga `Permission.Read`. El propietario del bucket no es necesariamente el mismo que el propietario del objeto.
`LogDeliveryWrite`	Especifica que al propietario se le otorga `Permission.FullControl` y al beneficiario del grupo `GroupGrantee.LogDelivery` se le otorga acceso `Permission.Write`, de modo que se puedan enviar los registros de acceso.

ACL predefinidas	Descripción
`Private`	Especifica que al propietario se le otorga `Permission.FullControl` .
`PublicRead`	Especifica que al usuario se le otorga `Permission.FullControl` y al beneficiario del grupo `GroupGrantee.AllUsers` se le otorga acceso `Permission.Read` .
`PublicReadWrite`	Especifica que al usuario se le otorga `Permission.FullControl` y al beneficiario del grupo `GroupGrantee.AllUsers` se le otorga acceso `Permission.Read` y `Permission.Write` .

Existen muchas formas de definir las opciones de configuración del clúster, en función del tipo de clúster que esté ejecutando. Los siguientes procedimientos muestran cómo definir la opción para casos comunes.

Para escribir archivos mediante las ACL predefinidas en Hive

- Desde el mensaje de comando de Hive, establezca la opción de configuración `fs.s3.canned.acl` en la ACL predefinida que desea que el clúster defina para los archivos que escriba en Amazon S3. Para acceder a la línea de comandos de Hive conéctese al nodo principal mediante SSH y escriba Hive en la línea de comandos de Hadoop. Para obtener más información, consulte Conectarse al nodo principal mediante SSH.

  El ejemplo siguiente establece la opción de configuración `fs.s3.canned.acl` en `BucketOwnerFullControl`, que proporciona al propietario del bucket de Amazon S3 un control completo sobre el archivo. Tenga en cuenta que el comando definido distingue entre mayúsculas y minúsculas y no contiene comillas ni espacios.

```
hive> set fs.s3.canned.acl=BucketOwnerFullControl;
create table acl (n int) location 's3://acltestbucket/acl/';
insert overwrite table acl select count(*) from acl;
```

Las dos últimas líneas del ejemplo crean una tabla que se almacena en Amazon S3 y escribe datos en la tabla.

Para escribir archivos mediante las ACL predefinidas en Pig

- Desde la línea de comandos de Pig, establezca la opción de configuración `fs.s3.canned.acl` en la ACL predefinida que desea que el clúster defina para los archivos que escriba en Amazon S3. Para acceder a la línea de comandos de Pig conéctese al nodo principal mediante SSH y escriba Pig en la línea de comandos de Hadoop. Para obtener más información, consulte Conectarse al nodo principal mediante SSH.

  El ejemplo siguiente establece la opción de configuración `fs.s3.canned.acl` en BucketOwnerFullControl, que proporciona al propietario del bucket de Amazon S3 un control completo sobre el archivo. Tenga en cuenta que el comando definido incluye un espacio antes del nombre de ACL predefinido y no contiene comillas.

```
pig> set fs.s3.canned.acl BucketOwnerFullControl;
store some data into 's3://acltestbucket/pig/acl';
```

Para escribir archivos mediante las ACL predefinidas en un JAR personalizado

- Defina la opción de configuración `fs.s3.canned.acl` utilizando Hadoop con la marca -D. Esto se muestra en el siguiente ejemplo.

```
hadoop jar hadoop-examples.jar wordcount
-Dfs.s3.canned.acl=BucketOwnerFullControl s3://mybucket/input s3://mybucket/output
```

Comprimir la salida de su clúster

Temas

- Compresión de datos de salida
- Compresión de datos intermedia

- Uso de la biblioteca Snappy con Amazon EMR

Compresión de datos de salida

Esto comprime la salida del trabajo de Hadoop. Si utiliza TextOutputFormat el resultado es un archivo de texto comprimido en gzip. Si está escribiendo en SequenceFiles entonces el resultado es un SequenceFile que está comprimido internamente. Esto se puede habilitar definiendo el ajuste de configuración mapred.output.compress en true.

Si está ejecutando un trabajo de streaming puede habilitarlo transfiriendo estos argumentos al trabajo de streaming.

```
-jobconf mapred.output.compress=true
```

También puede utilizar una acción de arranque para comprimir automáticamente todas las salidas de trabajo. Aquí se muestra cómo hacerlo con el cliente de Ruby.

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-hadoop \
--args "-s,mapred.output.compress=true"
```

Por último, si se escribe un JAR personalizado puede habilitar la compresión de salida con la siguiente línea al crear el trabajo.

```
FileOutputFormat.setCompressOutput(conf, true);
```

Compresión de datos intermedia

Si el trabajo reorganiza una cantidad de datos importante desde los mapeadores a los reductores, puede obtener una mejora del rendimiento habilitando la compresión intermedia. Comprima la salida de mapeado y descomprímala cuando llegue al nodo secundario. El ajuste de configuración es mapred.compress.map.output. Puede habilitarlo de manera similar a la compresión de salida.

Al escribir un archivo JAR personalizado, utilice el siguiente comando:

```
conf.setCompressMapOutput(true);
```

Uso de la biblioteca Snappy con Amazon EMR

Snappy es una biblioteca de compresión y descompresión optimizada para velocidad. Está disponible en las AMI de Amazon EMR versión 2.0 y posteriores, y se utiliza como predeterminada para compresión intermedia. Para obtener más información sobre Snappy, consulte http://code.google.com/p/snappy/.

Planificación y configuración de nodos principales

Cuando se lanza un clúster de Amazon EMR, es posible tener uno o tres nodos principales en el clúster. El lanzamiento de un clúster con tres nodos principales solo se admite en la versión 5.23.0 y posteriores de Amazon EMR. Amazon EMR puede aprovechar los grupos de ubicación de EC2 para garantizar que los nodos principales se coloquen en un hardware subyacente distinto a fin de mejorar aún más la disponibilidad del clúster. Para obtener más información, consulte Integración de Amazon EMR con grupos de ubicación de EC2.

Un clúster de Amazon EMR con varios nodos principales ofrece las siguientes ventajas clave:

- El nodo principal ya no es un punto de error único. Si uno de los nodos principales deja de funcionar, el clúster utiliza los otros dos nodos principales y se ejecuta sin interrupción. Mientras tanto, Amazon EMR sustituye automáticamente el nodo principal que ha dejado de funcionar por uno nuevo que tiene la misma configuración y las mismas acciones de arranque.
- Amazon EMR es compatible con las características de alta disponibilidad de Hadoop de HDFS NameNode y YARN ResourceManager, y admite la alta disponibilidad para algunas otras aplicaciones de código abierto.

  Para más información sobre cómo un clúster de Amazon EMR con varios nodos principales es compatible con aplicaciones de código abierto y otras características de Amazon EMR, consulte Aplicaciones y características compatibles.

> **Note**
>
> El clúster pueden encontrarse únicamente en una zona de disponibilidad o subred.

Esta sección proporciona información sobre las aplicaciones y características compatibles de un clúster de Amazon EMR, así como los detalles de configuración, prácticas recomendadas y consideraciones para lanzar el clúster.

Temas

- Aplicaciones y características compatibles
- Lanzar un clúster de Amazon EMR con varios nodos principales
- Integración de Amazon EMR con grupos de ubicación de EC2
- Consideraciones y prácticas recomendadas

Aplicaciones y características compatibles

Este tema contiene información sobre las características de alta disponibilidad de Hadoop de HDFS NameNode y YARN ResourceManager en un clúster de Amazon EMR y cómo funcionan las características de alta disponibilidad con las aplicaciones de código abierto y otras características de Amazon EMR.

HDFS de alta disponibilidad

Un clúster de Amazon EMR con varios nodos principales habilita la característica de alta disponibilidad de NameNode de HDFS en Hadoop. Para más información, consulte Alta disponibilidad del HDFS.

En un clúster de Amazon EMR, dos o más nodos independientes se configuran como NameNodes. Un NameNode está en el estado `active` y los otros en el estado `standby`. Si deja de funcionar el nodo que tiene NameNode en estado `active`, Amazon EMR inicia un proceso de conmutación por error automático de HDFS. Un nodo con NameNode en estado `standby` cambia al estado `active` y se encarga de todas las operaciones de cliente del clúster. Amazon EMR sustituye el nodo que ha dejado de funcionar por uno nuevo que se une al clúster en el estado `standby`.

> **Note**
>
> En las versiones comprendidas entre la 5.23.0 y la 5.30.1 (incluida) de Amazon EMR, solo dos de los tres nodos principales ejecutan NameNode de HDFS.

Si necesita saber qué NameNode tiene el estado `active`, puede utilizar SSH para conectarse a cualquier nodo principal del clúster y ejecutar el comando siguiente:

```
hdfs haadmin -getAllServiceState
```

La salida muestra los nodos en que está instalado NameNode y su estado. Por ejemplo, .

```
ip-##-#-#-##1.ec2.internal:8020 active
ip-##-#-#-##2.ec2.internal:8020 standby
ip-##-#-#-##3.ec2.internal:8020 standby
```

YARN ResourceManager de alta disponibilidad

Un clúster de Amazon EMR con varios nodos principales habilita la característica de alta disponibilidad ResourceManager de YARN en Hadoop. Para más información, consulte Alta disponibilidad de ResourceManager.

En un clúster de Amazon EMR con varios nodos principales, ResourceManager de YARN se ejecuta en los tres nodos principales. Un ResourceManager está en estado `active` y los otros dos en estado `standby`. Si deja de funcionar el nodo principal que tiene el ResourceManager en estado `active`, Amazon EMR comienza un proceso de conmutación por error automática. Un nodo principal que tiene un ResourceManager en estado `standby` se hace cargo todas las operaciones. Amazon EMR sustituye el nodo principal que ha dejado de funcionar por uno nuevo que se une al cuórum de ResourceManager con el estado `standby`.

Puede conectarse a "http://*`master-public-dns-name`*:8088/cluster" para cualquier nodo principal, lo que le dirigirá automáticamente al administrador de recursos en estado `active`. Para saber qué administrador de recursos tiene el estado `active`, utilice SSH para conectarse a cualquier nodo principal del clúster. A continuación, ejecute el comando siguiente para obtener una lista de los tres nodos principales y su estado:

```
yarn rmadmin -getAllServiceState
```

Aplicaciones admitidas en un clúster de Amazon EMR con varios nodos principales

Puede instalar y ejecutar las siguientes aplicaciones en un clúster de Amazon EMR con varios nodos principales. Para cada aplicación, el proceso de conmutación por error del nodo principal varía.

Aplicación	Disponibilidad durante la conmutación por error del nodo principal	Notas
Flink	La conmutación por error del nodo principal no afecta a la disponibilidad	Los trabajos de Flink en Amazon EMR se ejecutan como aplicaciones YARN. Los JobManager de Flink se ejecutan como ApplicationMasters de YARN en los nodos principales. JobManager no se ve afectado por el proceso de conmutación por error del nodo principal. Si utiliza Amazon EMR versión 5.27.0 o anteriores, JobManager es el único punto de error. Cuando se produce un error en JobManager, pierde todos los estados de trabajo y no reanudará los trabajos en ejecución. Puede habilitar la alta disponibi lidad de JobManager configurando el recuento de intentos de la aplicación, los puntos de comprobación y habilitando ZooKeeper como almacenamiento de estado para Flink. Para más información, consulte Configuración de Flink en un clúster de Amazon EMR con varios nodos principales. A partir de la versión 5.28.0 de Amazon EMR, no se necesita ninguna configuración manual para activar la alta disponibilidad de JobManager.
Ganglia	La conmutación por error del nodo principal no afecta a la disponibilidad	Ganglia está disponible en todos los nodos principales, por lo que se puede seguir ejecutando durante el proceso de conmutación por error del nodo principal.
Hadoop	Alta disponibilidad	

Aplicación	Disponibilidad durante la conmutación por error del nodo principal	Notas
		HDFS NameNode y YARN ResourceManager realizan una conmutación por error automátic a al nodo en espera cuando el nodo principal activo deja de funcionar.
HBase	Alta disponibilidad	HBase realiza una conmutación por error automática al nodo en espera cuando el nodo principal activo deja de funcionar. Si la conexión a HBase se realiza través de un servidor REST o Thrift, se debe cambiar a otro nodo principal cuando el nodo principal activo deje de funcionar.
HCatalog	La conmutación por error del nodo principal no afecta a la disponibilidad	HCatalog está basado en un metaalmacén de Hive, que existe fuera del clúster. HCatalog permanece disponible durante el proceso de conmutación por error del nodo principal.
JupyterHub	Alta disponibilidad	JupyterHub se ha instalado en las tres instancias principales. Se recomienda configura r la persistencia de los cuadernos con el fin de evitar su pérdida en caso de error del nodo principal. Para más información, consulte Configuración de la persistencia de los cuadernos en Amazon S3.
Livy	Alta disponibilidad	Livy se ha instalado en los tres nodos principal es. Cuando deja de funcionar el nodo principal activo, se pierde el acceso a la sesión actual de Livy y es necesario crear una nueva sesión de Livy en otro nodo principal o en el nodo de sustitución nuevo.

Aplicación	Disponibilidad durante la conmutación por error del nodo principal	Notas
Mahout	La conmutación por error del nodo principal no afecta a la disponibilidad	Dado que Mahout no tiene daemon, no se ve afectado por el proceso de conmutación por error del nodo principal.
MXNet	La conmutación por error del nodo principal no afecta a la disponibilidad	Dado que MXNet no tiene daemon, no se ve afectado por el proceso de conmutación por error del nodo principal.
Phoenix	Alta disponibilidad	QueryServer de Phoenix se ejecuta únicament e en uno de los tres nodos principales. En los tres nodos principales, Phoenix está configura do para conectar el QueryServer de Phoenix. Puede encontrar la IP privada del servidor de consultas de Phoenix mediante el archivo `/etc/phoenix/conf/phoenix-env.sh`
Pig	La conmutación por error del nodo principal no afecta a la disponibilidad	Dado que Pig no tiene daemon, no se ve afectado por el proceso de conmutación por error del nodo principal.
Spark	Alta disponibilidad	Todas las aplicaciones de Spark se ejecutan en contenedores de YARN y pueden reacciona r ante una conmutación por error del nodo principal de la misma forma que las caracterí sticas de alta disponibilidad de YARN.

Aplicación	Disponibilidad durante la conmutación por error del nodo principal	Notas
Sqoop	Alta disponibilidad	De forma predeterminada, sqoop-job y sqoop-metastore almacenan los datos (descripciones de trabajos) en el disco local del nodo principal que ejecuta el comando. Si desea guardar los datos del metaalmacén en una base de datos externa, consulte la documentación de Apache Sqoop.
Tez	Alta disponibilidad	Dado que los contenedores de Tez se ejecutan en YARN, Tez se comporta de la misma forma que YARN durante el proceso de conmutación por error del nodo principal.
TensorFlow	La conmutación por error del nodo principal no afecta a la disponibilidad	Dado que TensorFlow no tiene daemon, no se ve afectado por el proceso de conmutación por error del nodo principal.
Zeppelin	Alta disponibilidad	Zeppelin se ha instalado en los tres nodos principales. Zeppelin almacena las notas y las configuraciones de intérprete en HDFS de forma predeterminada para evitar que se pierdan datos. Las sesiones de intérpret e están completamente aisladas en las tres instancias principales. Los datos de la sesión se perderán en caso de error de la instancia principal. Se recomienda no modificar la misma nota simultáneamente en diferentes instancias principales.

Aplicación	Disponibilidad durante la conmutación por error del nodo principal	Notas
ZooKeeper	Alta disponibilidad	ZooKeeper es la base de la característica de conmutación por error automática de HDFS. ZooKeeper proporciona un servicio de alta disponibilidad para mantener los datos de coordinación, notificar a los clientes los cambios en esos datos y monitorizar los clientes para ver si se producen errores. Para más información, consulte Conmutación por error automática del HDFS.

Para ejecutar las siguientes aplicaciones en un clúster de Amazon EMR con varios nodos principales, debe configurar una base de datos externa. La base de datos externa se encuentra fuera del clúster y hace que persistan los datos durante el proceso de conmutación por error del nodo principal. Para las siguientes aplicaciones, los componentes de servicio se recuperarán automáticamente durante el proceso de conmutación por error del nodo principal, pero se pueden producir errores en los trabajos activos, en cuyo caso deberán repetirse.

Aplicación	Disponibilidad durante la conmutación por error del nodo principal	Notas
Hive	Alta disponibilidad únicamente para los componentes de servicio	Se requiere un metaalmacén externo para Hive. Debe ser un metaalmacén externo de MySQL, ya que PostgreSQL no es compatible con clústeres multimaestros. Para más información, consulte Configuración de un metaalmacén externo para Hive.
Hue	Alta disponibilidad únicamente para los componentes de servicio	Se requiere una base de datos externa para Hue. Para más información, consulte Uso de

Aplicación	Disponibilidad durante la conmutación por error del nodo principal	Notas
		Hue con una base de datos remota en Amazon RDS.
Oozie	Alta disponibilidad únicamente para los componentes de servicio	Se requiere una base de datos externa para Oozie. Para más información, consulte Uso de Oozie con una base de datos remota en Amazon RDS. Oozie-server y oozie-client se han instalado en los tres nodos principales. Los oozie-client están configurados para conectarse al oozie-server correcto de forma predeterminada.
PrestoDB o PrestoSQL/ Trino	Alta disponibilidad únicamente para los componentes de servicio	Se requiere un metaalmacén de Hive externo para PrestoDB (PrestoSQL en Amazon EMR 6.1.0-6.3.0 o Trino en Amazon EMR 6.4.0 y versiones posteriores). Puede utilizar Presto con el Catálogo de datos de AWS Glue o utilizar una base de datos MySQL externa para Hive. La CLI de Presto se ha instalado en los tres nodos principales a fin de que pueda utilizarla para acceder al coordinador de Presto desde cualquiera de los nodos principales. El coordinador de Presto se ha instalado en un solo nodo principal. Puede encontrar el nombre de DNS del nodo principal en el que se ha instalado el coordinador de Presto llamando a la API `describe-cluster` de Amazon EMR y leyendo el valor devuelto del campo `MasterPublicDnsName` en la respuesta.

> **Note**
>
> Cuando un nodo principal deja de funcionar, la conectividad de bases de datos Java (JDBC) o la conectividad de bases de datos abiertas (ODBC) termina su conexión con el nodo principal. Puede conectarse a cualquiera de los demás nodos principales para continuar su trabajo porque el daemon del metaalmacén de Hive se ejecuta en todos los nodos principales. También puede esperar a que se sustituya el nodo principal que ha dejado de funcionar.

Cómo funcionan las características de Amazon EMR en un clúster con varios nodos principales

Conexión con los nodos principales mediante SSH

Puede conectarse con cualquiera de los tres nodos principales de un clúster de Amazon EMR utilizando SSH de la misma forma que se conecta con un nodo principal único. Para más información, consulte Conectarse al nodo principal mediante SSH.

Si un nodo principal deja de funcionar, la conexión SSH a ese nodo principal finaliza. Para continuar con su trabajo, puede conectarse a uno de los otros dos nodos principales. De manera alternativa, puede tener acceso al nuevo nodo principal después de que Amazon EMR sustituya el que ha dejado de funcionar por uno nuevo.

> **Note**
>
> La dirección IP privada del nodo principal de sustitución es la misma que la del anterior. La dirección IP pública del nodo principal de sustitución puede cambiar. Puede recuperar las nuevas direcciones IP en la consola o mediante el comando `describe-cluster` de la CLI de AWS.
>
> NameNode solo se ejecuta en dos de los nodos principales. Sin embargo, puede ejecutar comandos `hdfs` en la CLI y ejecutar trabajos para tener acceso a HDFS en los tres nodos principales.

Cómo trabajar con pasos en un clúster de Amazon EMR con varios nodos principales

Puede enviar pasos a un clúster de Amazon EMR con varios nodos principales del mismo modo que trabaja con pasos en un clúster de un solo nodo principal. Para más información, consulte Enviar trabajo a un clúster.

A continuación se indican consideraciones para trabajar con los pasos de un clúster de Amazon EMR con varios nodos principales:

- Si un nodo principal deja de funcionar, los pasos que se están ejecutando en el nodo principal se marcan como FAILED. Los datos que se hayan escrito localmente se pierden. Sin embargo, el estado FAILED puede que no refleje el estado real de los pasos.
- Si un paso en ejecución ha iniciado una aplicación de YARN cuando el nodo principal deja de funcionar, el paso puede continuar y finalizar correctamente debido a la conmutación por error automática del nodo principal.
- Se recomienda que compruebe el estado de los pasos consultando la salida de los trabajos. Por ejemplo, los trabajos de MapReduce utilizan un archivo `_SUCCESS` para determinar si el trabajo finaliza correctamente.
- Se recomienda que establezca el parámetro ActionOnFailure en CONTINUE o CANCEL_AND_WAIT, en lugar de establecerlo en TERMINATE_JOB_FLOW o TERMINATE_CLUSTER.

Protección automática contra la terminación

Amazon EMR habilita automáticamente la protección contra la terminación para todos los clústeres con varios nodos principales y anula cualquier configuración de ejecución de pasos que proporcione al crear el clúster. Puede deshabilitar la protección contra la terminación después de que se haya lanzado el clúster. Consulte Configuración de la protección de terminación para ejecutar clústeres. Para cerrar un clúster con varios nodos principales, primero debe modificar los atributos del clúster para deshabilitar la protección contra la terminación. Para obtener instrucciones, consulte Terminar un clúster de Amazon EMR con varios nodos principales.

Para más información sobre la protección de terminación, consulte Uso de la protección de terminación.

Características no admitidas en un clúster de Amazon EMR con varios nodos principales

Las siguientes características de Amazon EMR no están disponibles actualmente en un clúster de Amazon EMR con varios nodos principales:

- EMR Notebooks
- Flotas de instancias
- Acceso de un clic al servidor del historial de Spark persistente
- Interfaces de usuario de aplicaciones persistentes
- El acceso con un solo clic a las interfaces de usuario de aplicaciones persistentes no está disponible actualmente para los clústeres de Amazon EMR con varios nodos principales ni para los clústeres de Amazon EMR integrados con AWS Lake Formation.

> **Note**
>
> Para utilizar la autenticación de Kerberos en un clúster, debe configurar un KDC externo. A partir de la versión 5.27.0 de Amazon EMR, puede configurar el cifrado transparente de HDFS en un clúster de Amazon EMR con varios nodos principales. Para más información, consulte Cifrado transparente en el HDFS en Amazon EMR.

Lanzar un clúster de Amazon EMR con varios nodos principales

En este tema se proporcionan detalles de configuración y ejemplos para lanzar un clúster de Amazon EMR con varios nodos principales.

> **Note**
>
> Amazon EMR habilita automáticamente la protección de terminación para todos los clústeres con varios nodos principales y anula cualquier configuración de terminación automática que proporcione al crear el clúster. Para cerrar un clúster con varios nodos principales, primero debe modificar los atributos del clúster para deshabilitar la protección contra la terminación. Para obtener instrucciones, consulte Terminar un clúster de Amazon EMR con varios nodos principales.

Requisitos previos

- Puede lanzar un clúster de Amazon EMR con varios nodos principales en subredes públicas y privadas de la VPC. EC2-Classic no se admite. Para lanzar un clúster de Amazon EMR con varios nodos principales en una subred pública, debe habilitar las instancias de dicha subred para que

reciban una dirección IP pública seleccionando Asignar automáticamente IPv4 en la consola o ejecutando del comando siguiente. Sustituya *22XXXX01* por su ID de subred.

```
aws ec2 modify-subnet-attribute --subnet-id subnet-22XXXX01 --map-public-ip-on-launch
```

- Para ejecutar Hive, Hue u Oozie en un clúster de Amazon EMR con varios nodos principales, debe crear un metaalmacén externo. Para más información, consulte Configuración de un metaalmacén externo para Hive, Uso de Hue con una base de datos remota en Amazon RDS o Apache Oozie.
- Para utilizar la autenticación de Kerberos en un clúster, debe configurar un KDC externo. Para más información, consulte Configuración de Kerberos en Amazon EMR.

Lanzar un clúster de Amazon EMR con varios nodos principales

Debe especificar un valor de recuento de instancias de tres para el grupo de instancias del nodo principal al lanzar un clúster de Amazon EMR con varios nodos principales. En los ejemplos siguientes, se muestra cómo lanzar el clúster utilizando la AMI predeterminada o una AMI personalizada.

> **Note**
>
> Debe especificar el ID de subred cuando lance un clúster de Amazon EMR con varios nodos principales mediante AWS CLI. Sustituya *22XXXX01* por su ID de subred en los ejemplos siguientes.

Example – Lanzamiento de un clúster de Amazon EMR con varios nodos principales mediante una AMI predeterminada

```
aws emr create-cluster \
--name "ha-cluster" \
--release-label emr-5.36.1 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

Example – Lanzamiento de un clúster de Amazon EMR con varios nodos principales mediante una AMI personalizada

```
aws emr create-cluster \
--name "custom-ami-ha-cluster" \
--release-label emr-5.36.1 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark \
--custom-ami-id ami-MyAmiID
```

Example – Lanzamiento de un clúster de Amazon EMR con varios nodos principales con un metaalmacén de Hive externo

Para ejecutar Hive en un clúster de Amazon EMR con varios nodos principales, debe especificar un metaalmacén externo para Hive, tal y como se muestra en el ejemplo siguiente,

1. Cree un archivo hiveConfiguration.json temporal que contenga las credenciales del metaalmacén de Hive.

   ```
   [
     {
       "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"
       }
     }
   ]
   ```

2. Lance el clúster con el metaalmacén de Hive.

   ```
   aws emr create-cluster \
   --name "ha-cluster-with-hive-metastore" \
   --release-label emr-5.36.1 \
   ```

```
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-
22XXXX01 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name= Spark Name=Hive \
--configurations ./hiveConfiguration.json
```

Terminar un clúster de Amazon EMR con varios nodos principales

Para terminar un clúster de Amazon EMR con varios nodos principales, debe desactivar la protección de terminación antes de terminar el clúster, tal y como se muestra en el ejemplo siguiente. Sustituya *j-3KVTXXXXXX7UG* por su ID de clúster.

```
aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-
protected
aws emr terminate-clusters --cluster-id j-3KVTXXXXXX7UG
```

Integración de Amazon EMR con grupos de ubicación de EC2

Al lanzar un clúster de Amazon EMR con varios nodos principales en Amazon EC2, tiene la opción de utilizar estrategias de grupos de ubicación para especificar cómo desea que se implementen las instancias de nodos principales para protegerlas contra los errores de hardware.

Las estrategias de grupos de ubicación se admiten a partir de la versión 5.23.0 de Amazon EMR como opción para clústeres con varios nodos principales. Actualmente, la estrategia de grupo de ubicación solo admite los tipos de nodos principales y la estrategia `SPREAD` se aplica a esos nodos. La estrategia `SPREAD` coloca un pequeño grupo de instancias en un hardware subyacente independiente para evitar la pérdida de varios nodos principales en caso de que se produzca un error de hardware. Tenga en cuenta que una solicitud de lanzamiento de instancia podría presentar errores si no hay suficiente hardware único para cumplir con la solicitud. Para más información sobre las estrategias y limitaciones de ubicación de EC2, consulte Grupos de ubicación en la Guía del usuario de EC2 para instancias de Linux.

Existe un límite inicial de Amazon EC2 de 500 clústeres habilitados para las estrategias de grupos de ubicación que se pueden lanzar por región de AWS. Póngase en contacto con la asistencia técnica de AWS para solicitar un aumento del número de grupos de ubicación permitidos. Puede identificar

los grupos de ubicación de EC2 que Amazon EMR crea mediante el seguimiento del par de clave-valor que Amazon EMR asocia a la estrategia de grupos de ubicación de Amazon EMR. Para más información acerca de las etiquetas de instancias de clúster de EC2, consulte Ver instancias del clúster en Amazon EC2.

Asociación de la política administrada del grupo de ubicación al rol de Amazon EMR

La estrategia de grupos de ubicación requiere una política administrada llamada `AmazonElasticMapReducePlacementGroupPolicy`, que permite a Amazon EMR crear, eliminar y describir grupos de ubicación en Amazon EC2. Debe asociar `AmazonElasticMapReducePlacementGroupPolicy` al rol de servicio de Amazon EMR antes de lanzar un clúster maestro múltiple de Amazon EMR.

También puede asociar la política administrada `AmazonEMRServicePolicy_v2` al rol de servicio de Amazon EMR en lugar de la política administrada del grupo de ubicación. `AmazonEMRServicePolicy_v2` permite el mismo acceso a los grupos de ubicación en Amazon EC2 que `AmazonElasticMapReducePlacementGroupPolicy`. Para obtener más información, consulte Rol de servicio para Amazon EMR (rol de EMR).

La política administrada `AmazonElasticMapReducePlacementGroupPolicy` es el siguiente texto JSON creado y administrado por Amazon EMR.

> **Note**
>
> Dado que la política administrada `AmazonElasticMapReducePlacementGroupPolicy` se actualiza automáticamente, la política aquí mostrada podría estar obsoleta. Utilice la consola de administración de AWS para ver la política actual.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Resource":"*",
         "Effect":"Allow",
         "Action":[
            "ec2:DeletePlacementGroup",
            "ec2:DescribePlacementGroups"
         ]
```

```
      },
      {
         "Resource":"arn:aws:ec2:*:*:placement-group/pg-*",
         "Effect":"Allow",
         "Action":[
            "ec2:CreatePlacementGroup"
         ]
      }
   ]
}
```

Lanzamiento de un clúster maestro múltiple de Amazon EMR con una estrategia de grupos de ubicación

Para lanzar un clúster maestro múltiple de Amazon EMR con una estrategia de grupos de ubicación, asocie la política administrada del grupo de ubicación `AmazonElasticMapReducePlacementGroupPolicy` al rol de Amazon EMR. Para obtener más información, consulte Asociación de la política administrada del grupo de ubicación al rol de Amazon EMR.

Cada vez que utiliza este rol para iniciar un clúster maestro múltiple de Amazon EMR, Amazon EMR intenta lanzar un clúster con una estrategia `SPREAD` aplicada a sus nodos principales. Si utiliza un rol que no tiene la política administrada del grupo de ubicación `AmazonElasticMapReducePlacementGroupPolicy` asociada a él, Amazon EMR intentará lanzar un clúster maestro múltiple de Amazon EMR sin una estrategia de grupo de ubicación.

Si lanza un clúster maestro múltiple de Amazon EMR con el parámetro `placement-group-configs` mediante la API o la CLI de Amazon EMR, Amazon EMR solo lanza el clúster si el rol de Amazon EMR tiene asociada la política administrada del grupo de ubicación `AmazonElasticMapReducePlacementGroupPolicy`. Si el rol de Amazon EMR no tiene la política asociada, se produce un error en el inicio del clúster maestro múltiple de Amazon EMR.

Example — Lanzamiento de un clúster maestro múltiple de Amazon EMR con una estrategia de grupos de ubicación mediante la API de Amazon EMR.

Cuando utilice la acción RunJobFlow para crear un clúster maestro múltiple de Amazon EMR, defina la propiedad `PlacementGroupConfigs` en lo siguiente. Actualmente, el rol de instancia `MASTER` utiliza automáticamente `SPREAD` como estrategia de grupos de ubicación.

```
{
```

```
   "Name":"ha-cluster",
   "PlacementGroupConfigs":[
      {
         "InstanceRole":"MASTER"
      }
   ],
   "ReleaseLabel":"emr-5.30.1",
   "Instances":{
      "ec2SubnetId":"subnet-22XXXX01",
      "ec2KeyName":"ec2_key_pair_name",
      "InstanceGroups":[
         {
            "InstanceCount":3,
            "InstanceRole":"MASTER",
            "InstanceType":"m5.xlarge"
         },
         {
            "InstanceCount":4,
            "InstanceRole":"CORE",
            "InstanceType":"m5.xlarge"
         }
      ]
   },
   "JobFlowRole":"EMR_EC2_DefaultRole",
   "ServiceRole":"EMR_DefaultRole"
}
```

- Sustituya *ha-cluster* por el nombre de su clúster de alta disponibilidad.
- Sustituya *subnet-22XXXX01* con su ID de subred.
- Sustituya *ec2_key_pair_name* con el nombre de su par de claves de EC2 para este clúster. El par de claves de EC2 es opcional y solo se precisa si desea utilizar SSH para acceder a su clúster.

Example — Lanzamiento de un clúster con varios nodos principales con una estrategia de grupos de ubicación mediante la CLI de Amazon EMR.

```
aws emr create-cluster \
--name "ha-cluster" \
--placement-group-configs InstanceRole=MASTER \
--release-label emr-5.30.1 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
```

```
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

- Sustituya *`ha-cluster`* por el nombre de su clúster de alta disponibilidad.
- Sustituya *`subnet-22XXXX01`* con su ID de subred.
- Sustituya *`ec2_key_pair_name`* con el nombre de su par de claves de EC2 para este clúster. El par de claves de EC2 es opcional y solo se precisa si desea utilizar SSH para acceder a su clúster.

Lanzamiento de un clúster con varios nodos principales sin una estrategia de grupos de ubicación

Para que un clúster con varios nodos principales lance nodos principales sin la estrategia de grupos de ubicación, debe realizar una de las siguientes acciones:

- Eliminar la política administrada `AmazonElasticMapReducePlacementGroupPolicy` del grupo de ubicación del rol de Amazon EMR o
- Lanzar un clúster con varios nodos principales con el parámetro `placement-group-configs` mediante la API o la CLI de Amazon EMR y con `NONE` como estrategia de grupos de ubicación.

Example — Lanzamiento de un clúster con varios nodos principales sin una estrategia de grupos de ubicación mediante la API de Amazon EMR.

Cuando utilice la acción RunJobFlow para crear un clúster con varios nodos principales, defina la propiedad `PlacementGroupConfigs` en lo siguiente.

```
{
   "Name":"ha-cluster",
   "PlacementGroupConfigs":[
      {
         "InstanceRole":"MASTER",
         "PlacementStrategy":"NONE"
      }
   ],
   "ReleaseLabel":"emr-5.30.1",
   "Instances":{
      "ec2SubnetId":"subnet-22XXXX01",
```

```
      "ec2KeyName":"ec2_key_pair_name",
      "InstanceGroups":[
         {
            "InstanceCount":3,
            "InstanceRole":"MASTER",
            "InstanceType":"m5.xlarge"
         },
         {
            "InstanceCount":4,
            "InstanceRole":"CORE",
            "InstanceType":"m5.xlarge"
         }
      ]
   },
   "JobFlowRole":"EMR_EC2_DefaultRole",
   "ServiceRole":"EMR_DefaultRole"
}
```

- Sustituya *ha-cluster* por el nombre de su clúster de alta disponibilidad.
- Sustituya *subnet-22XXXX01* con su ID de subred.
- Sustituya *ec2_key_pair_name* con el nombre de su par de claves de EC2 para este clúster. El par de claves de EC2 es opcional y solo se precisa si desea utilizar SSH para acceder a su clúster.

Example — Lanzamiento de un clúster con varios nodos principales sin una estrategia de grupos de ubicación mediante la CLI de Amazon EMR.

```
aws emr create-cluster \
--name "ha-cluster" \
--placement-group-configs InstanceRole=MASTER,PlacementStrategy=NONE \
--release-label emr-5.30.1 \
--instance-groups InstanceGroupType=MASTER,InstanceCount=3,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=4,InstanceType=m5.xlarge \
--ec2-attributes
 KeyName=ec2_key_pair_name,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=subnet-22XXXX01
 \
--service-role EMR_DefaultRole \
--applications Name=Hadoop Name=Spark
```

- Sustituya *ha-cluster* por el nombre de su clúster de alta disponibilidad.
- Sustituya *subnet-22XXXX01* con su ID de subred.

- Sustituya *ec2_key_pair_name* con el nombre de su par de claves de EC2 para este clúster. El par de claves de EC2 es opcional y solo se precisa si desea utilizar SSH para acceder a su clúster.

Comprobación de la configuración de la estrategia de grupos de ubicación asociada al clúster con varios nodos principales

Puede utilizar la API de descripción del clúster de Amazon EMR para ver la configuración de la estrategia de grupos de ubicación asociada al clúster con varios nodos principales.

Example

```
aws emr describe-cluster --cluster-id "j-xxxxx"
{
   "Cluster":{
      "Id":"j-xxxxx",
      ...
      ...
      "PlacementGroups":[
         {
            "InstanceRole":"MASTER",
            "PlacementStrategy":"SPREAD"
         }
      ]
   }
}
```

Consideraciones y prácticas recomendadas

Limitaciones de un clúster de Amazon EMR con varios nodos principales:

- No puede utilizar un clúster de Amazon EMR con varios nodos principales con flotas de instancias. Para más información sobre las capacidades de Amazon EMR que funcionan con varios nodos principales, consulte Aplicaciones y características compatibles.
- Si dos nodos principales dejan de funcionar simultáneamente, Amazon EMR no puede recuperar el clúster.
- Los clústeres de Amazon EMR con varios nodos principales no toleran errores de zonas de disponibilidad. En el caso de que deje de funcionar una zona de disponibilidad, se perderá el acceso al clúster de EMR.

- Amazon EMR no garantiza las características de alta disponibilidad de aplicaciones de código abierto distintas de las especificadas en Aplicaciones admitidas en un clúster de Amazon EMR con varios nodos principales.
- En las versiones 5.23.0 hasta la 5.30.1 de Amazon EMR, solo dos de los tres nodos principales ejecutan HDFS NameNode.

Consideraciones para la configuración la subred:

- Un clúster de Amazon EMR con varios nodos principales solo puede residir en una zona de disponibilidad o subred. Amazon EMR no puede sustituir un nodo principal que ha dejado de funcionar si la subred se utiliza en su totalidad o por encima de su capacidad en caso de que se produzca una conmutación por error. Para evitar este escenario, se recomienda que dedique una subred completa a un clúster de Amazon EMR. Además, debe asegurarse de que haya suficientes direcciones IP privadas disponibles en la subred.

Consideraciones para configurar los nodos principales:

- Para garantizar que el grupo de instancias del nodo principal también tenga un alto nivel de disponibilidad, se recomienda que lance al menos cuatro nodos principales. Si decide lanzar un clúster más pequeño con tres o menos nodos básicos, configure `dfs.replication parameter` para un mínimo de 2 a fin de que HDFS tenga suficiente replicación DFS. Para más información, consulte Configuración de HDFS.

> ⚠ Warning
>
> 1. Establecer `dfs.replication` en 1 en clústeres con menos de cuatro nodos puede conllevar la pérdida de datos del HDFS si un solo nodo deja de funcionar. Se recomienda que utilice un clúster con al menos cuatro nodos principales para las cargas de trabajo de producción.
> 2. Amazon EMR no permitirá que los clústeres escalen los nodos principales por debajo de `dfs.replication`. Por ejemplo, si `dfs.replication = 2`, el número mínimo de nodos principales es 2.

3. Cuando utiliza el escalado administrado, el escalado automático o decide cambiar el tamaño del clúster manualmente, se recomienda que establezca `dfs.replication` en 2 o más.

Consideraciones para configurar alarmas de métricas:

- Amazon EMR no proporciona métricas específicas para las aplicaciones relacionadas con HDFS o YARN. Se recomienda que configure alarmas para monitorizar el recuento de instancias del nodo principal. Configure las alarmas utilizando las siguientes métricas de Amazon CloudWatch: `MultiMasterInstanceGroupNodesRunning`, `MultiMasterInstanceGroupNodesRunningPercentage` o `MultiMasterInstanceGroupNodesRequested`. CloudWatch le notificará en el caso de que un nodo principal deje de funcionar y sea sustituido.
  - Si el `MultiMasterInstanceGroupNodesRunningPercentage` es inferior a 1,0 y superior a 0,5, es posible que el clúster haya perdido un nodo principal. En esta situación, Amazon EMR intentará sustituir un nodo principal.
  - Si el `MultiMasterInstanceGroupNodesRunningPercentage` cae por debajo de 0,5, es posible que hayan dejado de funcionar dos nodos principales. En esta situación, se pierde el cuórum y el clúster no se puede recuperar. Debe migrar manualmente los datos de este clúster.

  Para más información, consulte Configuración de alarmas en métricas.

Clústeres de EMR en AWS Outposts

A partir de la versión 5.28.0 de Amazon EMR, puede crear y ejecutar clústeres de EMR en AWS Outposts. AWS Outposts habilita los servicios, la infraestructura y los modelos operativos nativos de AWS en las instalaciones. En los entornos de AWS Outposts, puede utilizar las mismas API, herramientas e infraestructura de AWS que utiliza en la nube de AWS. Amazon EMR en AWS Outposts es ideal para cargas de trabajo de baja latencia que deben ejecutarse cerca de los datos y las aplicaciones en las instalaciones. Para obtener más información sobre AWS Outposts, consulte la Guía del usuario de AWS Outposts.

Requisitos previos

Estos son los requisitos previos para utilizar Amazon EMR en AWS Outposts:

- Debe haber instalado y configurado AWS Outposts en el centro de datos local.
- Debe tener una conexión de red fiable entre su entorno de Outpost y una región de AWS.
- Debe tener capacidad suficiente para los tipos de instancias compatibles con EMR disponibles en su Outpost.

Limitaciones

A continuación, se indican las limitaciones de uso de Amazon EMR en AWS Outposts:

- Las instancias bajo demanda son la única opción admitida para las instancias de Amazon EC2. Las instancias de spot no están disponibles para Amazon EMR en AWS Outposts.
- Si necesita volúmenes de almacenamiento de Amazon EBS adicionales, solo se admite SSD de uso general (GP2).
- Los buckets de S3 que almacenan objetos en una Región de AWS que usted especifique son la única opción de S3 compatible con Amazon EMR en Outposts. S3 en Outposts no es compatible con Amazon EMR en AWS Outposts.
- Únicamente los siguientes tipos de instancias son admitidas por Amazon EMR en AWS Outposts:

Clase de instancia	Tipos de instancias
Uso general	`m5.xlarge |m5.2xlarge | m5.4xlarge |m5.12xlarge | m5.24xlarge |m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge |m5d.24xlarge`
Optimizada para computación	`c5.xlarge |c5.2xlarge | c5.4xlarge |c5.18xlarge | c5d.xlarge |c5d.2xlarge | c5d.4xlarge | c5d.18xlarge`
Optimizada para memoria	`r5.xlarge |r5.2xlarge | r5.4xlarge |r5.12xlarge | r5d.xlarge |r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge`
Con optimización del almacenamiento	`i3en.xlarge |i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge |i3en.24xlarge`

Clase de instancia	Tipos de instancias

Consideraciones sobre la conectividad de red

- Si se pierde la conectividad de red entre su Outpost y su región de AWS, sus clústeres se seguirán ejecutando. Sin embargo, no podrá crear nuevos clústeres ni realizar nuevas acciones en clústeres existentes hasta que se restablezca la conectividad. En caso de errores en la instancia, la instancia no se reemplazará automáticamente. Además, se retrasarán algunas acciones, como agregar pasos a un clúster en ejecución, comprobar el estado de ejecución de pasos o enviar métricas y eventos de CloudWatch.
- Se recomienda que proporcione una conectividad de red de confianza y de alta disponibilidad entre la instancia de Outpost y la región de AWS. Si la conectividad de red entre su Outpost y su región de AWS se pierde durante un tiempo superior a algunas horas, los clústeres que tienen habilitada la protección de terminación seguirán ejecutándose y aquellos que la tengan desactivada se terminarán.
- Si la conectividad de red se ve afectada debido a un mantenimiento rutinario, se recomienda habilitar proactivamente la protección de terminación. De manera más general, la interrupción de la conectividad significa que no se podrá acceder a ninguna dependencia externa que no esté accesible localmente en la instancia de Outpost o en la red del cliente. Esto incluye Amazon S3, DynamoDB (utilizado con la vista de coherencia de EMRFS) y Amazon RDS, si se utiliza una instancia regional para un clúster de Amazon EMR con varios nodos principales.

Creación de un clúster de Amazon EMR en AWS Outposts

La creación de un clúster de Amazon EMR en AWS Outposts es similar a la creación de un clúster de Amazon EMR en la nube de AWS. Al crear un clúster de Amazon EMR en AWS Outposts, debe especificar una subred de Amazon EC2 asociada a su Outpost.

Una VPC de Amazon puede abarcar todas las zonas de disponibilidad de una región de AWS. AWS Outposts son extensiones de las zonas de disponibilidad; por ello, puede ampliar una VPC de Amazon de una cuenta para que abarque varias zonas de disponibilidad y ubicaciones de Outpost asociadas. Al configurar su Outpost, le asocia un grupo de subredes para ampliar su entorno de VPC regional a sus instalaciones. Las instancias de Outpost y los servicios relacionados aparecen como parte de su VPC regional, de manera similar a una zona de disponibilidad con subredes asociadas. Para obtener más información, consulte la Guía del usuario de AWS Outposts.

Consola

Para crear un nuevo clúster de Amazon EMR en AWS Outposts con la AWS Management Console, especifique una subred de Amazon EC2 asociada a la instancia de Outpost.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para crear un clúster en AWS Outposts con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Configuración del clúster, seleccione Grupos de instancias o Flotas de instancias. A continuación, elija un tipo de instancia en el menú desplegable Elegir tipo de instancia de EC2 o seleccione Acciones y elija Agregar volúmenes de EBS. Amazon EMR en AWS Outposts admite tipos limitados de volúmenes e instancias de Amazon EBS.
4. En Redes, seleccione una subred de EC2 con un ID de Outposts en este formato: op-123456789.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Old console

Para crear un clúster en AWS Outposts con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.

3. Elija Go to advanced options (Ir a las opciones avanzadas).
4. En Software Configuration (Configuración de software), elija 5.28.0 o una versión posterior en Release (Versión).
5. En Configuración de hardware, para la Subred de EC2, seleccione una subred de EC2 con un ID de Outpost en este formato: op-123456789.
6. Elija el tipo de instancia o agregue volúmenes de almacenamiento de Amazon EBS para grupos de instancias uniformes o flotas de instancias. Amazon EMR en AWS Outposts admite tipos limitados de volúmenes e instancias de Amazon EBS.

CLI

Para crear un clúster en AWS Outposts con la AWS CLI

- Para crear un nuevo clúster de Amazon EMR en AWS Outposts con la AWS CLI, especifique una subred de EC2 que esté asociada a su Outpost, como en el ejemplo siguiente. Sustituya *subnet-22XXXX01* con su ID de subred de EC2.

```
aws emr create-cluster \
--name "Outpost cluster" \
--release-label emr-5.36.1 \
--applications Name=Spark \
--ec2-attributes KeyName=myKey SubnetId=subnet-22XXXX01 \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Clústeres de EMR en Zonas locales de AWS

A partir de la versión 5.28.0 de Amazon EMR, podrá crear y ejecutar clústeres de Amazon EMR en una subred de Zonas locales de AWS como una extensión lógica de una región de AWS compatible con zonas locales. Una zona local permite que las características de Amazon EMR y un subconjunto de servicios de AWS, como los servicios informáticos y de almacenamiento, se encuentren más cerca de los usuarios, lo que proporciona un acceso de latencia muy baja a las aplicaciones que se ejecutan localmente. Para obtener una lista de las zonas locales disponibles, consulte Zonas locales de AWS. Para obtener información acerca del acceso a zonas locales de AWS disponibles, consulte Regiones, zonas de disponibilidad y zonas locales.

Tipos de instancias admitidas

Los siguientes tipos de instancias están disponibles para los clústeres de Amazon EMR en zonas locales. La disponibilidad de los distintos tipos de instancias puede variar según la región.

Clase de instancia	Tipos de instancias
Uso general	m5.xlarge | m5.2xlarge | m5.4xlarge | m5.12xlarge | m5.24xlarge | m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge | m5d.24xlarge
Optimizada para computación	c5.xlarge | c5.2xlarge | c5.4xlarge | c5.9xlarge | c5.18xlarge | c5d.xlarge | c5d.2xlarge | c5d.4xlarge| c5d.9xlarge | c5d.18xlarge
Optimizada para memoria	r5.xlarge | r5.2xlarge | r5.4xlarge | r5.12xlarge | r5d.xlarge | r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge
Con optimización del almacenamiento	i3en.xlarge | i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge | i3en.24xlarge

Creación de un clúster de Amazon EMR en zonas locales

Cree un clúster de Amazon EMR en Zonas locales de AWS; para ello, lance el clúster de Amazon EMR en una subred de Amazon VPC asociada a una zona local. Puede acceder al clúster utilizando el nombre de la zona local; por ejemplo, us-west-2-lax-1a en la consola de Oeste de EE. UU. (Oregón).

Las zonas locales no son compatibles con cuadernos de Amazon EMR ni admiten conexiones directamente a Amazon EMR utilizando el punto de conexión de VPC de la interfaz (AWS PrivateLink).

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para crear un clúster en una zona local con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Redes, seleccione una subred de EC2 con un ID de zona local en este formato: subnet 123abc | us-west-2-lax-1a.
4. Elija un tipo de instancia o agregue volúmenes de almacenamiento de Amazon EBS para grupos de instancias uniformes o flotas de instancias.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Old console

Para crear un clúster en una zona local con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Go to advanced options (Ir a las opciones avanzadas).
4. En Software Configuration (Configuración de software), elija 5.28.0 o una versión posterior en Release (Versión).
5. En Configuración de hardware, para la Subred de EC2, seleccione una subred de EC2 con un ID de zona local en este formato: subnet 123abc | us-west-2-lax-1a.
6. Agregue volúmenes de almacenamiento de Amazon EBS para grupos de instancias uniformes o flotas de instancias y elija un tipo de instancia.

CLI

Para crear un clúster en una zona local con la AWS CLI

- Utilice el comando create-cluster, junto con el SubnetId de la zona local como se muestra en el ejemplo siguiente. Sustituya subnet-22XXXX1234567 por el SubnetId de la zona local y sustituya otras opciones según sea necesario. Para obtener más información, consulte https://docs.aws.amazon.com/cli/latest/reference/emr/create-cluster.html.

```
aws emr create-cluster \
--name "Local Zones cluster" \
--release-label emr-5.29.0 \
--applications Name=Spark \
--ec2-attributes KeyName=myKey,SubnetId=subnet-22XXXX1234567 \
--instance-type m5.xlarge --instance-count 3 --use-default-roles
```

Configuración de Docker

Amazon EMR 6.x es compatible con Hadoop 3, que permite que YARN NodeManager lance contenedores, ya sea directamente en el clúster de Amazon EMR o bien dentro de un contenedor de Docker. Los contenedores de Docker proporcionan entornos de ejecución personalizados en los que se ejecuta el código de aplicación. El entorno de ejecución personalizado está aislado del entorno de ejecución de YARN NodeManager y otras aplicaciones.

Los contenedores de Docker pueden incluir bibliotecas especiales utilizadas por la aplicación y proporcionar diferentes versiones de las herramientas y las bibliotecas nativas, como R y Python. Puede utilizar las herramientas de Docker conocidas para definir dependencias en tiempo de ejecución y bibliotecas para sus aplicaciones.

Los clústeres de Amazon EMR 6.x se han configurado de forma predeterminada para permitir que las aplicaciones de YARN, como Spark, se ejecuten utilizando contenedores de Docker. Para personalizar la configuración del contenedor, edite las opciones de soporte de Docker definidas en los archivos `container-executor.cfg` y `yarn-site.xml` disponibles en el directorio `/etc/hadoop/conf`. Para obtener información detallada sobre cada opción de configuración y cómo se utiliza, consulte Lanzamiento de aplicaciones mediante contenedores de Docker.

Si lo desea, puede usar Docker cuando envíe un trabajo. Use las siguientes variables para especificar el tiempo de ejecución de Docker y la imagen de Docker.

- `YARN_CONTAINER_RUNTIME_TYPE=docker`
- `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`

Cuando usa contenedores de Docker para ejecutar sus aplicaciones de YARN, YARN descarga la imagen de Docker que se ha especificado al enviar el trabajo. Para que YARN resuelva esta imagen de Docker, debe configurarse con un registro de Docker. Las opciones de configuración de un registro de Docker dependen de si el clúster se ha implementado mediante una subred pública o privada.

Registros de Docker

Un registro de Docker es un sistema de almacenamiento y distribución para imágenes de Docker. En el caso de Amazon EMR, se recomienda usar Amazon ECR, que es un registro de contenedores de Docker completamente administrado que le permite crear sus propias imágenes personalizadas y alojarlas en una arquitectura escalable y de alta disponibilidad.

Consideraciones sobre la implementación

Los registros de Docker requieren acceso a la red desde cada host del clúster. Esto se debe a que cada host descarga imágenes del registro de Docker cuando la aplicación de YARN se ejecuta en el clúster. Estos requisitos de conectividad de red pueden limitar la elección del registro de Docker, dependiendo de si el clúster de Amazon EMR se ha implementado en una subred pública o privada.

Public subnet (Subred pública)

Cuando los clústeres de EMR se implementan en una subred pública, los nodos que ejecutan YARN NodeManager pueden acceder directamente a cualquier registro disponible a través de Internet.

Subred privada

Cuando los clústeres de EMR se implementan en una subred privada, los nodos que ejecutan YARN NodeManager no tienen acceso directo a Internet. Las imágenes de Docker se pueden alojar en Amazon ECR y se puede acceder a ellas a través de AWS PrivateLink.

Para más información acerca de cómo utilizar AWS PrivateLink para permitir el acceso a Amazon ECR en un escenario de subred privada, consulte Configuración de AWS PrivateLink para Amazon ECS y Amazon ECR.

Configuración de registros de Docker

Para utilizar registros de Docker con Amazon EMR, debe configurar Docker de tal forma que confíe en el registro específico que desea utilizar para resolver las imágenes de Docker. Los registros de confianza predeterminados son locales (privados) y de Centos. Para utilizar otros repositorios públicos o Amazon ECR, puede anular la configuración de `docker.trusted.registries` en `/etc/hadoop/conf/container-executor.cfg` mediante la API de clasificación de EMR con la clave de clasificación `container-executor`.

En el ejemplo siguiente se muestra cómo configurar el clúster de tal forma que confíe tanto en un repositorio público denominado `your-public-repo`, como en un punto de enlace de registro de ECR, `123456789123.dkr.ecr.us-east-1.amazonaws.com`. Si utiliza ECR, sustituya este punto de enlace por su punto de enlace de ECR específico.

```
[
  {
    "Classification": "container-executor",
    "Configurations": [
        {
            "Classification": "docker",
            "Properties": {
                "docker.trusted.registries": "local,centos,your-public-
repo,123456789123.dkr.ecr.us-east-1.amazonaws.com",
                "docker.privileged-containers.registries": "local,centos,your-public-
repo,123456789123.dkr.ecr.us-east-1.amazonaws.com"
            }
        }
    ]
  }
]
```

Para lanzar un clúster de Amazon EMR 6.0.0 con esta configuración mediante la AWS Command Line Interface (AWS CLI), cree un archivo llamado `container-executor.json` con el contenido de la configuración container-executor de JSON anterior. A continuación, utilice los siguientes comandos para lanzar el clúster.

```
export KEYPAIR=<Name of your Amazon EC2 key-pair>
export SUBNET_ID=<ID of the subnet to which to deploy the cluster>
export INSTANCE_TYPE=<Name of the instance type to use>
export REGION=<Region to which to deploy the cluster>
```

```
aws emr create-cluster \
    --name "EMR-6.0.0" \
    --region $REGION \
    --release-label emr-6.0.0 \
    --applications Name=Hadoop Name=Spark \
    --service-role EMR_DefaultRole \
    --ec2-attributes KeyName=$KEYPAIR,InstanceProfile=EMR_EC2_DefaultRole,SubnetId=
$SUBNET_ID \
    --instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=
$INSTANCE_TYPE InstanceGroupType=CORE,InstanceCount=2,InstanceType=$INSTANCE_TYPE \
    --configuration file://container-executor.json
```

Configuración de YARN para acceder a Amazon ECR en EMR 6.0.0 y versiones anteriores

Si es la primera vez que usa Amazon ECR, siga las instrucciones de Introducción a Amazon ECR y compruebe que dispone de acceso a Amazon ECR desde cada instancia del clúster de Amazon EMR.

En EMR 6.0.0 y versiones anteriores, para acceder a Amazon ECR mediante el comando de Docker, antes debe generar unas credenciales. Para verificar que YARN pueda acceder a las imágenes desde Amazon ECR, utilice la variable de entorno contenedora `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` para pasar una referencia a las credenciales que ha generado.

Ejecute el siguiente comando en uno de los nodos principales para obtener la línea de inicio de sesión de su cuenta de ECR.

```
aws ecr get-login --region us-east-1 --no-include-email
```

El comando `get-login` genera el comando correcto de la CLI de Docker que hay que ejecutar para crear las credenciales. Copie y ejecute la salida desde `get-login`.

```
sudo docker login -u AWS -p <password> https://<account-id>.dkr.ecr.us-
east-1.amazonaws.com
```

Este comando genera un archivo `config.json` en la carpeta `/root/.docker`. Copie este archivo en HDFS para que los trabajos enviados al clúster puedan usarlo para autenticarse en Amazon ECR.

Ejecute los siguientes comandos para copiar el archivo `config.json` en su directorio de inicio.

```
mkdir -p ~/.docker
sudo cp /root/.docker/config.json ~/.docker/config.json
sudo chmod 644 ~/.docker/config.json
```

Ejecute los siguientes comandos para poner el archivo config.json en HDFS, de tal forma que puedan usarlo los trabajos que se ejecutan en el clúster.

```
hadoop fs -put ~/.docker/config.json /user/hadoop/
```

YARN puede acceder a ECR como un registro de imágenes de Docker y extraer contenedores durante la ejecución del trabajo.

Después de configurar los registros de Docker y YARN, puede ejecutar las aplicaciones de YARN utilizando contenedores de Docker. Para más información, consulte Ejecutar aplicaciones de Spark con Docker mediante Amazon EMR 6.0.0.

En EMR 6.1.0 y versiones posteriores, no es necesario configurar manualmente la autenticación en Amazon ECR. Si se detecta un registro de Amazon ECR en la clave de clasificación `container-executor`, se activa la característica de autenticación automática de Amazon ECR y YARN se encarga del proceso de autenticación cuando se envía un trabajo de Spark con una imagen de ECR. Para confirmar si la autenticación automática está habilitada, compruebe `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` en yarn-site. La autenticación automática está habilitada y la configuración de autenticación de YARN se establece en `true` si `docker.trusted.registries` contiene una URL de registro de ECR.

Requisitos previos para utilizar la autenticación automática en Amazon ECR

- EMR versión 6.1.0 o posterior
- El registro de ECR incluido en la configuración se encuentra en la misma región que el clúster
- Rol de IAM con permisos para obtener el token de autorización y extraer cualquier imagen

Consulte Configuración con Amazon ECR para más información.

Cómo habilitar la autenticación automática

Siga Configuración de registros de Docker para establecer un registro de Amazon ECR como registro de confianza y asegúrese de que el repositorio de Amazon ECR y el clúster estén en la misma región.

Para habilitar esta característica incluso cuando el registro de ECR no esté establecido en el registro de confianza, utilice la clasificación de configuración para establecer `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` en `true`.

Cómo desactivar la autenticación automática

De forma predeterminada, la autenticación automática se deshabilita si no se detecta ningún registro de Amazon ECR en el registro de confianza.

Para desactivar la autenticación automática, incluso cuando el registro ECR se haya establecido en el registro de confianza, utilice la clasificación de configuración para establecer `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` en `false`.

Cómo comprobar si la autenticación automática está habilitada en un clúster

En el nodo maestro, utilice un editor de texto como `vi` para ver el contenido del archivo: `vi /etc/hadoop/conf.empty/yarn-site.xml`. Compruebe el valor de `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled`.

Control de la terminación de los clústeres

En esta sección se describen las opciones para cerrar los clústeres de Amazon EMR. Abarca la terminación automática y la protección contra la terminación, y la forma en que interactúan con otras características de Amazon EMR.

Puede cerrar un clúster de Amazon EMR de las siguientes maneras:

- Terminación tras la ejecución del último paso: cree un clúster transitorio que se cierre una vez completados todos los pasos.
- Terminación automática (después de la inactividad): cree un clúster con una política de terminación automática que se cierre después de un tiempo de inactividad específico. Para obtener más información, consulte Uso de una política de terminación automática.
- Terminación manual: cree un clúster de ejecución prolongada que continúe ejecutándose hasta que lo termine deliberadamente. Para obtener información sobre cómo terminar un clúster de forma manual, consulte Terminar un clúster.

También puede configurar la protección contra la terminación en un clúster para evitar el cierre de las instancias de EC2 por accidente o error.

Cuando Amazon EMR cierra el clúster, se cierran todas las instancias de Amazon EC2 del clúster. Los datos del almacén de instancias y de los volúmenes de EBS ya no están disponibles ni se pueden recuperar. La comprensión y administración de la terminación de los clústeres es fundamental para desarrollar una estrategia que permita administrar y preservar los datos escribiéndolos en Amazon S3 y equilibrando los costos.

Temas

- Configuración de un clúster para que continúe o termine después de la ejecución de pasos
- Uso de una política de terminación automática
- Uso de la protección de terminación

Configuración de un clúster para que continúe o termine después de la ejecución de pasos

En este tema se explican las diferencias entre utilizar un clúster de ejecución prolongada y crear un clúster transitorio que se cierre después de ejecutar el último paso. También se explica cómo configurar la ejecución de pasos de un clúster.

Crear un clúster en ejecución prolongada

De forma predeterminada, los clústeres que se crean con la consola o AWS CLI son de ejecución prolongada. Los clústeres de ejecución prolongada siguen funcionando, aceptando trabajo y acumulando gastos hasta que tome medidas para cerrarlos.

Un clúster de ejecución prolongada es eficaz en las siguientes situaciones:

- Cuando necesite consultar datos de forma interactiva o automática.
- Cuando necesite interactuar con aplicaciones de macrodatos alojadas en el clúster de forma continua.
- Cuando procesa periódicamente un conjunto de datos tan grande o con tanta frecuencia que resulta poco eficiente lanzar clústeres nuevos y cargar los datos cada vez.

También puede configurar la protección de terminación en un clúster de larga duración para evitar el cierre de las instancias de EC2 por accidente o error. Para obtener más información, consulte Uso de la protección de terminación.

> **Note**
>
> Amazon EMR habilita automáticamente la protección contra la terminación para todos los clústeres con varios nodos principales y anula cualquier configuración de ejecución de pasos que proporcione al crear el clúster. Puede deshabilitar la protección contra la terminación después de que se haya lanzado el clúster. Consulte Configuración de la protección de terminación para ejecutar clústeres. Para cerrar un clúster con varios nodos principales, primero debe modificar los atributos del clúster para deshabilitar la protección contra la terminación. Para obtener instrucciones, consulte Terminar un clúster de Amazon EMR con varios nodos principales.

Configuración de un clúster para que termine tras la ejecución de pasos

Al configurar la terminación tras la ejecución de pasos, el clúster se inicia, ejecuta las acciones de arranque y, a continuación, ejecuta los pasos que especifique. En cuanto se complete el último paso, Amazon EMR termina las instancias de Amazon EC2 del clúster. Los clústeres que lance con la API de Amazon EMR tienen habilitada la ejecución de pasos de forma predeterminada.

La terminación posterior a la ejecución de pasos es eficaz para los clústeres que realizan una tarea de procesamiento periódica, como un procesamiento de datos diario. La ejecución de pasos también lo ayuda a garantizar que solo se le facture el tiempo necesario para procesar sus datos. Para obtener más información acerca de los pasos, consulte Enviar trabajo a un clúster.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para activar la ejecución de pasos con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Pasos, seleccione Agregar paso. En el cuadro de diálogo Agregar paso, introduzca los valores en los campos adecuados. Las opciones varían según el tipo de paso. Para agregar el paso y salir del cuadro de diálogo, elija Agregar paso.
4. En Terminación del clúster, seleccione la casilla Terminar el clúster tras completar el último paso.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Old console

Para activar la ejecución de pasos con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Step execution (Ejecución de pasos).
4. Elija las demás opciones que sean necesarias para la aplicación y, a continuación, elija Create cluster (Crear clúster).

AWS CLI

Para activar la ejecución de pasos con AWS CLI

- Especifique el parámetro `--auto-terminate` cuando utilice el comando `create-cluster` para crear un clúster transitorio.

  El siguiente ejemplo muestra el uso del parámetro `--auto-terminate`. Puede escribir el comando siguiente y sustituir *myKey* por el nombre de su par de claves de EC2.

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 \
--applications Name=Hive Name=Pig --use-default-roles --ec2-attributes
 KeyName=myKey \
--steps Type=PIG,Name="Pig Program",ActionOnFailure=CONTINUE,\
Args=[-f,s3://mybucket/scripts/pigscript.pig,-p,\
INPUT=s3://mybucket/inputdata/,-p,OUTPUT=s3://mybucket/outputdata/,\
$INPUT=s3://mybucket/inputdata/,$OUTPUT=s3://mybucket/outputdata/]
--instance-type m5.xlarge --instance-count 3 --auto-terminate
```

API

Para desactivar la ejecución de pasos con la API de Amazon EMR

- Cuando utilice la acción RunJobFlow para crear un clúster, establezca la propiedad KeepJobFlowAliveWhenNoSteps en `true`.

Uso de una política de terminación automática

Una política de terminación automática le permite orquestar la limpieza de los clústeres sin necesidad de supervisar ni terminar manualmente los clústeres no utilizados. Cuando agrega una política de terminación automática a un clúster, debe especificar la cantidad de tiempo de inactividad tras el cual el clúster debe cerrarse automáticamente.

Según la versión de lanzamiento, Amazon EMR utiliza diferentes criterios para marcar un clúster como inactivo. En la siguiente tabla se describe cómo Amazon EMR determina la inactividad del clúster.

Cuando utiliza…	Un clúster se considera inactivo cuando…
Amazon EMR, versiones 5.34.0 y posteriores, y 6.4.0 y posteriores	• No hay aplicaciones de YARN activas • La utilización del HDFS es inferior al 10 % • No hay conexiones activas a cuadernos de EMR o a EMR Studio • No se utilizan interfaces de usuario de aplicaciones en el clúster
Amazon EMR, versiones 5.30.0 a 5.33.0 y 6.1.0 a 6.3.0	• No hay aplicaciones de YARN activas • El clúster no tiene trabajos de Spark activos **Note** Amazon EMR marca un clúster como inactivo y puede terminarlo automáticamente incluso si tiene un kernel de Python3 activo. Esto se debe a que al ejecutar un kernel de Python3 no se envía ningún trabajo de Spark al clúster. Para utilizar la terminación automática con un kernel de Python3, le recomendamos que utilice la versión 6.4.0 o posterior de Amazon EMR.

Note

Las versiones 6.4.0 y posteriores de Amazon EMR admiten un archivo en el clúster para detectar la actividad en el nodo principal: `/emr/metricscollector/isbusy`. Cuando

utiliza un clúster para ejecutar scripts de intérprete de comandos o aplicaciones que no son de YARN, puede tocar o actualizar periódicamente `isbusy` para indicar a Amazon EMR que el clúster no está inactivo.

Puede asociar una política de terminación automática al crear un clúster o agregar una política a un clúster existente. Para cambiar o deshabilitar la terminación automática, puede actualizar o eliminar la política.

Consideraciones

Tenga en cuenta las siguientes características y limitaciones antes de utilizar una política de terminación automática:

- En Asia-Pacífico (Yakarta), la terminación automática de Amazon EMR está disponible con Amazon EMR 6.14.0 y versiones posteriores.
- La terminación automática de Amazon EMR está disponible con Amazon EMR 5.30.0, 6.1.0 y versiones posteriores en las siguientes Regiones de AWS:

  Este de EE. UU. (Norte de Virginia y Ohio), Oeste de EE. UU. (Oregón y Norte de California), América del Sur (São Paulo), Europa (Fráncfort, Irlanda, Londres, Milán, París y Estocolmo), Canadá (centro), Asia-Pacífico (Hong Kong, Bombay, Seúl, Singapur, Sídney y Tokio), Medio Oriente (Baréin), África (Ciudad del Cabo), AWS GovCloud (Este de EE. UU.), AWS GovCloud (Oeste de EE. UU.), China (Pekín) operada por Sinnet y China (Ningxia) operada por NWCD.
- El tiempo de espera de inactividad se establece de forma predeterminada en 60 minutos (una hora) si no se especifica una cantidad. Puede especificar un tiempo de espera de inactividad mínimo de un minuto y un tiempo de espera de inactividad máximo de 7 días.
- Con las versiones 6.4.0 y posteriores de Amazon EMR, la terminación automática se habilita de forma predeterminada al crear un clúster nuevo con la consola de Amazon EMR.
- Amazon EMR publica métricas de Amazon CloudWatch de alta resolución cuando habilita la terminación automática de un clúster. Puede usar estas métricas para realizar un seguimiento de la actividad y la inactividad del clúster. Para obtener más información, consulte Métricas de capacidad del clúster.
- La terminación automática no se admite cuando se utilizan aplicaciones que no están basadas en YARN, como Presto, Trino o HBase.
- Para utilizar la terminación automática, el proceso de recopilación de métricas debe poder conectarse al punto de conexión de la API pública para la terminación automática en API Gateway.

Si utiliza un nombre de DNS privado con Amazon Virtual Private Cloud, la terminación automática no funcionará correctamente. Para garantizar que la terminación automática funcione, se recomienda que realice una de las siguientes acciones:

- Elimine el punto de conexión de VPC de la interfaz de API Gateway de su Amazon VPC.
- Siga las instrucciones de ¿Por qué aparece el error HTTP 403 Prohibido al conectarme a mis API de API Gateway desde una VPC? para deshabilitar la configuración del nombre de DNS privado.
- En su lugar, lance su clúster en una subred privada. Para más información, consulte el tema sobre Subredes privadas.

- (EMR 5.30.0 y versiones posteriores) Si quita la regla de salida predeterminada Permitir todo a 0.0.0.0/ para el grupo de seguridad principal, debe agregar una regla que permita la conectividad TCP de salida a su grupo de seguridad para el acceso al servicio en el puerto 9443. El grupo de seguridad para el acceso al servicio también debe permitir el tráfico TCP entrante en el puerto 9443 desde el grupo de seguridad principal. Para más información sobre la configuración de grupos de seguridad, consulte Grupo de seguridad administrado por Amazon EMR para la instancia principal (subredes privadas).

Permisos para utilizar la terminación automática

Antes de poder aplicar y administrar políticas de terminación automática para Amazon EMR, debe asociar los permisos que se enumeran en el siguiente ejemplo de política de permisos de IAM a los recursos de IAM que administran su clúster de EMR.

```
{
  "Version": "2012-10-17",
  "Statement": {
      "Sid": "AllowAutoTerminationPolicyActions",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:PutAutoTerminationPolicy",
        "elasticmapreduce:GetAutoTerminationPolicy",
        "elasticmapreduce:RemoveAutoTerminationPolicy"
      ],
      "Resource": "<your-resources>"
    }
}
```

Asociación, actualización o eliminación de una política de terminación automática

En esta sección se incluyen instrucciones que le ayudarán a asociar, actualizar o eliminar una política de terminación automática de un clúster de Amazon EMR. Antes de trabajar con políticas de terminación automática, asegúrese de tener los permisos de IAM necesarios. Consulte Permisos para utilizar la terminación automática.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para asociar una política de terminación automática al crear un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Terminación del clúster, seleccione Terminar el clúster después del tiempo de inactividad.
4. Especifique el número de horas y minutos de inactividad que pueden transcurrir antes de que el clúster termine automáticamente. El tiempo de espera de inactividad predeterminado es de 1 hora.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Para asociar, actualizar o eliminar una política de terminación automática en un clúster en ejecución con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar.

3. En la pestaña Propiedades de la página de detalles del clúster, busque Terminación del clúster y seleccione Editar.
4. Seleccione o anule la selección de Habilitar la terminación automática para activar o desactivar la característica. Si activa la terminación automática, especifique el número de horas y minutos de inactividad que pueden transcurrir antes de que el clúster termine automáticamente. A continuación, seleccione Guardar cambios para confirmar.

Old console

Para asociar una política de terminación automática al crear un clúster con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. En Configuración de hardware, seleccione Terminación automática.
4. Especifique el número de horas y minutos de inactividad después de los cuales el clúster se debe terminar automáticamente. El tiempo de espera de inactividad predeterminado es de una hora.
5. Elija las demás opciones que sean necesarias para la aplicación y, a continuación, elija Create cluster (Crear clúster).

Para asociar, actualizar o eliminar una política de terminación automática en un clúster en ejecución con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Clústeres y elija el clúster que desee actualizar.
3. Elija la pestaña Hardware en la página de detalles del clúster.
4. Seleccione o anule la selección de Habilitar la terminación automática para activar o desactivar la característica. Si activa la terminación automática, especifique el número de horas y minutos de inactividad tras los cuales el clúster debe terminar automáticamente.

AWS CLI

Antes de comenzar

Antes de trabajar con políticas de terminación automática, se recomienda que actualice a la versión más reciente de la AWS CLI. Para obtener instrucciones, consulte Instalación, actualización y desinstalación de la AWS CLI.

Para asociar o actualizar una política de terminación automática mediante AWS CLI

- Puede utilizar el comando `aws emr put-auto-termination-policy` para asociar o actualizar una política de terminación automática en un clúster.

  El siguiente ejemplo especifica 3600 segundos para *`IdleTimeout`*. Si no especifica *`IdleTimeout`*, el valor predeterminado es de una hora.

  ```
  aws emr put-auto-termination-policy \
  --cluster-id <your-cluster-id> \
  --auto-termination-policy IdleTimeout=3600
  ```

  > **Note**
  >
  > Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

  También puede especificar un valor para `--auto-termination-policy` cuando utilice el comando `aws emr create-cluster`. Para obtener más información acerca del uso de los comandos de Amazon EMR en la AWS CLI, consulte la Referencia de los comandos de la AWS CLI.

Para eliminar una política de terminación automática con AWS CLI

- Utilice el comando `aws emr remove-auto-termination-policy` para eliminar una política de terminación automática de un clúster. Para obtener más información acerca del uso de los comandos de Amazon EMR en la AWS CLI, consulte la Referencia de los comandos de la AWS CLI.

```
aws emr remove-auto-termination-policy --cluster-id <your-cluster-id>
```

Uso de la protección de terminación

Si la protección de terminación está habilitada en un clúster de ejecución prolongada, puede seguir terminando el clúster, pero primero debe quitar de forma explícita la protección de terminación de este. Esto ayuda a garantizar que las instancias EC2 no se cierren por accidente o por error. La protección de terminación resulta especialmente útil si el clúster contiene datos almacenados en discos locales que se deben recuperar antes de que las instancias se terminen. Puede habilitar la protección de terminación al crear un clúster y también cambiar este ajuste en un clúster en ejecución.

Con la protección de terminación habilitada, la acción `TerminateJobFlows` de la API de Amazon EMR no funciona. Los usuarios no pueden terminar el clúster con esta API ni con el comando `terminate-clusters` desde la AWS CLI. La API devuelve un error y la CLI se cierra con un código de devolución distinto de cero. Cuando utilice la consola de Amazon EMR para terminar un clúster, se le pedirá que realice un paso adicional para desactivar la protección de terminación.

> **Warning**
>
> La protección contra la terminación no garantiza que los datos se conserven en caso de un error humano o de una solución alternativa; por ejemplo, si se emite un comando de reinicio desde la línea de comandos mientras se está conectado a la instancia mediante SSH, si una aplicación o un script que se ejecuta en la instancia emite un comando de reinicio o si se utiliza la API de Amazon EC2 o de Amazon EMR para deshabilitar la protección contra la terminación. Incluso con la protección contra la terminación habilitada, los datos guardados en el almacenamiento de instancias, incluidos los datos del HDFS, se pueden perder. Escriba salidas de datos en las ubicaciones de Amazon S3 y cree estrategias de copia de seguridad adecuadas para sus requisitos de continuidad empresarial.

La protección de terminación no afecta a la capacidad para escalar los recursos del clúster con cualquiera de las siguientes acciones:

- Cambiar manualmente el tamaño de un clúster con la AWS Management Console o la AWS CLI. Para obtener más información, consulte Cambiar manualmente el tamaño de un clúster en ejecución.
- Eliminar instancias de un grupo de instancias secundarias o de tareas mediante una política de escalado descendente con escalado automático. Para obtener más información, consulte Uso del escalado automático con una política personalizada para grupos de instancias .
- Eliminar instancias de una flota de instancias reduciendo la capacidad de destino. Para obtener más información, consulte Opciones de flota de instancias.

Protección contra la terminación y Amazon EC2

Un clúster de Amazon EMR con protección de terminación habilitada tiene el atributo `disableAPITermination` definido para todas las instancias de Amazon EC2 en el clúster. Si una solicitud de terminación se origina en Amazon EMR y las configuraciones de Amazon EMR y Amazon EC2 de una instancia entran en conflicto, la configuración de Amazon EMR anula la configuración de Amazon EC2. Por ejemplo, si utiliza la consola de Amazon EC2 para habilitar la protección de terminación en una instancia de Amazon EC2 en un clúster que tenga la protección de terminación desactivada, al utilizar la consola de Amazon EMR, los comandos de la AWS CLI de Amazon EMR o la API de Amazon EMR para terminar el clúster, Amazon EMR establece `DisableApiTermination` en `false` y termina la instancia junto con las demás.

> **Important**
>
> Si se crea una instancia como parte de un clúster de Amazon EMR con la protección de terminación y se utilizan la API de Amazon EC2 o los comandos de la AWS CLI para modificar dicha instancia de forma que `DisableApiTermination` sea `false` y, a continuación, la API de Amazon EC2 o los comandos de la AWS CLI ejecutan la acción `TerminateInstances`, la instancia de Amazon EC2 se termina.

Protección de terminación y nodos de YARN en mal estado

Amazon EMR comprueba periódicamente el estado de Apache Hadoop YARN de los nodos que se ejecutan en las instancias de Amazon EC2 básicas y de tareas de un clúster. El estado se notifica mediante el servicio de comprobación de estado NodeManager. Si un nodo informa de que su estado es `UNHEALTHY`, el controlador de instancias de Amazon EMR lo agrega a la lista negra y no le asigna contenedores YARN hasta que vuelva a estar en buen estado. Por lo general, el mal estado de los

nodos se debe a que la utilización del disco es superior al 90 %. Para obtener más información sobre cómo identificar y recuperar los nodos en mal estado, consulte Errores de recursos.

Si el nodo permanece en estado UNHEALTHY durante más de 45 minutos, Amazon EMR realiza la siguiente acción en función del estado de la protección de terminación.

Termination protection	Resultado
Habilitada (recomendado)	Las instancias principales de Amazon EC2 permanecen en un estado de lista de denegación y siguen contabilizándose para la capacidad del clúster. Puede conectarse a una instancia básica de Amazon EC2 para la configuración y la recuperación de datos, así como cambiar el tamaño del clúster para agregar capacidad. Para obtener más informaci ón, consulte Errores de recursos. Los nodos de tarea en mal estado están exentos de la protección contra la terminación y se terminarán.
Deshabilitado	La instancia de Amazon EC2 se ha terminado . Amazon EMR aprovisiona una instancia nueva en función del número especificado de instancias del grupo de instancias o la capacidad de destino para las flotas de instancias. Si todos los nodos secundarios tienen el estado UNHEALTHY durante más de 45 minutos, el clúster se termina e informa de un estado NO_SLAVES_LEFT . **Important** Es posible que se pierdan datos de HDFS si una instancia secundari a se termina en mal estado. Si el

Termination protection	Resultado
	nodo almacenaba bloques que no se replicaron a otros nodos, estos bloques se pierden, lo que podría dar lugar a pérdida de datos. Recomendamos utilizar la protección de terminación para que pueda conectarse a las instancias y recuperar los datos cuando sea necesario.

Protección de terminación y ejecución de pasos

Cuando habilita la ejecución de pasos y también habilita la protección contra la terminación, Amazon EMR ignora la protección contra la terminación.

Al enviar pasos a un clúster, puede establecer la propiedad `ActionOnFailure` para determinar qué sucede si el paso no puede completar su ejecución debido a un error. Los valores posibles para esta propiedad son `TERMINATE_CLUSTER` (`TERMINATE_JOB_FLOW` con versiones anteriores), `CANCEL_AND_WAIT` y `CONTINUE`. Para obtener más información, consulte Enviar trabajo a un clúster.

Si se produce un error en un paso que se ha configurado con el valor `ActionOnFailure` en `CANCEL_AND_WAIT`, y la ejecución de pasos se ha habilitado, el clúster termina sin ejecutar los pasos siguientes.

Si se produce un error en un paso que tiene la propiedad `ActionOnFailure` establecida en `TERMINATE_CLUSTER`, utilice la tabla de configuración mostrada a continuación para determinar el resultado.

ActionOnFailure	Ejecución de pasos	Termination protection	Resultado
`TERMINATE_CLUSTER`	Habilitado	Deshabilitado	El clúster termina
	Habilitado	Habilitado	El clúster termina

ActionOnFailure	Ejecución de pasos	Termination protection	Resultado
	Deshabilitado	Habilitado	El clúster continúa
	Deshabilitado	Deshabilitado	El clúster termina

Protección de terminación e instancias de spot

La protección de terminación de Amazon EMR no impide que una instancia de spot de Amazon EC2 termine cuando el precio de spot supera el precio de spot máximo.

Configuración de la protección de terminación al lanzar un clúster

Puede habilitar o deshabilitar la protección de terminación al lanzar un clúster con la consola, la AWS CLI o la API.

La configuración predeterminada de la protección de terminación depende de cómo se lance el clúster:

- Consola de Amazon EMR (nueva): la protección contra la terminación está habilitada de forma predeterminada.
- Opciones rápidas de la consola de Amazon EMR (antigua): la protección contra la terminación está deshabilitada de forma predeterminada.
- Opciones avanzadas de la consola de Amazon EMR (antigua): la protección de terminación está habilitada de forma predeterminada.
- AWS CLI `aws emr create-cluster`: la protección contra la terminación está deshabilitada a menos que se especifique `--termination-protected`.
- Comando RunJobFlow de la API de Amazon EMR: la protección contra la terminación está deshabilitada a menos que el valor booleano `TerminationProtected` esté establecido en `true`.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para activar o desactivar la protección contra la terminación al crear un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En el caso de la versión de lanzamiento de EMR, elija emr-6.6.0 o posterior.
4. En Terminación del clúster, asegúrese de que la opción Use la protección contra la terminación esté preseleccionada o anule la selección para desactivarla.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Old console

Para activar o desactivar la protección contra la terminación al crear un clúster con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Go to advanced options (Ir a las opciones avanzadas).
4. En Step 3: General Cluster Settings (Paso 3: Configuración general del clúster), en General Options (Opciones generales), asegúrese de que Termination protection (Protección de terminación) está seleccionada para habilitarla o anule la selección para deshabilitarla.
5. Elija las demás opciones que sean necesarias para la aplicación, elija Next (Siguiente) y, a continuación, termine de configurar el clúster.

AWS CLI

Para activar o desactivar la protección de terminación al crear un clúster con la AWS CLI

- Con la AWS CLI, puede lanzar un clúster con la protección de terminación habilitada mediante el comando `create-cluster` con el parámetro `--termination-protected`. La protección de terminación está deshabilitada de forma predeterminada.

  En el siguiente ejemplo, se crea un clúster con la protección de terminación habilitada:

  > **Note**
  >
  > Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name "TerminationProtectedCluster" --release-
label emr-5.36.1 \
--applications Name=Hadoop Name=Hive Name=Pig \
--use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge \
--instance-count 3 --termination-protected
```

  Para más información sobre el uso de comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Configuración de la protección de terminación para ejecutar clústeres

Puede configurar la protección de terminación para un clúster en ejecución utilizando la consola o la AWS CLI.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para activar o desactivar la protección contra la terminación de un clúster en ejecución con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar.
3. En la pestaña Propiedades de la página de detalles del clúster, busque Terminación del clúster y seleccione Editar.
4. Seleccione o anule la selección de la casilla Use la protección contra la terminación para activar o desactivar la característica. A continuación, seleccione Guardar cambios para confirmar.

Old console

Para activar o desactivar la protección contra la terminación de un clúster en ejecución con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. En la página Clusters (Clústeres), elija el Name (Nombre) del clúster.
3. En la pestaña Summary (Resumen), en Termination protection (Protección de terminación), elija Change (Cambiar).
4. Para habilitar la protección de terminación, elija On (Activada). Para deshabilitar la protección de terminación, elija Off (Desactivada). A continuación, elija la marca de verificación verde para confirmar los cambios.

AWS CLI

Para activar o desactivar la protección de terminación en un clúster en ejecución mediante la AWS CLI

- Para habilitar la protección de terminación en un clúster en ejecución mediante la AWS CLI, utilice el comando `modify-cluster-attributes` con el parámetro `--termination-protected`. Para deshabilitarla, utilice el parámetro `--no-termination-protected`.

  En el siguiente ejemplo, se habilita la protección de terminación en el clúster que tiene el ID *j-3KVTXXXXXX7UG*:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --termination-protected
  ```

  En el siguiente ejemplo, se deshabilita la protección de terminación en el mismo clúster:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
  ```

Uso de las AMI de Amazon Linux en Amazon EMR

imágenes de máquina de Amazon (AMI) de Amazon Linux

Amazon EMR utiliza una imagen de máquina de Amazon (AMI) de Amazon Linux para inicializar instancias de Amazon EC2 cuando crea y lanza un clúster. La AMI contiene el sistema operativo Amazon Linux, además del software y las configuraciones necesarias para que cada instancia aloje las aplicaciones del clúster.

De forma predeterminada, cuando se crea un clúster, Amazon EMR utiliza una AMI de Amazon Linux predeterminada que se ha creado específicamente para la versión de Amazon EMR que se esté utilizando. Para más información sobre la AMI predeterminada de Amazon Linux, consulte Uso de la AMI de Amazon Linux predeterminada para Amazon EMR. Si utiliza la versión 5.7.0 o superior de Amazon EMR, puede optar por especificar una AMI de Amazon Linux personalizada en lugar de la AMI de Amazon Linux predeterminada para Amazon EMR. Una AMI personalizada le permite cifrar el volumen de dispositivo raíz y personalizar las aplicaciones y las configuraciones como alternativa al uso de acciones de arranque. Puede especificar una AMI personalizada para cada tipo de instancia en la configuración del grupo de instancias o de la flota de instancias de un clúster de Amazon EMR.

La compatibilidad con varias AMI personalizadas le brinda la flexibilidad de utilizar más de un tipo de arquitectura en un clúster. Consulte Uso de una AMI personalizada.

Amazon EMR asocia automáticamente un volumen SSD de uso general de Amazon EBS como dispositivo raíz para todas las AMI. El uso de una AMI respaldada por EBS mejora el rendimiento. Los costos de Amazon EBS se prorratean por hora en función de los cargos de Amazon EBS mensuales para volúmenes gp2 en la región donde se ejecuta el clúster. Por ejemplo, el costo por hora del volumen raíz en cada instancia de clúster en una región en la que se cobran 0,10 USD/GB al mes sería aproximadamente de 0,00139 USD por hora (0,10 USD/GB al mes dividido entre 30 días dividido entre 24 horas y multiplicado por 10 GB). Tanto si utiliza la AMI predeterminada como una personalizada de Amazon Linux, puede especificar el tamaño del volumen de dispositivo raíz de Amazon EBS desde 10 hasta 100 GiB.

Para más información sobre las AMI de Amazon Linux, consulte Imágenes de máquina de Amazon (AMI). Para más información sobre el almacenamiento de instancias para las instancias de Amazon EMR, consulte Almacenamiento de la instancia.

Uso de la AMI de Amazon Linux predeterminada para Amazon EMR

Cada versión de lanzamiento de Amazon EMR utiliza una AMI de Amazon Linux predeterminada para Amazon EMR a menos que se especifique una AMI personalizada. A partir de las versiones 5.36 y 6.6 de Amazon EMR, el comportamiento predeterminado para actualizar Amazon Linux 2 (AL2) en una AMI predeterminada de Amazon EMR consiste en aplicar automáticamente la versión más reciente de AL2 a la AMI de Amazon EMR predeterminada.

Actualizaciones automáticas de Amazon Linux 2 para las versiones de Amazon EMR

Al lanzar un clúster con la versión de revisión más reciente de Amazon EMR 5.36 o posterior, o 6.6 o posterior, Amazon EMR utiliza la versión más reciente de Amazon Linux 2 para la AMI de Amazon EMR predeterminada. Por ejemplo:

- Cuando hay una versión `x.x.0` y una versión `x.x.1`, la versión `x.x.0` deja de recibir actualizaciones de la AMI cuando se lanza `x.x.1`.
- Del mismo modo, `x.x.1` deja de recibir actualizaciones de la AMI cuando se lanza `x.x.2`.
- Más adelante, cuando se lanza `x.y.0`, `x.x.[latest]` sigue recibiendo actualizaciones de la AMI al mismo tiempo que `x.y.[latest]`.

Para comprobar si está utilizando la versión de revisión más reciente indicada por el número que aparece después del segundo punto decimal (`6.8.1`) de una versión de Amazon EMR, consulte las versiones disponibles en la Guía de versiones de Amazon EMR, consulte el menú desplegable Versión de Amazon EMR al crear un clúster en la consola o utilice la API `ListReleaseLabels` o la acción `list-release-labels` de la CLI. Para recibir actualizaciones cuando lancemos una nueva versión de Amazon EMR, suscríbase a la fuente RSS de la página Novedades de la Guía de versiones.

Si lo desea, puede optar por lanzar el clúster con la versión de Amazon Linux con la que se envió por primera vez la versión de Amazon EMR. Para obtener información sobre cómo especificar la versión de Amazon Linux para su clúster, consulte Cambio de versión de Amazon Linux al crear un clúster.

Versiones de Amazon Linux predeterminadas

En la siguiente tabla se muestra la información de Amazon Linux sobre la versión de revisión más reciente de Amazon EMR 6.x, versiones 6.6 y posteriores.

OsRelea Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2023 727.0	4.14.320	14 de agosto de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 (6.10+), eu-south-1 , eu-south-2 (6.10+), ap-east-1 , ap-south-1 , ap-south-2 (6.10+), ap-southeast-3 , ap-southeast-4 (6.8+ y 5.36.1), ap-northe

OsRelea Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
			ast-1 , ap-northe ast-2 , ap-northe ast-3 , ap-southe ast-1 , ap-southe ast-2 , af-south-1 , sa-east-1 , me-centra l-1 (6.10+), me-south- 1 , ca-central-1 , il-central-1 (6.9+ y 5.36.1)

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230719.0	4.14.320	2 de agosto de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 (6.10+), eu-south-1 , eu-south-2 (6.10+), ap-east-1 , ap-south-1 , ap-south-2 (6.10+), ap-southeast-3 , ap-southeast-4 (6.8+ y 5.36.1), ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 (6.10+), me-south-1 , ca-central-1 , il-central-1 (6.9+ y 5.36.1)

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230628.0	4.14.318	12 de julio de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 (6.10+), eu-south-1 , eu-south-2 (6.10+), ap-east-1 , ap-south-1 , ap-south-2 (6.10+), ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 (6.10+), me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230612.0	4.14.314	23 de junio de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 (6.10+), eu-south-1 , eu-south-2 (6.10+), ap-east-1 , ap-south-1 , ap-south-2 (6.10+), ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 (6.10+), me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230504.1	4.14.313	16 de mayo de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 (6.10+), eu-south-1 , eu-south-2 (6.10+), ap-east-1 , ap-south-1 , ap-south-2 (6.10+), ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230418.0	4.14.311	3 de mayo de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 (solo 6,10), eu-south-1 , eu-south-2 (solo 6,10), ap-east-1 , ap-south-1 , ap-south-2 (solo 6,10), ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230404.1	4.14.311	18 de abril de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 , eu-south-1 , eu-south-2 , ap-east-1 , ap-south-1 , ap-south-2 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 , me-south-1 , ca-central-1
2.0.20230404.0	4.14.311	10 de abril de 2023	us-east-1 , eu-west-3

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230320.0	4.14.309	30 de marzo de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 , eu-south-1 , eu-south-2 , ap-east-1 , ap-south-1 , ap-south-2 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230307.0	4.14.305	15 de marzo de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 , eu-south-1 , eu-south-2 , ap-east-1 , ap-south-1 , ap-south-2 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230207.0	4.14.304	3 de marzo de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-central-2 , eu-south-1 , eu-south-2 , ap-east-1 , ap-south-1 , ap-south-2 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-central-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20230119.1	4.14.301	9 de febrero de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20221210.1	4.14.301	12 de enero de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20221103.3	4.14.296	5 de diciembre de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20221004.0	4.14.294	2 de noviembre de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20220912.1	4.14.291	7 de octubre de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1
2.0.20220805.0	4.14.287	30 de agosto de 2022	us-west-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20220719.0	4.14.287	10 de agosto de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OsRelease Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.20220426.0	4.14.281	10 de junio de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OsRelea Label (Versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2022 406.1	4.14.275	2 de mayo de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

En la siguiente tabla se muestra la información de Amazon Linux sobre la versión de revisión más reciente de Amazon EMR 5.x, versiones 5.36 y posteriores.

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2023 504.1	4.14.313	16 de mayo de 2023	Este de EE. UU. (Norte de Virginia), Este de EE. UU. (Ohio), Oeste de EE. UU. (Norte de Californi a), Oeste de EE. UU. (Oregón), Canadá (centro), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (París), Europa (Fráncfort), Europa (Milán), Asia-Pacífico (Hong Kong), Asia-Pací fico (Bombay), Asia-Pací fico (Yakarta), Asia-Pací fico (Tokio), Asia-Pací fico (Seúl), Asia-Pacífico (Osaka), Asia-Pacífico (Singapur), Asia-Pacífico (Sídney), África (Ciudad del Cabo), América del Sur (São Paulo), Medio Oriente (Baréin), Medio Oriente (EAU)
2.0.2023 418.0	4.14.311	3 de mayo de 2023	Este de EE. UU. (Norte de Virginia), Este de EE. UU. (Ohio), Oeste de EE. UU. (Norte de Californi a), Oeste de EE. UU.

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
			(Oregón), Canadá (centro), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (París), Europa (Fráncfort), Europa (Milán), Asia-Pacífico (Hong Kong), Asia-Pacífico (Bombay), Asia-Pacífico (Yakarta), Asia-Pacífico (Tokio), Asia-Pacífico (Seúl), Asia-Pacífico (Osaka), Asia-Pacífico (Singapur), Asia-Pacífico (Sídney), África (Ciudad del Cabo), América del Sur (São Paulo), Medio Oriente (Baréin), Medio Oriente (EAU)

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2023 404.1	4.14.311	18 de abril de 2023	Este de EE. UU. (Norte de Virginia), Este de EE. UU. (Ohio), Oeste de EE. UU. (Norte de Californi a), Oeste de EE. UU. (Oregón), Canadá (centro), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (París), Europa (Fráncfort), Europa (Milán), Asia-Pacífico (Hong Kong), Asia-Pací fico (Bombay), Asia-Pací fico (Yakarta), Asia-Pací fico (Tokio), Asia-Pací fico (Seúl), Asia-Pacífico (Osaka), Asia-Pacífico (Singapur), Asia-Pacífico (Sídney), África (Ciudad del Cabo), América del Sur (São Paulo), Medio Oriente (Baréin)
2.0.2023 404.0	4.14.311	10 de abril de 2023	Este de EE. UU. (Norte de Virginia), Europa (París)

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2023 320.0	4.14.309	30 de marzo de 2023	Este de EE. UU. (Norte de Virginia), Este de EE. UU. (Ohio), Oeste de EE. UU. (Norte de Californi a), Oeste de EE. UU. (Oregón), Canadá (centro), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (París), Europa (Fráncfort), Europa (Milán), Asia-Pacífico (Hong Kong), Asia-Pací fico (Bombay), Asia-Pací fico (Yakarta), Asia-Pací fico (Tokio), Asia-Pací fico (Seúl), Asia-Pacífico (Osaka), Asia-Pacífico (Singapur), Asia-Pacífico (Sídney), África (Ciudad del Cabo), América del Sur (São Paulo), Medio Oriente (Baréin)

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2023 307.0	4.14.305	15 de marzo de 2023	Este de EE. UU. (Norte de Virginia), Este de EE. UU. (Ohio), Oeste de EE. UU. (Norte de Californi a), Oeste de EE. UU. (Oregón), Canadá (centro), Europa (Estocolmo), Europa (Irlanda), Europa (Londres), Europa (París), Europa (Fráncfort), Europa (Milán), Asia-Pacífico (Hong Kong), Asia-Pací fico (Bombay), Asia-Pací fico (Yakarta), Asia-Pací fico (Tokio), Asia-Pací fico (Seúl), Asia-Pacífico (Osaka), Asia-Pacífico (Singapur), Asia-Pacífico (Sídney), África (Ciudad del Cabo), América del Sur (São Paulo), Medio Oriente (Baréin)

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2023 207.0	4.14.304	3 de marzo de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2022 210.1	4.14.301	12 de enero de 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-centra l-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southe ast-3 , ap-northe ast-1 , ap-northe ast-2 , ap-northe ast-3 , ap-southe ast-1 , ap-southe ast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2022 103.3	4.14.296	5 de diciembre de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2022004.0	4.14.294	2 de noviembre de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2022 912.1	4.14.291	7 de octubre de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1
2.0.2022 719.0	4.14.287	10 de agosto de 2022	us-west-1 , eu-west-3 , eu-north-1 , eu-central-1 , ap-south-1 , me-south-1

OSRelea Label (versión de Amazon Linux)	Versión de kernel de Amazon Linux	Fecha de disponibilidad	Regiones admitidas
2.0.2022 426.0	4.14.281	14 de junio de 2022	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 , eu-west-2 , eu-west-3 , eu-central-1 , eu-south-1 , ap-east-1 , ap-south-1 , ap-southeast-3 , ap-northeast-1 , ap-northeast-2 , ap-northeast-3 , ap-southeast-1 , ap-southeast-2 , af-south-1 , sa-east-1 , me-south-1 , ca-central-1

Consideraciones sobre las actualizaciones de software

Estos son los comportamientos predeterminados de las actualizaciones de software que se deben tener en cuenta:

Amazon EMR 5.35.0 y versiones anteriores, y 6.5.0 y versiones anteriores: la AMI de Amazon Linux está “bloqueada” en la versión de lanzamiento de Amazon EMR

En el caso de Amazon EMR 5.35.0 y versiones anteriores, y 6.5.0 y versiones anteriores, la AMI predeterminada se basa en la AMI de Amazon Linux más actualizada que esté disponible en el momento del lanzamiento de Amazon EMR. La AMI se prueba para determinar su compatibilidad con las aplicaciones de macrodatos y las características de Amazon EMR incluidas en esa versión.

Cada versión de Amazon EMR 5.35.0 y versiones anteriores, y 6.5.0 y versiones anteriores de Amazon EMR está “bloqueada” en su respectiva versión de la AMI de Amazon Linux asignada para mantener la compatibilidad. Esto significa que las versiones anteriores de la AMI de Amazon Linux se utilizan para las versiones anteriores de Amazon EMR, incluso cuando las AMI de Amazon Linux más nuevas estén disponibles. Por este motivo, se recomienda utilizar la versión más reciente de Amazon EMR, a menos que necesite una versión anterior por motivos de compatibilidad y no pueda realizar la migración. Si tiene que utilizar una versión anterior de Amazon EMR por motivos de compatibilidad, se recomienda que utilice la versión más reciente de una serie. Por ejemplo, si tiene que utilizar la serie 5.12, utilice la 5.12.2 en lugar de la 5.12.0 o la 5.12.1. Si aparece una versión nueva de una serie, considere la posibilidad de migrar las aplicaciones a dicha versión.

Para más información sobre el comportamiento de actualización automática introducido en las versiones 5.36.0 y 6.6.0 de Amazon EMR, y sus respectivas versiones superiores, consulte Actualizaciones automáticas de Amazon Linux 2 para las versiones de Amazon EMR.

El comportamiento de arranque predeterminado excluye las actualizaciones del kernel

Cuando una instancia de Amazon EC2 de un clúster que está basada en la AMI de Amazon Linux predeterminada para EMR arranca por primera vez, comprueba los repositorios de paquetes habilitados para Amazon Linux y Amazon EMR con el fin de determinar si existen actualizaciones de software para la versión de la AMI. Al igual que con otras instancias de Amazon EC2, las actualizaciones de seguridad críticas e importantes de estos repositorios se instalan automáticamente. Sin embargo, si utiliza una versión anterior de la AMI de Amazon Linux, es posible que la actualización de seguridad más reciente no se instale automáticamente. Esto se debe a que los repositorios a los que hace referencia EMR son fijos para cada versión de la AMI de Amazon Linux. Tenga en cuenta también que la configuración de red debe permitir la salida de HTTP y HTTPS a los repositorios de Amazon Linux en Amazon S3; de lo contrario, las actualizaciones de seguridad no se realizarán correctamente. Para más información, consulte Repositorio de paquetes en la Guía del usuario de Amazon EC2 para instancias de Linux. De forma predeterminada, otros paquetes de software y actualizaciones del kernel que requieren un reinicio, incluidos NVIDIA y CUDA, no se descargan automáticamente en el primer arranque.

> **Important**
>
> Los clústeres de Amazon EMR que ejecutan las AMI (imágenes de máquina de Amazon) de Amazon Linux o Amazon Linux 2 utilizan el comportamiento predeterminado de Amazon Linux y no descargan ni instalan automáticamente actualizaciones importantes y críticas del kernel que requieren un reinicio. Este comportamiento es el mismo que el de otras instancias

de Amazon EC2 que ejecutan la AMI de Amazon Linux predeterminada. Si aparecen nuevas actualizaciones de software de Amazon Linux que requieren un reinicio (por ejemplo, actualizaciones del kernel, NVIDIA y CUDA) tras el lanzamiento de una versión de Amazon EMR, las instancias de clúster de Amazon EMR que ejecutan la AMI predeterminada no descargan ni instalan automáticamente esas actualizaciones. Para obtener actualizaciones del kernel, puede personalizar la AMI de Amazon EMR para que utilice la AMI de Amazon Linux más reciente.

El clúster se lanza con o sin actualizaciones

Tenga en cuenta que, si las actualizaciones de software no se pueden instalar porque no se puede acceder a los repositorios de paquetes la primera vez que se inicia el clúster, la instancia del clúster sigue completando su lanzamiento. Por ejemplo, es posible que no se pueda acceder a los repositorios porque S3 no está disponible temporalmente o puede que tenga reglas de VPC o firewall configuradas para bloquear el acceso.

No se ejecuta **`sudo yum update`**

Cuando se conecta a una instancia de clúster mediante SSH, las primeras líneas de la salida en pantalla proporcionan un enlace a las notas de la versión de la AMI de Amazon Linux que utiliza la instancia, un aviso de la versión más reciente de la AMI de Amazon Linux, un aviso del número de paquetes disponibles para la actualización desde los repositorios habilitados y una directiva para ejecutar `sudo yum update`.

> **Important**
>
> Se recomienda no ejecutar `sudo yum update` en las instancias de clúster, ni mientras está conectado con SSH ni mediante una acción de arranque. Esto puede causar incompatibilidades, ya que todos los paquetes se instalan de forma indiscriminada.

Prácticas recomendadas de actualización de software

Prácticas recomendadas para administrar las actualizaciones de software

- Si utiliza una versión anterior de Amazon EMR, es conveniente que pruebe una migración a la versión más reciente antes de actualizar los paquetes de software.

- Si migra a una versión posterior o actualiza los paquetes de software, pruebe primero la implementación en un entorno que no sea de producción. La opción para clonar clústeres mediante la consola de Amazon EMR resulta muy útil para esto.
- Evalúe las actualizaciones de software para las aplicaciones y para la versión de la AMI de Amazon Linux por separado. Únicamente pruebe e instale en entornos de producción los paquetes que considere absolutamente necesarios para el nivel de seguridad requerido y la funcionalidad o el rendimiento de las aplicaciones.
- Visite el Centro de seguridad de Amazon Linux para comprobar si existen actualizaciones.
- Evite instalar paquetes conectándose a las instancias de clúster individuales mediante SSH. En su lugar, utilice una acción de arranque para instalar y actualizar los paquetes en todas las instancias de clúster que lo necesiten. Para ello, es necesario que termine un clúster y lo vuelva a lanzar. Para obtener más información, consulte Crear acciones de arranque para instalar software adicional.

Uso de una AMI personalizada

Si utiliza la versión 5.7.0 o superior de Amazon EMR, puede optar por especificar una AMI de Amazon Linux personalizada en lugar de la AMI de Amazon Linux predeterminada para Amazon EMR. Una AMI personalizada resulta útil si desea hacer lo siguiente:

- Preinstalar aplicaciones y realizar otras personalizaciones en lugar de utilizar acciones de arranque. Esto puede mejorar el tiempo de inicio del clúster y simplificar el flujo de trabajo de inicio. Para obtener más información y un ejemplo, consulte Creación de una AMI de Amazon Linux personalizada a partir de una instancia preconfigurada.
- Implementar configuraciones de clúster y de nodo más sofisticadas de lo que permiten las acciones de arranque.
- Cifre los volúmenes del dispositivo raíz de EBS (volúmenes de arranque) de instancias de EC2 en su clúster si está utilizando una versión de Amazon EMR anterior a la 5.24.0. Al igual que con la AMI predeterminada, el tamaño mínimo del volumen raíz de una AMI personalizada es de 10 GiB. Para obtener más información, consulte Creación de una AMI personalizada con un volumen de dispositivo raíz de Amazon EBS cifrado.

> **Note**
>
> A partir de la versión 5.24.0 de Amazon EMR, podrá utilizar una opción de configuración de seguridad para cifrar los volúmenes de almacenamiento y el dispositivo raíz de

EBS cuando especifique AWS KMS como su proveedor de claves. Para obtener más información, consulte Cifrado de disco local.

Una AMI personalizada debe existir en la misma región de AWS donde se crea el clúster. También debe coincidir con la arquitectura de la instancia de EC2. Por ejemplo, una instancia m5.xlarge tiene una arquitectura x86_64. Por lo tanto, para aprovisionar una instancia m5.xlarge mediante una AMI personalizada, la AMI también debe tener una arquitectura x86_64. Del mismo modo, para aprovisionar una instancia m6g.xlarge, que tiene una arquitectura arm64, la AMI personalizada debe tener una arquitectura arm64. Para más información acerca de la identificación de una AMI de Linux para su tipo de instancia, consulte Buscar una AMI de Linux en la Guía del usuario de Amazon EC2 para instancias de Linux.

Important

Los clústeres de Amazon EMR que ejecutan las AMI (imágenes de máquina de Amazon) de Amazon Linux o Amazon Linux 2 utilizan el comportamiento predeterminado de Amazon Linux y no descargan ni instalan automáticamente actualizaciones importantes y críticas del kernel que requieren un reinicio. Este comportamiento es el mismo que el de otras instancias de Amazon EC2 que ejecutan la AMI de Amazon Linux predeterminada. Si aparecen nuevas actualizaciones de software de Amazon Linux que requieren un reinicio (por ejemplo, actualizaciones del kernel, NVIDIA y CUDA) tras el lanzamiento de una versión de Amazon EMR, las instancias de clúster de Amazon EMR que ejecutan la AMI predeterminada no descargan ni instalan automáticamente esas actualizaciones. Para obtener actualizaciones del kernel, puede personalizar la AMI de Amazon EMR para que utilice la AMI de Amazon Linux más reciente.

Creación de una AMI de Amazon Linux personalizada a partir de una instancia preconfigurada

Los pasos básicos para preinstalar el software y realizar otras configuraciones para crear una AMI de Amazon Linux personalizada para Amazon EMR son los siguientes:

- Lance una instancia desde la AMI de Amazon Linux de base.
- Conéctese a la instancia para instalar software y realizar otras personalizaciones.
- Cree una nueva imagen (instantánea de AMI) de la instancia que haya configurado.

Después de crear la imagen en función de su instancia personalizada, puede copiar dicha imagen a un destino cifrado tal y como se describe en Creación de una AMI personalizada con un volumen de dispositivo raíz de Amazon EBS cifrado.

Tutorial: Creación de una AMI desde una instancia con software personalizado Instalado

Para lanzar una instancia EC2 en función de la AMI de Amazon Linux más reciente

1. Utilice la AWS CLI para ejecutar el siguiente comando, que crea una instancia a partir de una AMI existente. Sustituya *MyKeyName* por el par de claves que utiliza para conectarse a la instancia y *MyAmiId* por el ID de una AMI de Amazon Linux adecuada. Para conocer los ID de AMI más recientes, consulte Amazon Linux AMI.

   > **Note**
   >
   > Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

   ```
   aws ec2 run-instances --image-id MyAmiID \
   --count 1 --instance-type m5.xlarge \
   --key-name MyKeyName --region us-west-2
   ```

   El valor de salida `InstanceId` se utiliza como *MyInstanceId* en el siguiente paso.
2. Ejecute el siguiente comando:

   ```
   aws ec2 describe-instances --instance-ids MyInstanceId
   ```

   El valor de salida `PublicDnsName` se utiliza para conectarse a la instancia en el siguiente paso.

Para conectar a la instancia e instalar software

1. Utilice una conexión SSH que le permita ejecutar comandos de shell en su instancia de Linux. Para más información, consulte Conexión a la instancia de Linux mediante SSH en la Guía del usuario de Amazon EC2 para instancias de Linux.
2. Realice las personalizaciones necesarias. Por ejemplo:

```
sudo yum install MySoftwarePackage
sudo pip install MySoftwarePackage
```

Para crear una instantánea desde su imagen personalizada

- Después de personalizar la instancia, utilice el comando `create-image` para crear una AMI desde la instancia.

  ```
  aws ec2 create-image --no-dry-run --instance-id MyInstanceId --name MyEmrCustomAmi
  ```

  El valor de salida `imageID` se utiliza al lanzar el clúster o crear una instantánea cifrada. Para obtener más información, consulte Utilice una sola AMI personalizada en un clúster de EMR y Creación de una AMI personalizada con un volumen de dispositivo raíz de Amazon EBS cifrado.

Cómo utilizar una AMI personalizada en un clúster de Amazon EMR

Puede utilizar una AMI personalizada para aprovisionar un clúster de Amazon EMR de dos maneras:

- Utilice una sola AMI personalizada para todas las instancias de EC2 del clúster.
- Utilice diferentes AMI personalizadas para los distintos tipos de instancias de EC2 que se utilizan en el clúster.

Puede usar solo una de las dos opciones al aprovisionar un clúster de EMR y no puede cambiarla una vez que se haya iniciado el clúster.

Consideraciones sobre el uso de una sola AMI personalizada en lugar de varias en un clúster de Amazon EMR

Consideración	AMI personalizada única	Múltiples AMI personalizadas
Utilice los procesadores x86 y Graviton2 con AMI personalizadas en el mismo clúster	✗ No compatible	✓ Compatible

Consideración	AMI personalizada única	Múltiples AMI personalizadas
La personalización de la AMI varía según el tipo de instancia	✗ No compatible	✓ Compatible
Cambie las AMI personalizadas al agregar nuevos grupos o flotas de instancias de tareas a un clúster en ejecución. Nota: No puede cambiar la AMI personalizada de los grupos o flotas de instancias existentes.	✗ No compatible	✓ Compatible
Utilice la Consola de AWS para iniciar un clúster	✓ Compatible	✗ No compatible
Utilice AWS CloudFormation para iniciar un clúster	✓ Compatible	✓ Compatible

Utilice una sola AMI personalizada en un clúster de EMR

Para especificar el ID de una AMI personalizada al crear un clúster, utilice una de las siguientes opciones:

- AWS Management Console
- AWS CLI
- SDK de Amazon EMR
- API de Amazon EMR RunJobFlow
- AWS CloudFormation (consulte la propiedad `CustomAmiID` en Clúster InstanceGroupConfig, Clúster InstanceTypeConfig, Recurso InstanceGroupConfig o Recurso InstanceFleetConfig-InstanceTypeConfig)

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para especificar una sola AMI personalizada utilizando la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Nombre y aplicaciones, busque Opciones del sistema operativo. Elija AMI personalizada e ingrese el ID de su AMI en el campo AMI personalizada.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para especificar una sola AMI personalizada utilizando la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Crear clúster e Ir a las opciones avanzadas.
3. En Configuración de software, en Versión, elija emr-5.7.0 o una versión posterior y, a continuación, elija las demás opciones que necesite para la aplicación. Elija Next (Siguiente).
4. Seleccione valores en Hardware Configuration (Configuración de hardware) que sean adecuados para la aplicación y elija Next (Siguiente).
5. En Opciones adicionales, en ID de AMI personalizada, ingrese un valor y asegúrese de que esté seleccionada la opción Actualizar todos los paquetes instalados que se están reiniciando. Para obtener más información acerca del cambio de la opción de actualización, consulte Administración de actualizaciones de repositorio de paquetes de la AMI.

6. Para iniciar el clúster, elija Next (Siguiente) y complete las demás opciones de configuración.

AWS CLI

Para especificar una sola AMI personalizada con la AWS CLI

- Utilice el parámetro `--custom-ami-id` para especificar el ID de AMI al ejecutar el comando `aws emr create-cluster`.

  El siguiente ejemplo especifica un clúster que utiliza una sola AMI personalizada con un volumen de arranque de 20 GiB. Para obtener más información, consulte Especificación del tamaño de volumen de dispositivo raíz de Amazon EBS.

  > **Note**
  >
  > Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name "Cluster with My Custom AMI" \
--custom-ami-id MyAmiID --ebs-root-volume-size 20 \
--release-label emr-5.7.0 --use-default-roles \
--instance-count 2 --instance-type m5.xlarge
```

Uso de varias AMI personalizadas en un clúster de Amazon EMR

Para crear un clúster mediante varias AMI personalizadas, utilice una de las siguientes opciones:

- CLI de AWS versión 1.20.21 o superior
- SDK de AWS
- Amazon EMR RunJobFlow en la referencia de la API de Amazon EMR
- AWS CloudFormation (consulte la propiedad `CustomAmiID` en Clúster InstanceGroupConfig, Clúster InstanceTypeConfig, Recurso InstanceGroupConfig o Recurso InstanceFleetConfig-InstanceTypeConfig)

La Consola de administración de AWS actualmente no admite la creación de un clúster con varias AMI personalizadas.

Example - Uso de AWS CLI para crear un clúster de grupos de instancias mediante varias AMI personalizadas

Con la versión 1.20.21 o posterior de AWS CLI, puede asignar una sola AMI personalizada a todo el clúster o puede asignar varias AMI personalizadas a cada nodo de instancia del clúster.

En el siguiente ejemplo, se muestra un clúster de grupos de instancias uniforme creado con dos tipos de instancias (m5.xlarge) que se utilizan en todos los tipos de nodos (maestro, principal y de tarea). Cada nodo tiene varias AMI personalizadas. En el ejemplo se ilustran varias características de la configuración de varias AMI personalizadas:

- No hay ninguna AMI personalizada asignada en el nivel de clúster. Esto sirve para evitar conflictos entre varias AMI personalizadas y una sola AMI personalizada, lo que provocaría un error en el lanzamiento del clúster.
- El clúster puede tener varias AMI personalizadas en los nodos maestros, principales y de tareas individuales. Esto permite realizar personalizaciones de AMI individuales, como aplicaciones preinstaladas, configuraciones de clústeres sofisticadas y volúmenes de dispositivos raíz de Amazon EBS cifrados.
- El nodo principal del grupo de instancias solo puede tener un tipo de instancia y la AMI personalizada correspondiente. Del mismo modo, el nodo maestro solo puede tener un tipo de instancia y la AMI personalizada correspondiente.
- El clúster puede tener varios nodos de tarea.

```
aws emr create-cluster --instance-groups
InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-234567
InstanceGroupType=TASK,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-345678
InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-456789
```

Example - Uso de la versión 1.20.21 o posterior de AWS CLI para agregar un nodo de tarea a un clúster de grupos de instancias en ejecución con varios tipos de instancias y varias AMI personalizadas

Con la versión 1.20.21 o posterior de AWS CLI, puede agregar varias AMI personalizadas a un grupo de instancias que agrega a un clúster en ejecución. El argumento `CustomAmiId` se puede usar con

el comando `add-instance-groups`, como se muestra en el siguiente ejemplo. Observe que se utiliza el mismo ID de AMI personalizada múltiple (ami-123456) en más de un nodo.

```
aws emr create-cluster --instance-groups
InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-234567

{
    "ClusterId": "j-123456",
    ...
}

aws emr add-instance-groups --cluster-id j-123456 --instance-groups
 InstanceGroupType=Task,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-345678
```

Example - Uso de la versión 1.20.21 o posterior de AWS CLI para crear un clúster de flota de instancias, varias AMI personalizadas, varios tipos de instancias, un nodo maestro bajo demanda, un nodo principal bajo demanda, varios nodos principales y de tarea

```
aws emr create-cluster --instance-fleets
InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge
 CustomAmiId=ami-123456}']
InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,C
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']
InstanceFleetType=TASK,TargetSpotCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,Custo
{InstanceType=m6g.xlarge, CustomAmiId=ami-567890}']
```

Example - Utilice la CLI de AWS versión 1.20.21 o superior para agregar un nodo de tarea a un clúster en ejecución con varios tipos de instancias y varias AMI personalizadas

```
aws emr create-cluster --instance-fleets
InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge
 CustomAmiId=ami-123456}']
InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,C
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']

{
    "ClusterId": "j-123456",
    ...
}
```

```
aws emr add-instance-fleet --cluster-id j-123456 --instance-fleet
InstanceFleetType=TASK,TargetSpotCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge,Custo
{InstanceType=m6g.xlarge, CustomAmiId=ami-345678}']
```

Administración de actualizaciones de repositorio de paquetes de la AMI

Durante el primer arranque, de forma predeterminada, las AMI de Amazon Linux se conectan a repositorios de paquetes para instalar las actualizaciones de seguridad antes de que inicien otros servicios. En función de sus requisitos, puede optar por desactivar estas actualizaciones al especificar una AMI personalizada para Amazon EMR. La opción de deshabilitar esta característica solo está disponible cuando se utiliza una AMI personalizada. De forma predeterminada, las actualizaciones del kernel de Amazon Linux y otros paquetes de software que requieren un reinicio no se actualizan. Tenga en cuenta que la configuración de red debe permitir la salida de HTTP y HTTPS a los repositorios de Amazon Linux en Amazon S3; de lo contrario, las actualizaciones de seguridad no se realizarán correctamente.

> **Warning**
>
> Recomendamos encarecidamente que elija actualizar todos los paquetes instalados durante el reinicio al especificar una AMI personalizada. Si se elige no actualizar paquetes se crean riesgos de seguridad adicionales.

Con la AWS Management Console, puede seleccionar la opción de desactivar las actualizaciones al elegir AMI personalizada.

Con la AWS CLI, puede especificar `--repo-upgrade-on-boot NONE` junto con `--custom-ami-id` al utilizar el comando create-cluster.

Con la API de Amazon EMR, puede especificar `NONE` para el parámetro RepoUpgradeOnBoot.

Creación de una AMI personalizada con un volumen de dispositivo raíz de Amazon EBS cifrado

Para cifrar el volumen de dispositivo raíz de Amazon EBS de una AMI de Amazon Linux para Amazon EMR, copie una imagen de instantánea desde una AMI sin cifrar a un destino cifrado. Para más información sobre la creación de volúmenes cifrados de EBS, consulte Cifrado de Amazon

EBS en la Guía del usuario de Amazon EC2 para instancias de Linux. La AMI de origen para la instantánea puede ser la AMI base de Amazon Linux o puede copiar una instantánea de una AMI derivada de la AMI base de Amazon Linux que haya personalizado.

> **Note**
>
> A partir de la versión 5.24.0 de Amazon EMR, podrá utilizar una opción de configuración de seguridad para cifrar los volúmenes de almacenamiento y el dispositivo raíz de EBS cuando especifique AWS KMS como su proveedor de claves. Para obtener más información, consulte Cifrado de disco local.

Puede utilizar un proveedor de claves externas o una clave de AWS KMS para cifrar el volumen raíz de EBS. Se debe permitir al rol de servicio que utiliza Amazon EMR (normalmente, el `EMR_DefaultRole` predeterminado) que cifre y descifre el volumen, como mínimo, para que Amazon EMR pueda crear un clúster con la AMI. Cuando se utiliza AWS KMS como proveedor de claves, esto significa que se deben permitir las siguientes acciones:

- `kms:encrypt`
- `kms:decrypt`
- `kms:ReEncrypt*`
- `kms:CreateGrant`
- `kms:GenerateDataKeyWithoutPlaintext"`
- `kms:DescribeKey"`

La forma más sencilla de hacerlo consiste en agregar el rol como usuario clave, tal y como se describe en el siguiente tutorial. Se proporciona la siguiente instrucción de política de ejemplo por si tiene que personalizar políticas de rol.

```
{
"Version": "2012-10-17",
"Statement": [
   {
   "Sid": "EmrDiskEncryptionPolicy",
   "Effect": "Allow",
   "Action": [
      "kms:Encrypt",
```

```
      "kms:Decrypt",
      "kms:ReEncrypt*",
      "kms:CreateGrant",
      "kms:GenerateDataKeyWithoutPlaintext",
      "kms:DescribeKey"
      ],
   "Resource": [
      "*"
      ]
   }
]
}
```

Tutorial: Creación de una AMI personalizada con un volumen de dispositivo raíz cifrado utilizando una clave de KMS

El primer paso de este ejemplo consiste en encontrar el ARN de una clave de KMS o crear una nueva. Para más información sobre la creación de claves, consulte Creación de claves en la Guía para desarrolladores de AWS Key Management Service. El siguiente procedimiento muestra cómo añadir el rol de servicio predeterminado, `EMR_DefaultRole`, como un usuario clave a la política de claves. Anote el valor ARN para la clave en el momento de crearla o editarla. Utilice el ARN más superior al crear la AMI.

Para agregar el rol de servicio para Amazon EC2 a la lista de usuarios de claves de cifrado utilizando la consola

1. Inicie sesión en la AWS Management Console y abra la consola AWS Key Management Service (AWS KMS) en https://console.aws.amazon.com/kms.
2. Para cambiar la Región de AWS, utilice el Selector de regiones ubicado en la esquina superior derecha de la página.
3. Elija el alias de la clave de KMS que vaya a utilizar.
4. En la página de detalles de la clave, en Key Users (Usuarios de claves), seleccione Add (Añadir).
5. En el cuadro de diálogo Asociar, elija el rol de servicio de Amazon EMR. El nombre del rol predeterminado es `EMR_DefaultRole`.
6. Elija Attach (Adjuntar).

Para crear una AMI cifrada usando la AWS CLI

- Utilice el comando `aws ec2 copy-image` desde la AWS CLI para crear una AMI con un volumen de dispositivo raíz de EBS cifrado y la clave que ha modificado. Sustituya el valor `--kms-key-id` especificado con todo el ARN de la clave que ha creado o modificado antes.

  > **Note**
  >
  > Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws ec2 copy-image --source-image-id MyAmiId \
--source-region us-west-2 --name MyEncryptedEMRAmi \
--encrypted --kms-key-id arn:aws:kms:us-west-2:12345678910:key/xxxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx
```

La salida del comando proporciona el ID de la AMI que ha creado, que puede especificar al crear un clúster. Para obtener más información, consulte Utilice una sola AMI personalizada en un clúster de EMR. También puede elegir personalizar esta AMI instalando software y realizando otras configuraciones. Para obtener más información, consulte Creación de una AMI de Amazon Linux personalizada a partir de una instancia preconfigurada.

Prácticas recomendadas y consideraciones

Cuando cree una AMI personalizada para Amazon EMR, tenga en cuenta lo siguiente:

- La versión 5.30.0 y posteriores de Amazon EMR y la serie 6.x de Amazon EMR se basan en Amazon Linux 2. Para estas versiones de Amazon EMR, debe utilizar imágenes basadas en Amazon Linux 2 para AMI personalizadas. Para buscar una AMI personalizada base, consulte Búsqueda de una AMI de Linux.
- Para las versiones de Amazon EMR anteriores a 5.30.0 y 6.x, no se admiten las AMI de Amazon Linux 2.
- Debe utilizar una AMI de Amazon Linux de 64 bits. No se admite una AMI de 32 bits.
- No se admiten las AMI de Amazon Linux con varios volúmenes de Amazon EBS.

- Base la personalización en la AMI de Amazon Linux respaldada por EBS más reciente. Para obtener una lista de las AMI de Amazon Linux y los ID de AMI correspondientes, consulte AMI de Amazon Linux.
- No copie una instantánea de una instancia de Amazon EMR existente para crear una AMI personalizada. Esto provoca errores.
- Solo son compatibles el tipo de virtualización HVM y las instancias compatibles con Amazon EMR. Asegúrese de seleccionar la imagen de HVM y un tipo de instancia compatible con Amazon EMR a medida que realice el proceso de personalización de la AMI. Para conocer las instancias y tipos de virtualización compatibles, consulte Tipos de instancias admitidas.
- El rol de servicio deben tener permisos de lanzamiento en la AMI, por lo que la AMI debe ser pública o debe ser el propietario de la AMI o que el propietario la haya compartido con usted.
- La creación de usuarios en la AMI con el mismo nombre que las aplicaciones provoca errores (por ejemplo, `hadoop`, `hdfs`, `yarn` o `spark`).
- El contenido de `/tmp`, `/var` y `/emr` (si se encuentran en la AMI), se traslada a `/mnt/tmp`, `/mnt/var` y `/mnt/emr` respectivamente durante el inicio. Los archivos se conservan, pero si hay una gran cantidad de datos, el startup puede tardar más de lo esperado.
- Si utiliza una AMI personalizada de Amazon Linux basada en una AMI de Amazon Linux con una fecha de creación del 11 de agosto de 2018, el servidor de Oozie no podrá iniciarse. Si utiliza Oozie, cree una AMI personalizada basada en un ID de AMI de Amazon Linux con una fecha de creación diferente. Puede utilizar el siguiente comando de la AWS CLI para obtener una lista de los ID de imagen de todas las AMI de HVM de Amazon Linux con la versión 2018.03, junto con la fecha de lanzamiento, de modo que pueda elegir una AMI de Amazon Linux adecuada como base. Sustituya MyRegion por el identificador de su región, como 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
```

- En los casos en los que utilice una VPC con un nombre de dominio no estándar y AmazonProvidedDNS, no debe utilizar la opción `rotate` en la configuración de DNS del sistema operativo.

Para más información, consulte Creación de una AMI de Linux con respaldo en Amazon EBS en la Guía del usuario de Amazon EC2 para instancias de Linux.

Cambio de versión de Amazon Linux al crear un clúster

Cuando lanza un clúster con Amazon EMR 6.6.0 o una versión posterior, se utiliza automáticamente la versión más reciente de Amazon Linux 2 que se haya validado para la AMI de Amazon EMR predeterminada. Puede especificar una versión de Amazon Linux diferente para su clúster con la consola Amazon EMR o AWS CLI.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para cambiar la versión de Amazon Linux al crear un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Versión de EMR, elija emr-6.6.0 o una superior.
4. En Opciones del sistema operativo, seleccione Versión de Amazon Linux y seleccione la casilla Aplicar automáticamente las últimas actualizaciones de Amazon Linux.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Old console

Para cambiar la versión de Amazon Linux al crear un clúster con la consola antigua

1. Abra la consola de Amazon EMR en https://console.aws.amazon.com/elasticmapreduce.
2. Elija Crear clúster y vaya a Opciones avanzadas.
3. En Configuración de software, en Versión, elija emr-6.6.0 o una versión posterior y, a continuación, elija las demás opciones que necesite para la aplicación. Elija Next (Siguiente).

4. Seleccione valores en Hardware Configuration (Configuración de hardware) que sean adecuados para la aplicación y elija Next (Siguiente).
5. En Opciones adicionales, seleccione la versión de Amazon Linux y, a continuación, la versión de lanzamiento de Amazon Linux para su clúster.
6. Para iniciar el clúster, elija Next (Siguiente) y complete las demás opciones de configuración.

AWS CLI

Para cambiar la versión de Amazon Linux al crear un clúster con la AWS CLI

- Utilice el parámetro `--os-release-label` para especificar la versión de Amazon Linux al ejecutar el comando `aws emr` create-cluster.

```
aws emr create-cluster --name "Cluster with Different Amazon Linux Release" \
--os-release-label 2.0.20210312.1 \
--release-label emr-6.6.0 --use-default-roles \
--instance-count 2 --instance-type m5.xlarge
```

Especificación del tamaño de volumen de dispositivo raíz de Amazon EBS

Esta opción solo está disponible en las versiones 4.x y posteriores de Amazon EMR. Puede especificar el tamaño del volumen desde 10 GiB (el valor predeterminado) hasta 100 GiB al crear un clúster con la AWS Management Console, la AWS CLI o la API de Amazon EMR. Este tamaño se aplica únicamente al volumen de dispositivo raíz de EBS a todas las instancias del clúster. No se aplica a volúmenes de almacenamiento, que se especifican de forma independiente para cada tipo de instancia al crear el clúster.

Para más información acerca de Amazon EBS, consulte Volumen de dispositivo raíz de Amazon EC2.

> **Note**
>
> Si utiliza la AMI predeterminada, Amazon EMR adjunta SSD de uso general (gp2) como el tipo de volumen de dispositivo raíz. Una AMI personalizada puede tener otro tipo de volumen de dispositivo raíz distinto. Sin embargo, el tamaño mínimo del volumen raíz para una AMI personalizada también es de 10 GiB. Para obtener más información, consulte Uso de una AMI personalizada.

El costo del volumen de dispositivo raíz de EBS se prorratea por hora, en función de los cargos mensuales de EBS para ese tipo de volumen en la región en la que se ejecuta el clúster. Lo mismo ocurre con los volúmenes de almacenamiento. Los cargos se muestran en GB, pero debe especificar el tamaño del volumen raíz en GiB, por lo es posible que desee tener en cuenta esto en las estimaciones (1 GB es 0,931323 GiB). Para estimar los cargos asociados con los volúmenes de dispositivo raíz de EBS en su clúster, utilice la siguiente fórmula:

($EBS GB/mes)×0,931323÷30÷24×EMR_EBSRootGiB * InstanceCount

Por ejemplo, tome un clúster que tenga un nodo maestro, un nodo principal y que utilice la AMI base de Amazon Linux con el volumen de dispositivo raíz de 10 GiB predeterminado. Si el costo de EBS en la región es de 0,10 USD/GB mensuales, resulta ser aproximadamente 0,00129 USD por instancia por hora y 0,00258 USD por hora para el clúster (0,10 GB al mes dividido entre 30 días, dividido entre 24 horas, multiplicado por 10 GB, multiplicado por 2 instancias de clúster).

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para especificar el tamaño de volumen de dispositivo raíz de Amazon EBS utilizando la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Configuración de clúster, vaya a la sección Opción de aprovisionamiento y escalado de clústeres. Expanda la flecha del volumen raíz de EBS e introduzca un valor entre 10 GiB y 100 GiB.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para especificar el tamaño de volumen de dispositivo raíz de Amazon EBS utilizando la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Go to advanced options (Ir a las opciones avanzadas).
4. En Configuración de software, en Versión, elija la versión 4.x o versiones posteriores de Amazon EMR. Elija otras opciones que sean necesarias para su aplicación y seleccione Siguiente.
5. En Hardware Configuration (Configuración de hardware), en Root device EBS volume size (Tamaño del volumen de EBS del dispositivo raíz), introduzca un valor entre 10 GiB y 100 GiB.

CLI

Para especificar el tamaño de volumen de dispositivo raíz de Amazon EBS utilizando la AWS CLI

- Utilice el parámetro `--ebs-root-volume-size` del comando create-cluster, tal y como se muestra en el siguiente ejemplo.

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --release-label emr-5.7.0 \
--ebs-root-volume-size 20 --instance-groups InstanceGroupType=MASTER,\
InstanceCount=1,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=2,InstanceType=m5.xlarge
```

Configuración de software del clúster

Cuando selecciona una versión de software, Amazon EMR utiliza una imagen de máquina de Amazon (AMI) con Amazon Linux para instalar el software que usted elija al lanzar un clúster, como Hadoop, Spark y Hive. Amazon EMR ofrece nuevas versiones de forma periódica, agregando nuevas características y aplicaciones, además de actualizaciones generales. Le recomendamos que utilice la última versión para lanzar el clúster siempre que sea posible. La última versión es la opción predeterminada al lanzar un clúster desde la consola.

Para más información sobre las versiones de Amazon EMR y las versiones de software disponibles en cada lanzamiento, consulte la Guía de versiones de Amazon EMR. Para más información sobre cómo editar las configuraciones predeterminadas de aplicaciones y software instalados en su clúster, consulte Configuración de aplicaciones en la Guía de versiones de Amazon EMR. Algunas versiones de los componentes del ecosistema de Hadoop y Spark de código abierto que se incluyen en las versiones de Amazon EMR tienen parches y mejoras, que se documentan en la Guía de versiones de Amazon EMR.

Además del software y las aplicaciones estándar que están disponibles para su instalación en su clúster, puede utilizar acciones de arranque para instalar software personalizado. Las acciones de arranque son scripts que se ejecutan en las instancias cuando se lanza el clúster y que se ejecutan en nuevos nodos que se añaden a su clúster cuando se crean. Las acciones de arranque también son útiles para invocar comandos de la AWS CLI en cada nodo para copiar objetos desde Amazon S3 a cada nodo del clúster.

> **Note**
>
> Las acciones de arranque se utilizan de forma distinta en Amazon EMR versión 4.x y posteriores. Para más información sobre estas diferencias respecto a las versiones 2.x y 3.x de la AMI de Amazon EMR, consulte Diferencias introducidas en la versión 4.x en la Guía de versiones de Amazon EMR.

Crear acciones de arranque para instalar software adicional

Puede utilizar una acción de arranque para instalar software adicional o personalizar la configuración de las instancias de clúster. Las acciones de arranque son scripts que se ejecutan en el clúster después de que Amazon EMR lanza la instancia mediante la imagen de máquina de Amazon (AMI)

de Amazon Linux. Las acciones de arranque se ejecutan antes de que Amazon EMR instale las aplicaciones que se especifican al crear el clúster y antes de que los nodos del clúster comiencen a procesar los datos. Si añade nodos a un clúster en ejecución, las acciones de arranque también se ejecutan en esos nodos de la misma manera. Puede crear acciones de arranque personalizadas y especificarlas al crear el clúster.

La mayoría de las acciones de arranque predefinidas para las versiones 2.x y 3.x de la AMI de Amazon EMR no se admiten en sus versiones 4.x. Por ejemplo, `configure-Hadoop` y `configure-daemons` no se admiten en la versión 4.x de Amazon EMR. En su lugar, Amazon EMR versión 4.x proporciona esta funcionalidad de forma nativa. Para más información sobre cómo migrar las acciones de arranque de las versiones 2.x y 3.x de la AMI de Amazon EMR a la versión 4.x de Amazon EMR, consulte Personalización de la configuración de clústeres y aplicaciones con versiones anteriores de la AMI de Amazon EMR en la Guía de versiones de Amazon EMR.

Conceptos básicos de las acciones de arranque

Las acciones de arranque se ejecutan como usuario de Hadoop de forma predeterminada. Puede ejecutar una acción de arranque con privilegios raíz utilizando `sudo`.

Todas las interfaces de administración de Amazon EMR admiten acciones de arranque. Puede especificar hasta 16 acciones de arranque por clúster proporcionando varios parámetros `bootstrap-actions` desde la consola, la AWS CLI o la API.

Desde la consola de Amazon EMR, puede especificar opcionalmente una acción de arranque al crear un clúster.

Cuando utilice la CLI, puede pasar referencias a scripts de acciones de arranque a Amazon EMR si agrega el parámetro `--bootstrap-actions` al crear el clúster mediante el comando `create-cluster`.

```
--bootstrap-actions Path="s3://mybucket/filename",Args=[arg1,arg2]
```

Si la acción de arranque devuelve un código de error distinto de cero, Amazon EMR lo trata como un error y termina la instancia. Si hay demasiadas instancias que generen error en sus acciones de arranque, entonces Amazon EMR termina el clúster. Si solo unas pocas instancias generan error, Amazon EMR intenta reasignar las instancias erróneas y continuar. Utilice el código de error `lastStateChangeReason` del clúster para identificar los errores provocados por una acción de arranque.

Ejecución condicional de una acción de arranque

Para ejecutar solo una acción de arranque en el nodo maestro, puede usar una acción de arranque personalizada con un poco de lógica para determinar si el nodo es maestro.

```
#!/bin/bash
if grep isMaster /mnt/var/lib/info/instance.json | grep false;
then
    echo "This is not master node, do nothing,exiting"
    exit 0
fi
echo "This is master, continuing to execute script"
# continue with code logic for master node below
```

El siguiente resultado se imprimirá desde un nodo principal.

```
This is not master node, do nothing, exiting
```

El siguiente resultado se imprimirá desde el nodo maestro.

```
This is master, continuing to execute script
```

Para utilizar esta lógica, cargue su acción de arranque, incluido el código anterior, en su bucket de Amazon S3. En AWS CLI, agregue el parámetro `--bootstrap-actions` a la llamada a la API `aws emr create-cluster` y especifique la ubicación del script de arranque como valor de `Path`.

Acciones de apagado

Una script de acción de arranque puede crear una o más acciones de apagado escribiendo scripts en el directorio `/mnt/var/lib/instance-controller/public/shutdown-actions/`. Cuando un clúster se termina, todos los scripts en este directorio se ejecutan en paralelo. Cada script se debe ejecutar y completar en un plazo de 60 segundos.

No se garantiza la ejecución de los scripts de acción de apagado si el nodo termina con un error.

> **Note**
>
> Cuando se utiliza la versión 4.0 y posteriores de Amazon EMR, debe crear manualmente el directorio `/mnt/var/lib/instance-controller/public/shutdown-actions/` en el nodo maestro. No existe de forma predeterminada; sin embargo, después de crearlos,

los scripts en este directorio se ejecutan a pesar de todo antes del apagado. Para obtener más información acerca de la conexión al nodo principal para crear directorios, consulte Conectarse al nodo principal mediante SSH.

Usar acciones de arranque personalizadas

Puede crear un script personalizado para realizar una acción de arranque personalizada. Cualquiera de las interfaces de Amazon EMR puede hacer referencia a una acción de arranque personalizada.

> **Note**
>
> Para obtener el mejor rendimiento, se recomienda que almacene las acciones de arranque personalizadas, los scripts y otros archivos que desee utilizar con Amazon EMR en un bucket de Amazon S3 que se encuentre en la misma Región de AWS que su clúster.

Contenido

- Agregar acciones de arranque personalizadas
- Usar una acción de arranque personalizada para copiar un objeto de Amazon S3 en cada nodo

Agregar acciones de arranque personalizadas

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para crear un clúster con una acción de arranque personalizada mediante la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.

3. En Acciones de arranque, elija Agregar para especificar un nombre, una ubicación del script y argumentos opcionales para la acción. Seleccione Agregar acción de arranque.
4. Opcionalmente, agregue más acciones de arranque.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Old console

Para crear un clúster con una acción de arranque personalizada mediante la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Haga clic en Go to advanced options (Ir a las opciones avanzadas).
4. En los pasos 1 y 2 de Create Cluster - Advanced Options (Crear clúster: opciones avanzadas), elija las opciones que desee y continúe en el Step 3: General Cluster Settings (Paso 3: Configuración general del clúster).
5. En Bootstrap Actions (Acciones de arranque) seleccione Configure and add (Configurar y añadir) para especificar el nombre, la ubicación del JAR y los argumentos de la acción de arranque. Elija Add (Agregar).
6. Opcionalmente, añada más acciones de arranque como desee.
7. Continúe para crear el clúster. Las acciones de arranque se realizarán después de que el clúster se haya aprovisionado e inicializado.

   Mientras el nodo principal del clúster está en ejecución, puede conectarse al nodo principal y ver los archivos de registro que el script de la acción de arranque generó en el directorio `/mnt/var/log/bootstrap-actions/1`.

CLI

Para crear un clúster con una acción de arranque personalizada mediante la AWS CLI

Cuando utilice la AWS CLI para incluir una acción de arranque, especifique `Path` y `Args` como una lista separada por comas. El siguiente ejemplo no utiliza una lista de argumentos.

- Para lanzar un clúster con una acción de arranque personalizada, escriba el comando siguiente y sustituya *myKey* por el nombre del par de claves de EC2. Incluya `--bootstrap-actions` como parámetro y especifique la ubicación del script de arranque como valor de `Path`.

  - Usuarios de Linux, UNIX y Mac OS X:

    ```
    aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 \
    --use-default-roles --ec2-attributes KeyName=myKey \
    --applications Name=Hive Name=Pig \
    --instance-count 3 --instance-type m5.xlarge \
    --bootstrap-actions Path="s3://elasticmapreduce/bootstrap-actions/download.sh"
    ```

  - Usuarios de Windows:

    ```
    aws emr create-cluster --name "Test cluster" --release-label emr-4.2.0 --use-
    default-roles --ec2-attributes KeyName=myKey --applications Name=Hive Name=Pig
     --instance-count 3 --instance-type m5.xlarge --bootstrap-actions Path="s3://
    elasticmapreduce/bootstrap-actions/download.sh"
    ```

  Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos básicos. Todos los nodos utilizarán el tipo de instancia especificado en el comando.

  > **Note**
  >
  > Si no ha creado con anterioridad el rol de servicio de Amazon EMR predeterminado y el perfil de instancia de EC2, escriba `aws emr create-default-roles` para crearlos antes de escribir el subcomando `create-cluster`.

  Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Usar una acción de arranque personalizada para copiar un objeto de Amazon S3 en cada nodo

Puede utilizar una acción de arranque para copiar objetos desde Amazon S3 en cada nodo de un clúster antes de que se instalen las aplicaciones. La AWS CLI se instala en cada nodo del clúster, por lo que la acción de arranque puede llamar a los comandos de la AWS CLI.

En el siguiente ejemplo, se muestra un script sencillo de acción de arranque que copia el archivo `myfile.jar` desde Amazon S3 en la carpeta local `/mnt1/myfolder` de cada nodo del clúster. El script se guarda en Amazon S3 con el nombre de archivo `copymyfile.sh` y contiene lo siguiente.

```
#!/bin/bash
aws s3 cp s3://mybucket/myfilefolder/myfile.jar /mnt1/myfolder
```

Al lanzar el clúster, debe especificar el script. En el siguiente ejemplo de la AWS CLI se muestra cómo hacerlo:

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 \
--use-default-roles --ec2-attributes KeyName=myKey \
--applications Name=Hive Name=Pig \
--instance-count 3 --instance-type m5.xlarge \
--bootstrap-actions Path="s3://mybucket/myscriptfolder/copymyfile.sh"
```

Configuración del hardware y las redes de los clústeres

Una consideración importante al crear un clúster de Amazon EMR es cómo configurar instancias de Amazon EC2 y las opciones de red. En este capítulo, se abordan las siguientes opciones y, a continuación, se combinan todas ellas con las directrices y prácticas recomendadas.

- Tipos de nodos: Las instancias de EC2 de un clúster de EMR se organizan en tipos de nodos. Existen tres tipos: el nodo principal, el nodo básico y los nodos de tarea. Cada tipo de nodo realiza un conjunto de roles definidos por las aplicaciones distribuidas que se instalan en el clúster. Por ejemplo, durante un trabajo de Hadoop MapReduce o Spark, los componentes de los nodos básicos y de tareas procesan los datos, transfieren la salida a Amazon S3 o HDFS y devuelven los metadatos de estado al nodo principal. Para un clúster de un solo nodo, todos los componentes se ejecutan en el nodo principal. Para obtener más información, consulte Descripción de los tipos de nodos: principales, básicos y de tarea.
- Instancias de EC2: cuando crea un clúster, toma decisiones sobre las instancias de Amazon EC2 en las que se ejecutará cada tipo de nodo. El tipo de instancia de EC2 determina el perfil de procesamiento y almacenamiento del nodo. La elección de la instancia de Amazon EC2 para los

nodos es importante porque determina el perfil de rendimiento de los tipos de nodos individuales del clúster. Para obtener más información, consulte Configuración de instancias de Amazon EC2.

- Redes: puede lanzar su clúster de Amazon EMR en una VPC mediante una subred pública, una subred privada o una subred compartida. Su configuración de red determina cómo los clientes y los servicios pueden conectarse a los clústeres para realizar su trabajo, cómo se conectan los clústeres a los almacenes de datos y otros recursos de AWS, y las opciones de las que dispone para controlar el tráfico en esas conexiones. Para obtener más información, consulte Configurar redes.
- Agrupación de instancias: el conjunto de instancias de EC2 que aloja cada tipo de nodo se denomina flota de instancias o grupo de instancias uniforme. La configuración de agrupación de instancias es una elección que se toma al crear un clúster. Esta opción determina cómo puede agregar nodos al clúster mientras está en ejecución. La configuración se aplica a todos los tipos de nodos. No se puede cambiar más adelante. Para obtener más información, consulte Crear un clúster con flotas de instancias o grupos de instancias uniformes.

> **Note**
>
> La configuración de las flotas de instancias está disponible solo en las versiones 4.8.0 y posteriores de Amazon EMR, excluidas las versiones 5.0.0 y 5.0.3.

Descripción de los tipos de nodos: principales, básicos y de tarea

Utilice esta sección para entender cómo Amazon EMR utiliza cada uno de estos tipos de nodos y como base para la planificación de capacidad de los clústeres.

Nodo principal

El nodo principal administra el clúster y normalmente ejecuta los componentes principales de las aplicaciones distribuidas. Por ejemplo, el nodo principal ejecuta el servicio YARN ResourceManager para administrar recursos de aplicaciones. También ejecuta el servicio HDFS NameNode, hace un seguimiento del estado de los trabajos enviados al clúster y supervisa la salud de los grupos de instancias.

Para supervisar el progreso de un clúster e interactuar directamente con las aplicaciones, puede conectar con el nodo principal a través de SSH como usuario de Hadoop. Para obtener más información, consulte Conectarse al nodo principal mediante SSH. La conexión con el nodo principal le permite acceder directamente a los directorios y los archivos, como los archivos de registro de

Hadoop. Para obtener más información, consulte Ver archivos de registro de . También puede ver las interfaces de usuario que publican las aplicaciones como sitios web que se ejecutan en el nodo principal. Para obtener más información, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

> **Note**
>
> Con la versión 5.23.0 y posteriores de Amazon EMR, puede lanzar un clúster con tres nodos principales para admitir una alta disponibilidad de aplicaciones como YARN Resource Manager, HDFS NameNode, Spark, Hive y Ganglia. El nodo principal ya no es un posible punto único de error con esta característica. Si se produce un error en uno de los nodos principales, se produce un error automáticamente en Amazon EMR a través de un nodo principal en espera y se sustituye el nodo principal con error por uno nuevo con las mismas acciones de arranque y configuración. Para más información, consulte Planificación y configuración de nodos principales.

Nodos básicos

Los nodos básicos los administra el nodo principal. Los nodos secundarios ejecutan el daemon del nodo de datos para coordinar el almacenamiento de datos como parte del Sistema de archivos distribuido de Hadoop (HDFS). También ejecutan el daemon Task Tracker y realizan otras tareas de cálculo en paralelo en los datos que las aplicaciones instaladas requieren. Por ejemplo, un nodo secundario ejecuta daemons NodeManager de YARN, tareas MapReduce de Hadoop y ejecutores de Spark.

Solo hay un grupo de instancias principal o una flota de instancias por clúster, pero puede haber varios nodos ejecutándose en varias instancias de Amazon EC2 en el grupo o la flota de instancias. Con los grupos de instancia, puede agregar y eliminar instancias de Amazon EC2 mientras se está ejecutando el clúster. También puede configurar el escalado automático para agregar instancias en función del valor de una métrica. Para más información sobre cómo agregar y eliminar instancias de Amazon EC2 con la configuración de los grupos de instancias, consulte Usar el escalado de clústeres.

Con las flotas de instancias, puede agregar y eliminar instancias de forma eficaz modificando las capacidades de destino de la flota de instancias para las opciones bajo demanda y de spot según corresponda. Para obtener más información acerca de las capacidades de destino, consulte Opciones de flota de instancias.

> **Warning**
>
> La eliminación de daemons de HDFS de un nodo secundario de ejecución o la terminación de nodos secundarios conlleva el riesgo de pérdida de datos. Tenga cuidado al configurar nodos secundarios que utilizan instancias de spot. Para obtener más información, consulte ¿Cuándo se deben utilizar las instancias de spot?.

Nodos de tarea

Puede utilizar nodos de tarea para agregar potencia para llevar a cabo tareas de computación paralela en los datos, como las tareas de Hadoop MapReduce y ejecutores de Spark. Los nodos de tareas no ejecutan el daemon de nodo de datos, ni tampoco almacenan datos en HDFS. Al igual que con los nodos básicos, puede agregar nodos de tarea a un clúster agregando instancias de EC2 en un grupo de instancias uniforme existente o modificando las capacidades de destino para una tarea de la flota de instancias.

Con la configuración de grupo de instancias uniforme, puede tener hasta un total de 48 grupos de instancias de tarea. La capacidad de agregar grupos de instancias de este modo le permite combinar tipos de instancia de Amazon EC2 y opciones de precio, como las instancias bajo demanda e instancias de spot. Esto le ofrece flexibilidad para responder a los requisitos de carga de trabajo de forma rentable.

Con la configuración de flota de instancias, la capacidad de combinar tipos de instancia y opciones de compra está integrada, por lo que solo hay una flota de instancias de tarea.

Dado que las instancias de spot se utilizan a menudo para ejecutar nodos de tarea, Amazon EMR tiene una funcionalidad predeterminada para programar trabajos de YARN, de modo que los trabajos en ejecución no presenten errores cuando los nodos de tarea que se ejecutan en las instancias de spot se terminen. Para ello, Amazon EMR permite que los procesos maestros de la aplicación se ejecuten únicamente en los nodos principales. El proceso maestro de la aplicación controla los trabajos en ejecución y debe mantenerse activo durante toda la vida del trabajo.

La versión 5.19.0 y posteriores de Amazon EMR utilizan la característica integrada de etiquetas de nodo YARN para lograrlo. (Las versiones anteriores utilizaban una revisión de código). Las propiedades en las clasificaciones de configuración `yarn-site` y `capacity-scheduler` se ajustan de forma predeterminada para que capacity-scheduler y fair-scheduler de YARN utilicen las etiquetas de nodo. Amazon EMR etiqueta automáticamente los nodos principales con la etiqueta `CORE` y establece las propiedades para que los maestros de la aplicación se programen únicamente

en los nodos con la etiqueta CORE. La modificación manual de las propiedades relacionadas en las clasificaciones de configuración yarn-site y capacity-scheduler o directamente en los archivos XML asociados podría interrumpir esta característica o modificar esta funcionalidad.

A partir de la serie de versiones 6.x de Amazon EMR, la característica de etiquetas de nodo YARN está desactivada de forma predeterminada. De forma predeterminada, los procesos principales de la aplicación se pueden ejecutar tanto en nodos básicos como en nodos de tarea. Puede habilitar la función de etiquetas de nodo YARN configurando las siguientes propiedades:

- `yarn.node-labels.enabled: true`
- `yarn.node-labels.am.default-node-label-expression: 'CORE'`

Para obtener información sobre propiedades específicas, consulte Configuración de Amazon EMR para evitar errores en los trabajos debido a la terminación de instancias de spot de los nodos de tarea.

Configuración de instancias de Amazon EC2

Las instancias EC2 aparecen en diferentes configuraciones conocidas como tipos de instancia. Cada tipo de instancia tiene una CPU, entrada/salida y capacidad de almacenamiento diferentes. Además del tipo de instancia, puede elegir distintas opciones de compra para instancias de Amazon EC2. Puede especificar diferentes tipos de instancia y opciones de compra dentro de grupos o flotas de instancias. Para obtener más información, consulte Crear un clúster con flotas de instancias o grupos de instancias uniformes. Para obtener recomendaciones sobre cómo elegir los tipos de instancias y las opciones de compra para su aplicación, consulte Prácticas recomendadas para la configuración del clúster.

> **Important**
>
> Al elegir un tipo de instancia mediante la AWS Management Console, el número de vCPU que se muestra para cada tipo de instancia es el número de núcleos virtuales de YARN para ese tipo de instancia, no el número de vCPU de EC2 para ese tipo de instancia. Para más información sobre el número de vCPU para cada tipo de instancia, consulte Tipos de instancias de Amazon EC2.

Temas

- Tipos de instancias admitidas

- Configurar redes
- Crear un clúster con flotas de instancias o grupos de instancias uniformes

Tipos de instancias admitidas

En esta sección se describen los tipos de instancias que admite Amazon EMR, organizados por Región de AWS. Para más información sobre los tipos de instancias, consulte Instancias de Amazon EC2 y Matriz de tipo de instancia de la AMI de Amazon Linux.

No todos los tipos de instancias están disponibles en todas las regiones y la disponibilidad de las instancias está sujeta a la disponibilidad y la demanda en la región y la zona de disponibilidad especificadas. La zona de disponibilidad de una instancia viene determinada por la subred que utilice para lanzar el clúster.

Consideraciones

Tenga en cuenta lo siguiente al elegir los tipos de instancias del clúster de Amazon EMR.

> **Important**
>
> Al elegir un tipo de instancia mediante la AWS Management Console, el número de vCPU que se muestra para cada tipo de instancia es el número de núcleos virtuales de YARN para ese tipo de instancia, no el número de vCPU de EC2 para ese tipo de instancia. Para más información sobre el número de vCPU para cada tipo de instancia, consulte Tipos de instancias de Amazon EC2.

- Si crea un clúster con un tipo de instancia que no está disponible en la región y la zona de disponibilidad especificadas, es posible que no se pueda aprovisionar el clúster o que se bloquee durante el aprovisionamiento. Para obtener información sobre la disponibilidad de las instancias, consulte la página de precios de Amazon EMR o consulte las tablas de Tipos de instancias admitidos por Región de AWS en esta página.
- A partir de la versión de 5.13.0 de Amazon EMR, todas las instancias utilizan virtualización HVM y almacenamiento respaldado por EBS para los volúmenes raíz. Cuando se utilizan versiones de Amazon EMR anteriores a la 5.13.0, algunas instancias de generaciones anteriores utilizan virtualización PVM. Para más información, consulte Tipos de virtualización de AMI de Linux.
- Algunos tipos de instancia son compatibles con redes mejoradas. Para obtener más información, consulte Redes mejoradas en Linux.

- Los controladores NVIDIA y CUDA se instalan en los tipos de instancia GPU de forma predeterminada.

Tipos de instancias admitidos por Región de AWS

En la siguiente tabla, se describen los tipos de instancias que admite Amazon EMR, organizados por Región de AWS. En las tablas también se enumeran las primeras versiones de Amazon EMR de las series 4.x, 5.x o 6.x que admiten cada tipo de instancia. Por ejemplo, EMR admite instancias m4.16xlarge en Este de EE. UU. (Norte de Virginia) en las versiones 4.8.3, 5.2.1 y 6.0.0 y sus respectivas versiones posteriores.

EE. UU. Este (Norte de Virginia): us-east-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6in.2xlarge	emr-5.36.1, emr-6.10.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0
	m7i.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m7i.2xlarge	emr-5.36.1, emr-6.10.0
	m7i.4xlarge	emr-5.36.1, emr-6.10.0
	m7i.8xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.2xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.4xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6id.24xlarge	emr-5.36.1, emr-6.8.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
	p5.48xlarge	emr-6.14.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

EE.UU. Este (Ohio) (us-east-2)

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0
	m7i.xlarge	emr-5.36.1, emr-6.10.0
	m7i.2xlarge	emr-5.36.1, emr-6.10.0
	m7i.4xlarge	emr-5.36.1, emr-6.10.0
	m7i.8xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.2xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.4xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6id.2xlarge	emr-5.36.1, emr-6.8.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Optimización de almacenamiento	d2.xlarge	emr-4.6.0, emr-5.0.3, emr-6.0.0
	d2.2xlarge	emr-4.6.0, emr-5.0.3, emr-6.0.0
	d2.4xlarge	emr-4.6.0, emr-5.0.3, emr-6.0.0
	d2.8xlarge	emr-4.6.0, emr-5.0.3, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Oeste de EE. UU. (Norte de California): us-west-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

EE.UU. Oeste (Oregón) (us-west-2)

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6id.8xlarge	emr-5.36.1, emr-6.8.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0
	m7i.xlarge	emr-5.36.1, emr-6.10.0
	m7i.2xlarge	emr-5.36.1, emr-6.10.0
	m7i.4xlarge	emr-5.36.1, emr-6.10.0
	m7i.8xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.2xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.4xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
	p5.48xlarge	emr-6.14.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6a.12xlarge	emr-5.36.1, emr-6.8.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0
	i4g.8xlarge	emr-5.36.1, emr-6.10.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

AWS GovCloud (Oeste de EE. UU): us-gov-west-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

AWS GovCloud (Este de EE. UU): us-gov-east-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.16.0, emr-6.0.0
	m5.2xlarge	emr-5.16.0, emr-6.0.0
	m5.4xlarge	emr-5.16.0, emr-6.0.0
	m5.8xlarge	emr-5.16.0, emr-6.0.0
	m5.12xlarge	emr-5.16.0, emr-6.0.0
	m5.16xlarge	emr-5.16.0, emr-6.0.0
	m5.24xlarge	emr-5.16.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.16.0, emr-6.0.0
	m5d.2xlarge	emr-5.16.0, emr-6.0.0
	m5d.4xlarge	emr-5.16.0, emr-6.0.0
	m5d.8xlarge	emr-5.16.0, emr-6.0.0
	m5d.12xlarge	emr-5.16.0, emr-6.0.0
	m5d.16xlarge	emr-5.16.0, emr-6.0.0
	m5d.24xlarge	emr-5.16.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.16.0, emr-6.0.0
	c5.2xlarge	emr-5.16.0, emr-6.0.0
	c5.4xlarge	emr-5.16.0, emr-6.0.0
	c5.9xlarge	emr-5.16.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.12xlarge	emr-5.16.0, emr-6.0.0
	c5.18xlarge	emr-5.16.0, emr-6.0.0
	c5.24xlarge	emr-5.16.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.16.0, emr-6.0.0
	c5d.2xlarge	emr-5.16.0, emr-6.0.0
	c5d.4xlarge	emr-5.16.0, emr-6.0.0
	c5d.9xlarge	emr-5.16.0, emr-6.0.0
	c5d.18xlarge	emr-5.16.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.16.0, emr-6.0.0
	r5.2xlarge	emr-5.16.0, emr-6.0.0
	r5.4xlarge	emr-5.16.0, emr-6.0.0
	r5.8xlarge	emr-5.16.0, emr-6.0.0
	r5.12xlarge	emr-5.16.0, emr-6.0.0
	r5.16xlarge	emr-5.16.0, emr-6.0.0
	r5.24xlarge	emr-5.16.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.16.0, emr-6.0.0
	r5d.2xlarge	emr-5.16.0, emr-6.0.0
	r5d.4xlarge	emr-5.16.0, emr-6.0.0
	r5d.8xlarge	emr-5.16.0, emr-6.0.0
	r5d.12xlarge	emr-5.16.0, emr-6.0.0
	r5d.16xlarge	emr-5.16.0, emr-6.0.0
	r5d.24xlarge	emr-5.16.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-5.16.0, emr-6.0.0
	d2.2xlarge	emr-5.16.0, emr-6.0.0
	d2.4xlarge	emr-5.16.0, emr-6.0.0
	d2.8xlarge	emr-5.16.0, emr-6.0.0
	i3.xlarge	emr-5.16.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.2xlarge	emr-5.16.0, emr-6.0.0
	i3.4xlarge	emr-5.16.0, emr-6.0.0
	i3.8xlarge	emr-5.16.0, emr-6.0.0
	i3.16xlarge	emr-5.16.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

África (Ciudad del Cabo): af-south-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.29.0, emr-6.0.0
	m5.2xlarge	emr-5.29.0, emr-6.0.0
	m5.4xlarge	emr-5.29.0, emr-6.0.0
	m5.8xlarge	emr-5.29.0, emr-6.0.0
	m5.12xlarge	emr-5.29.0, emr-6.0.0
	m5.16xlarge	emr-5.29.0, emr-6.0.0
	m5.24xlarge	emr-5.29.0, emr-6.0.0
	m5d.xlarge	emr-5.29.0, emr-6.0.0
	m5d.2xlarge	emr-5.29.0, emr-6.0.0
	m5d.4xlarge	emr-5.29.0, emr-6.0.0
	m5d.8xlarge	emr-5.29.0, emr-6.0.0
	m5d.12xlarge	emr-5.29.0, emr-6.0.0
	m5d.16xlarge	emr-5.29.0, emr-6.0.0
	m5d.24xlarge	emr-5.29.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.29.0, emr-6.0.0
	c5.2xlarge	emr-5.29.0, emr-6.0.0
	c5.4xlarge	emr-5.29.0, emr-6.0.0
	c5.9xlarge	emr-5.29.0, emr-6.0.0
	c5.12xlarge	emr-5.29.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.18xlarge	emr-5.29.0, emr-6.0.0
	c5.24xlarge	emr-5.29.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.29.0, emr-6.0.0
	c5d.2xlarge	emr-5.29.0, emr-6.0.0
	c5d.4xlarge	emr-5.29.0, emr-6.0.0
	c5d.9xlarge	emr-5.29.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5d.12xlarge	emr-5.29.0, emr-6.0.0
	c5d.18xlarge	emr-5.29.0, emr-6.0.0
	c5d.24xlarge	emr-5.29.0, emr-6.0.0
	c5n.xlarge	emr-5.29.0, emr-6.0.0
	c5n.2xlarge	emr-5.29.0, emr-6.0.0
	c5n.4xlarge	emr-5.29.0, emr-6.0.0
	c5n.9xlarge	emr-5.29.0, emr-6.0.0
	c5n.18xlarge	emr-5.29.0, emr-6.0.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.29.0, emr-6.0.0
	r5.2xlarge	emr-5.29.0, emr-6.0.0
	r5.4xlarge	emr-5.29.0, emr-6.0.0
	r5.8xlarge	emr-5.29.0, emr-6.0.0
	r5.12xlarge	emr-5.29.0, emr-6.0.0
	r5.16xlarge	emr-5.29.0, emr-6.0.0
	r5.24xlarge	emr-5.29.0, emr-6.0.0
	r5d.xlarge	emr-5.29.0, emr-6.0.0
	r5d.2xlarge	emr-5.29.0, emr-6.0.0
	r5d.4xlarge	emr-5.29.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.8xlarge	emr-5.29.0, emr-6.0.0
	r5d.12xlarge	emr-5.29.0, emr-6.0.0
	r5d.16xlarge	emr-5.29.0, emr-6.0.0
	r5d.24xlarge	emr-5.29.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Optimización de almacenamiento	d2.xlarge	emr-5.29.0, emr-6.0.0
	d2.2xlarge	emr-5.29.0, emr-6.0.0
	d2.4xlarge	emr-5.29.0, emr-6.0.0
	d2.8xlarge	emr-5.29.0, emr-6.0.0
	i3.xlarge	emr-5.29.0, emr-6.0.0
	i3.2xlarge	emr-5.29.0, emr-6.0.0
	i3.4xlarge	emr-5.29.0, emr-6.0.0
	i3.8xlarge	emr-5.29.0, emr-6.0.0
	i3.16xlarge	emr-5.29.0, emr-6.0.0
	i3en.xlarge	emr-5.29.0, emr-6.0.0
	i3en.2xlarge	emr-5.29.0, emr-6.0.0
	i3en.3xlarge	emr-5.29.0, emr-6.0.0
	i3en.6xlarge	emr-5.29.0, emr-6.0.0
	i3en.12xlarge	emr-5.29.0, emr-6.0.0
	i3en.24xlarge	emr-5.29.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia-Pacífico (Hong Kong): ap-east-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.20.0, emr-6.0.0
	m5.2xlarge	emr-5.20.0, emr-6.0.0
	m5.4xlarge	emr-5.20.0, emr-6.0.0
	m5.8xlarge	emr-5.20.0, emr-6.0.0
	m5.12xlarge	emr-5.20.0, emr-6.0.0
	m5.16xlarge	emr-5.20.0, emr-6.0.0
	m5.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.20.0, emr-6.0.0
	m5d.2xlarge	emr-5.20.0, emr-6.0.0
	m5d.4xlarge	emr-5.20.0, emr-6.0.0
	m5d.8xlarge	emr-5.20.0, emr-6.0.0
	m5d.12xlarge	emr-5.20.0, emr-6.0.0
	m5d.16xlarge	emr-5.20.0, emr-6.0.0
	m5d.24xlarge	emr-5.20.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.20.0, emr-6.0.0
	c5.2xlarge	emr-5.20.0, emr-6.0.0
	c5.4xlarge	emr-5.20.0, emr-6.0.0
	c5.9xlarge	emr-5.20.0, emr-6.0.0
	c5.12xlarge	emr-5.20.0, emr-6.0.0
	c5.18xlarge	emr-5.20.0, emr-6.0.0
	c5.24xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.20.0, emr-6.0.0
	c5d.2xlarge	emr-5.20.0, emr-6.0.0
	c5d.4xlarge	emr-5.20.0, emr-6.0.0
	c5d.9xlarge	emr-5.20.0, emr-6.0.0
	c5d.18xlarge	emr-5.20.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.20.0, emr-6.0.0
	r5.2xlarge	emr-5.20.0, emr-6.0.0
	r5.4xlarge	emr-5.20.0, emr-6.0.0
	r5.8xlarge	emr-5.20.0, emr-6.0.0
	r5.12xlarge	emr-5.20.0, emr-6.0.0
	r5.16xlarge	emr-5.20.0, emr-6.0.0
	r5.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.20.0, emr-6.0.0
	r5d.2xlarge	emr-5.20.0, emr-6.0.0
	r5d.4xlarge	emr-5.20.0, emr-6.0.0
	r5d.8xlarge	emr-5.20.0, emr-6.0.0
	r5d.12xlarge	emr-5.20.0, emr-6.0.0
	r5d.16xlarge	emr-5.20.0, emr-6.0.0
	r5d.24xlarge	emr-5.20.0, emr-6.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Optimización de almacenamiento	d2.xlarge	emr-5.20.0, emr-6.0.0
	d2.2xlarge	emr-5.20.0, emr-6.0.0
	d2.4xlarge	emr-5.20.0, emr-6.0.0
	d2.8xlarge	emr-5.20.0, emr-6.0.0
	i3.xlarge	emr-5.20.0, emr-6.0.0
	i3.2xlarge	emr-5.20.0, emr-6.0.0
	i3.4xlarge	emr-5.20.0, emr-6.0.0
	i3.8xlarge	emr-5.20.0, emr-6.0.0
	i3.16xlarge	emr-5.20.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia-Pacífico (Yakarta): ap-southeast-3

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.30.2, emr-6.0.1
	m5.2xlarge	emr-5.30.2, emr-6.0.1
	m5.4xlarge	emr-5.30.2, emr-6.0.1
	m5.8xlarge	emr-5.30.2, emr-6.0.1
	m5.12xlarge	emr-5.30.2, emr-6.0.1
	m5.16xlarge	emr-5.30.2, emr-6.0.1
	m5.24xlarge	emr-5.30.2, emr-6.0.1
	m5a.xlarge	emr-5.30.2, emr-6.0.1
	m5a.2xlarge	emr-5.30.2, emr-6.0.1
	m5a.4xlarge	emr-5.30.2, emr-6.0.1
	m5a.8xlarge	emr-5.30.2, emr-6.0.1
	m5a.12xlarge	emr-5.30.2, emr-6.0.1
	m5a.16xlarge	emr-5.30.2, emr-6.0.1
	m5a.24xlarge	emr-5.30.2, emr-6.0.1
	m5d.xlarge	emr-5.30.2, emr-6.0.1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.2xlarge	emr-5.30.2, emr-6.0.1
	m5d.4xlarge	emr-5.30.2, emr-6.0.1
	m5d.8xlarge	emr-5.30.2, emr-6.0.1
	m5d.12xlarge	emr-5.30.2, emr-6.0.1
	m5d.16xlarge	emr-5.30.2, emr-6.0.1
	m5d.24xlarge	emr-5.30.2, emr-6.0.1
	m6g.xlarge	emr-5.30.2, emr-6.1.1
	m6g.2xlarge	emr-5.30.2, emr-6.1.1
	m6g.4xlarge	emr-5.30.2, emr-6.1.1
	m6g.8xlarge	emr-5.30.2, emr-6.1.1
	m6g.12xlarge	emr-5.30.2, emr-6.1.1
	m6g.16xlarge	emr-5.30.2, emr-6.1.1
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.30.2, emr-6.0.1
	c5.2xlarge	emr-5.30.2, emr-6.0.1
	c5.4xlarge	emr-5.30.2, emr-6.0.1
	c5.9xlarge	emr-5.30.2, emr-6.0.1
	c5.12xlarge	emr-5.30.2, emr-6.0.1
	c5.18xlarge	emr-5.30.2, emr-6.0.1
	c5.24xlarge	emr-5.30.2, emr-6.0.1
	c5d.xlarge	emr-5.30.2, emr-6.0.1
	c5d.2xlarge	emr-5.30.2, emr-6.0.1
	c5d.4xlarge	emr-5.30.2, emr-6.0.1
	c5d.9xlarge	emr-5.30.2, emr-6.0.1
	c5d.12xlarge	emr-5.30.2, emr-6.0.1
	c5d.18xlarge	emr-5.30.2, emr-6.0.1
	c5d.24xlarge	emr-5.30.2, emr-6.0.1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.xlarge	emr-5.30.2, emr-6.0.1
	c5n.2xlarge	emr-5.30.2, emr-6.0.1
	c5n.4xlarge	emr-5.30.2, emr-6.0.1
	c5n.9xlarge	emr-5.30.2, emr-6.0.1
	c5n.18xlarge	emr-5.30.2, emr-6.0.1
	c6g.xlarge	emr-5.31.1, emr-6.1.1
	c6g.2xlarge	emr-5.31.1, emr-6.1.1
	c6g.4xlarge	emr-5.31.1, emr-6.1.1
	c6g.8xlarge	emr-5.31.1, emr-6.1.1
	c6g.12xlarge	emr-5.31.1, emr-6.1.1
	c6g.16xlarge	emr-5.31.1, emr-6.1.1
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
Optimizada para memoria	r5.xlarge	emr-5.30.2, emr-6.0.1
	r5.2xlarge	emr-5.30.2, emr-6.0.1
	r5.4xlarge	emr-5.30.2, emr-6.0.1
	r5.8xlarge	emr-5.30.2, emr-6.0.1
	r5.12xlarge	emr-5.30.2, emr-6.0.1
	r5.16xlarge	emr-5.30.2, emr-6.0.1
	r5.24xlarge	emr-5.30.2, emr-6.0.1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.xlarge	emr-5.30.2, emr-6.0.1
	r5d.2xlarge	emr-5.30.2, emr-6.0.1
	r5d.4xlarge	emr-5.30.2, emr-6.0.1
	r5d.8xlarge	emr-5.30.2, emr-6.0.1
	r5d.12xlarge	emr-5.30.2, emr-6.0.1
	r5d.16xlarge	emr-5.30.2, emr-6.0.1
	r5d.24xlarge	emr-5.30.2, emr-6.0.1
	r6g.xlarge	emr-5.31.1, emr-6.1.1
	r6g.2xlarge	emr-5.31.1, emr-6.1.1
	r6g.4xlarge	emr-5.31.1, emr-6.1.1
	r6g.8xlarge	emr-5.31.1, emr-6.1.1
	r6g.12xlarge	emr-5.31.1, emr-6.1.1
	r6g.16xlarge	emr-5.31.1, emr-6.1.1
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-5.30.2, emr-6.0.1
	d2.2xlarge	emr-5.30.2, emr-6.0.1
	d2.4xlarge	emr-5.30.2, emr-6.0.1
	d2.8xlarge	emr-5.30.2, emr-6.0.1
	i3.xlarge	emr-5.30.2, emr-6.0.1
	i3.2xlarge	emr-5.30.2, emr-6.0.1
	i3.4xlarge	emr-5.30.2, emr-6.0.1
	i3.8xlarge	emr-5.30.2, emr-6.0.1
	i3.16xlarge	emr-5.30.2, emr-6.0.1
	i3en.xlarge	emr-5.30.2, emr-6.0.1
	i3en.2xlarge	emr-5.30.2, emr-6.0.1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3en.3xlarge	emr-5.30.2, emr-6.0.1
	i3en.6xlarge	emr-5.30.2, emr-6.0.1
	i3en.12xlarge	emr-5.30.2, emr-6.0.1
	i3en.24xlarge	emr-5.30.2, emr-6.0.1
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia-Pacífico (Bombay): ap-south-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.6.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia-Pacífico (Hyderabad): ap-south-2

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.36.0, emr-6.7.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.12xlarge	emr-5.36.0, emr-6.7.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0
Computación optimizada	c5.xlarge	emr-5.36.0, emr-6.7.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.4xlarge	emr-5.36.0, emr-6.7.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0
	c6i.xlarge	emr-5.36.0, emr-6.7.0
	c6i.2xlarge	emr-5.36.0, emr-6.7.0
	c6i.4xlarge	emr-5.36.0, emr-6.7.0
	c6i.8xlarge	emr-5.36.0, emr-6.7.0
	c6i.12xlarge	emr-5.36.0, emr-6.7.0
	c6i.16xlarge	emr-5.36.0, emr-6.7.0
	c6i.24xlarge	emr-5.36.0, emr-6.7.0
	c6i.32xlarge	emr-5.36.0, emr-6.7.0
Optimizada para memoria	r5.xlarge	emr-5.36.0, emr-6.7.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.2xlarge	emr-5.36.0, emr-6.7.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	i3.xlarge	emr-5.36.0, emr-6.7.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3en.24xlarge	emr-5.36.0, emr-6.7.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia-Pacífico (Osaka): ap-northeast-3

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-5.10.0, emr-6.0.0
	d2.2xlarge	emr-5.10.0, emr-6.0.0
	d2.4xlarge	emr-5.10.0, emr-6.0.0
	d2.8xlarge	emr-5.10.0, emr-6.0.0
	i3.xlarge	emr-5.10.0, emr-6.0.0
	i3.2xlarge	emr-5.10.0, emr-6.0.0
	i3.4xlarge	emr-5.10.0, emr-6.0.0
	i3.8xlarge	emr-5.10.0, emr-6.0.0
	i3.16xlarge	emr-5.10.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia-Pacífico (Seúl): ap-northeast-2

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.1.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.1.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.1.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.1.0, emr-5.0.0, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia-Pacífico (Singapur): ap-southeast-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6in.24xlarge	emr-5.36.1, emr-6.10.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d3en.4xlarge	emr-5.33.0, emr-6.3.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Asia Pacífico (Sídney) (ap-southeast-2)

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Asia Pacífico (Tokio): ap-northeast-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d3en.6xlarge	emr-5.33.0, emr-6.3.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Canadá (centro): ca-central-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-4.8.2, emr-5.1.0, emr-6.0.0
	d2.2xlarge	emr-4.8.2, emr-5.1.0, emr-6.0.0
	d2.4xlarge	emr-4.8.2, emr-5.1.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d2.8xlarge	emr-4.8.2, emr-5.1.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4g.8xlarge	emr-5.36.1, emr-6.10.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

China (Ningxia): cn-northwest-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.9.1, emr-5.5.3, emr-6.0.0
	d2.2xlarge	emr-4.9.1, emr-5.5.3, emr-6.0.0
	d2.4xlarge	emr-4.9.1, emr-5.5.3, emr-6.0.0
	d2.8xlarge	emr-4.9.1, emr-5.5.3, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0

China (Pekín): cn-north-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0

Europa (Fráncfort): eu-central-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6a.24xlarge	emr-5.36.1, emr-6.8.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0
	d3en.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Europa (Zúrich): eu-central-2

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.36.0, emr-6.7.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.12xlarge	emr-5.36.0, emr-6.7.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.16xlarge	emr-5.36.0, emr-6.7.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0
Computación optimizada	c5.xlarge	emr-5.36.0, emr-6.7.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0
	c5.18xlarge	emr-5.36.0, emr-6.7.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.8xlarge	emr-5.36.0, emr-6.7.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Optimizada para memoria	r5.xlarge	emr-5.36.0, emr-6.7.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.4xlarge	emr-5.36.0, emr-6.7.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	i3.xlarge	emr-5.36.0, emr-6.7.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0

Europa (Irlanda): eu-west-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5dn.xlarge	emr-5.31.0, emr-6.1.0
	m5dn.2xlarge	emr-5.31.0, emr-6.1.0
	m5dn.4xlarge	emr-5.31.0, emr-6.1.0
	m5dn.8xlarge	emr-5.31.0, emr-6.1.0
	m5dn.12xlarge	emr-5.31.0, emr-6.1.0
	m5dn.16xlarge	emr-5.31.0, emr-6.1.0
	m5dn.24xlarge	emr-5.31.0, emr-6.1.0
	m5n.xlarge	emr-5.31.0, emr-6.1.0
	m5n.2xlarge	emr-5.31.0, emr-6.1.0
	m5n.4xlarge	emr-5.31.0, emr-6.1.0
	m5n.8xlarge	emr-5.31.0, emr-6.1.0
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0
	m5n.24xlarge	emr-5.31.0, emr-6.1.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
	m6id.xlarge	emr-5.36.1, emr-6.8.0
	m6id.2xlarge	emr-5.36.1, emr-6.8.0
	m6id.4xlarge	emr-5.36.1, emr-6.8.0
	m6id.8xlarge	emr-5.36.1, emr-6.8.0
	m6id.12xlarge	emr-5.36.1, emr-6.8.0
	m6id.16xlarge	emr-5.36.1, emr-6.8.0
	m6id.24xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0
	m6idn.2xlarge	emr-5.36.1, emr-6.10.0
	m6idn.4xlarge	emr-5.36.1, emr-6.10.0
	m6idn.8xlarge	emr-5.36.1, emr-6.10.0
	m6idn.12xlarge	emr-5.36.1, emr-6.10.0
	m6idn.16xlarge	emr-5.36.1, emr-6.10.0
	m6idn.24xlarge	emr-5.36.1, emr-6.10.0
	m6idn.32xlarge	emr-5.36.1, emr-6.10.0
	m6in.xlarge	emr-5.36.1, emr-6.10.0
	m6in.2xlarge	emr-5.36.1, emr-6.10.0
	m6in.4xlarge	emr-5.36.1, emr-6.10.0
	m6in.8xlarge	emr-5.36.1, emr-6.10.0
	m6in.12xlarge	emr-5.36.1, emr-6.10.0
	m6in.16xlarge	emr-5.36.1, emr-6.10.0
	m6in.24xlarge	emr-5.36.1, emr-6.10.0
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0
	m7g.2xlarge	emr-5.36.1, emr-6.10.0
	m7g.4xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
	m7gd.xlarge	emr-5.36.1, emr-6.10.0
	m7gd.2xlarge	emr-5.36.1, emr-6.10.0
	m7gd.4xlarge	emr-5.36.1, emr-6.10.0
	m7gd.8xlarge	emr-5.36.1, emr-6.10.0
	m7gd.12xlarge	emr-5.36.1, emr-6.10.0
	m7gd.16xlarge	emr-5.36.1, emr-6.10.0
	m7i.xlarge	emr-5.36.1, emr-6.10.0
	m7i.2xlarge	emr-5.36.1, emr-6.10.0
	m7i.4xlarge	emr-5.36.1, emr-6.10.0
	m7i.8xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.2xlarge	emr-5.36.1, emr-6.10.0
	m7i-flex.4xlarge	emr-5.36.1, emr-6.10.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6id.xlarge	emr-5.36.1, emr-6.8.0
	c6id.2xlarge	emr-5.36.1, emr-6.8.0
	c6id.4xlarge	emr-5.36.1, emr-6.8.0
	c6id.8xlarge	emr-5.36.1, emr-6.8.0
	c6id.12xlarge	emr-5.36.1, emr-6.8.0
	c6id.16xlarge	emr-5.36.1, emr-6.8.0
	c6id.24xlarge	emr-5.36.1, emr-6.8.0
	c6id.32xlarge	emr-5.36.1, emr-6.8.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0
	c7g.4xlarge	emr-5.36.1, emr-6.7.0
	c7g.8xlarge	emr-5.36.1, emr-6.7.0
	c7g.12xlarge	emr-5.36.1, emr-6.7.0
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
	c7gd.xlarge	emr-5.36.1, emr-6.10.0
	c7gd.2xlarge	emr-5.36.1, emr-6.10.0
	c7gd.4xlarge	emr-5.36.1, emr-6.10.0
	c7gd.8xlarge	emr-5.36.1, emr-6.10.0
	c7gd.12xlarge	emr-5.36.1, emr-6.10.0
	c7gd.16xlarge	emr-5.36.1, emr-6.10.0
	c7gn.xlarge	emr-5.36.1, emr-6.10.0
	c7gn.2xlarge	emr-5.36.1, emr-6.10.0
	c7gn.4xlarge	emr-5.36.1, emr-6.10.0
	c7gn.8xlarge	emr-5.36.1, emr-6.10.0
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p2.xlarge	emr-5.10.0, emr-6.0.0
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6a.xlarge	emr-5.36.1, emr-6.8.0
	r6a.2xlarge	emr-5.36.1, emr-6.8.0
	r6a.4xlarge	emr-5.36.1, emr-6.8.0
	r6a.8xlarge	emr-5.36.1, emr-6.8.0
	r6a.12xlarge	emr-5.36.1, emr-6.8.0
	r6a.16xlarge	emr-5.36.1, emr-6.8.0
	r6a.24xlarge	emr-5.36.1, emr-6.8.0
	r6a.32xlarge	emr-5.36.1, emr-6.8.0
	r6a.48xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	r6idn.xlarge	emr-5.36.1, emr-6.10.0
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0
	r6idn.8xlarge	emr-5.36.1, emr-6.10.0
	r6idn.12xlarge	emr-5.36.1, emr-6.10.0
	r6idn.16xlarge	emr-5.36.1, emr-6.10.0
	r6idn.24xlarge	emr-5.36.1, emr-6.10.0
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0
	r6in.xlarge	emr-5.36.1, emr-6.10.0
	r6in.2xlarge	emr-5.36.1, emr-6.10.0
	r6in.4xlarge	emr-5.36.1, emr-6.10.0
	r6in.8xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6in.12xlarge	emr-5.36.1, emr-6.10.0
	r6in.16xlarge	emr-5.36.1, emr-6.10.0
	r6in.24xlarge	emr-5.36.1, emr-6.10.0
	r6in.32xlarge	emr-5.36.1, emr-6.10.0
	r7g.xlarge	emr-5.36.1, emr-6.10.0
	r7g.2xlarge	emr-5.36.1, emr-6.10.0
	r7g.4xlarge	emr-5.36.1, emr-6.10.0
	r7g.8xlarge	emr-5.36.1, emr-6.10.0
	r7g.12xlarge	emr-5.36.1, emr-6.10.0
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0
	r7gd.2xlarge	emr-5.36.1, emr-6.10.0
	r7gd.4xlarge	emr-5.36.1, emr-6.10.0
	r7gd.8xlarge	emr-5.36.1, emr-6.10.0
	r7gd.12xlarge	emr-5.36.1, emr-6.10.0
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2gd.xlarge	emr-5.36.1, emr-6.10.0
	x2gd.2xlarge	emr-5.36.1, emr-6.10.0
	x2gd.4xlarge	emr-5.36.1, emr-6.10.0
	x2gd.8xlarge	emr-5.36.1, emr-6.10.0
	x2gd.12xlarge	emr-5.36.1, emr-6.10.0
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.2xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.4xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d2.8xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.xlarge	emr-5.33.0, emr-6.3.0
	d3en.2xlarge	emr-5.33.0, emr-6.3.0
	d3en.4xlarge	emr-5.33.0, emr-6.3.0
	d3en.6xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d3en.8xlarge	emr-5.33.0, emr-6.3.0
	d3en.12xlarge	emr-5.33.0, emr-6.3.0
	h1.2xlarge	emr-5.17.0, emr-6.0.0
	h1.4xlarge	emr-5.17.0, emr-6.0.0
	h1.8xlarge	emr-5.17.0, emr-6.0.0
	h1.16xlarge	emr-5.17.0, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4g.xlarge	emr-5.36.1, emr-6.10.0
	i4g.2xlarge	emr-5.36.1, emr-6.10.0
	i4g.4xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4g.8xlarge	emr-5.36.1, emr-6.10.0
	i4g.16xlarge	emr-5.36.1, emr-6.10.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Europa (Londres): eu-west-2

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g3.4xlarge	emr-5.18.0, emr-6.0.0
	g3.8xlarge	emr-5.18.0, emr-6.0.0
	g3.16xlarge	emr-5.18.0, emr-6.0.0
	g3s.xlarge	emr-5.19.0, emr-6.0.0
	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
	p3.2xlarge	emr-5.10.0, emr-6.0.0
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	r6id.xlarge	emr-5.36.1, emr-6.8.0
	r6id.2xlarge	emr-5.36.1, emr-6.8.0
	r6id.4xlarge	emr-5.36.1, emr-6.8.0
	r6id.8xlarge	emr-5.36.1, emr-6.8.0
	r6id.12xlarge	emr-5.36.1, emr-6.8.0
	r6id.16xlarge	emr-5.36.1, emr-6.8.0
	r6id.24xlarge	emr-5.36.1, emr-6.8.0
	r6id.32xlarge	emr-5.36.1, emr-6.8.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
	z1d.xlarge	emr-5.17.0, emr-6.0.0
	z1d.2xlarge	emr-5.17.0, emr-6.0.0
	z1d.3xlarge	emr-5.17.0, emr-6.0.0
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Optimización de almacenamiento	d2.xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0
	d2.2xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0
	d2.4xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	d2.8xlarge	emr-4.8.2, emr-5.0.3, emr-6.0.0
	d3.xlarge	emr-5.33.0, emr-6.3.0
	d3.2xlarge	emr-5.33.0, emr-6.3.0
	d3.4xlarge	emr-5.33.0, emr-6.3.0
	d3.8xlarge	emr-5.33.0, emr-6.3.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Europa (Milán): eu-south-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.29.0, emr-6.0.0
	m5.2xlarge	emr-5.29.0, emr-6.0.0
	m5.4xlarge	emr-5.29.0, emr-6.0.0
	m5.8xlarge	emr-5.29.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5.12xlarge	emr-5.29.0, emr-6.0.0
	m5.16xlarge	emr-5.29.0, emr-6.0.0
	m5.24xlarge	emr-5.29.0, emr-6.0.0
	m5a.xlarge	emr-5.29.0, emr-6.0.0
	m5a.2xlarge	emr-5.29.0, emr-6.0.0
	m5a.4xlarge	emr-5.29.0, emr-6.0.0
	m5a.8xlarge	emr-5.29.0, emr-6.0.0
	m5a.12xlarge	emr-5.29.0, emr-6.0.0
	m5a.16xlarge	emr-5.29.0, emr-6.0.0
	m5a.24xlarge	emr-5.29.0, emr-6.0.0
	m5d.xlarge	emr-5.29.0, emr-6.0.0
	m5d.2xlarge	emr-5.29.0, emr-6.0.0
	m5d.4xlarge	emr-5.29.0, emr-6.0.0
	m5d.8xlarge	emr-5.29.0, emr-6.0.0
	m5d.12xlarge	emr-5.29.0, emr-6.0.0
	m5d.16xlarge	emr-5.29.0, emr-6.0.0
	m5d.24xlarge	emr-5.29.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.29.0, emr-6.0.0
	c5.2xlarge	emr-5.29.0, emr-6.0.0
	c5.4xlarge	emr-5.29.0, emr-6.0.0
	c5.9xlarge	emr-5.29.0, emr-6.0.0
	c5.12xlarge	emr-5.29.0, emr-6.0.0
	c5.18xlarge	emr-5.29.0, emr-6.0.0
	c5.24xlarge	emr-5.29.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.29.0, emr-6.0.0
	c5d.2xlarge	emr-5.29.0, emr-6.0.0
	c5d.4xlarge	emr-5.29.0, emr-6.0.0
	c5d.9xlarge	emr-5.29.0, emr-6.0.0
	c5d.12xlarge	emr-5.29.0, emr-6.0.0
	c5d.18xlarge	emr-5.29.0, emr-6.0.0
	c5d.24xlarge	emr-5.29.0, emr-6.0.0
	c5n.xlarge	emr-5.29.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.2xlarge	emr-5.29.0, emr-6.0.0
	c5n.4xlarge	emr-5.29.0, emr-6.0.0
	c5n.9xlarge	emr-5.29.0, emr-6.0.0
	c5n.18xlarge	emr-5.29.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.29.0, emr-6.0.0
	r5.2xlarge	emr-5.29.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.4xlarge	emr-5.29.0, emr-6.0.0
	r5.8xlarge	emr-5.29.0, emr-6.0.0
	r5.12xlarge	emr-5.29.0, emr-6.0.0
	r5.16xlarge	emr-5.29.0, emr-6.0.0
	r5.24xlarge	emr-5.29.0, emr-6.0.0
	r5a.xlarge	emr-5.29.0, emr-6.0.0
	r5a.2xlarge	emr-5.29.0, emr-6.0.0
	r5a.4xlarge	emr-5.29.0, emr-6.0.0
	r5a.8xlarge	emr-5.29.0, emr-6.0.0
	r5a.12xlarge	emr-5.29.0, emr-6.0.0
	r5a.16xlarge	emr-5.29.0, emr-6.0.0
	r5a.24xlarge	emr-5.29.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.29.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.2xlarge	emr-5.29.0, emr-6.0.0
	r5d.4xlarge	emr-5.29.0, emr-6.0.0
	r5d.8xlarge	emr-5.29.0, emr-6.0.0
	r5d.12xlarge	emr-5.29.0, emr-6.0.0
	r5d.16xlarge	emr-5.29.0, emr-6.0.0
	r5d.24xlarge	emr-5.29.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-5.29.0, emr-6.0.0
	d2.2xlarge	emr-5.29.0, emr-6.0.0
	d2.4xlarge	emr-5.29.0, emr-6.0.0
	d2.8xlarge	emr-5.29.0, emr-6.0.0
	i3.xlarge	emr-5.29.0, emr-6.0.0
	i3.2xlarge	emr-5.29.0, emr-6.0.0
	i3.4xlarge	emr-5.29.0, emr-6.0.0
	i3.8xlarge	emr-5.29.0, emr-6.0.0
	i3.16xlarge	emr-5.29.0, emr-6.0.0
	i3en.xlarge	emr-5.29.0, emr-6.0.0
	i3en.2xlarge	emr-5.29.0, emr-6.0.0
	i3en.3xlarge	emr-5.29.0, emr-6.0.0
	i3en.6xlarge	emr-5.29.0, emr-6.0.0
	i3en.12xlarge	emr-5.29.0, emr-6.0.0
	i3en.24xlarge	emr-5.29.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Europa (España): eu-south-2

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.36.0, emr-6.7.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.12xlarge	emr-5.36.0, emr-6.7.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0
	m6g.16xlarge	emr-5.36.0, emr-6.7.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0
Computación optimizada	c5.xlarge	emr-5.36.0, emr-6.7.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.18xlarge	emr-5.36.0, emr-6.7.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0
Optimizada para memoria	r5.xlarge	emr-5.36.0, emr-6.7.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5.16xlarge	emr-5.36.0, emr-6.7.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0
	r5d.4xlarge	emr-5.36.0, emr-6.7.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	i3.xlarge	emr-5.36.0, emr-6.7.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.4xlarge	emr-5.36.0, emr-6.7.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0

Europa (París): eu-west-3

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Optimización de almacenamiento	d2.xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0
	d2.2xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0
	d2.4xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0
	d2.8xlarge	emr-4.9.2, emr-5.5.3, emr-6.0.0
	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0
	im4gn.xlarge	emr-5.35.0, emr-6.6.0
	im4gn.2xlarge	emr-5.35.0, emr-6.6.0
	im4gn.4xlarge	emr-5.35.0, emr-6.6.0
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0
	im4gn.16xlarge	emr-5.35.0, emr-6.6.0
	is4gen.xlarge	emr-5.35.0, emr-6.6.0
	is4gen.2xlarge	emr-5.35.0, emr-6.6.0
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Europa (Estocolmo): eu-north-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.17.0, emr-6.0.0
	m5.2xlarge	emr-5.17.0, emr-6.0.0
	m5.4xlarge	emr-5.17.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5.8xlarge	emr-5.17.0, emr-6.0.0
	m5.12xlarge	emr-5.17.0, emr-6.0.0
	m5.16xlarge	emr-5.17.0, emr-6.0.0
	m5.24xlarge	emr-5.17.0, emr-6.0.0
	m5d.xlarge	emr-5.17.0, emr-6.0.0
	m5d.2xlarge	emr-5.17.0, emr-6.0.0
	m5d.4xlarge	emr-5.17.0, emr-6.0.0
	m5d.8xlarge	emr-5.17.0, emr-6.0.0
	m5d.12xlarge	emr-5.17.0, emr-6.0.0
	m5d.16xlarge	emr-5.17.0, emr-6.0.0
	m5d.24xlarge	emr-5.17.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6gd.xlarge	emr-5.33.0, emr-6.3.0
	m6gd.2xlarge	emr-5.33.0, emr-6.3.0
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6gd.8xlarge	emr-5.33.0, emr-6.3.0
	m6gd.12xlarge	emr-5.33.0, emr-6.3.0
	m6gd.16xlarge	emr-5.33.0, emr-6.3.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.17.0, emr-6.0.0
	c5.2xlarge	emr-5.17.0, emr-6.0.0
	c5.4xlarge	emr-5.17.0, emr-6.0.0
	c5.9xlarge	emr-5.17.0, emr-6.0.0
	c5.12xlarge	emr-5.17.0, emr-6.0.0
	c5.18xlarge	emr-5.17.0, emr-6.0.0
	c5.24xlarge	emr-5.17.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5d.xlarge	emr-5.17.0, emr-6.0.0
	c5d.2xlarge	emr-5.17.0, emr-6.0.0
	c5d.4xlarge	emr-5.17.0, emr-6.0.0
	c5d.9xlarge	emr-5.17.0, emr-6.0.0
	c5d.12xlarge	emr-5.17.0, emr-6.0.0
	c5d.18xlarge	emr-5.17.0, emr-6.0.0
	c5d.24xlarge	emr-5.17.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
Optimizada para memoria	r5.xlarge	emr-5.17.0, emr-6.0.0
	r5.2xlarge	emr-5.17.0, emr-6.0.0
	r5.4xlarge	emr-5.17.0, emr-6.0.0
	r5.8xlarge	emr-5.17.0, emr-6.0.0
	r5.12xlarge	emr-5.17.0, emr-6.0.0
	r5.16xlarge	emr-5.17.0, emr-6.0.0
	r5.24xlarge	emr-5.17.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.17.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.2xlarge	emr-5.17.0, emr-6.0.0
	r5d.4xlarge	emr-5.17.0, emr-6.0.0
	r5d.8xlarge	emr-5.17.0, emr-6.0.0
	r5d.12xlarge	emr-5.17.0, emr-6.0.0
	r5d.16xlarge	emr-5.17.0, emr-6.0.0
	r5d.24xlarge	emr-5.17.0, emr-6.0.0
	r5dn.xlarge	emr-5.31.0, emr-6.1.0
	r5dn.2xlarge	emr-5.31.0, emr-6.1.0
	r5dn.4xlarge	emr-5.31.0, emr-6.1.0
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0
	r5dn.12xlarge	emr-5.31.0, emr-6.1.0
	r5dn.16xlarge	emr-5.31.0, emr-6.1.0
	r5dn.24xlarge	emr-5.31.0, emr-6.1.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	d2.xlarge	emr-5.17.0, emr-6.0.0
	d2.2xlarge	emr-5.17.0, emr-6.0.0
	d2.4xlarge	emr-5.17.0, emr-6.0.0
	d2.8xlarge	emr-5.17.0, emr-6.0.0
	i3.xlarge	emr-5.17.0, emr-6.0.0
	i3.2xlarge	emr-5.17.0, emr-6.0.0
	i3.4xlarge	emr-5.17.0, emr-6.0.0
	i3.8xlarge	emr-5.17.0, emr-6.0.0
	i3.16xlarge	emr-5.17.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Medio Oriente (Baréin): me-south-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.24.0, emr-6.0.0
	m5.2xlarge	emr-5.24.0, emr-6.0.0
	m5.4xlarge	emr-5.24.0, emr-6.0.0
	m5.8xlarge	emr-5.24.0, emr-6.0.0
	m5.12xlarge	emr-5.24.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5.16xlarge	emr-5.24.0, emr-6.0.0
	m5.24xlarge	emr-5.24.0, emr-6.0.0
	m5d.xlarge	emr-5.24.0, emr-6.0.0
	m5d.2xlarge	emr-5.24.0, emr-6.0.0
	m5d.4xlarge	emr-5.24.0, emr-6.0.0
	m5d.8xlarge	emr-5.24.0, emr-6.0.0
	m5d.12xlarge	emr-5.24.0, emr-6.0.0
	m5d.16xlarge	emr-5.24.0, emr-6.0.0
	m5d.24xlarge	emr-5.24.0, emr-6.0.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.24.0, emr-6.0.0
	c5.2xlarge	emr-5.24.0, emr-6.0.0
	c5.4xlarge	emr-5.24.0, emr-6.0.0
	c5.9xlarge	emr-5.24.0, emr-6.0.0
	c5.12xlarge	emr-5.24.0, emr-6.0.0
	c5.18xlarge	emr-5.24.0, emr-6.0.0
	c5.24xlarge	emr-5.24.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.24.0, emr-6.0.0
	c5d.2xlarge	emr-5.24.0, emr-6.0.0
	c5d.4xlarge	emr-5.24.0, emr-6.0.0
	c5d.9xlarge	emr-5.24.0, emr-6.0.0
	c5d.18xlarge	emr-5.24.0, emr-6.0.0
	c5n.xlarge	emr-5.24.0, emr-6.0.0
	c5n.2xlarge	emr-5.24.0, emr-6.0.0
	c5n.4xlarge	emr-5.24.0, emr-6.0.0
	c5n.9xlarge	emr-5.24.0, emr-6.0.0
	c5n.18xlarge	emr-5.24.0, emr-6.0.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Optimizada para memoria	r5.xlarge	emr-5.24.0, emr-6.0.0
	r5.2xlarge	emr-5.24.0, emr-6.0.0
	r5.4xlarge	emr-5.24.0, emr-6.0.0
	r5.8xlarge	emr-5.24.0, emr-6.0.0
	r5.12xlarge	emr-5.24.0, emr-6.0.0
	r5.16xlarge	emr-5.24.0, emr-6.0.0
	r5.24xlarge	emr-5.24.0, emr-6.0.0
	r5d.xlarge	emr-5.24.0, emr-6.0.0
	r5d.2xlarge	emr-5.24.0, emr-6.0.0
	r5d.4xlarge	emr-5.24.0, emr-6.0.0
	r5d.8xlarge	emr-5.24.0, emr-6.0.0
	r5d.12xlarge	emr-5.24.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.16xlarge	emr-5.24.0, emr-6.0.0
	r5d.24xlarge	emr-5.24.0, emr-6.0.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
Optimización de almacenamiento	d2.xlarge	emr-5.24.0, emr-6.0.0
	d2.2xlarge	emr-5.24.0, emr-6.0.0
	d2.4xlarge	emr-5.24.0, emr-6.0.0
	d2.8xlarge	emr-5.24.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i3.xlarge	emr-5.24.0, emr-6.0.0
	i3.2xlarge	emr-5.24.0, emr-6.0.0
	i3.4xlarge	emr-5.24.0, emr-6.0.0
	i3.8xlarge	emr-5.24.0, emr-6.0.0
	i3.16xlarge	emr-5.24.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Medio Oriente (EAU): me-central-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.36.0, emr-6.7.0
	m5.2xlarge	emr-5.36.0, emr-6.7.0
	m5.4xlarge	emr-5.36.0, emr-6.7.0
	m5.8xlarge	emr-5.36.0, emr-6.7.0
	m5.12xlarge	emr-5.36.0, emr-6.7.0
	m5.16xlarge	emr-5.36.0, emr-6.7.0
	m5.24xlarge	emr-5.36.0, emr-6.7.0
	m5d.xlarge	emr-5.36.0, emr-6.7.0
	m5d.2xlarge	emr-5.36.0, emr-6.7.0
	m5d.4xlarge	emr-5.36.0, emr-6.7.0
	m5d.8xlarge	emr-5.36.0, emr-6.7.0
	m5d.12xlarge	emr-5.36.0, emr-6.7.0
	m5d.16xlarge	emr-5.36.0, emr-6.7.0
	m5d.24xlarge	emr-5.36.0, emr-6.7.0
	m6g.xlarge	emr-5.36.0, emr-6.7.0
	m6g.2xlarge	emr-5.36.0, emr-6.7.0
	m6g.4xlarge	emr-5.36.0, emr-6.7.0
	m6g.8xlarge	emr-5.36.0, emr-6.7.0
	m6g.12xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6g.16xlarge	emr-5.36.0, emr-6.7.0
	m6gd.xlarge	emr-5.36.0, emr-6.7.0
	m6gd.2xlarge	emr-5.36.0, emr-6.7.0
	m6gd.4xlarge	emr-5.36.0, emr-6.7.0
	m6gd.8xlarge	emr-5.36.0, emr-6.7.0
	m6gd.12xlarge	emr-5.36.0, emr-6.7.0
	m6gd.16xlarge	emr-5.36.0, emr-6.7.0
	m6i.xlarge	emr-5.36.0, emr-6.7.0
	m6i.2xlarge	emr-5.36.0, emr-6.7.0
	m6i.4xlarge	emr-5.36.0, emr-6.7.0
	m6i.8xlarge	emr-5.36.0, emr-6.7.0
	m6i.12xlarge	emr-5.36.0, emr-6.7.0
	m6i.16xlarge	emr-5.36.0, emr-6.7.0
	m6i.24xlarge	emr-5.36.0, emr-6.7.0
	m6i.32xlarge	emr-5.36.0, emr-6.7.0
Computación optimizada	c5.xlarge	emr-5.36.0, emr-6.7.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0
	c5.4xlarge	emr-5.36.0, emr-6.7.0
	c5.9xlarge	emr-5.36.0, emr-6.7.0
	c5.12xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.18xlarge	emr-5.36.0, emr-6.7.0
	c5.24xlarge	emr-5.36.0, emr-6.7.0
	c5d.xlarge	emr-5.36.0, emr-6.7.0
	c5d.2xlarge	emr-5.36.0, emr-6.7.0
	c5d.4xlarge	emr-5.36.0, emr-6.7.0
	c5d.9xlarge	emr-5.36.0, emr-6.7.0
	c5d.12xlarge	emr-5.36.0, emr-6.7.0
	c5d.18xlarge	emr-5.36.0, emr-6.7.0
	c5d.24xlarge	emr-5.36.0, emr-6.7.0
	c6g.xlarge	emr-5.36.0, emr-6.7.0
	c6g.2xlarge	emr-5.36.0, emr-6.7.0
	c6g.4xlarge	emr-5.36.0, emr-6.7.0
	c6g.8xlarge	emr-5.36.0, emr-6.7.0
	c6g.12xlarge	emr-5.36.0, emr-6.7.0
	c6g.16xlarge	emr-5.36.0, emr-6.7.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
Optimizada para memoria	r5.xlarge	emr-5.36.0, emr-6.7.0
	r5.2xlarge	emr-5.36.0, emr-6.7.0
	r5.4xlarge	emr-5.36.0, emr-6.7.0
	r5.8xlarge	emr-5.36.0, emr-6.7.0
	r5.12xlarge	emr-5.36.0, emr-6.7.0
	r5.16xlarge	emr-5.36.0, emr-6.7.0
	r5.24xlarge	emr-5.36.0, emr-6.7.0
	r5d.xlarge	emr-5.36.0, emr-6.7.0
	r5d.2xlarge	emr-5.36.0, emr-6.7.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.4xlarge	emr-5.36.0, emr-6.7.0
	r5d.8xlarge	emr-5.36.0, emr-6.7.0
	r5d.12xlarge	emr-5.36.0, emr-6.7.0
	r5d.16xlarge	emr-5.36.0, emr-6.7.0
	r5d.24xlarge	emr-5.36.0, emr-6.7.0
	r6g.xlarge	emr-5.36.0, emr-6.7.0
	r6g.2xlarge	emr-5.36.0, emr-6.7.0
	r6g.4xlarge	emr-5.36.0, emr-6.7.0
	r6g.8xlarge	emr-5.36.0, emr-6.7.0
	r6g.12xlarge	emr-5.36.0, emr-6.7.0
	r6g.16xlarge	emr-5.36.0, emr-6.7.0
	r6i.xlarge	emr-5.36.0, emr-6.7.0
	r6i.2xlarge	emr-5.36.0, emr-6.7.0
	r6i.4xlarge	emr-5.36.0, emr-6.7.0
	r6i.8xlarge	emr-5.36.0, emr-6.7.0
	r6i.12xlarge	emr-5.36.0, emr-6.7.0
	r6i.16xlarge	emr-5.36.0, emr-6.7.0
	r6i.24xlarge	emr-5.36.0, emr-6.7.0
	r6i.32xlarge	emr-5.36.0, emr-6.7.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	i3.xlarge	emr-5.36.0, emr-6.7.0
	i3.2xlarge	emr-5.36.0, emr-6.7.0
	i3.4xlarge	emr-5.36.0, emr-6.7.0
	i3.8xlarge	emr-5.36.0, emr-6.7.0
	i3.16xlarge	emr-5.36.0, emr-6.7.0
	i3en.xlarge	emr-5.36.0, emr-6.7.0
	i3en.2xlarge	emr-5.36.0, emr-6.7.0
	i3en.3xlarge	emr-5.36.0, emr-6.7.0
	i3en.6xlarge	emr-5.36.0, emr-6.7.0
	i3en.12xlarge	emr-5.36.0, emr-6.7.0
	i3en.24xlarge	emr-5.36.0, emr-6.7.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

América del Sur (São Paulo): sa-east-1

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
Uso general	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0
	m5.4xlarge	emr-5.13.0, emr-6.0.0
	m5.8xlarge	emr-5.13.0, emr-6.0.0
	m5.12xlarge	emr-5.13.0, emr-6.0.0
	m5.16xlarge	emr-5.13.0, emr-6.0.0
	m5.24xlarge	emr-5.13.0, emr-6.0.0
	m5a.xlarge	emr-5.20.0, emr-6.0.0
	m5a.2xlarge	emr-5.20.0, emr-6.0.0
	m5a.4xlarge	emr-5.20.0, emr-6.0.0
	m5a.8xlarge	emr-5.20.0, emr-6.0.0
	m5a.12xlarge	emr-5.20.0, emr-6.0.0
	m5a.16xlarge	emr-5.20.0, emr-6.0.0
	m5a.24xlarge	emr-5.20.0, emr-6.0.0
	m5ad.xlarge	emr-5.20.0, emr-6.0.0
	m5ad.2xlarge	emr-5.20.0, emr-6.0.0
	m5ad.4xlarge	emr-5.20.0, emr-6.0.0
	m5ad.8xlarge	emr-5.33.0, emr-6.3.0
	m5ad.12xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0
	m5d.xlarge	emr-5.13.0, emr-6.0.0
	m5d.2xlarge	emr-5.13.0, emr-6.0.0
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0
	m5d.12xlarge	emr-5.13.0, emr-6.0.0
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0
	m5zn.xlarge	emr-5.33.0, emr-6.3.0
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0
	m5zn.6xlarge	emr-5.33.0, emr-6.3.0
	m5zn.12xlarge	emr-5.33.0, emr-6.3.0
	m6a.xlarge	emr-5.36.1, emr-6.8.0
	m6a.2xlarge	emr-5.36.1, emr-6.8.0
	m6a.4xlarge	emr-5.36.1, emr-6.8.0
	m6a.8xlarge	emr-5.36.1, emr-6.8.0
	m6a.12xlarge	emr-5.36.1, emr-6.8.0
	m6a.16xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	m6a.24xlarge	emr-5.36.1, emr-6.8.0
	m6a.32xlarge	emr-5.36.1, emr-6.8.0
	m6a.48xlarge	emr-5.36.1, emr-6.8.0
	m6g.xlarge	emr-5.30.0, emr-6.1.0
	m6g.2xlarge	emr-5.30.0, emr-6.1.0
	m6g.4xlarge	emr-5.30.0, emr-6.1.0
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0
	m6g.16xlarge	emr-5.30.0, emr-6.1.0
	m6i.xlarge	emr-5.35.0, emr-6.6.0
	m6i.2xlarge	emr-5.35.0, emr-6.6.0
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0
	m6i.12xlarge	emr-5.35.0, emr-6.6.0
	m6i.16xlarge	emr-5.35.0, emr-6.6.0
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Computación optimizada	c5.xlarge	emr-5.13.0, emr-6.0.0
	c5.2xlarge	emr-5.13.0, emr-6.0.0
	c5.4xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5.9xlarge	emr-5.13.0, emr-6.0.0
	c5.12xlarge	emr-5.13.0, emr-6.0.0
	c5.18xlarge	emr-5.13.0, emr-6.0.0
	c5.24xlarge	emr-5.13.0, emr-6.0.0
	c5a.xlarge	emr-5.31.0, emr-6.1.0
	c5a.2xlarge	emr-5.31.0, emr-6.1.0
	c5a.4xlarge	emr-5.31.0, emr-6.1.0
	c5a.8xlarge	emr-5.31.0, emr-6.1.0
	c5a.12xlarge	emr-5.31.0, emr-6.1.0
	c5a.16xlarge	emr-5.31.0, emr-6.1.0
	c5a.24xlarge	emr-5.31.0, emr-6.1.0
	c5ad.xlarge	emr-5.33.0, emr-6.3.0
	c5ad.2xlarge	emr-5.33.0, emr-6.3.0
	c5ad.4xlarge	emr-5.33.0, emr-6.3.0
	c5ad.8xlarge	emr-5.33.0, emr-6.3.0
	c5ad.12xlarge	emr-5.33.0, emr-6.3.0
	c5ad.16xlarge	emr-5.33.0, emr-6.3.0
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0
	c5d.xlarge	emr-5.13.0, emr-6.0.0
	c5d.2xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c5d.4xlarge	emr-5.13.0, emr-6.0.0
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0
	c5d.24xlarge	emr-5.13.0, emr-6.0.0
	c5n.xlarge	emr-5.20.0, emr-6.0.0
	c5n.2xlarge	emr-5.20.0, emr-6.0.0
	c5n.4xlarge	emr-5.20.0, emr-6.0.0
	c5n.9xlarge	emr-5.20.0, emr-6.0.0
	c5n.18xlarge	emr-5.20.0, emr-6.0.0
	c6a.xlarge	emr-5.36.1, emr-6.8.0
	c6a.2xlarge	emr-5.36.1, emr-6.8.0
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0
	c6a.12xlarge	emr-5.36.1, emr-6.8.0
	c6a.16xlarge	emr-5.36.1, emr-6.8.0
	c6a.24xlarge	emr-5.36.1, emr-6.8.0
	c6a.32xlarge	emr-5.36.1, emr-6.8.0
	c6a.48xlarge	emr-5.36.1, emr-6.8.0
	c6g.xlarge	emr-5.31.0, emr-6.1.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6g.2xlarge	emr-5.31.0, emr-6.1.0
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0
	c6g.12xlarge	emr-5.31.0, emr-6.1.0
	c6g.16xlarge	emr-5.31.0, emr-6.1.0
	c6gd.xlarge	emr-5.33.0, emr-6.3.0
	c6gd.2xlarge	emr-5.33.0, emr-6.3.0
	c6gd.4xlarge	emr-5.33.0, emr-6.3.0
	c6gd.8xlarge	emr-5.33.0, emr-6.3.0
	c6gd.12xlarge	emr-5.33.0, emr-6.3.0
	c6gd.16xlarge	emr-5.33.0, emr-6.3.0
	c6gn.xlarge	emr-5.33.0, emr-6.3.0
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0
	c6gn.8xlarge	emr-5.33.0, emr-6.3.0
	c6gn.12xlarge	emr-5.33.0, emr-6.3.0
	c6gn.16xlarge	emr-5.33.0, emr-6.3.0
	c6i.xlarge	emr-5.35.0, emr-6.6.0
	c6i.2xlarge	emr-5.35.0, emr-6.6.0
	c6i.4xlarge	emr-5.35.0, emr-6.6.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	c6i.8xlarge	emr-5.35.0, emr-6.6.0
	c6i.12xlarge	emr-5.35.0, emr-6.6.0
	c6i.16xlarge	emr-5.35.0, emr-6.6.0
	c6i.24xlarge	emr-5.35.0, emr-6.6.0
	c6i.32xlarge	emr-5.35.0, emr-6.6.0
	c6in.xlarge	emr-5.36.1, emr-6.10.0
	c6in.2xlarge	emr-5.36.1, emr-6.10.0
	c6in.4xlarge	emr-5.36.1, emr-6.10.0
	c6in.8xlarge	emr-5.36.1, emr-6.10.0
	c6in.12xlarge	emr-5.36.1, emr-6.10.0
	c6in.16xlarge	emr-5.36.1, emr-6.10.0
	c6in.24xlarge	emr-5.36.1, emr-6.10.0
	c6in.32xlarge	emr-5.36.1, emr-6.10.0
Computación acelerada	g4dn.xlarge	emr-5.30.0, emr-6.0.0
	g4dn.2xlarge	emr-5.30.0, emr-6.0.0
	g4dn.4xlarge	emr-5.30.0, emr-6.0.0
	g4dn.8xlarge	emr-5.30.0, emr-6.0.0
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
	g5.xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	g5.2xlarge	emr-5.36.1, emr-6.9.0
	g5.4xlarge	emr-5.36.1, emr-6.9.0
	g5.8xlarge	emr-5.36.1, emr-6.9.0
	g5.12xlarge	emr-5.36.1, emr-6.9.0
	g5.16xlarge	emr-5.36.1, emr-6.9.0
	g5.24xlarge	emr-5.36.1, emr-6.9.0
	g5.48xlarge	emr-5.36.1, emr-6.9.0
Optimizada para memoria	r5.xlarge	emr-5.13.0, emr-6.0.0
	r5.2xlarge	emr-5.13.0, emr-6.0.0
	r5.4xlarge	emr-5.13.0, emr-6.0.0
	r5.8xlarge	emr-5.13.0, emr-6.0.0
	r5.12xlarge	emr-5.13.0, emr-6.0.0
	r5.16xlarge	emr-5.13.0, emr-6.0.0
	r5.24xlarge	emr-5.13.0, emr-6.0.0
	r5a.xlarge	emr-5.20.0, emr-6.0.0
	r5a.2xlarge	emr-5.20.0, emr-6.0.0
	r5a.4xlarge	emr-5.20.0, emr-6.0.0
	r5a.8xlarge	emr-5.20.0, emr-6.0.0
	r5a.12xlarge	emr-5.20.0, emr-6.0.0
	r5a.16xlarge	emr-5.20.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5a.24xlarge	emr-5.20.0, emr-6.0.0
	r5ad.xlarge	emr-5.20.0, emr-6.0.0
	r5ad.2xlarge	emr-5.20.0, emr-6.0.0
	r5ad.4xlarge	emr-5.20.0, emr-6.0.0
	r5ad.8xlarge	emr-5.33.0, emr-6.3.0
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0
	r5ad.16xlarge	emr-5.33.0, emr-6.3.0
	r5ad.24xlarge	emr-5.20.0, emr-6.0.0
	r5b.xlarge	emr-5.33.0, emr-6.3.0
	r5b.2xlarge	emr-5.33.0, emr-6.3.0
	r5b.4xlarge	emr-5.33.0, emr-6.3.0
	r5b.8xlarge	emr-5.33.0, emr-6.3.0
	r5b.12xlarge	emr-5.33.0, emr-6.3.0
	r5b.16xlarge	emr-5.33.0, emr-6.3.0
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0
	r5d.2xlarge	emr-5.13.0, emr-6.0.0
	r5d.4xlarge	emr-5.13.0, emr-6.0.0
	r5d.8xlarge	emr-5.13.0, emr-6.0.0
	r5d.12xlarge	emr-5.13.0, emr-6.0.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r5d.16xlarge	emr-5.13.0, emr-6.0.0
	r5d.24xlarge	emr-5.13.0, emr-6.0.0
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0
	r5n.8xlarge	emr-5.31.0, emr-6.1.0
	r5n.12xlarge	emr-5.31.0, emr-6.1.0
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0
	r6g.xlarge	emr-5.31.0, emr-6.1.0
	r6g.2xlarge	emr-5.31.0, emr-6.1.0
	r6g.4xlarge	emr-5.31.0, emr-6.1.0
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0
	r6g.16xlarge	emr-5.31.0, emr-6.1.0
	r6gd.xlarge	emr-5.33.0, emr-6.3.0
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0
	r6gd.4xlarge	emr-5.33.0, emr-6.3.0
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	r6gd.16xlarge	emr-5.33.0, emr-6.3.0
	r6i.xlarge	emr-5.35.0, emr-6.6.0
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0
	r6i.8xlarge	emr-5.35.0, emr-6.6.0
	r6i.12xlarge	emr-5.35.0, emr-6.6.0
	r6i.16xlarge	emr-5.35.0, emr-6.6.0
	r6i.24xlarge	emr-5.35.0, emr-6.6.0
	r6i.32xlarge	emr-5.35.0, emr-6.6.0
	x1.16xlarge	emr-5.36.1, emr-6.9.0
	x1.32xlarge	emr-5.36.1, emr-6.9.0
	x1e.xlarge	emr-5.36.1, emr-6.10.0
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0
	x1e.8xlarge	emr-5.36.1, emr-6.10.0
	x1e.16xlarge	emr-5.36.1, emr-6.10.0
	x1e.32xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	x2iedn.xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.2xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.8xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.16xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Optimización de almacenamiento	i3.xlarge	emr-5.9.0, emr-6.0.0
	i3.2xlarge	emr-5.9.0, emr-6.0.0
	i3.4xlarge	emr-5.9.0, emr-6.0.0
	i3.8xlarge	emr-5.9.0, emr-6.0.0
	i3.16xlarge	emr-5.9.0, emr-6.0.0
	i3en.xlarge	emr-5.25.0, emr-6.0.0
	i3en.2xlarge	emr-5.25.0, emr-6.0.0
	i3en.3xlarge	emr-5.25.0, emr-6.0.0
	i3en.6xlarge	emr-5.25.0, emr-6.0.0
	i3en.12xlarge	emr-5.25.0, emr-6.0.0
	i3en.24xlarge	emr-5.25.0, emr-6.0.0
	i4i.xlarge	emr-5.36.1, emr-6.8.0
	i4i.2xlarge	emr-5.36.1, emr-6.8.0

Clase de instancia	Tipo de instancia	Versión mínima admitida de Amazon EMR
	i4i.4xlarge	emr-5.36.1, emr-6.8.0
	i4i.8xlarge	emr-5.36.1, emr-6.8.0
	i4i.16xlarge	emr-5.36.1, emr-6.8.0
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Instancias de generaciones anteriores

Amazon EMR admite las Instancias de generaciones anteriores para admitir las aplicaciones que están optimizadas para estas instancias y que todavía no se han actualizado. Para obtener más información acerca de estos tipos de instancias y sobre las rutas de actualización, consulte Instancias de generaciones anteriores.

Clase de instancia	Tipos de instancias
General Purpose	m1.small[1] | m1.medium[1] | m1.large[1] | m1.xlarge[1] | m3.xlarge[1] | m3.2xlarge[1] | m4.large | m4.xlarge | m4.2xlarge | m4.4xlarge | m4.10xlarge | m4.16xlarge
Compute Optimized	c1.medium[1] [2] | c1.xlarge[1] | c3.xlarge[1] | c3.2xlarge[1] | c3.4xlarge[1] | c3.8xlarge[1] | c4.large | c4.xlarge | c4.2xlarge | c4.4xlarge | c4.8xlarge
GPU Optimized	g2.2xlarge
Memory Optimized	m2.xlarge[1] | m2.2xlarge[1] | m2.4xlarge[1] | r3.xlarge | r3.2xlarge | r3.4xlarge | r3.8xlarge | r4.xlarge | r4.2xlarge | r4.4xlarge | r4.8xlarge | r4.16xlarge
Storage Optimized	i2.xlarge | i2.2xlarge | i2.4xlarge | i2.8xlarge

[1] Utiliza la AMI de virtualización PVM con las versiones de Amazon EMR anteriores a la 5.13.0. Para obtener más información, consulte Tipos de virtualización de AMI de Linux.

² No se admite en la versión 5.15.0.

Opciones de compra de instancias

A la hora de configurar un clúster, puede elegir una opción de compra para instancias de Amazon EC2. Puede elegir utilizar instancias bajo demanda, instancias de spot o ambas. Los precios varían en función del tipo de instancia y de la región. El precio de Amazon EMR se suma al precio de Amazon EC2 (el precio de los servidores subyacentes) y al precio de Amazon EBS (si se adjuntan volúmenes de Amazon EBS). Para conocer los precios actuales, consulte Precios de Amazon EMR.

La opción elegida de utilizar grupos de instancias o flotas de instancias en el clúster determina cómo puede cambiar las opciones de compra de instancias mientras se ejecuta un clúster. Si decide utilizar grupos de instancias uniformes, solo puede especificar la opción de compra de un grupo de instancias cuando lo crea y el tipo de instancia y la opción de compra se aplican a todas las instancias de Amazon EC2 en cada grupo de instancias. Si decide utilizar flotas de instancias, puede cambiar las opciones de compra después de haber creado cada flota de instancias y puede combinar opciones de compra para satisfacer el objetivo de capacidad que especifique. Para obtener más información acerca de estas configuraciones, consulte Crear un clúster con flotas de instancias o grupos de instancias uniformes.

On-Demand Instances

Con las instancias bajo demanda, paga la capacidad de cómputo por segundo. De forma opcional, puede hacer que estas instancias bajo demanda utilicen las opciones de compra de instancia reservada o instancia dedicada. Con las instancias reservadas, realiza un pago único para una instancia a fin de reservar capacidad. Las instancias dedicadas están físicamente aisladas en el nivel de hardware de host de las instancias que pertenecen a otras cuentas de AWS. Para más información sobre las opciones de compra, consulte Opciones de compra de instancias en la Guía del usuario de Amazon EC2 para instancias de Linux.

Uso de instancias reservadas

Para utilizar instancias reservadas en Amazon EMR, debe utilizar Amazon EC2 para adquirir la instancia reservada y especificar los parámetros de la reserva, incluido el ámbito de la reserva en lo que respecta a una región o una zona de disponibilidad. Para más información, consulte Instancias reservadas de Amazon EC2 y Compra de instancias reservadas en la Guía del usuario de Amazon EC2 para instancias de Linux. Cuando haya adquirido una instancia reservada, y si se dan todas las condiciones siguientes, Amazon EMR la utilizará cuando se lance un clúster:

- Una instancia bajo demanda se especifica en la configuración del clúster que coincide con la especificación de instancia reservada.
- El clúster se ha lanzado dentro del ámbito de la reserva de la instancia (la zona de disponibilidad o región).
- La capacidad de instancia reservada sigue estando disponible

Por ejemplo, supongamos que adquiere una instancia reservada `m5.xlarge` con la reserva de instancia en el ámbito de la región Este de EE. UU. A continuación lanza un clúster de Amazon EMR en el Este de EE. UU. que utiliza dos instancias `m5.xlarge`. La primera instancia se factura con la tarifa de instancia reservada y la otra se factura con la tarifa de la opción bajo demanda. La capacidad de instancias reservadas se utiliza antes de crear las instancias bajo demanda.

Uso de instancias dedicadas

Para utilizar instancias dedicadas, adquiera las instancias dedicadas utilizando Amazon EC2 y, a continuación, cree una VPC con el atributo de tenencia Dedicada. Dentro de Amazon EMR, puede especificar que un clúster se debe lanzar en esta VPC. Cualquier instancia bajo demanda del clúster que se ajuste a la especificación de instancia dedicada usará las instancias dedicadas disponibles cuando se lanza el clúster.

> **Note**
>
> Amazon EMR no admite la configuración del atributo `dedicated` en instancias individuales.

Spot Instances

Las instancias de spot en Amazon EMR proporcionan una opción para adquirir capacidad de instancias de Amazon EC2 a un costo reducido en comparación con la compra bajo demanda. La desventaja de utilizar instancias de spot es que las instancias pueden terminar si la capacidad de spot deja de estar disponible para el tipo de instancia que está ejecutando. Para obtener más información sobre si el uso de instancias de spot podría ser adecuado para su aplicación, consulte ¿Cuándo se deben utilizar las instancias de spot?.

Cuando Amazon EC2 tiene capacidad no utilizada, ofrece instancias de EC2 a un precio reducido, denominado precio de spot. El precio fluctúa en función de la disponibilidad y la demanda y se establece por región y zona de disponibilidad. Al elegir las instancias de spot, debe especificar el

precio de spot máximo que está dispuesto a pagar para cada tipo de instancia EC2. Cuando el precio de spot de la zona de disponibilidad del clúster se encuentra por debajo del precio de spot máximo especificado para dicho tipo de instancia, se lanzan las instancias. Mientras se ejecutan las instancias, se le factura el precio de spot actual, no el precio de spot máximo.

> **Note**
>
> Las instancias de spot con una duración definida (también conocidas como bloques de spot) dejarán de estar disponibles para los nuevos clientes a partir del 1 de julio de 2021. En el caso de los clientes que hayan utilizado previamente la característica, se continuará ofreciendo soporte a las instancias de spot con una duración definida hasta el 31 de diciembre de 2022.

Para conocer los precios actuales, consulte Precios de instancias de spot de Amazon EC2. Para obtener más información, consulte Instancias de spot en la Guía del usuario de Amazon EC2 para instancias de Linux. Al crear y configurar un clúster, especifica las opciones de red que en última instancia determinan la zona de disponibilidad donde se lanza el clúster. Para obtener más información, consulte Configurar redes.

> **Tip**
>
> Puede ver el precio de spot en tiempo real en la consola al pasar el ratón sobre la ayuda contextual situada junto a la opción de compra Spot al crear un clúster utilizando las Advanced Options (Opciones avanzadas). Se muestran los precios de cada zona de disponibilidad en la región seleccionada. Los precios más bajos se muestran en filas de color verde. Debido a las fluctuaciones de los precios de spot entre las zonas de disponibilidad, la selección de la zona de disponibilidad con el menor precio inicial podría no coincidir con el precio más bajo durante la vida útil del clúster. Para obtener resultados óptimos, estudie el historial de precios de la zona de disponibilidad antes elegir. Para más información, consulte Historial de precios para las instancias de spot en la Guía del usuario de Amazon EC2 para instancias de Linux.

Las opciones de instancias de spot dependen de si utiliza flotas de instancias o grupos de instancias uniformes en la configuración del clúster.

Instancias de subasta en grupos de instancias uniformes

Al utilizar las instancias de spot en un grupo de instancias uniforme, todas las instancias de un grupo de instancias deben ser instancias de spot. Especifique una única subred o zona de disponibilidad para el clúster. Para cada grupo de instancias, especifique una sola instancia de spot y un precio de spot máximo. Las instancias de spot de ese tipo se lanzan si el precio de spot de la zona de disponibilidad y la región del clúster está por debajo del precio de spot máximo. Las instancias se terminan si el precio de spot está por encima de su precio de spot máximo. Solo puede establecer el precio de spot al configurar un grupo de instancias. No se puede cambiar más adelante. Para obtener más información, consulte Crear un clúster con flotas de instancias o grupos de instancias uniformes.

Instancias de subasta en flotas de instancias

Al utilizar la configuración de flotas de instancias, las opciones adicionales le ofrecen mayor control sobre cómo lanzar y terminar instancias de spot. Fundamentalmente, las flotas de instancias utilizan un método de diferente a los grupos de instancias uniformes para lanzar instancias. La forma de funcionamiento consiste en establecer una capacidad de destino para las instancias de spot (e instancias bajo demanda) y hasta cinco tipos de instancias. También puede especificar una capacidad ponderada para cada tipo de instancia o utilizar el vCPU (vcores de YARN) del tipo de instancia como capacidad ponderada. Esta capacidad ponderada se tiene en cuenta en la capacidad de destino cuando se provisiona una instancia de ese tipo. Amazon EMR aprovisiona instancias con ambas opciones de compra hasta que se alcanza la capacidad de destino para cada destino. Además, puede definir una serie de zonas de disponibilidad de las cuales Amazon EMR pueda elegir al lanzar instancias. También proporciona opciones de spot adicionales para cada flota, incluido un tiempo de espera de aprovisionamiento. Para obtener más información, consulte Configurar flotas de instancias.

Almacenamiento de la instancia

El almacén de instancias o el almacenamiento de volúmenes de Amazon EBS se utiliza para los datos de HDFS, así como para los búferes, cachés, datos de pruebas y otro contenido temporal que algunas aplicaciones pueden "verter" en el sistema de archivos local.

Amazon EBS funciona de forma distinta dentro de Amazon EMR que con las instancias de Amazon EC2 normales. Los volúmenes de Amazon EBS asociados a clústeres de Amazon EMR son efímeros: los volúmenes se eliminan al terminar el clúster y las instancias (por ejemplo, al reducir grupos de instancias), por lo que es importante que no prevea la persistencia de datos. Aunque los datos son efímeros, es posible que los datos en HDFS se puedan replicar en función del número y de la especialización de los nodos del clúster. Al agregar volúmenes de almacenamiento de Amazon EBS, estos se montan como volúmenes adicionales. No forman parte del volumen raíz. YARN está

configurado para utilizar todos los volúmenes adicionales, pero usted es responsable de asignar los volúmenes adicionales como almacenamiento local (para archivos de registro locales por ejemplo).

Otras advertencias para el uso de Amazon EBS con clústeres de Amazon EMR son:

- No puede realizar una instantánea de un volumen de Amazon EBS y posteriormente restaurarlo en Amazon EMR. Para crear configuraciones personalizadas reutilizables, utilice una AMI personalizada (disponible en Amazon EMR versión 5.7.0 y posteriores). Para obtener más información, consulte Uso de una AMI personalizada.
- Solo se admite un volumen de dispositivo raíz cifrado de Amazon EBS cuando se utiliza una AMI personalizada. Para obtener más información, consulte Creación de una AMI personalizada con un volumen de dispositivo raíz de Amazon EBS cifrado.
- Si aplica etiquetas con la API de Amazon EMR, dichas operaciones se aplicarán a volúmenes de EBS.
- Existe un límite de 25 volúmenes por instancia.
- Los volúmenes de Amazon EBS en los nodos principales no pueden ser inferiores a 5 GB.

Almacenamiento predeterminado de Amazon EBS para instancias

Amazon EMR adjunta automáticamente un volumen SSD de uso general (gp2) de 10 GB de Amazon EBS como dispositivo raíz de su AMI para mejorar el rendimiento. Además, para instancias de EC2 con almacenamiento exclusivo para EBS, Amazon EMR asigna los volúmenes de almacenamiento gp2 de Amazon EBS a las instancias. Al crear un clúster utilizando la versión 5.22.0 y posteriores de Amazon EMR, la cantidad predeterminada de almacenamiento de Amazon EBS aumenta en función del tamaño de la instancia. Se puede dividir un mayor almacenamiento en varios volúmenes, lo que ofrece un mayor rendimiento de IOPS y, a su vez, un mejor rendimiento para algunas cargas de trabajo estandarizadas. Si desea utilizar una configuración de almacenamiento para instancias gp2 de Amazon EBS diferente, puede especificarlo al crear un clúster de Amazon EMR o agregar nodos a un clúster existente. En este momento, los volúmenes gp3 de Amazon EBS no se pueden utilizar como volúmenes raíz en un clúster de Amazon EMR. Solo puede utilizar los volúmenes gp2 de Amazon EBS como volúmenes raíz y agregar volúmenes gp3 como volúmenes adicionales. En la siguiente tabla se identifica el número predeterminado de volúmenes de almacenamiento gp2 de Amazon EBS, los tamaños y los tamaños totales por tipo de instancia.

Los costos de Amazon EBS se prorratean por hora en función de los cargos mensuales de los volúmenes gp2 en la Región de AWS donde se ejecuta el clúster. Por ejemplo, el costo por hora de Amazon EBS para el volumen raíz en cada nodo de clúster de una región en la que se cobran

0,10 USD/GB al mes sería aproximadamente de 0,00139 USD por hora (0,10 USD/GB al mes dividido entre 30 días dividido entre 24 horas y multiplicado por 10 GB).

Volúmenes de almacenamiento gp2 de Amazon EBS predeterminados y tamaño por tipo de instancia para la versión 5.22.0 y posteriores de Amazon EMR

Tamaño de instancia	Número de volúmenes	Tamaño del volumen (GiB)	Tamaño total (GiB)
*.large	1	32	32
*.xlarge	2	32	64
*.2xlarge	4	32	128
*.4xlarge	4	64	256
*.8xlarge	4	128	512
9xlarge	4	144	576
*.10xlarge	4	160	640
*.12xlarge	4	192	768
*.16xlarge	4	256	1024
*.18xlarge	4	288	1152
*.24xlarge	4	384	1536

Especificación de volúmenes de almacenamiento adicionales de EBS

Al configurar los tipos de instancia en Amazon EMR, puede especificar volúmenes de EBS adicionales, lo que agrega capacidad más allá del almacén de instancias (en caso de incluirse) y el volumen de EBS predeterminado. Amazon EBS ofrece los siguientes tipos de volúmenes: uso general (SSD), de IOPS aprovisionadas (SSD), de rendimiento optimizado (HDD), en frío (HDD) y magnéticos. Se diferencian en las características de rendimiento y en el precio, por tanto puede personalizar su almacenamiento en función de las necesidades empresariales y de análisis de sus aplicaciones. Por ejemplo, algunas aplicaciones podrían tener la necesidad de volcar en el disco, mientras que otras pueden trabajar de forma segura en la memoria o utilizando Amazon S3.

Solo puede asociar volúmenes de Amazon EBS a instancias durante el tiempo de inicio del clúster y cuando agrega un grupo de instancias de nodos de tarea adicional. Si una instancia en un clúster de Amazon EMR presenta errores, tanto la instancia como los volúmenes de Amazon EBS asociados se sustituirán con volúmenes nuevos. Por lo tanto, si separa manualmente un volumen de Amazon EBS, Amazon EMR lo trata como un error y sustituye tanto los almacenamientos de la instancia (si procede) como los almacenes de volumen.

Amazon EMR no le permite modificar el tipo de volumen de gp2 a gp3 para un clúster de EMR existente. Si quiere utilizar gp3 para sus cargas de trabajo o casos de uso, debe lanzar un nuevo clúster de EMR. Además, no se recomienda actualizar el rendimiento y las IOPS de un clúster que esté en uso o que se esté aprovisionando, ya que Amazon EMR utiliza los valores de rendimiento y de IOPS que especificó en el momento de lanzar el clúster para cualquier instancia nueva que se agregue durante el escalado vertical del clúster. Consulte Comparación de los tipos de volúmenes gp2 y gp3 de Amazon EBS y Selección de las IOPS y el rendimiento al migrar a gp3.

> **Important**
>
> Para utilizar un volumen gp3 con su clúster de EMR, lance un nuevo clúster de EMR mediante la API, el SDK o la CLI.

Comparación de los tipos de volúmenes gp2 y gp3 de Amazon EBS

A continuación se incluye una comparación del costo entre los volúmenes gp2 y gp3 en la región us-east-1 (Norte de Virginia)

Tipo de volumen	gp3	gp2
Tamaño del volumen	1 GiB – 16 TiB	1 GiB – 16 TiB
IOPS predeterminadas o de referencia	3 000	3 IOPS/GiB (mínimo 100 IOPS) hasta un máximo de 16 000 IOPS. Los volúmenes de menos de 1 TiB también pueden alcanzar ráfagas de hasta 3000 IOPS.
Máximo de IOPS por volumen	16,000	16,000
Rendimiento predeterminado o de referencia	125 MiB/s	El límite de rendimiento está comprendido entre 128 MiB/s y 250 MiB/s, en función del tamaño del volumen.
Rendimiento máximo por volumen	1000 MiB/s	250 MiB/s
Precio	0,08 USD/GiB al mes, 3000 IOPS gratuitas y 0,005 USD/IOPS aprovisionadas al mes, más de 3000; 125 MiB/s gratuitas y 0,04 USD/MiB/s aprovisionadas al mes, más de 125 MiB/s	0,10 USD/GiB al mes

Selección de las IOPS y el rendimiento al migrar a gp3

Al aprovisionar un volumen de gp2, debe calcular el tamaño del volumen para obtener las IOPS y el rendimiento proporcionales. Con gp3, no es necesario aprovisionar un volumen mayor para obtener un mayor rendimiento. Puede elegir el tamaño y el rendimiento que desee según las necesidades de la aplicación. Si selecciona el tamaño y los parámetros de rendimiento adecuados (IOPS, rendimiento), podrá reducir al máximo los costos sin que ello afecte al rendimiento.

A continuación se incluye una tabla que lo ayudará a seleccionar las opciones de configuración de gp3:

Tamaño del volumen	IOPS	Rendimiento
1–170 GiB	3 000	125 MiB/s
170–334 GiB	3 000	125 MiB/s si el tipo de instancia de EC2 elegido admite 125 MiB/s o menos, utilice una cantidad mayor según el uso, 250 MiB/s* como máximo.
334–1000 GiB	3 000	125 MiB/s si el tipo de instancia de EC2 elegido admite 125 MiB/s o menos, utilice una cantidad mayor según el uso, 250 MiB/s* como máximo.
1000 GiB	Haga coincidir las IOPS de gp2 (tamaño en GiB x 3) o las IOPS máximas basadas en el volumen gp2 actual	125 MiB/s si el tipo de instancia de EC2 elegido admite 125 MiB/s o menos, utilice una cantidad mayor según el uso, 250 MiB/s* como máximo.

* Gp3 tiene la capacidad de proporcionar un rendimiento de hasta 1000 MiB/s. Como gp2 proporciona un rendimiento máximo de 250 MiB/s, es posible que no necesite superar este límite cuando utilice gp3.

Configurar redes

La mayoría de los clústeres se lanzan en una red virtual mediante Amazon Virtual Private Cloud (Amazon VPC). Una VPC es una red virtual dentro de AWS que está aislada de forma lógica en su cuenta de AWS. Puede configurar aspectos como los intervalos de direcciones IP privadas, las

subredes, las tablas de enrutamiento y las puertas de enlace de red. Para obtener más información, consulte la Guía del usuario de Amazon VPC.

VPC ofrece las siguientes capacidades:

- Procesamiento de información confidencial

  El lanzamiento de un clúster en una VPC es similar al lanzamiento del clúster en una red privada con herramientas adicionales, como, por ejemplo, tablas de enrutamiento y ACL de red, para definir quién tiene acceso a la red. Si va a procesar información confidencial en su clúster, es posible que desee el control de acceso adicional que ofrece el lanzamiento del clúster en una VPC. Además, puede elegir lanzar sus recursos en una subred privada en la que ninguno de ellos tenga conectividad directa a Internet.
- Acceso a los recursos de una red interna

  Si el origen de datos se encuentra en una red privada, puede ser poco práctico o indeseable cargar dichos datos en AWS para importarlos en Amazon EMR, ya sea debido a la cantidad de datos que deba transferir o debido a la naturaleza confidencial de los datos. En su lugar, puede lanzar el clúster en una VPC y conectar su centro de datos a su VPC a través de una conexión de VPN, habilitando el clúster para que acceda a recursos de su red interna. Por ejemplo, si tiene una base de datos Oracle en su centro de datos, el lanzamiento del clúster en una VPC conectada a dicha red por VPN permite que el clúster acceda a la base de datos Oracle.

Subredes públicas y privadas

Puede lanzar clústeres de Amazon EMR tanto en subredes de VPC públicas y privadas. Esto significa que no necesita conectividad a Internet para ejecutar un clúster de Amazon EMR. Sin embargo, es posible que tenga que configurar la traducción de direcciones de red (NAT) y las puertas de enlace de VPN para acceder a servicios o recursos ubicados fuera de la VPC; por ejemplo, en una intranet corporativa o en puntos de conexión de servicio de AWS públicos como AWS Key Management Service.

> **Important**
>
> Amazon EMR solo es compatible con el lanzamiento de clústeres en subredes privadas en su versión 4.2 y posteriores.

Para obtener más información sobre Amazon VPC, consulte la Guía del usuario de Amazon VPC.

Temas

- Opciones de Amazon VPC
- Configurar una VPC para alojar clústeres
- Lanzar clústeres en una VPC
- Política de Amazon S3 mínima para subred privada
- Más recursos para obtener información sobre VPC

Opciones de Amazon VPC

Cuando se lanza un clúster de Amazon EMR en una VPC, puede lanzarse en una subred compartida, privada o pública. Existen ligeras pero importantes diferencias en la configuración, en función del tipo de subred que elija para un clúster.

Subredes públicas

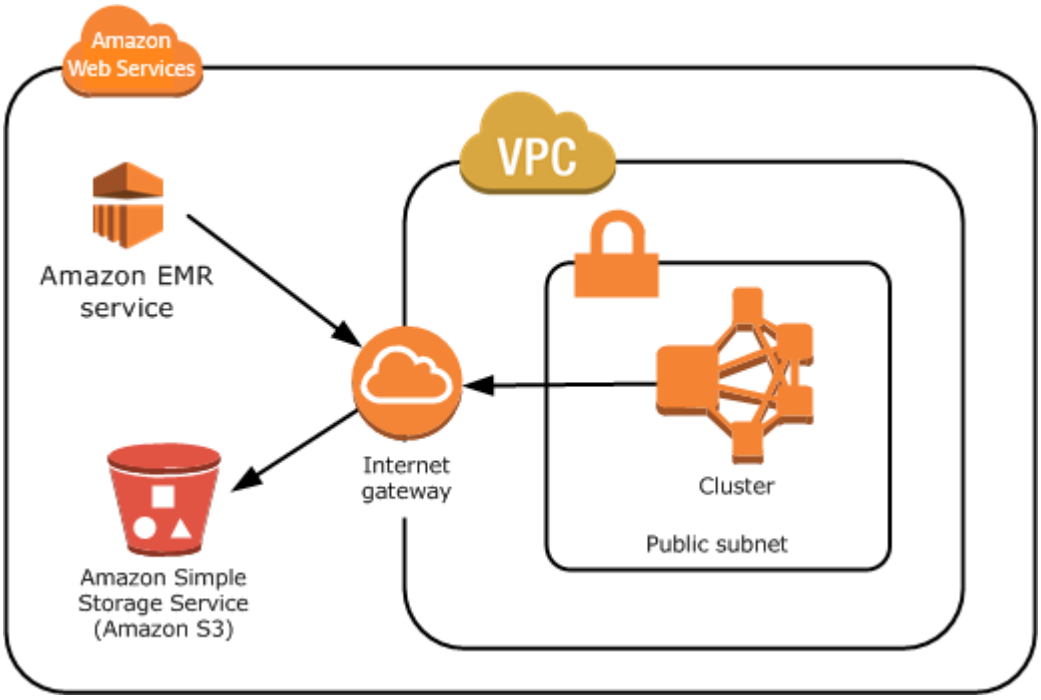
Los clústeres de EMR de una subred pública requieren una gateway de Internet conectada. Esto se debe a que los clústeres de Amazon EMR deben tener acceso a los servicios de AWS y a Amazon EMR. Si un servicio, como Amazon S3, ofrece la posibilidad de crear un punto de conexión de VPC, puede acceder a dichos servicios mediante el punto de conexión, en lugar de acceder a un punto de conexión público a través de una puerta de enlace de Internet. Además, Amazon EMR no puede comunicarse con clústeres en subredes públicas a través de un dispositivo de traducción de direcciones de red (NAT). Para este fin se necesita una gateway de Internet, pero puede utilizar una instancia NAT o una gateway para otro tráfico en situaciones más complejas.

Todas las instancias de un clúster se conectan a Amazon S3 a través de un punto de conexión de VPC o una puerta de enlace de Internet. Otros servicios de AWS que actualmente no admiten puntos de conexión de la VPC utilizan solo una puerta de enlace de Internet.

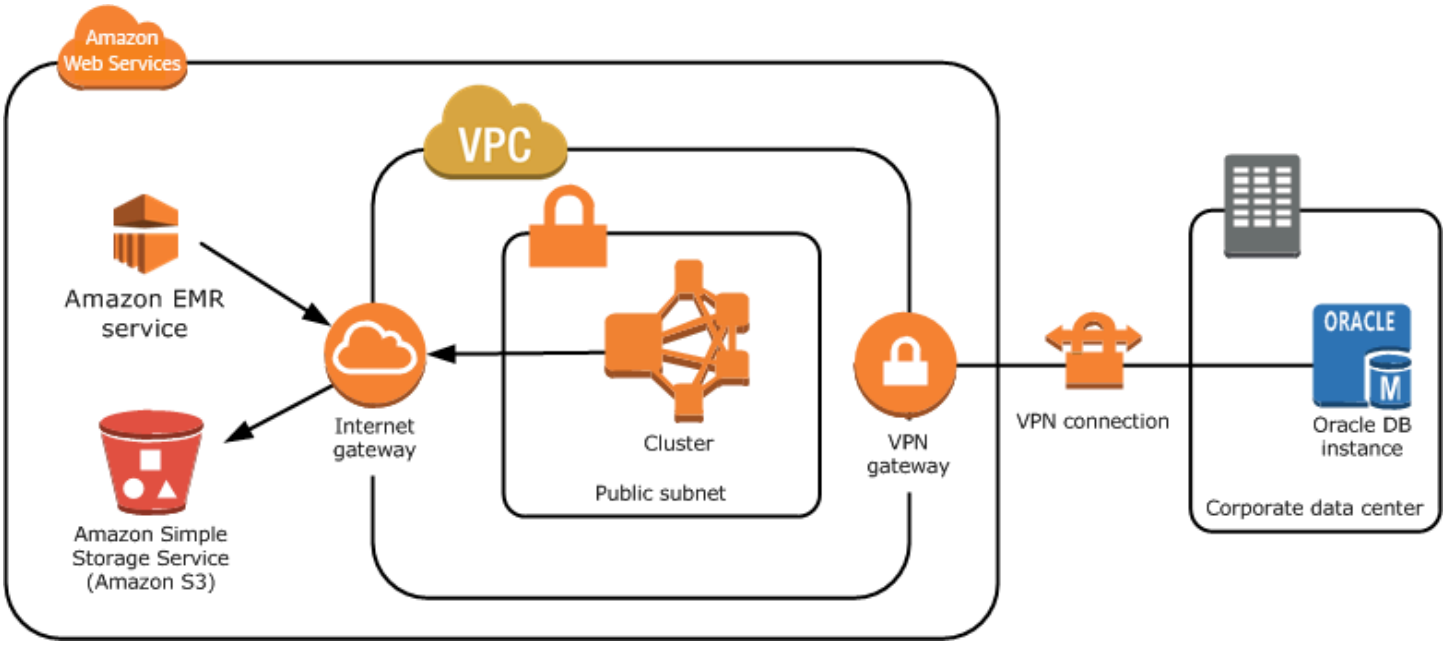
Si tiene recursos de AWS adicionales que no quiera conectar a la puerta de enlace de Internet, puede lanzar esos componentes en una subred privada que cree dentro de la VPC.

Los clústeres que se ejecutan en una subred pública utilizan dos grupos de seguridad: uno para el nodo principal y otro para los nodos básicos y de tarea. Para obtener más información, consulte Control del tráfico de red con grupos de seguridad.

El siguiente diagrama muestra cómo se ejecuta un clúster de Amazon EMR en una VPC mediante una subred pública. El clúster puede conectarse a otros recursos de AWS, como buckets de Amazon S3, a través de una puerta de enlace de Internet.



En el siguiente diagrama muestra cómo configurar una VPC para que un clúster en la VPC pueda acceder a los recursos en su propia red, como, por ejemplo, una base de datos Oracle.



Subredes privadas

Una subred privada le permite lanzar recursos de AWS sin necesidad de que la subred tenga asociada una puerta de enlace de Internet. Amazon EMR es compatible con el lanzamiento de clústeres en subredes privadas en su versión 4.2.0 y posteriores.

> **Note**
>
> Al configurar un clúster de Amazon EMR en una subred privada, se recomienda que también configure puntos de conexión de VPC para Amazon S3. Si su clúster de EMR se encuentra en una subred privada sin puntos de conexión de VPC para Amazon S3, incurrirá en cargos adicionales de puerta de enlace de NAT asociados al tráfico de S3, ya que el tráfico entre su clúster de EMR y S3 no permanecerá dentro de su VPC.

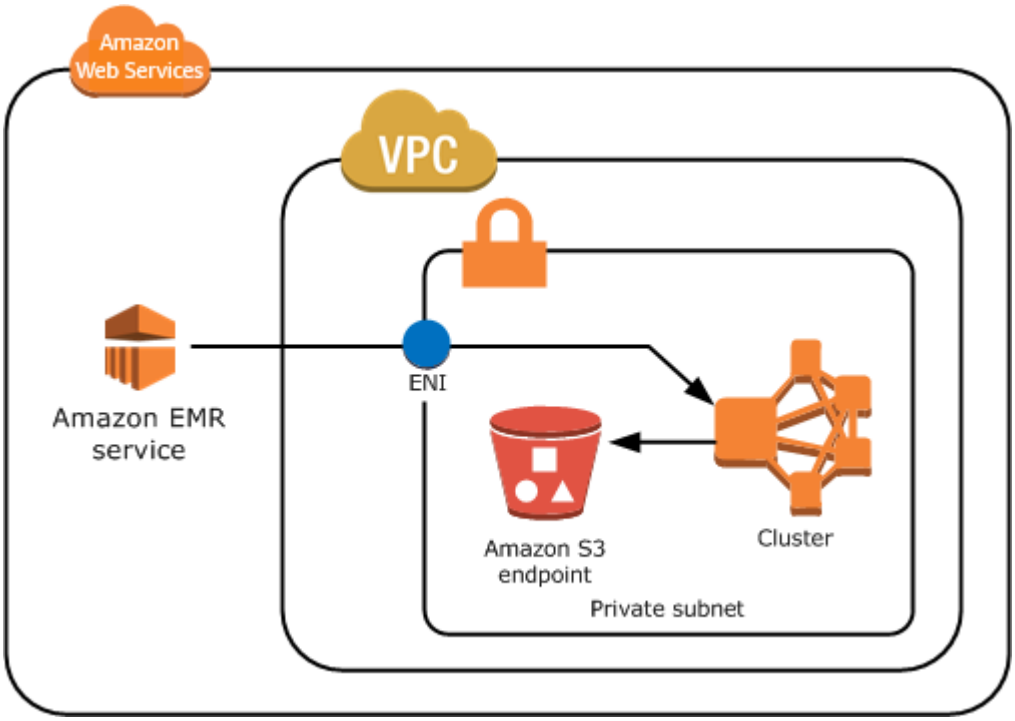
Las subredes privadas se diferencian de las públicas en los siguientes aspectos:

- Para acceder a los servicios de AWS que no proporcionan un punto de conexión de VPC, debe seguir utilizando una instancia de NAT o una puerta de enlace de Internet.
- Como mínimo, deberá proporcionar una ruta al bucket de registros de servicio de Amazon EMR y al repositorio de Amazon Linux en Amazon S3. Para obtener más información, consultar Política de Amazon S3 mínima para subred privada
- Si utiliza las características de EMRFS, debe disponer de un punto de conexión de VPC de Amazon S3 y de una ruta desde la subred privada a DynamoDB.
- La depuración solo funciona si proporciona una ruta desde la subred privada a un punto de conexión de Amazon SQS público.
- La creación de una configuración de subred privada con una instancia de NAT o puerto de enlace en una subred pública solo es compatible con la AWS Management Console. La forma más sencilla de agregar y configurar instancias NAT y puntos de conexión de VPC de Amazon S3 para los clústeres de Amazon EMR consiste en utilizar la página Lista de subredes de la VPC de la consola de Amazon EMR. Para configurar puertas de enlace de NAT, consulte Puertas de enlace de NAT en la Guía del usuario de Amazon VPC.
- No puede cambiar una subred con un clúster de Amazon EMR existente de pública a privada o viceversa. Para localizar un clúster de Amazon EMR dentro de una subred privada, el clúster debe iniciarse en dicha subred privada.

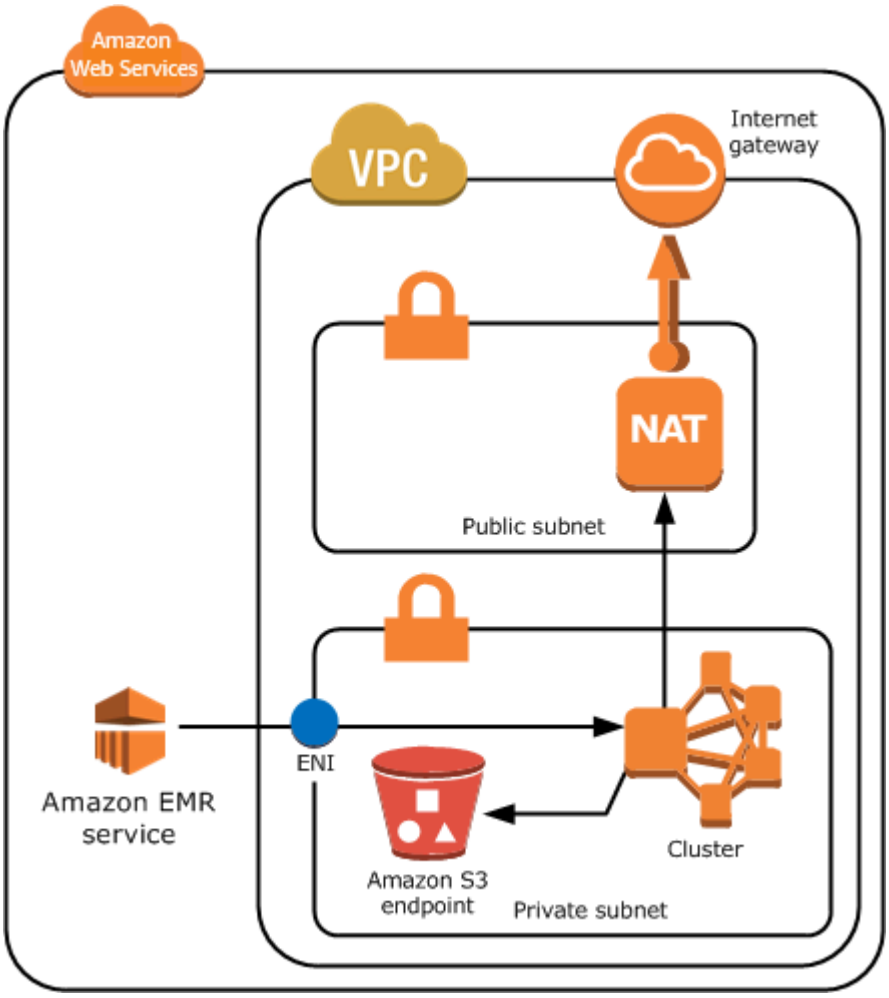
Amazon EMR crea y utiliza distintos grupos de seguridad predeterminados para los clústeres en una subred privada: ElasticMapReduce-Master-Private, ElasticMapReduce-Slave-Private y ElasticMapReduce-ServiceAccess. Para obtener más información, consulte Control del tráfico de red con grupos de seguridad.

Para obtener una lista completa de NACL de su clúster, elija Grupos de seguridad para nodo principal y Grupos de seguridad para nodos básicos y de tareas en la página Detalles del clúster de la consola de Amazon EMR.

La imagen siguiente muestra cómo se configura un clúster de Amazon EMR dentro de una subred privada. La única comunicación fuera de la subred es a Amazon EMR.



La imagen siguiente muestra una configuración de ejemplo para un clúster de Amazon EMR dentro de una subred privada conectada a una instancia de NAT que reside en una subred pública.



Subredes compartidas

El uso compartido de la VPC permite a los clientes compartir subredes con otras cuentas de AWS dentro de la misma organización de AWS. Puede lanzar clústeres de Amazon EMR en subredes compartidas públicas o privadas, con las siguientes advertencias.

El propietario de la subred debe compartir una subred con usted antes de lanzar un clúster de Amazon EMR en ella. Sin embargo, las subredes compartidas se pueden dejar de compartir más adelante. Para obtener más información, consulte Uso de VPC compartidas. Cuando un clúster se lanza en una subred compartida y esa subred se deja de compartir, es posible que observe comportamientos específicos basados en el estado del clúster de Amazon EMR cuando la subred se deja de compartir.

- La subred se deja de compartir antes de que el clúster se haya lanzado correctamente: si el propietario deja de compartir la VPC o la subred de Amazon mientras que el participante se está

lanzando un clúster, es posible que el clúster no se inicie o que se inicialice parcialmente sin aprovisionar todas las instancias solicitadas.

- La subred se deja de compartir después de que el clúster se haya lanzado correctamente: cuando el propietario deja de compartir una subred o VPC de Amazon con el participante, los clústeres del participante no podrán cambiar de tamaño para añadir nuevas instancias o para sustituir instancias en mal estado.

Cuando se lanza un clúster de Amazon EMR, se crean varios grupos de seguridad. En una subred compartida, el participante de la subred controla estos grupos de seguridad. El propietario de la subred pueden ver estos grupos de seguridad, pero no puede realizar ninguna acción en ellos. Si el propietario de la subred quiere eliminar o modificar el grupo de seguridad, el participante que ha creado el grupo de seguridad debe realizar la acción.

Control de permisos de VPC con IAM

De forma predeterminada, todos los usuarios pueden ver todas las subredes de la cuenta y cualquier usuario puede lanzar un clúster en cualquier subred.

Al lanzar un clúster en una VPC, puede utilizar AWS Identity and Access Management (IAM) para controlar el acceso a los clústeres y restringir acciones mediante políticas, de la misma forma que lo haría con los clústeres lanzados en Amazon EC2 Classic. Para más información acerca de IAM, consulte la Guía del usuario de IAM.

También puede utilizar IAM para controlar quién puede crear y administrar subredes. Por ejemplo, puede crear una cuenta para administrar subredes y una segunda cuenta que pueda lanzar clústeres, pero no pueda modificar la configuración de Amazon VPC. Para obtener más información sobre la administración de políticas y acciones en Amazon EC2 y Amazon VPC, consulte Políticas de IAM para Amazon EC2 en la Guía del usuario de Amazon EC2 para instancias de Linux.

Configurar una VPC para alojar clústeres

Antes de poder lanzar clústeres en una VPC, debe crear una VPC y una subred. Para subredes públicas, debe crear una gateway de Internet y asociarla a la subred. Las siguientes instrucciones describen cómo crear una VPC capaz de alojar clústeres de Amazon EMR.

Para crear una VPC con subredes para un clúster de Amazon EMR

1. Abra la consola de Amazon VPC en https://console.aws.amazon.com/vpc/.
2. En la parte superior derecha de la página, elija la Región de AWS para su VPC.

3. Seleccione Create VPC (Crear VPC).
4. En la página Crear VPC, seleccione VPC y más.
5. En Generación automática de etiquetas de nombre, habilite la generación automática e ingrese un nombre para la VPC. Esto le ayudará a identificar la VPC y la subred en la consola de Amazon VPC después de crearlas.
6. En el campo bloque CIDR IPv4, introduzca un espacio de dirección IP privada para que su VPC garantice una resolución de nombres de host de DNS adecuada; de lo contrario, podría experimentar errores de clúster de Amazon EMR. Esto incluye los siguientes rangos de direcciones IP:
   - 10.0.0.0 - 10.255.255.255
   - 172.16.0.0 - 172.31.255.255
   - 192.168.0.0 - 192.168.255.255
7. En Number of Availability Zones (AZs) (Número de zonas de disponibilidad), seleccione el número de zonas de disponibilidad que quiere usar con las subredes.
8. En Número de subredes públicas, elija una sola subred pública para agregar a la VPC. Si los datos que usa el clúster están disponibles en Internet (por ejemplo, en Amazon S3 o Amazon RDS), solo necesita utilizar una subred pública y no necesita agregar una subred privada.
9. En Number of private subnets (Número de subredes privadas), elija el número de subredes públicas que desea agregar a la VPC. Seleccione una o más si los datos de la aplicación se almacenan en su propia red (por ejemplo, en una base de datos de Oracle). Para una VPC en una subred privada, todas las instancias EC2 deben tener, como mínimo, una ruta a Amazon EMR a través de la interfaz de red elástica. En la consola, esto se configura automáticamente.
10. De manera opcional, en Puertas de enlace de NAT, puede optar por agregar puertas de enlace de NAT. Solo son necesarias si tiene subredes privadas que necesitan comunicarse con Internet.
11. En Puntos de conexión de la VPC, elija de manera opcional agregar puntos de conexión para Amazon S3 en sus subredes.
12. Compruebe que las opciones Habilitar nombres de host DNS y Habilitar la resolución de DNS estén marcadas. Para obtener más información, consulte Utilización de DNS con su VPC.
13. Seleccione Create VPC (Crear VPC).
14. En una ventana de estado se muestra el trabajo en curso. Cuando finalice el trabajo, elija Ver VPC para ir a la página Sus VPC, donde se muestran su VPC predeterminada y la VPC que

acaba de crear. La VPC que acaba de crear es una VPC no predeterminada. Por lo tanto, la columna Default VPC muestra No.

15. Si desea asociar su VPC a una entrada de DNS que no incluya un nombre de dominio, vaya a Conjuntos de opciones de DHCP, elija Crear conjunto de opciones de DHCP y omita un nombre de dominio. Después de crear el conjunto de opciones, vaya a la nueva VPC, elija Editar conjunto de opciones de DHCP en el menú Acciones y seleccione el nuevo conjunto de opciones. No puede editar el nombre de dominio utilizando la consola después de haber creado el conjunto de opciones de DNS.

    Es una práctica recomendada con Hadoop y aplicaciones relacionadas garantizar la resolución del nombre de dominio completo (FQDN) para los nodos. Para garantizar una resolución de DNS adecuada, configure una VPC que incluya un conjunto de opciones de DHCP cuyos parámetros se definan con los valores siguientes:

    - domain-name = **`ec2.internal`**

      Use **`ec2.internal`** si su región es Este de EE. UU. (Norte de Virginia). Para las demás regiones, utilice *`region-name`***`.compute.internal`**. Para ver ejemplos de `us-west-2`, utilice **`us-west-2.compute.internal`**. Para la región AWS GovCloud (Oeste de EE. UU), utilice **`us-gov-west-1.compute.internal`**.
    - domain-name-servers = **`AmazonProvidedDNS`**

    Para más información, consulte Conjuntos de opciones de DHCP en la Guía del usuario de Amazon VPC.
16. Después de haber creado la VPC, vaya a la página Subredes y anote el ID de subred de una de las subredes de su nueva VPC. Utilice esta información al lanzar el clúster de Amazon EMR en la VPC.

Lanzar clústeres en una VPC

Después de tener una subred configurada para alojar clústeres de Amazon EMR, lance el clúster en dicha subred especificando el identificador de subred asociado al crear el clúster.

> **Note**
>
> Amazon EMR admite subredes privadas en su versión 4.2 y superiores.

Cuando se lanza el clúster, Amazon EMR agrega grupos de seguridad en función de si el clúster se lanza en subredes privadas o públicas de VPC. Todos los grupos de seguridad permiten la entrada en el puerto 8443 para comunicarse con el servicio de Amazon EMR, pero los intervalos de direcciones IP varían en subredes públicas y privadas. Amazon EMR administra todos estos grupos de seguridad y es posible que tenga que agregar direcciones IP adicionales al intervalo de AWS a lo largo del tiempo. Para obtener más información, consulte Control del tráfico de red con grupos de seguridad.

Para administrar el clúster en una VPC, Amazon EMR asocia un dispositivo de red al nodo principal y lo administra a través de este dispositivo. Puede ver este dispositivo mediante la acción `DescribeInstances` de la API de Amazon EC2. Si modifica este dispositivo de algún modo, el clúster podría fallar.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para lanzar un clúster en una VPC utilizando la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Redes, vaya al campo Nube privada virtual (VPC). Ingrese el nombre de la VPC o elija Buscar para seleccionar la VPC. También puede elegir Crear VPC para crear una VPC que pueda utilizar para su clúster.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para lanzar un clúster en una VPC utilizando la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Go to advanced options (Ir a las opciones avanzadas).
4. En la sección Hardware Configuration (Configuración de hardware), para Network (Red), seleccione el ID de una red de VPC que haya creado con anterioridad.
5. En EC2 Subnet (Subred de EC2), seleccione el ID de una subred que haya creado con anterioridad.

   a. Si la subred privada está configurada correctamente con opciones de punto de enlace de S3 e instancia NAT, se muestra (EMR Ready) (Lista para EMR) encima de los identificadores y nombres de subred.

   b. Si la subred privada no tiene una instancia NAT o un punto de enlace de S3, puede configurarlo eligiendo Add S3 endpoint and NAT instance (Añadir punto de enlace de S3 e instancia NAT), Add S3 endpoint (Añadir punto de enlace de S3) o Add NAT instance (Añadir instancia NAT). Seleccione las opciones que desee para la instancia NAT y el punto de enlace de S3 y elija Configure (Configurar).

   > **Important**
   >
   > Para crear una instancia NAT desde Amazon EMR, necesita los permisos ec2:CreateRoute, `ec2:RevokeSecurityGroupEgress`, `ec2:AuthorizeSecurityGroupEgress`, `cloudformation:DescribeStackEvents` y `cloudformation:CreateStack`.

   > **Note**
   >
   > No existe un costo adicional por lanzar una instancia de Amazon EC2 para su dispositivo NAT.

6. Continúe con la creación del clúster.

AWS CLI

Para lanzar un clúster en una VPC utilizando la AWS CLI

> **Note**
>
> La AWS CLI no proporciona una forma de crear una instancia NAT automáticamente y conectarla a su subred privada. Sin embargo, para crear un punto de conexión de S3 en su subred, puede utilizar los comandos de la CLI de Amazon VPC. Utilice la consola para crear instancias NAT y lanzar clústeres en una subred privada.

Una vez que la VPC se ha configurado, puede lanzar clústeres de Amazon EMR en ella utilizando el subcomando `create-cluster` con el parámetro `--ec2-attributes`. Utilice el parámetro `--ec2-attributes` para especificar la subred de VPC para el clúster.

- Para crear un clúster en una subred específica, escriba el siguiente comando, sustituya *myKey* por el nombre del par de claves de Amazon EC2 y sustituya *77XXXX03* por el ID de subred.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.2.0 --
applications Name=Hadoop Name=Hive Name=Pig --use-default-roles --ec2-attributes
 KeyName=myKey,SubnetId=subnet-77XXXX03 --instance-type m5.xlarge --instance-
count 3
```

  Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos básicos. Todos los nodos utilizan el tipo de instancia especificado en el comando.

  > **Note**
  >
  > Si no ha creado con anterioridad el rol de servicio de Amazon EMR predeterminado y el perfil de instancia de EC2, escriba `aws emr create-default-roles` para crearlos antes de escribir el subcomando `create-cluster`.

Política de Amazon S3 mínima para subred privada

En el caso de las subredes privadas, como mínimo debe proporcionar a Amazon EMR la capacidad de acceder a los repositorios de Amazon Linux. Esta política de subred privada forma parte de las políticas de puntos de conexión de VPC para el acceso a Amazon S3. Con Amazon EMR 5.25.0 o posterior, para habilitar el acceso de un clic al servidor del historial de Spark persistente, debe permitir a Amazon EMR el acceso al bucket del sistema que recopila los registros de eventos de Spark. Si habilita el registro, proporcione permisos PUT a un bucket `aws157-logs-*`. Para más información, consulte Acceso de un clic al servidor del historial de Spark persistente.

Usted debe determinar las restricciones de política que satisfacen sus necesidades empresariales. Por ejemplo, puede especificar la región `packages.us-east-1.amazonaws.com` para evitar un nombre de bucket de Amazon S3 ambiguo. En el siguiente ejemplo de política se proporcionan permisos para acceder a los repositorios de Amazon Linux y al bucket del sistema de Amazon EMR para recopilar registros de eventos de Spark. Sustituya *`MyRegion`* por la región en la que residen sus buckets de registros, por ejemplo `us-east-1`.

Para más información acerca del uso de políticas de IAM con puntos de conexión de Amazon VPC, consulte Políticas de punto de conexión para Amazon S3.

```
{
   "Version": "2008-10-17",
   "Statement": [
       {
           "Sid": "AmazonLinuxAMIRepositoryAccess",
           "Effect": "Allow",
           "Principal": "*",
           "Action": "s3:GetObject",
           "Resource": [
            "arn:aws:s3:::packages.MyRegion.amazonaws.com/*",
            "arn:aws:s3:::repo.MyRegion.amazonaws.com/*",
            "arn:aws:s3:::repo.MyRegion.emr.amazonaws.com/*"
           ]
       },
       {
           "Sid": "EnableApplicationHistory",
           "Effect": "Allow",
           "Principal": "*",
           "Action": [
               "s3:Put*",
               "s3:Get*",
```

```
                "s3:Create*",
                "s3:Abort*",
                "s3:List*"
            ],
            "Resource": [
             "arn:aws:s3:::prod.MyRegion.appinfo.src/*"
            ]
        }
    ]
}
```

El siguiente ejemplo de política proporciona los permisos necesarios para acceder a los repositorios de Amazon Linux 2. La AMI de Amazon Linux 2 es el valor predeterminado.

```
{
    "Statement": [
        {
            "Sid": "AmazonLinux2AMIRepositoryAccess",
            "Effect": "Allow",
            "Principal": "*",
            "Action": "s3:GetObject",
            "Resource": [
             "arn:aws:s3:::amazonlinux.MyRegion.amazonaws.com/*",
             "arn:aws:s3:::amazonlinux-2-repos-MyRegion/*"
            ]
        }
    ]
}
```

Más recursos para obtener información sobre VPC

Utilice los siguientes temas para obtener más información sobre las VPC y las subredes.

- Subredes privadas en una VPC
  - Escenario 2: VPC con subredes públicas y privadas (NAT)
  - Instancias NAT
  - Alta disponibilidad para instancias NAT de Amazon VPC: un ejemplo
- Subredes públicas en una VPC
  - Escenario 1: VPC con una única subred pública
- Información de VPC general

- Guía del usuario de Amazon VPC
- Interconexión de VPC
- Uso de interfaces de red elásticas con una VPC
- Securely connect to Linux instances running in a private VPC

Crear un clúster con flotas de instancias o grupos de instancias uniformes

Al crear un clúster y especificar la configuración del nodo principal y de los nodos básicos y de tareas, tiene dos opciones de configuración. Puede utilizar flotas de instancias o grupos de instancias uniformes. La opción de configuración que elija se aplica a todos los nodos, se aplica durante toda la vida útil del clúster y las flotas de instancias y los grupos de instancias no pueden coexistir en un clúster. La configuración de las flotas de instancias está disponible en la versión 4.8.0 y posteriores de Amazon EMR, excluyendo las versiones 5.0.x.

Puede utilizar la consola de Amazon EMR, la AWS CLI o la API de Amazon EMR para crear clústeres con cualquier configuración. Al utilizar el comando `create-cluster` desde la AWS CLI, puede utilizar los parámetros `--instance-fleets` para crear el clúster utilizando las flotas de instancias o, como opción, utilizar los parámetros `--instance-groups` para crearlo usando grupos de instancias uniformes.

Lo mismo sucede utilizando la API de Amazon EMR. Puede utilizar la configuración `InstanceGroups` para especificar una matriz de objetos `InstanceGroupConfig` o bien utilizar la configuración `InstanceFleets` para especificar una matriz de objetos `InstanceFleetConfig`.

En la nueva consola de Amazon EMR, puede elegir utilizar grupos de instancias o flotas de instancias al crear un clúster y tiene la opción de utilizar instancias de spot con cada uno de ellos. En la consola anterior de Amazon EMR, si utiliza la configuración predeterminada de las Opciones rápidas al crear un clúster, Amazon EMR aplica la configuración de los grupos de instancias uniformes al clúster y utiliza instancias bajo demanda. Para utilizar las instancias de spot con grupos de instancias uniformes o para configurar las flotas de instancias y otras personalizaciones, elija Advanced Options (Opciones avanzadas).

Flotas de instancias

La configuración de flotas de instancias ofrece la más amplia variedad de opciones de aprovisionamiento para instancias de Amazon EC2. Cada tipo de nodo tiene una sola flota de instancia y la flota de instancia de tarea es opcional. Puede especificar hasta cinco tipos de instancias de EC2 por flota o 30 tipos de instancias de EC2 por flota al crear un clúster mediante

AWS CLI o la API de Amazon EMR y una estrategia de asignación para instancias de spot y bajo demanda. En el caso de las flotas de instancia básica y de tarea, se asigna una capacidad de destino para instancias bajo demanda y otra para instancias de spot. Amazon EMR elige cualquier combinación de los tipos de instancia especificados para satisfacer las capacidades de destino, aprovisionando tanto las instancias bajo demanda como las instancias de spot.

En el caso del tipo de nodo principal, Amazon EMR elige un solo tipo de instancia de la lista de instancias y especifica si está aprovisionado como instancia bajo demanda o instancia de spot. Las flotas de instancias también ofrecen opciones adicionales para las compras de instancias de spot y bajo demanda. Las opciones de instancias de spot incluyen un tiempo de espera que especifica la acción que se debe realizar en caso de que no se pueda aprovisionar la capacidad de spot y una estrategia de asignación preferida (optimizada en función de la capacidad) para lanzar flotas de instancias de spot. Las flotas de instancias bajo demanda también se pueden lanzar mediante la opción de estrategia de asignación (precio más bajo). Si utiliza un rol de servicio que no es el predeterminado de EMR o utiliza una política administrada de EMR en su rol de servicio, debe agregar permisos adicionales al rol de servicio de clúster personalizado para habilitar la opción de estrategia de asignación. Para obtener más información, consulte Rol de servicio para Amazon EMR (rol de EMR).

Para más información acerca de la configuración de las flotas de instancias, consulte Configurar flotas de instancias.

Grupos de instancias uniformes

Los grupos de instancias uniformes ofrecen una configuración más sencilla que las flotas de instancias. Cada clúster de Amazon EMR puede incluir hasta 50 grupos de instancias: un grupo de instancias principales que contiene una instancia de Amazon EC2, un grupo de instancias básicas que contiene una o varias instancias de EC2 y hasta 48 grupos de instancias de tarea opcionales. Cada grupo de instancias básicas y de tarea puede contener cualquier número de instancias de Amazon EC2. Puede escalar cada grupo de instancias agregando y eliminando instancias de Amazon EC2 manualmente o puede configurar el escalado automático. Para obtener información sobre la adición y eliminación de instancias, consulte Usar el escalado de clústeres.

Para obtener más información acerca de la configuración de grupos de instancias uniformes, consulte Configuración de grupos de instancias uniformes.

Uso de flotas de instancias y grupos de instancias

Temas

- Configurar flotas de instancias
- Utilizar reservas de capacidad con flotas de instancias
- Configuración de grupos de instancias uniformes
- Prácticas recomendadas para la flexibilidad de las instancias y las zonas de disponibilidad
- Prácticas recomendadas para la configuración del clúster

Configurar flotas de instancias

> **Note**
>
> La configuración de las flotas de instancias está disponible solo en las versiones 4.8.0 y posteriores de Amazon EMR, excluidas las versiones 5.0.0 y 5.0.3.

La configuración de la flota de instancias para los clústeres de Amazon EMR le permite seleccionar una amplia variedad de opciones de aprovisionamiento para las instancias de Amazon EC2 y lo ayuda a desarrollar una estrategia de recursos flexible y elástica para cada tipo de nodo del clúster.

En una configuración de flota de instancias, debe especificar una capacidad de destino para las instancias bajo demanda y las instancias de spot dentro de cada flota. Cuando se lanza el clúster, Amazon EMR aprovisiona instancias hasta que se atienden los destinos. Cuando Amazon EC2 reclama una instancia de spot en un clúster en ejecución debido a un incremento de los precios, o una instancia devuelve un error, Amazon EMR intenta sustituir la instancia por cualquiera de los tipos de instancias que especifique. De ese modo, resulta más sencillo recuperar la capacidad durante un pico de los precios de spot.

Puede especificar hasta cinco tipos de instancias de Amazon EC2 por flota para que Amazon EMR los utilice al atender los destinos o un máximo de 30 tipos de instancias de Amazon EC2 por flota al crear un clúster mediante la AWS CLI o la API de Amazon EMR y una estrategia de asignación para instancias de spot y bajo demanda.

También puede seleccionar varias subredes para distintas zonas de disponibilidad. Cuando Amazon EMR lanza el clúster, busca en esas subredes para encontrar las instancias y las opciones de compra que especifique. Si Amazon EMR detecta un evento de AWS a gran escala en una o varias zonas de disponibilidad, Amazon EMR intenta automáticamente enrutar el tráfico de las zonas de disponibilidad afectadas e intenta lanzar nuevos clústeres que cree en zonas de disponibilidad alternativas de acuerdo con sus selecciones. Tenga en cuenta que la selección de la zona de

disponibilidad del clúster solo se realiza en el momento de crear el clúster. Los nodos del clúster existentes no se vuelven a lanzar automáticamente en una nueva zona de disponibilidad en el caso de que se produzca una interrupción en la zona de disponibilidad.

Consideraciones

Tenga en cuenta los siguientes aspectos al utilizar flotas de instancias con Amazon EMR.

- Puede tener una flota de instancias, y solo una, por tipo de nodo (principal, básico, de tarea). Puede especificar hasta cinco tipos de instancias de Amazon EC2 para cada flota en la AWS Management Console (o un máximo de 30 tipos por flota de instancias al crear un clúster mediante AWS CLI o la API de Amazon EMR y una Estrategia de asignación para flotas de instancias).
- Amazon EMR elige cualquiera o todos los tipos de instancia de EC2 especificados para aprovisionar con las opciones de compra de spot y bajo demanda.
- Establezca las capacidades de destino para instancias de spot y bajo demanda para la flota de instancias básicas y de tarea. Utilice vCPU o una unidad genérica asignada a cada instancia de Amazon EC2 que se tiene en cuenta para los destinos. Amazon EMR aprovisiona instancias hasta que se ha satisfecho por completo cada capacidad de destino. Para la flota principal, el destino es siempre uno.
- Puede elegir una subred (zona de disponibilidad) o un intervalo. Si elige un intervalo, Amazon EMR aprovisiona capacidad en la zona de disponibilidad que sea la mejor opción.
- Cuando se especifica una capacidad de destino para las instancias de spot:
  - Para cada tipo de instancia, especifique un precio de spot máximo. Amazon EMR aprovisiona las instancias de spot si el precio de spot está por debajo del precio de spot máximo. Solo tiene que pagar el precio de spot, que no es necesariamente el precio de spot máximo.
  - Para cada flota, defina un período de tiempo de espera para aprovisionamiento de instancias de spot. Si Amazon EMR no puede aprovisionar la capacidad de spot, puede terminar el clúster o en su lugar cambiarlo a aprovisionamiento de capacidad bajo demanda. Esto solo se aplica al aprovisionamiento de clústeres, no a su cambio de tamaño. Si el tiempo de espera finaliza durante el proceso de cambio de tamaño del clúster, las solicitudes de spot no aprovisionadas se anularán sin transferirse a la capacidad bajo demanda.
- En el caso de cada flota, puede especificar una de las siguientes estrategias de asignación para sus instancias de spot: optimizadas en función de la relación precio-capacidad, optimizadas en función de la capacidad, con el precio más bajo o diversificadas en todos los grupos.
- En el caso de cada flota, puede aplicar la estrategia de asignación con el precio más bajo para sus instancias bajo demanda; no puede personalizar dicha estrategia para estas instancias.

- En el caso de cada flota con `allocation strategy - lowest-price` bajo demanda, puede optar por aplicar opciones de reserva de capacidad.
- Compruebe el tamaño de la subred antes de lanzar el clúster. Cuando aprovisiona un clúster con una flota de tareas y no hay suficientes direcciones IP disponibles en la subred correspondiente, la flota pasará a un estado de suspensión en lugar de terminar el clúster con un error. Para evitar este problema, se recomienda aumentar la cantidad de direcciones IP en las subredes.

Opciones de flota de instancias

Utilice las siguientes directrices para comprender las opciones de flota de instancia.

Temas

- Configuración de las capacidades de destino
- Opciones de lanzamiento
- Opciones para varias subredes (zonas de disponibilidad)
- Configuración del nodo maestro

Configuración de las capacidades de destino

Especifique las capacidades de destino que desee para la flota de instancias secundarias y la flota de instancias de tareas. Cuando lo haga, eso determina el número de instancias bajo demanda e instancias de spot que aprovisiona Amazon EMR. Cuando se especifica una instancia, decide la cantidad de cada instancia que se tiene en cuenta para el destino. Cuando se aprovisiona una instancia bajo demanda, se tiene en cuenta para el destino bajo demanda. Lo mismo sucede para las instancias de spot. A diferencia de las flotas de instancias básicas y de tarea, la flota de instancias principales siempre es una instancia. Por tanto, la capacidad de destino para esta flota es siempre uno.

Cuando se utiliza la consola, las vCPU del tipo de instancia de Amazon EC2 se utilizan como recuento para las capacidades de destino de forma predeterminada. Puede cambiar esto a Generic units (Unidades genéricas) y, a continuación, especificar el recuento de cada tipo de instancia EC2. Cuando se utiliza la AWS CLI, asigna manualmente unidades genéricas para cada tipo de instancia.

> **Important**
>
> Al elegir un tipo de instancia mediante la AWS Management Console, el número de vCPU que se muestra para cada tipo de instancia es el número de núcleos virtuales de YARN

para ese tipo de instancia, no el número de vCPU de EC2 para ese tipo de instancia. Para más información sobre el número de vCPU para cada tipo de instancia, consulte Tipos de instancias de Amazon EC2.

Para cada flota, se especifican hasta cinco tipos de instancia de Amazon EC2. Si usa una Estrategia de asignación para flotas de instancias y crea un clúster con AWS CLI o la API de Amazon EMR, puede especificar hasta 30 tipos de instancias de EC2 por flota de instancias. Amazon EMR elige cualquier combinación de estos tipos de instancias de EC2 para satisfacer sus capacidades de destino. Dado que Amazon EMR desea rellenar la capacidad de destino por completo, podría producirse un sobreuso. Por ejemplo, si hay dos unidades no atendidas y Amazon EMR solo puede aprovisionar una instancia con un recuento de cinco unidades, la instancia aún se aprovisionará, lo que significa que la capacidad de destino se excederá en tres unidades.

Si reduce la capacidad de destino para cambiar el tamaño de un clúster en ejecución, Amazon EMR intenta completar tareas de aplicación y termina instancias para satisfacer el nuevo destino. Para obtener más información, consulte Terminación al completar las tareas.

Opciones de lanzamiento

En el caso de cada grupo de instancias, especifique un precio de spot máximo por cada tipo de instancia en una flota. Puede configurar este precio como un porcentaje del precio bajo demanda o como un importe específico en dólares. Amazon EMR aprovisiona instancias de spot si el precio de spot actual en una zona de disponibilidad está por debajo de su precio de spot máximo. Solo tiene que pagar el precio de spot, que no es necesariamente el precio de spot máximo.

> **Note**
>
> Las instancias de spot con una duración definida (también conocidas como bloques de spot) dejarán de estar disponibles para los nuevos clientes a partir del 1 de julio de 2021. En el caso de los clientes que hayan utilizado previamente la característica, se continuará ofreciendo soporte a las instancias de spot con una duración definida hasta el 31 de diciembre de 2022.

Disponible en Amazon EMR 5.12.1 y versiones posteriores, tiene la opción de lanzar flotas de instancias de spot y bajo demanda con una asignación de capacidad optimizada. Esta opción de estrategia de asignación se puede configurar en la antigua AWS Management Console o mediante

la API `RunJobFlow`. Tenga en cuenta que no puede personalizar la estrategia de asignación en la nueva consola. El uso de la opción de estrategia de asignación requiere permisos de rol de servicio adicionales. Si utiliza el rol de servicio predeterminado de Amazon EMR y la política administrada (`EMR_DefaultRole` y `AmazonEMRServicePolicy_v2`) para el clúster, los permisos para la opción de estrategia de asignación ya están incluidos. Si no utiliza el rol de servicio ni la política administrada de Amazon EMR predeterminados, debe agregarlos para utilizar esta opción. Consulte Rol de servicio para Amazon EMR (rol de EMR).

Para más información acerca de las instancias de spot, consulte Instancias de spot en la Guía del usuario de Amazon EC2 para instancias de Linux. Para más información acerca de las instancias bajo demanda, consulte Instancias bajo demanda en la Guía del usuario de Amazon EC2 para instancias de Linux.

Si opta por lanzar flotas de instancias bajo demanda con la estrategia de asignación con el precio más bajo, tiene la opción de utilizar las reservas de capacidad. Las opciones de reserva de capacidad se pueden establecer mediante la API `RunJobFlow` de Amazon EMR. Las reservas de capacidad requieren permisos de rol de servicio adicionales que debe agregar para utilizar estas opciones. Consulte Permisos de estrategia de asignación. Tenga en cuenta que no puede personalizar las reservas de capacidad en la nueva consola.

Opciones para varias subredes (zonas de disponibilidad)

Cuando utilice flotas de instancias, puede especificar varias subredes de Amazon EC2 dentro de una VPC, cada una de ellas correspondiente a una zona de disponibilidad diferente. Si utiliza EC2-Classic, las zonas de disponibilidad se especifican de forma explícita. Amazon EMR identifica la mejor zona de disponibilidad para lanzar instancias de acuerdo con sus especificaciones de flota. Las instancias se aprovisionan siempre en una única zona de disponibilidad. Puede seleccionar subredes privadas o subredes públicas, pero no puede combinarlas y las subredes que especifique deben estar dentro de la misma VPC.

Configuración del nodo maestro

Dado que la flota de instancias principales es únicamente una sola instancia, su configuración es ligeramente distinta de las flotas de instancias básicas y de tarea. Solo seleccione bajo demanda o de spot para la flota de instancias principales, ya que se compone de una única instancia. Si utiliza la consola para crear la flota de instancias, la capacidad de destino para la opción de compra que seleccione se define en 1. Si utiliza la AWS CLI, defina siempre `TargetSpotCapacity` o `TargetOnDemandCapacity` en 1, según proceda. Aún puede elegir hasta cinco tipos de instancias para la flota de instancias principal (o un máximo de 30 si utiliza la opción de estrategia

de asignación para las instancias de spot o bajo demanda). Sin embargo, a diferencia de las flotas de instancias básicas y de tarea, donde Amazon EMR podría aprovisionar varias instancias de tipos distintos, Amazon EMR selecciona un único tipo de instancia para aprovisionar la flota de instancias principales.

Estrategia de asignación para flotas de instancias

Con las versiones 5.12.1 y posteriores de Amazon EMR, puede utilizar la opción de estrategia de asignación con instancias de spot y bajo demanda para cada nodo del clúster. Al crear un clúster mediante AWS CLI, la API de Amazon EMR o la consola de Amazon EMR con una estrategia de asignación, puede especificar hasta 30 tipos de instancias de Amazon EC2 por flota. Con la configuración predeterminada de flota de instancias del clúster de Amazon EMR, puede tener hasta 5 tipos de instancias por flota. Se recomienda que utilice la opción de estrategia de asignación para un aprovisionamiento de clústeres más rápido, una asignación de instancias de spot más precisa y menos interrupciones de dichas instancias.

Temas

- Estrategia de asignación con instancias bajo demanda
- Estrategia de asignación con instancias de spot
- Permisos de estrategia de asignación
- Permisos de IAM necesarios para una estrategia de asignación

Estrategia de asignación con instancias bajo demanda

Cuando utiliza una estrategia de asignación, sus instancias bajo demanda utilizan la estrategia de precio más bajo. Esto lanza primero las instancias con el precio más bajo. Al lanzar instancias bajo demanda, puede utilizar reservas de capacidad abiertas o específicas en sus cuentas. Puede usar reservas de capacidad abiertas para los nodos principales, básicos y de tarea. Es posible que la capacidad de las instancias bajo demanda con una estrategia de asignación para flotas de instancias sea insuficiente. Se recomienda que especifique un número mayor de tipos de instancias para diversificar y reducir la posibilidad de que la capacidad sea insuficiente. Para obtener más información, consulte Utilizar reservas de capacidad con flotas de instancias.

Estrategia de asignación con instancias de spot

En el caso de las instancias de spot, puede elegir una de las siguientes estrategias de asignación:

`price-capacity-optimized` (recomendado)

La estrategia de asignación optimizada en función de la relación precio-capacidad lanza instancias de spot desde los grupos de instancias de spot que tienen la mayor capacidad disponible y el precio más bajo para la cantidad de instancias que se están lanzando. Como resultado, la estrategia optimizada en función de la relación precio-capacidad suele tener más probabilidades de conseguir capacidad de spot y ofrece tasas de interrupción más bajas.

`capacity-optimized`

La estrategia de asignación optimizada en función de la capacidad lanza las instancias de spot en los grupos más disponibles con la menor probabilidad de interrupción a corto plazo. Esta es una buena opción para las cargas de trabajo que podrían tener un costo de interrupción más alto asociado con el trabajo que se reinicia. Esta es la estrategia predeterminada para las versiones 6.9.0 y anteriores de Amazon EMR.

`diversified`

Con la estrategia de asignación diversificada, Amazon EC2 distribuye las instancias de spot entre todos los grupos de capacidad de spot.

`lowest-price`

La estrategia de asignación de precios más bajos lanza instancias de spot desde el grupo de precios más bajos que tiene capacidad disponible. Si el grupo con el precio más bajo no tiene capacidad disponible, las instancias de spot provienen del siguiente grupo con el precio más bajo que tenga capacidad disponible. Si un grupo se queda sin capacidad antes de cubrir la capacidad deseada, la flota de Amazon EC2 extraerá capacidad del siguiente grupo con el precio más bajo para cumplir su solicitud. Para garantizar que se logre la capacidad deseada, es posible que reciba instancias de spot de varios grupos. Dado que esta estrategia solo tiene en cuenta el precio de la instancia y no la disponibilidad de capacidad, podría generar tasas de interrupción elevadas.

Permisos de estrategia de asignación

La opción de estrategia de asignación requiere varios permisos de IAM que se incluyen automáticamente en el rol de servicio predeterminado de Amazon EMR y en la política administrada de Amazon EMR (`EMR_DefaultRole` y `AmazonEMRServicePolicy_v2`). Si utiliza un rol de servicio o una política administrada personalizados para su clúster, debe agregar estos permisos antes de crear el clúster. Para obtener más información, consulte Permisos de estrategia de asignación.

Las reservas de capacidad bajo demanda (ODCR) opcionales están disponibles cuando se utiliza la opción de estrategia de asignación bajo demanda. Las opciones de reserva de capacidad le permiten especificar una preferencia para utilizar primero la capacidad reservada para los clústeres de Amazon EMR. Puede utilizarlas para asegurarse de que sus cargas de trabajo críticas utilicen la capacidad que ya ha reservado mediante ODCR abiertas o específicas. En el caso de las cargas de trabajo que no son críticas, las preferencias de reserva de capacidad permiten especificar si se debe consumir la capacidad reservada.

Las reservas de capacidad solo las pueden utilizar instancias que tengan atributos coincidentes (tipo de instancia, plataforma, zona de disponibilidad). De forma predeterminada, Amazon EMR utiliza automáticamente las reservas de capacidad abiertas al aprovisionar instancias bajo demanda que coinciden con los atributos de la instancia. Si no dispone de instancias en ejecución que coincidan con los atributos de reserva de capacidad, permanecen sin utilizar hasta que lanza una instancia que coincide con sus atributos. Si no desea utilizar ninguna reserva de capacidad al lanzar el clúster, debe establecer la preferencia de reserva de capacidad como ninguna en las opciones de lanzamiento.

Sin embargo, también puede dirigir una reserva de capacidad para cargas de trabajo específicas. Esto le permite controlar de manera explícita qué instancias pueden ejecutarse en esa capacidad reservada. Para más información sobre las reservas de capacidad bajo demanda, consulte Utilizar reservas de capacidad con flotas de instancias.

Permisos de IAM necesarios para una estrategia de asignación

El Rol de servicio para Amazon EMR (rol de EMR) necesita permisos adicionales para crear un clúster que utilice la opción de estrategia de asignación para flotas de instancias de spot o bajo demanda.

Se incluyen automáticamente estos permisos en el rol de servicio predeterminado de Amazon EMR `EMR_DefaultRole` y en la política administrada de Amazon EMR `AmazonEMRServicePolicy_v2`.

Si utiliza un rol de servicio personalizado o una política administrada para su clúster, debe agregar los siguientes permisos:

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
```

```
   "Effect": "Allow",
   "Action": [
    "ec2:DeleteLaunchTemplate",
    "ec2:CreateLaunchTemplate",
    "ec2:DescribeLaunchTemplates",
    "ec2:CreateLaunchTemplateVersion",
    "ec2:CreateFleet"
   ],
   "Resource": "*"
  }
}
```

Los siguientes permisos de rol de servicio son necesarios para crear un clúster que utilice reservas de capacidad abiertas o específicas. Debe incluir estos permisos además de los permisos necesarios para utilizar la opción de estrategia de asignación.

Example Documento de política para las reservas de capacidad de los roles de servicio

Para utilizar reservas de capacidad abiertas, debe incluir los siguientes permisos adicionales.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions"
            ],
            "Resource": "*"
        }
    ]
}
```

Example

Para utilizar reservas de capacidad específicas, debe incluir los siguientes permisos adicionales.

```
{
    "Version": "2012-10-17",
    "Statement": [
```

```
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions",
                "resource-groups:ListGroupResources"
            ],
            "Resource": "*"
        }
    ]
}
```

Configuración de flotas de instancias para su clúster

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para crear un clúster con flotas de instancias mediante la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y, a continuación, seleccione Crear clúster.
3. En Configuración del clúster, elija Flotas de instancias.
4. En el caso de cada grupo de nodos, seleccione Agregar tipo de instancia y elija hasta 5 tipos de instancias para las flotas de instancias principales y básicas y hasta quince tipos de instancias para las flotas de instancias de tarea. Amazon EMR podría aprovisionar cualquier combinación de estos tipos de instancia cuando se lance el clúster.
5. En cada tipo de grupo de nodos, seleccione el menú desplegable Acciones situado junto a cada instancia para cambiar estas configuraciones:

Agregar volúmenes de EBS

Especifique los volúmenes de EBS que desee asociar al tipo de instancia una vez que Amazon EMR lo aprovisione.

Edición de capacidad ponderada

En el caso del grupo de nodos principales, cambie este valor a cualquier número de unidades que se adapte a sus aplicaciones. Se utiliza el número de núcleos virtuales de YARN por cada tipo de instancias de flota como las unidades de capacidad ponderada predeterminadas. No puede editar la capacidad ponderada del nodo principal.

Edición del precio de spot máximo

Debe especificar un precio de spot máximo para cada tipo de instancia de una flota. Puede configurar este precio como un porcentaje del precio bajo demanda o como un importe específico en dólares. Si en una zona de disponibilidad el precio de spot actual está por debajo de su precio de spot máximo, Amazon EMR aprovisiona instancias de spot. Solo tiene que pagar el precio de spot, que no es necesariamente el precio de spot máximo.

6. De manera opcional, para agregar grupos de seguridad a sus nodos, amplíe Grupos de seguridad de EC2 (firewall) en la sección Redes y seleccione el grupo de seguridad para cada tipo de nodo.
7. De manera opcional, seleccione la casilla situada junto a Aplicar estrategia de asignación si desea utilizar la opción de estrategia de asignación y seleccione la estrategia que desee especificar para las instancias de spot. No debe seleccionar esta opción si su rol de servicio de Amazon EMR no tiene los permisos necesarios. Para obtener más información, consulte Estrategia de asignación para flotas de instancias.
8. Elija cualquier otra opción que se aplique a su clúster.
9. Para lanzar el clúster, elija Crear clúster.

Old console

Para crear un clúster con flotas de instancias mediante la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.

2. Elija Create cluster.
3. En la parte superior de la ventana de la consola, seleccione Ir a las opciones avanzadas, ingrese las opciones de Configuración del software y, a continuación, seleccione Siguiente.
4. En Composición del clúster, seleccione Flotas de instancias. Al seleccionar la opción de flotas de instancias, debería ver opciones para especificar la capacidad de destino de las instancias de spot y bajo demanda en la tabla Nodos e instancias del clúster.
5. En Network (Red), introduzca un valor. Si elige una VPC para Red, seleccione una sola subred de EC2 o utilice CTRL y clic para elegir varias subredes de Amazon EC2. Las subredes que seleccione deben ser del mismo tipo (públicas o privadas). Si elige solo una, el clúster se lanza en dicha subred. Si elige un grupo, se elige la subred más adecuada del grupo cuando se lanza el clúster.

   > **Note**
   >
   > La cuenta y la región podrían darle la opción de elegir Lanzar en EC2-Classic para Red. Si elige esta opción, seleccione una o varias en EC2 Availability Zones (Zonas de disponibilidad de EC2) en lugar de EC2 Subnets (Subredes de EC2). Para más información, consulte Amazon EC2 y Amazon VPC en la Guía del usuario de Amazon EC2 para instancias de Linux.

6. En Estrategia de asignación, seleccione la casilla para aplicar las estrategias de asignación si desea utilizar la opción de estrategia de asignación. Para obtener más información, consulte Estrategia de asignación para flotas de instancias.
7. En Tipo de nodo, si desea cambiar el nombre predeterminado de una flota de instancias, haga clic en el icono de lápiz y, a continuación, introduzca un nombre fácil de recordar. Si desea eliminar la flota de instancias de Tarea, elija el icono X situado en la parte derecha de la fila Tarea.
8. Elija Agregar o quitar tipos de instancias en la flota y elija hasta cinco tipos de instancias de la lista para las flotas de instancias principales y básicas; agregue hasta quince tipos de instancias para las flotas de instancias de tarea. Amazon EMR podría elegir aprovisionar cualquier combinación de estos tipos de instancia cuando se lance el clúster.
9. Para cada tipo de instancia básica y de tarea, elija cómo quiere definir la capacidad ponderada (Cada instancia cuenta como X unidades) de esa instancia. La cantidad de núcleos virtuales de YARN para cada tipo de instancia de la flota se utiliza como unidad de capacidad ponderada predeterminada, pero puede cambiar el valor por cualquier unidad que sea adecuada para sus aplicaciones.

10. En Capacidad de destino, defina la cantidad total de instancias de spot y bajo demanda que desea por flota. EMR garantiza que las instancias de la flota cumplan con las unidades solicitadas para la capacidad de destino de spot y bajo demanda. Si no se especifican unidades de spot o bajo demanda para una flota, no se aprovisiona capacidad para esa flota.
11. Si se ha configurado una flota con una capacidad de destino para spot, puede introducir su precio de spot máximo como un porcentaje del precio bajo demanda o puede introducir una cantidad en dólares ($).
12. Para que los volúmenes de EBS se asocien al tipo de instancia cuando esta se aprovisiona, elija el icono de lápiz situado junto al almacenamiento de EBS y, a continuación, introduzca las opciones de configuración de EBS.
13. Si ha establecido un recuento instantáneo para unidades de spot, establezca las opciones avanzadas de spot de acuerdo con las siguientes directrices:

    - Tiempo de espera de aprovisionamiento: utilice esta opción para controlar lo que hace Amazon EMR cuando no puede aprovisionar instancias de spot entre los Tipos de instancias de la flota especificados. Introduzca un periodo de tiempo de espera en minutos y, a continuación, elija si desea Terminate the cluster (Terminar el clúster) o Switch to provisioning On-Demand Instances (Cambiar al aprovisionamiento de instancias bajo demanda). Si decide cambiar a instancias bajo demanda, la capacidad asignada de las instancias bajo demanda se tiene en cuenta en la capacidad de destino para las instancias de spot y Amazon EMR aprovisiona las instancias bajo demanda hasta que se completa la capacidad de destino para las instancias de spot.
14. Elija Siguiente, modifique otras opciones del clúster y, a continuación, elija Siguiente.
15. Si ha seleccionado aplicar la nueva opción de estrategia de asignación, en la configuración Opciones de seguridad, seleccione un rol de EMR y un perfil de instancia de EC2 que contengan los permisos necesarios para la opción de estrategia de asignación. De lo contrario, se producirá un error durante la creación del clúster.
16. Elija Create Cluster (Crear clúster).

AWS CLI

Para crear y lanzar un clúster con flotas de instancias con AWS CLI, siga estas directrices:

- Para crear y lanzar un clúster con flotas de instancias, utilice el comando `create-cluster` junto con parámetros `--instance-fleet`.

- Para obtener los detalles de configuración de las flotas de instancias en un clúster, utilice el comando `list-instance-fleets`.
- Para agregar varias AMI de Amazon Linux personalizadas a un clúster que esté creando, utilice la opción `CustomAmiId` con cada especificación de `InstanceType`. Puede configurar los nodos de la flota de instancias con varios tipos de instancias y varias AMI personalizadas según sus necesidades. Consulte Ejemplos: creación de un clúster con la configuración de flotas de instancia.
- Para realizar cambios en la capacidad de destino de una flota de instancias de destino, utilice el comando `modify-instance-fleet`.
- Para añadir una flota de instancias de tareas a un clúster que no dispone de una, utilice el comando `add-instance-fleet`.
- Se pueden agregar varias AMI personalizadas a la flota de instancias de tarea mediante el argumento CustomAmiId con el comando add-instance-fleet. Consulte Ejemplos: creación de un clúster con la configuración de flotas de instancia.
- Para utilizar la opción de estrategia de asignación al crear una flota de instancias, actualice el rol de servicio para incluir el ejemplo del documento de política en la siguiente sección.
- Para utilizar las opciones de reserva de capacidad al crear una flota de instancias con la estrategia de asignación bajo demanda, actualice el rol de servicio para incluir el ejemplo del documento de política en la siguiente sección.
- Las flotas de instancias se incluyen automáticamente en el rol de servicio de EMR predeterminada y en la política administrada de Amazon EMR (`EMR_DefaultRole` y `AmazonEMRServicePolicy_v2`). Si utiliza un rol de servicio personalizado o una política administrada personalizada para su clúster, debe agregar los nuevos permisos para la estrategia de asignación en la siguiente sección.

Ejemplos: creación de un clúster con la configuración de flotas de instancia

Los siguientes ejemplos muestran comandos `create-cluster` con una variedad de opciones que puede combinar.

> **Note**
>
> Si no ha creado con anterioridad el rol de servicio de Amazon EMR predeterminado y el perfil de instancia de EC2, utilice `aws emr create-default-roles` para crearlos antes de utilizar el comando `create-cluster`.

Example Ejemplo: VPC predeterminada, principal bajo demanda, básica bajo demanda con tipo de instancia único

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarge}
```

Example Ejemplo: VPC predeterminada, principal de spot, spot básica con tipo de instancia único

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetSpotCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}'] \
    InstanceFleetType=CORE,TargetSpotCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}']
```

Example Ejemplo: subred de EC2 única, principal bajo demanda, básica combinada con tipo de instancia único

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-ab12345c'] \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge}'] \
    InstanceFleetType=CORE,TargetOnDemandCapacity=2,TargetSpotCapacity=6,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=2}']
```

Example Ejemplo: intervalo de subredes de EC2, principal bajo demanda, básicas de spot con varios tipos de instancias ponderados, tiempo de espera para spot

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-
ab12345c','subnet-de67890f'] \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
```

```
InstanceTypeConfigs=['{InstanceType=m5.xlarge}'] \
    InstanceFleetType=CORE,TargetSpotCapacity=11,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}',\
'{InstanceType=m4.2xlarge,BidPrice=0.9,WeightedCapacity=5}'],\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=120,TimeoutAction=SWITCH_TO_ON
```

Example Ejemplo: intervalo de subredes de EC2, principal bajo demanda, básicas y de tarea mixtas con varios tipos de instancias ponderados, tiempo de espera para instancias de spot básicas

```
aws emr create-cluster --release-label emr-5.3.1 --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,SubnetIds=['subnet-
ab12345c','subnet-de67890f'] \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=m5.xlarg
 \
    InstanceFleetType=CORE,TargetOnDemandCapacity=8,TargetSpotCapacity=6,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}',\
'{InstanceType=m4.2xlarge,BidPrice=0.9,WeightedCapacity=5}'],\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=120,TimeoutAction=SWITCH_TO_ON
 \
    InstanceFleetType=TASK,TargetOnDemandCapacity=3,TargetSpotCapacity=3,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,WeightedCapacity=3}']
```

Example Ejemplo: spot principal, sin básicos ni de tareas, configuración de Amazon EBS, VPC predeterminada

```
aws emr create-cluster --release-label Amazon EMR 5.3.1 --service-role EMR_DefaultRole
 \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetSpotCapacity=1,\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=60,TimeoutAction=TERMINATE_CLU
\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5,\
EbsConfiguration={EbsOptimized=true,EbsBlockDeviceConfigs=[{VolumeSpecification={VolumeType=gp2
\
SizeIn GB=100}},{VolumeSpecification={VolumeType=io1,SizeInGB=100,Iop
 s=100},VolumesPerInstance=4}]}}']
```

Example Ejemplo: varias AMI personalizadas, varios tipos de instancias, principal bajo demanda, básico bajo demanda

```
aws emr create-cluster --release-label Amazon EMR 5.3.1 --service-role EMR_DefaultRole
 \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \
    InstanceFleetType=MASTER,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456},
{InstanceType=m6g.xlarge, CustomAmiId=ami-234567}'] \
    InstanceFleetType=CORE,TargetOnDemandCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456},
{InstanceType=m6g.xlarge, CustomAmiId=ami-234567}']
```

Example Ejemplo: agregar un nodo de tarea a un clúster en ejecución con varios tipos de instancias y varias AMI personalizadas

```
aws emr add-instance-fleet --cluster-id j-123456 --release-label Amazon EMR 5.3.1 \
  --service-role EMR_DefaultRole \
  --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleet \
    InstanceFleetType=Task,TargetSpotCapacity=1,\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,CustomAmiId=ami-123456}',\
'{InstanceType=m6g.xlarge,CustomAmiId=ami-234567}']
```

Example Ejemplo: utilizar un archivo de configuración JSON

Puede configurar parámetros de flota de instancias en un archivo JSON y, a continuación, hacer referencia al archivo JSON como único parámetro para las flotas de instancias. Por ejemplo, el siguiente comando hace referencia a un archivo de configuración JSON, *my-fleet-config.json*:

```
aws emr create-cluster --release-label emr-5.30.0 --service-role EMR_DefaultRole \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole \
--instance-fleets file://my-fleet-config.json
```

El archivo *my-fleet-config.json* especifica flotas de instancias principales, básicas y de tarea como se muestra en el siguiente ejemplo. La flota de instancias básica utiliza un precio de spot máximo (`BidPrice`) como porcentaje del precio bajo demanda, mientras que las flotas de instancias de tarea y principales utilizan un precio de spot máximo (BidPriceAsPercentageofOnDemandPrice) como cadena en USD.

```
[
    {
        "Name": "Masterfleet",
        "InstanceFleetType": "MASTER",
        "TargetSpotCapacity": 1,
        "LaunchSpecifications": {
            "SpotSpecification": {
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "SWITCH_TO_ON_DEMAND"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPrice": "0.89"
            }
        ]
    },
    {
        "Name": "Corefleet",
        "InstanceFleetType": "CORE",
        "TargetSpotCapacity": 1,
        "TargetOnDemandCapacity": 1,
        "LaunchSpecifications": {
          "OnDemandSpecification": {
            "AllocationStrategy": "lowest-price",
            "CapacityReservationOptions":
            {
                "UsageStrategy": "use-capacity-reservations-first",
                "CapacityReservationResourceGroupArn": "String"
            }
        },
            "SpotSpecification": {
                "AllocationStrategy": "capacity-optimized",
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "TERMINATE_CLUSTER"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPriceAsPercentageOfOnDemandPrice": 100
            }
```

```
        ]
    },
    {
        "Name": "Taskfleet",
        "InstanceFleetType": "TASK",
        "TargetSpotCapacity": 1,
        "LaunchSpecifications": {
          "OnDemandSpecification": {
            "AllocationStrategy": "lowest-price",
            "CapacityReservationOptions":
            {
                "CapacityReservationPreference": "none"
            }
        },
            "SpotSpecification": {
                "TimeoutDurationMinutes": 120,
                "TimeoutAction": "TERMINATE_CLUSTER"
            }
        },
        "InstanceTypeConfigs": [
            {
                "InstanceType": "m5.xlarge",
                "BidPrice": "0.89"
            }
        ]
    }
]
```

Modificar capacidades de destino para una flota de instancias

Utilice el comando `modify-instance-fleet` para especificar nuevas capacidades de destino para una flota de instancias. Debe especificar el ID del clúster y el ID de la flota de instancias. Utilice el comando `list-instance-fleets` para recuperar los ID de la flota de instancias.

```
aws emr modify-instance-fleet --cluster-id <cluster-id> \
  --instance-fleet \
    InstanceFleetId='<instance-fleet-id>',TargetOnDemandCapacity=1,TargetSpotCapacity=1
```

Agregar una flota de instancias de tarea a un clúster

Si un clúster tiene únicamente flotas de instancias principales y básicas, puede utilizar el comando `add-instance-fleet` para agregar una flota de instancias de tarea. Solo puede utilizar esto para añadir las flotas de instancias de tareas.

```
aws emr add-instance-fleet --cluster-id <cluster-id>
  --instance-fleet \
    InstanceFleetType=TASK,TargetSpotCapacity=1,\
LaunchSpecifications={SpotSpecification='{TimeoutDurationMinutes=20,TimeoutAction=TERMINATE_CLU
\
InstanceTypeConfigs=['{InstanceType=m5.xlarge,BidPrice=0.5}']
```

Obtener detalles de configuración de flotas de instancias en un clúster

Utilice el comando `list-instance-fleets` para obtener detalles de configuración de las flotas de instancias en un clúster. El comando toma un ID de clúster como entrada. El siguiente ejemplo muestra el comando y su salida de un clúster que contiene un grupo de instancias de tarea principal y un grupo de instancias de tarea básico. Para ver la sintaxis de respuesta completa, consulte ListInstanceFleets en la Referencia de la API de Amazon EMR.

```
list-instance-fleets --cluster-id <cluster-id>
```

```
{
    "InstanceFleets": [
        {
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1488759094.637,
                    "CreationDateTime": 1488758719.817
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "ProvisionedSpotCapacity": 6,
            "Name": "CORE",
            "InstanceFleetType": "CORE",
            "LaunchSpecifications": {
                "SpotSpecification": {
                    "TimeoutDurationMinutes": 60,
```

```
                    "TimeoutAction": "TERMINATE_CLUSTER"
                }
            },
            "ProvisionedOnDemandCapacity": 2,
            "InstanceTypeSpecifications": [
                {
                    "BidPrice": "0.5",
                    "InstanceType": "m5.xlarge",
                    "WeightedCapacity": 2
                }
            ],
            "Id": "if-1ABC2DEFGHIJ3"
        },
        {
            "Status": {
                "Timeline": {
                    "ReadyDateTime": 1488759058.598,
                    "CreationDateTime": 1488758719.811
                },
                "State": "RUNNING",
                "StateChangeReason": {
                    "Message": ""
                }
            },
            "ProvisionedSpotCapacity": 0,
            "Name": "MASTER",
            "InstanceFleetType": "MASTER",
            "ProvisionedOnDemandCapacity": 1,
            "InstanceTypeSpecifications": [
                {
                    "BidPriceAsPercentageOfOnDemandPrice": 100.0,
                    "InstanceType": "m5.xlarge",
                    "WeightedCapacity": 1
                }
            ],
           "Id": "if-2ABC4DEFGHIJ4"
        }
    ]
}
```

Utilizar reservas de capacidad con flotas de instancias

Para lanzar flotas de instancias bajo demanda con opciones de reserva de capacidad, asocie los permisos de rol de servicio adicionales necesarios para utilizar las opciones de reserva de capacidad. Dado que las opciones de reserva de capacidad deben utilizarse junto con la estrategia de asignación bajo demanda, también debe incluir los permisos necesarios para la estrategia de asignación en su rol de servicio y en su política administrada. Para obtener más información, consulte Permisos de estrategia de asignación.

Amazon EMR admite reservas de capacidad abiertas y específicas. En los temas siguientes se muestran las configuraciones de las flotas de instancias que puede utilizar con la acción `RunJobFlow` o el comando `create-cluster` para lanzar flotas de instancias mediante reservas de capacidad bajo demanda.

Uso de reservas de capacidad abiertas en la medida de lo posible

Si las instancias bajo demanda del clúster coinciden con los atributos de las reservas de capacidad abiertas (tipo de instancia, plataforma, tenencia y zona de disponibilidad) disponibles en su cuenta, las reservas de capacidad se aplican automáticamente. Sin embargo, no se garantiza que se utilicen sus reservas de capacidad. Para aprovisionar el clúster, Amazon EMR evalúa todos los grupos de instancias especificados en la solicitud de lanzamiento y utiliza el que tenga el precio más bajo y la capacidad suficiente para lanzar todos los nodos principales solicitados. Las reservas de capacidad abiertas disponibles que coincidan con el grupo de instancias se aplican automáticamente. Si las reservas de capacidad abiertas disponibles no coinciden con el grupo de instancias, no se utilizarán.

Una vez aprovisionados los nodos principales, se selecciona y se fija la zona de disponibilidad. Amazon EMR aprovisiona los nodos de tarea en grupos de instancias, empezando por los de menor precio, en la zona de disponibilidad seleccionada hasta que se aprovisionen todos los nodos de este tipo. Las reservas de capacidad abiertas disponibles que coincidan con los grupos de instancias se aplican automáticamente.

A continuación se incluyen casos de uso de la lógica de asignación de capacidad de Amazon EMR para utilizar reservas de capacidad abiertas en la medida de lo posible.

Ejemplo 1: el grupo de instancias con el precio más bajo de la solicitud de lanzamiento tiene reservas de capacidad abiertas disponibles

En este caso, Amazon EMR lanza la capacidad en el grupo de instancias de menor precio con instancias bajo demanda. Las reservas de capacidad abiertas disponibles del grupo de instancias se usan automáticamente.

On-Demand Strategy	lowest-price		
Requested Capacity	100		
Instance Type	c5.xlarge	m5.xlarge	r5.xlarge
Available Open capacity reservations	150	100	100
On-Demand Price	$	$$	$$$
Instancias aprovisionadas	100	-	-
Reserva de capacidad abierta utilizada	100	-	-
Reservas de capacidad abiertas disponibles	50	100	100

Después de lanzar la flota de instancias, puede ejecutar `describe-capacity-reservations` para ver cuántas reservas de capacidad sin utilizar quedan.

Ejemplo 2: el grupo de instancias con el precio más bajo de la solicitud de lanzamiento no tiene reservas de capacidad abiertas disponibles

En este caso, Amazon EMR lanza la capacidad en el grupo de instancias de menor precio con instancias bajo demanda. Sin embargo, sus reservas de capacidad abiertas no se utilizan.

On-Demand Strategy	lowest-price		
Requested Capacity	100		
Instance Type	c5.xlarge	m5.xlarge	r5.xlarge
Reservas de capacidad abiertas disponibles	-	-	100

On-Demand Price	$	$$	$$$
Instancias aprovisionadas	100	-	-
Reserva de capacidad abierta utilizada	-	-	-
Reservas de capacidad abiertas disponibles	-	-	100

Configuración de flotas de instancias para utilizar las reservas de capacidad abiertas en la medida de lo posible

Cuando utilice la acción `RunJobFlow` para crear un clúster basado en una flota de instancias, defina la estrategia de asignación bajo demanda en `lowest-price` y `CapacityReservationPreference` para las opciones de reserva de capacidad en `open`. Como alternativa, si deja este campo en blanco, Amazon EMR establece de forma predeterminada la preferencia de reserva de capacidad de la instancia bajo demanda en `open`.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "open"
         }
        }
    }
```

También puede usar la CLI de Amazon EMR para crear un clúster basado en una flota de instancias mediante reservas de capacidad abiertas.

```
aws emr create-cluster \
 --name 'open-ODCR-cluster' \
 --release-label emr-5.30.0 \
 --service-role EMR_DefaultRole \
 --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
```

```
 --instance-fleets
 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
   LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={CapacityReservationPreference=open}}'}
```

Donde,

- `open-ODCR-cluster` se sustituye por el nombre del clúster que utiliza reservas de capacidad abiertas.
- `subnet-22XXXX01` se sustituye por el ID de la subred.

Uso de las reservas de capacidad abiertas en primer lugar

Puede optar por anular la estrategia de asignación de precios más bajos y priorizar el uso de las reservas de capacidad abiertas disponibles al aprovisionar un clúster de Amazon EMR. En este caso, Amazon EMR evalúa todos los grupos de instancias con reservas de capacidad especificadas en la solicitud de lanzamiento y utiliza el que tenga el precio más bajo y la capacidad suficiente para lanzar todos los nodos básicos solicitados. Si ninguno de los grupos de instancias con reservas de capacidad tiene capacidad suficiente para los nodos principales solicitados, Amazon EMR recurrirá al caso en el que se hace todo lo posible que se ha descrito en el tema anterior. Es decir, Amazon EMR reevalúa todos los grupos de instancias especificados en la solicitud de lanzamiento y utiliza el que tenga el precio más bajo y la capacidad suficiente para lanzar todos los nodos básicos solicitados. Las reservas de capacidad abiertas disponibles que coincidan con el grupo de instancias se aplican automáticamente. Si las reservas de capacidad abiertas disponibles no coinciden con el grupo de instancias, no se utilizarán.

Una vez aprovisionados los nodos principales, se selecciona y se fija la zona de disponibilidad. Amazon EMR aprovisiona los nodos de tarea en grupos de instancias con reservas de capacidad, empezando por los de menor precio, en la zona de disponibilidad seleccionada hasta que se aprovisionen todos los nodos de este tipo. Amazon EMR utiliza primero las reservas de capacidad abiertas disponibles en cada grupo de instancias de la zona de disponibilidad seleccionada y, solo si es necesario, utiliza la estrategia de menor precio para aprovisionar los nodos de tarea restantes.

A continuación se incluyen casos de uso de la lógica de asignación de capacidad de Amazon EMR para utilizar primero las reservas de capacidad abiertas.

Ejemplo 1: el grupo de instancias con reservas de capacidad abiertas disponibles en la solicitud de lanzamiento tiene capacidad suficiente para los nodos principales

En este caso, Amazon EMR lanza la capacidad en el grupo de instancias con las reservas de capacidad abiertas disponibles, independientemente del precio del grupo de instancias. Como resultado, las reservas de capacidad abiertas se utilizan siempre que es posible, hasta que se aprovisionen todos los nodos principales.

On-Demand Strategy	lowest-price		
Requested Capacity	100		
Usage Strategy	use-capacity-reservations-first		
Instance Type	c5.xlarge	m5.xlarge	r5.xlarge
Available Open capacity reservations	-	-	150
On-Demand Price	$	$$	$$$
Instancias aprovisionadas	-	-	100
Reserva de capacidad abierta utilizada	-	-	100
Reservas de capacidad abiertas disponibles	-	-	50

Ejemplo 2: el grupo de instancias con reservas de capacidad abiertas disponibles en la solicitud de lanzamiento no tiene capacidad suficiente para los nodos básicos

En este caso, Amazon EMR recurre al lanzamiento de nodos principales con la estrategia de menor precio y haciendo todo lo posible por utilizar las reservas de capacidad.

On-Demand Strategy	lowest-price

Requested Capacity	100		
Usage Strategy	use-capacity-reservations-first		
Instance Type	c5.xlarge	m5.xlarge	r5.xlarge
Available Open capacity reservations	10	50	50
On-Demand Price	$	$$	$$$
Instancias aprovisionadas	100	-	-
Reserva de capacidad abierta utilizada	10	-	-
Reservas de capacidad abiertas disponibles	-	50	50

Después de lanzar la flota de instancias, puede ejecutar `describe-capacity-reservations` para ver cuántas reservas de capacidad sin utilizar quedan.

Configuración de las flotas de instancias para que utilicen las reservas de capacidad abiertas en primer lugar

Cuando utilice la acción `RunJobFlow` para crear un clúster basado en una flota de instancias, defina la estrategia de asignación bajo demanda en `lowest-price` y `UsageStrategy` para `CapacityReservationOptions` en `use-capacity-reservations-first`.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first"
         }
       }
    }
```

También puede utilizar la CLI de Amazon EMR para crear un clúster basado en una flota de instancias mediante reservas de capacidad abiertas.

```
aws emr create-cluster \
  --name 'use-CR-first-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={UsageStrategy=use-capacity-reservations-first}}'}
```

Donde,

- `use-CR-first-cluster` se sustituye por el nombre del clúster que utiliza reservas de capacidad abiertas.
- `subnet-22XXXX01` se sustituye por el ID de la subred.

Uso de reservas de capacidad específicas en primer lugar

Al aprovisionar un clúster de Amazon EMR, puede optar por anular la estrategia de asignación de precios más bajos y priorizar el uso de las reservas de capacidad específicas disponibles. En este caso, Amazon EMR evalúa todos los grupos de instancias con reservas de capacidad específicas en la solicitud de lanzamiento y elige el que tenga el precio más bajo y la capacidad suficiente para lanzar todos los nodos básicos solicitados. Si ninguno de los grupos de instancias con reservas de capacidad específicas tiene capacidad suficiente para los nodos principales, Amazon EMR recurrirá al caso en el que se hace todo lo posible que se ha descrito previamente. Es decir, Amazon EMR reevalúa todos los grupos de instancias especificados en la solicitud de lanzamiento y selecciona el que tenga el precio más bajo y la capacidad suficiente para lanzar todos los nodos básicos solicitados. Las reservas de capacidad abiertas disponibles que coincidan con el grupo de instancias se aplican automáticamente. Sin embargo, sus reservas de capacidad específicas no se utilizan.

Una vez aprovisionados los nodos principales, se selecciona y se fija la zona de disponibilidad. Amazon EMR aprovisiona los nodos de tarea en grupos de instancias con reservas de capacidad

específicas, empezando por los de menor precio, en la zona de disponibilidad seleccionada hasta que se aprovisionen todos los nodos de este tipo. Amazon EMR intenta utilizar primero las reservas de capacidad específicas disponibles en cada grupo de instancias de la zona de disponibilidad seleccionada. Luego, solo si es necesario, Amazon EMR utiliza la estrategia de menor precio para aprovisionar los nodos de tarea restantes.

A continuación se indican casos de uso de la lógica de asignación de capacidad de Amazon EMR para utilizar primero las reservas de capacidad específicas.

Ejemplo 1: el grupo de instancias con reservas de capacidad específicas disponibles en la solicitud de lanzamiento tiene capacidad suficiente para los nodos principales

En este caso, Amazon EMR lanza la capacidad en el grupo de instancias con las reservas de capacidad específicas disponibles, independientemente del precio del grupo de instancias. Como resultado, las reservas de capacidad específicas se utilizan siempre que es posible, hasta que se aprovisionen todos los nodos básicos.

On-Demand Strategy	lowest-price		
Usage Strategy	use-capacity-reservations-first		
Requested Capacity	100		
Instance Type	c5.xlarge	m5.xlarge	r5.xlarge
Available targeted capacity reservations	-	-	150
On-Demand Price	$	$$	$$$
Instancias aprovisionadas	-	-	100
Reserva de capacidad específica utilizada	-	-	100
Reservas de capacidad específicas disponibles	-	-	50

Example Ejemplo 2: el grupo de instancias con reservas de capacidad específicas disponibles en la solicitud de lanzamiento no tiene capacidad suficiente para los nodos básicos

On-Demand Strategy	lowest-price		
Requested Capacity	100		
Usage Strategy	use-capacity-reservations-first		
Instance Type	c5.xlarge	m5.xlarge	r5.xlarge
Available targeted capacity reservations	10	50	50
On-Demand Price	$	$$	$$$
Instancias aprovisionadas	100	-	-
Reservas de capacidad específicas utilizadas	-	-	-
Reservas de capacidad específicas disponibles	10	50	50

Después de lanzar la flota de instancias, puede ejecutar `describe-capacity-reservations` para ver cuántas reservas de capacidad sin utilizar quedan.

Configuración de flotas de instancias para que utilicen primero las reservas de capacidad específicas

Cuando utilice la acción `RunJobFlow` para crear un clúster basado en una flota de instancias, defina la estrategia de asignación bajo demanda en `lowest-price`, `UsageStrategy` para `CapacityReservationOptions` en `use-capacity-reservations-first` y `CapacityReservationResourceGroupArn` para `CapacityReservationOptions` en `<your resource group ARN>`. Para más información, consulte Uso de reservas de capacidad en la Guía del usuario de Amazon EC2 para instancias de Linux.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first",
            "CapacityReservationResourceGroupArn": "arn:aws:resource-groups:sa-
east-1:123456789012:group/MyCRGroup"
         }
       }
    }
```

Donde `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup` se sustituye por el ARN del grupo de recursos.

También puede utilizar la CLI de Amazon EMR para crear un clúster basado en una flota de instancias mediante reservas de capacidad específicas.

```
aws emr create-cluster \
  --name 'targeted-CR-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets
 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \
    InstanceFleetType=CORE,TargetOnDemandCapacity=100,\
InstanceTypeConfigs=['{InstanceType=c5.xlarge},{InstanceType=m5.xlarge},
{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={UsageStrategy=use-capacity-reservations-
first,CapacityReservationResourceGroupArn=arn:aws:resource-groups:sa-
east-1:123456789012:group/MyCRGroup}}'}
```

Donde,

- `targeted-CR-cluster` se sustituye por el nombre del clúster que utiliza reservas de capacidad específicas.
- `subnet-22XXXX01` se sustituye por el ID de la subred.
- `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup` se sustituye por el ARN del grupo de recursos.

Evitar el uso de reservas de capacidad abiertas disponibles

Example

Si quiere evitar utilizar inesperadamente alguna de sus reservas de capacidad abiertas al lanzar un clúster de Amazon EMR, defina la estrategia de asignación bajo demanda en `lowest-price` y `CapacityReservationPreference` para `CapacityReservationOptions` en `none`. De lo contrario, Amazon EMR establece de forma predeterminada la preferencia de reserva de capacidad de la instancia bajo demanda en `open` e intenta utilizar las reservas de capacidad abiertas disponibles en la medida de lo posible.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "none"
         }
       }
    }
```

También puede usar la CLI de Amazon EMR para crear un clúster basado en una flota de instancias sin utilizar reservas de capacidad abiertas.

```
aws emr create-cluster \
  --name 'none-CR-cluster' \
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-fleets \

 InstanceFleetType=MASTER,TargetOnDemandCapacity=1,InstanceTypeConfigs=['{InstanceType=c4.xlarg
 \

 InstanceFleetType=CORE,TargetOnDemandCapacity=100,InstanceTypeConfigs=['{InstanceType=c5.xlarg
{InstanceType=m5.xlarge},{InstanceType=r5.xlarge}'],\
LaunchSpecifications={OnDemandSpecification='{AllocationStrategy=lowest-
price,CapacityReservationOptions={CapacityReservationPreference=none}}'}
```

Donde,

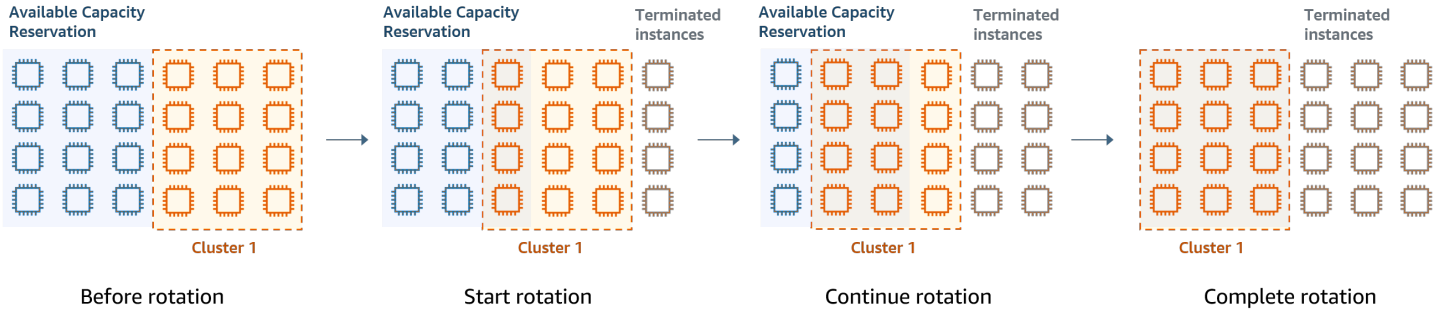
- `none-CR-cluster` se sustituye por el nombre del clúster que no utiliza reservas de capacidad abiertas.
- `subnet-22XXXX01` se sustituye por el ID de la subred.

Escenarios de uso de reservas de capacidad

Puede beneficiarse del uso de reservas de capacidad en los siguientes escenarios.

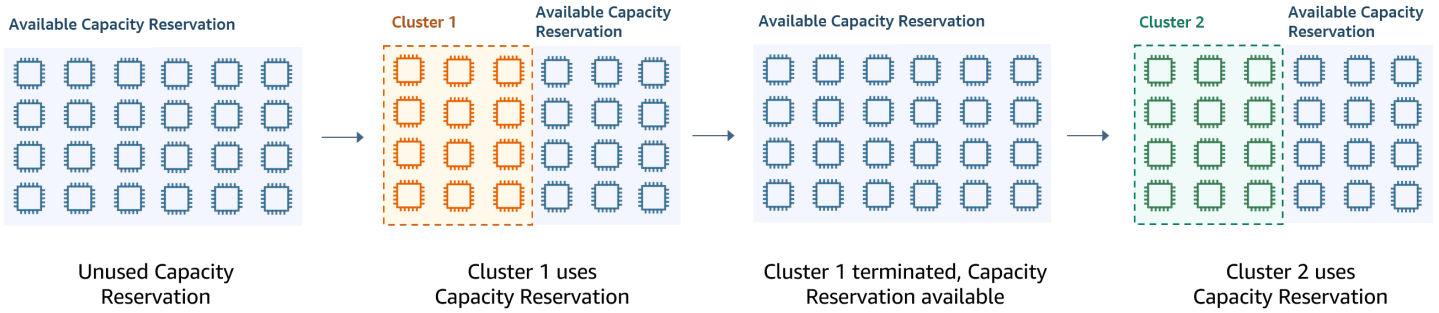
Escenario 1: rotar un clúster de ejecución prolongada con reservas de capacidad

Al rotar un clúster de ejecución prolongada, es posible que tenga requisitos estrictos en cuanto a los tipos de instancias y las zonas de disponibilidad para las nuevas instancias que aprovisione. Con las reservas de capacidad, puede utilizar la garantía de capacidad para completar la rotación del clúster sin interrupciones.



Escenario 2: aprovisionar clústeres sucesivos de corta duración con reservas de capacidad

También puede utilizar las reservas de capacidad para aprovisionar un grupo de clústeres sucesivos de corta duración para cargas de trabajo individuales, de modo que, cuando termine un clúster, el siguiente pueda utilizar las reservas de capacidad. Puede utilizar reservas de capacidad específicas para garantizar que solo los clústeres previstos utilicen las reservas de capacidad.



Configuración de grupos de instancias uniformes

Con la configuración de grupos de instancias, cada tipo de nodo (principal, secundario o de tareas) se compone del mismo tipo de instancia y las mismas opciones de compra para las instancias: bajo demanda o de spot. Estos ajustes se especifican al crear un grupo de instancias. No se pueden cambiar más adelante. Puede, sin embargo, añadir instancias del mismo tipo y opción de compra a los grupos instancias secundarias y de tareas. También puede eliminar instancias.

Si las instancias bajo demanda del clúster coinciden con los atributos de las reservas de capacidad abiertas (tipo de instancia, plataforma, tenencia y zona de disponibilidad) disponibles en su cuenta, las reservas de capacidad se aplican automáticamente. Puede usar reservas de capacidad abiertas para los nodos principales, básicos y de tarea. Sin embargo, no puede utilizar reservas de capacidad específicas ni impedir que las instancias se lancen en reservas de capacidad abiertas con atributos coincidentes al aprovisionar clústeres mediante grupos de instancias. Si quiere utilizar reservas de capacidad específicas o evitar que las instancias se lancen en reservas de capacidad abiertas, utilice flotas de instancias en su lugar. Para obtener más información, consulte Utilizar reservas de capacidad con flotas de instancias.

Para agregar distintos tipos de instancia una vez creado un clúster, puede agregar grupos de instancia de tareas adicionales. Puede elegir distintos tipos de instancia y opciones de compra para cada grupo de instancias. Para obtener más información, consulte Usar el escalado de clústeres.

Al lanzar instancias, la preferencia de reserva de capacidad de una instancia bajo demanda se configura de manera predeterminada en `open`, lo que le permite ejecutarse en cualquier reserva de capacidad abierta que tenga atributos coincidentes (tipo de instancia, plataforma, zona de disponibilidad). Para más información sobre las reservas de capacidad bajo demanda, consulte Utilizar reservas de capacidad con flotas de instancias.

En esta sección se explica cómo crear un clúster con grupos de instancias uniformes. Para obtener más información sobre la modificación de un grupo de instancias existente añadiendo o eliminando instancias de forma manual o con escalado automático, consulte Administrar clústeres.

Usar la consola para configurar grupos de instancias uniformes

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para crear un clúster con grupos de instancias mediante la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y, a continuación, seleccione Crear clúster.
3. En Configuración del clúster, elija Grupos de instancias.
4. En Grupos de nodos, hay una sección para cada tipo de grupo de nodos. Para el grupo de nodos principales, seleccione la casilla Usar varios nodos principales si desea tener 3 nodos principales. Seleccione la casilla Usar la opción de compra de spot si desea utilizar la opción de compra de spot.
5. Para los grupos de nodos principales y básicos, seleccione Agregar tipo de instancia y elija hasta 5 tipos de instancias. Para el grupo de tarea, seleccione Agregar tipo de instancia y elija hasta quince tipos de instancias. Amazon EMR podría aprovisionar cualquier combinación de estos tipos de instancia cuando se lance el clúster.
6. En cada tipo de grupo de nodos, seleccione el menú desplegable Acciones situado junto a cada instancia para cambiar estas configuraciones:

   Agregar volúmenes de EBS

   Especifique los volúmenes de EBS que desee asociar al tipo de instancia una vez que Amazon EMR lo aprovisione.

   Edición del precio de spot máximo

   Debe especificar un precio de spot máximo para cada tipo de instancia de una flota. Puede configurar este precio como un porcentaje del precio bajo demanda o como un importe específico en dólares. Si en una zona de disponibilidad el precio de spot actual está por debajo de su precio de spot máximo, Amazon EMR aprovisiona instancias de spot. Solo tiene que pagar el precio de spot, que no es necesariamente el precio de spot máximo.
7. De manera opcional, expanda Configuración de nodos para introducir una configuración de JSON o para cargar JSON desde Amazon S3.
8. Elija cualquier otra opción que se aplique a su clúster.
9. Para lanzar el clúster, elija Crear clúster.

Old console

El procedimiento siguiente trata sobre las Advanced options (Opciones avanzadas) que puede utilizar a la hora de crear un clúster. Mediante Quick options (Opciones rápidas) también se crea un clúster con la configuración de grupos de instancias.

Para crear un clúster con grupos de instancias uniformes mediante la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Go to advanced options (Ir a las opciones avanzadas), escriba las opciones de Software Configuration (Configuración de software) y, a continuación, elija Next (Siguiente).
4. En la pantalla Hardware Configuration (Configuración de hardware), deje seleccionado Uniform instance groups (Grupos de instancias uniformes).
5. Elija la Network (Red) y, a continuación, elija la EC2 Subnet (Subred de EC2) para el clúster. La subred que elija está asociada a un grupo de disponibilidad, que se indica con cada subred. Para obtener más información, consulte Configurar redes.

   > **Note**
   >
   > La cuenta y la región podrían darle la opción de elegir Lanzar en EC2-Classic para Red. Si elige dicha opción, seleccione una EC2 Availability Zone (Zona de disponibilidad de EC2) en lugar de una EC2 Subnet (Subred de EC2). Para más información, consulte Amazon EC2 y Amazon VPC en la Guía del usuario de Amazon EC2 para instancias de Linux.

6. Dentro de cada fila Node type (Tipo de nodo):

   - En Tipo de nodo, si desea cambiar el nombre predeterminado del grupo de instancias, haga clic en el icono de lápiz y, a continuación, introduzca un nombre fácil de recordar. Si desea eliminar el grupo de instancias De tareas, haga clic en el icono X. Elija Add task instance group (Añadir grupo de instancias de tareas) para añadir grupos de instancias Task (De tareas) adicionales.
   - En Tipo de instancia seleccione el icono de lápiz y, a continuación, elija el tipo de instancia que desea utilizar para dicho tipo de nodo.

> **Important**
>
> Al elegir un tipo de instancia mediante la AWS Management Console, el número de vCPU que se muestra para cada tipo de instancia es el número de núcleos virtuales de YARN para ese tipo de instancia, no el número de vCPU de EC2 para ese tipo de instancia. Para más información sobre el número de vCPU para cada tipo de instancia, consulte Tipos de instancias de Amazon EC2.

- En Tipo de instancia, haga clic en el icono del lápiz para Configuraciones y, a continuación, edite las configuraciones de aplicaciones para cada grupo de instancias.
- En Instance count (Recuento de instancias), escriba el número de instancias en que desea utilizar para cada tipo de nodo.
- En Opción de compra, elija Bajo demanda o Spot. Si elige Spot, seleccione una opción para el precio máximo de las instancias de spot. De forma predeterminada, está seleccionado Usar bajo demanda como precio máximo. Puede seleccionar Set max $/h (Establecer $/h máx.) y, a continuación, introducir el precio máximo. La zona de disponibilidad de la EC2 Subnet (Subred de EC2) que elija es inferior al Maximum Spot price (Precio de spot máximo).

> **Tip**
>
> Deténgase en la información contextual de Spot para ver el precio de spot actual de las zonas de disponibilidad de la región actual. El menor precio de spot se muestra en verde. Es posible que desee utilizar esta información para cambiar su selección de EC2 Subnet (Subred de EC2).

- En Auto Scaling for Core and Task node types (Escalado automático para los tipos de nodos secundarios y de tareas), haga clic en el icono de lápiz y, a continuación, configure las opciones de escalado automático. Para obtener más información, consulte Uso del escalado automático con una política personalizada para grupos de instancias .

7. Elija Add task instance group (Añadir grupo de instancias de tareas) como desee y configure los ajustes como se describe en el paso anterior.
8. Elija Next (Siguiente), modifique otras opciones del clúster y láncelo.

Usar la AWS CLI para crear un clúster con grupos de instancias uniformes

Para especificar la configuración de grupos de instancias para un clúster con la AWS CLI, utilice el comando `create-cluster` junto con el parámetro `--instance-groups`. Amazon EMR supone la opción de instancia bajo demanda, a menos que especifique el argumento `BidPrice` para un grupo de instancias. Para obtener ejemplos de comandos `create-cluster` que lanzan grupos de instancias uniformes con instancias bajo demanda y diversas opciones de clúster, escriba `aws emr create-cluster help` en la línea de comando o consulte create-cluster en la referencia de comandos de la AWS CLI.

Puede utilizar la AWS CLI para crear grupos de instancias uniformes en un clúster que utilice instancias de spot. El precio de spot que se ofrece depende de la zona de disponibilidad. Cuando se utiliza la CLI o la API, puede especificar la zona de disponibilidad bien con el argumento `AvailabilityZone` (si está utilizando una red EC2-Classic) o el argumento `SubnetID` del parámetro `--ec2-attributes` . La zona de disponibilidad o la subred que seleccione se aplica al clúster, por lo que se utiliza para todos los grupos de instancias. Si no especifica ninguna zona de disponibilidad o subred de forma explícita, Amazon EMR selecciona la zona de disponibilidad con el menor precio de spot al lanzar el clúster.

El siguiente ejemplo muestra un comando `create-cluster` que crea grupos de instancias principales, básicas y de tarea que utilizan instancias de spot. Sustituya *myKey* por el nombre de su par de claves de Amazon EC2.

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name "MySpotCluster" \
  --release-label emr-5.36.1 \
  --use-default-roles \
  --ec2-attributes KeyName=myKey \
  --instance-groups \
    InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,BidPrice=0.25 \
    InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=2,BidPrice=0.03 \
    InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=4,BidPrice=0.03 \
    InstanceGroupType=TASK,InstanceType=m5.xlarge,InstanceCount=2,BidPrice=0.04
```

Con la CLI, puede crear clústeres de grupos de instancias uniformes que especifiquen una AMI personalizada única para cada tipo de instancia del grupo de instancias. Esto le permite utilizar diferentes arquitecturas de instancias en el mismo grupo de instancias. Cada tipo de instancia debe usar una AMI personalizada con una arquitectura coincidente. Por ejemplo, se configuraría un tipo de instancia m5.xlarge con una AMI personalizada de arquitectura x86_64 y un tipo de instancia m6g.xlarge con la AMI personalizada de arquitectura `AWS  AARCH64` correspondiente (ARM).

En el siguiente ejemplo, se muestra un clúster de grupos de instancias uniforme creado con dos tipos de instancias, cada uno con su propia AMI personalizada. Tenga en cuenta que las AMI personalizadas se especifican solo en el nivel de tipo de instancia, no en el nivel de clúster. El objetivo es evitar conflictos entre las AMI de tipo de instancia y una AMI en el nivel de clúster, lo que provocaría un error en el lanzamiento del clúster.

```
aws emr create-cluster
  --release-label emr-5.30.0 \
  --service-role EMR_DefaultRole \
  --ec2-attributes SubnetId=subnet-22XXXX01,InstanceProfile=EMR_EC2_DefaultRole \
  --instance-groups \

 InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
 \

 InstanceGroupType=CORE,InstanceType=m6g.xlarge,InstanceCount=1,CustomAmiId=ami-234567
```

Puede agregar varias AMI personalizadas a un grupo de instancias que agrega a un clúster en ejecución. El argumento `CustomAmiId` se puede usar con el comando `add-instance-groups`, como se muestra en el siguiente ejemplo.

```
aws emr add-instance-groups --cluster-id j-123456 \
  --instance-groups \

 InstanceGroupType=Task,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
```

Uso del SDK de Java para crear un grupo de instancias

Puede instanciar un objeto `InstanceGroupConfig` que especifica la configuración de un grupo de instancias para un clúster. Para utilizar instancias de spot, define las propiedades `withBidPrice` y `withMarket` en el objeto `InstanceGroupConfig`. El código siguiente muestra cómo definir grupos de instancias principales, básicas y de tarea que ejecutan instancias de spot.

```
InstanceGroupConfig instanceGroupConfigMaster = new InstanceGroupConfig()
 .withInstanceCount(1)
 .withInstanceRole("MASTER")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.25");

InstanceGroupConfig instanceGroupConfigCore = new InstanceGroupConfig()
 .withInstanceCount(4)
 .withInstanceRole("CORE")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.03");

InstanceGroupConfig instanceGroupConfigTask = new InstanceGroupConfig()
 .withInstanceCount(2)
 .withInstanceRole("TASK")
 .withInstanceType("m4.large")
 .withMarket("SPOT")
 .withBidPrice("0.10");
```

Prácticas recomendadas para la flexibilidad de las instancias y las zonas de disponibilidad

Cada Región de AWS tiene varias ubicaciones aisladas conocidas como zonas de disponibilidad. Cuando lanza una instancia, tiene la opción de especificar una zona de disponibilidad (AZ) en la Región de AWS que está utilizando. La flexibilidad de las zonas de disponibilidad consiste en la distribución de las instancias en varias zonas de disponibilidad. Si una instancia produce un error, puede diseñar la aplicación para que una instancia en otra zona de disponibilidad pueda manejar las solicitudes. Para más información acerca de las zonas de disponibilidad, consulte la documentación sobre Regiones y zonas de disponibilidad en la Guía del usuario de Amazon EC2.

La flexibilidad de las instancias consiste en el uso de varios tipos de instancias para satisfacer los requisitos de capacidad. Al expresar la flexibilidad con las instancias, puede utilizar la capacidad agregada en todos los tamaños, familias y generaciones de instancias. Una mayor flexibilidad mejora las posibilidades de encontrar y asignar la cantidad necesaria de capacidad de computación en comparación con un clúster que utiliza un solo tipo de instancia.

La flexibilidad de las instancias y de las zonas de disponibilidad reduce los errores de capacidad insuficiente (ICE) y las interrupciones de spot en comparación con un clúster con un solo tipo de instancia o zona de disponibilidad. Utilice las prácticas recomendadas que se describen aquí para determinar qué instancias debe diversificar una vez que conozca la familia y el tamaño iniciales de

las instancias. Este enfoque maximiza la disponibilidad de los grupos de capacidad de Amazon EC2 con una variación mínima de rendimiento y costo.

Ser flexible en cuanto a las zonas de disponibilidad

Se recomienda que configure todas las zonas de disponibilidad para utilizarlas en su nube privada virtual (VPC) y que las seleccione para su clúster de EMR. Los clústeres deben existir en una sola zona de disponibilidad, pero con las flotas de instancias de Amazon EMR, puede seleccionar varias subredes para distintas zonas de disponibilidad. Cuando Amazon EMR lanza el clúster, busca en esas subredes para encontrar las instancias y las opciones de compra que especifique. Al aprovisionar un clúster de EMR para varias subredes, el clúster puede acceder a un conjunto de capacidades de Amazon EC2 más amplio en comparación con los clústeres de una sola subred.

Si debe priorizar un número determinado de zonas de disponibilidad para utilizarlas en su nube privada virtual (VPC) para su clúster de EMR, puede utilizar la capacidad de puntuación de ubicación de spot con Amazon EC2. Con la puntuación de ubicación de spot, especifica los requisitos de computación para sus instancias de spot y, a continuación, EC2 devuelve las diez Regiones de AWS o zonas de disponibilidad mejor puntuadas en una escala del 1 al 10. Una puntuación de 10 indica que es muy probable que su solicitud de spot se complete correctamente y una puntuación de 1 indica que es poco probable que esto suceda. Para más información sobre cómo utilizar la puntuación de asignación de spot, consulte Puntuación de ubicación de spot en la Guía del usuario de Amazon EC2 para instancias de Linux.

Ser flexible en cuanto a los tipos de instancias

La flexibilidad de las instancias consiste en el uso de varios tipos de instancias para satisfacer los requisitos de capacidad. La flexibilidad de las instancias beneficia al uso de las instancias de spot y bajo demanda de Amazon EC2. Con las instancias de spot, la flexibilidad de las instancias permite a Amazon EC2 lanzar instancias desde grupos de capacidad más amplios utilizando datos de capacidad en tiempo real. También predice qué instancias están más disponibles. Esto ofrece menos interrupciones y puede reducir el costo total de una carga de trabajo. Con las instancias bajo demanda, la flexibilidad de las instancias reduce los errores de capacidad insuficiente (ICE) cuando la capacidad total se aprovisiona en un mayor número de grupos de instancias.

En el caso de los clústeres de grupos de instancias, puede especificar hasta 50 tipos de instancias de EC2. En el caso de las flotas de instancias con estrategia de asignación, puede especificar hasta 30 tipos de instancias de EC2 para cada grupo de nodos principales, básicos y de tarea. Una gama más amplia de instancias mejora las ventajas de la flexibilidad de las instancias.

Expresar la flexibilidad de las instancias

Tenga en cuenta las siguientes prácticas recomendadas para expresar la flexibilidad de las instancias de su aplicación.

Temas

- Determinar la familia y el tamaño de las instancias
- Incluir instancias adicionales

Determinar la familia y el tamaño de las instancias

Amazon EMR admite varios tipos de instancias para distintos casos de uso. Estos tipos de instancias se enumeran en la documentación Tipos de instancias admitidas. Cada tipo de instancia pertenece a una familia de instancias que describe para qué aplicación se ha optimizado el tipo.

Para las cargas de trabajo nuevas, debe compararlas con los tipos de instancias de la familia de uso general, como `m5` o `c5`. Luego, supervise las métricas del sistema operativo y de YARN de Ganglia y Amazon CloudWatch para determinar los cuellos de botella del sistema en momentos de máxima carga. Los cuellos de botella incluyen la CPU, la memoria, el almacenamiento y las operaciones de E/S. Después de identificar los cuellos de botella, elija la familia de instancias optimizadas para la computación, optimizadas para memoria, optimizadas para el almacenamiento u otra familia de instancias adecuada para sus tipos de instancias. Para más información, consulte la página Determine right infrastructure for your Spark workloads en la guía de prácticas recomendadas de Amazon EMR en GitHub.

A continuación, identifique el contenedor de YARN o el ejecutor de Spark más pequeño que requiere su aplicación. Es el tamaño de instancia más pequeño que cabe en el contenedor y el tamaño mínimo de instancia para el clúster. Utilice esta métrica para determinar las instancias con las que puede diversificar aún más. Una instancia más pequeña permitirá una mayor flexibilidad de instancias.

Para obtener la máxima flexibilidad de instancias, debe utilizar tantas instancias como sea posible. Se recomienda que diversifique con instancias que tengan especificaciones de hardware similares. Esto maximiza el acceso de los grupos de capacidad de EC2 con una variación mínima de rendimiento y costo. Diversifique en todos los tamaños. Para ello, priorice primero a AWS Graviton y a las generaciones anteriores. Como regla general, intente ser flexible con al menos 15 tipos de instancias para cada carga de trabajo. Se recomienda que comience con instancias de uso general,

optimizadas para la computación u optimizadas para memoria. Estos tipos de instancias ofrecerán la mayor flexibilidad.

Incluir instancias adicionales

Para obtener la máxima diversidad, incluya tipos de instancias adicionales. Priorice primero la flexibilidad de tamaño de la instancia, Graviton y generación. Esto permite acceder a grupos de capacidad de EC2 adicionales con perfiles de costo y rendimiento similares. Si necesita más flexibilidad debido a los ICE o a interrupciones de spot, considere la flexibilidad de variantes y familias. Cada enfoque tiene sus ventajas y desventajas que dependen de su caso de uso y sus requisitos.

- Flexibilidad de tamaño: en primer lugar, diversifique con instancias de diferentes tamaños dentro de la misma familia. Las instancias de la misma familia ofrecen el mismo costo y rendimiento, pero pueden lanzar una cantidad diferente de contenedores en cada host. Por ejemplo, si el tamaño mínimo de ejecutor que necesita es de 2 vCPU y 8 GB de memoria, el tamaño mínimo de instancia es `m5.xlarge`. Para mayor flexibilidad de tamaño, incluya `m5.xlarge`, `m5.2xlarge`, `m5.4xlarge`, `m5.8xlarge`, `m5.12xlarge`, `m5.16xlarge` y `m5.24xlarge`.
- Flexibilidad de Graviton: además del tamaño, puede diversificar con las instancias de Graviton. Las instancias de Graviton funcionan con procesadores de AWS Graviton2 que ofrecen la mejor relación precio-rendimiento para las cargas de trabajo en la nube de Amazon EC2. Por ejemplo, con un tamaño de instancia mínimo de `m5.xlarge`, puede incluir `m6g.xlarge`, `m6g.2xlarge`, `m6g.4xlarge`, `m6g.8xlarge` y `m6g.16xlarge` para obtener flexibilidad de Graviton.
- Flexibilidad de generación: al igual que la flexibilidad de Graviton y de tamaño, las instancias de las familias de generaciones anteriores comparten las mismas especificaciones de hardware. Esto se traduce en un perfil de costo y rendimiento similar, con un aumento en el total del grupo de Amazon EC2 accesible. Para mayor flexibilidad de generación, incluya `m4.xlarge`, `m4.2xlarge`, `m4.10xlarge` y `m4.16xlarge`.
- Flexibilidad de familias y variantes
  - Capacidad: para optimizar la capacidad, se recomienda la flexibilidad de las instancias en todas las familias de instancias. Las instancias comunes de diferentes familias de instancias tienen grupos de instancias más amplios que pueden ayudar a cumplir los requisitos de capacidad. Sin embargo, las instancias de distintas familias tendrán diferentes proporciones de vCPU y memoria. Esto se traduce en una infrautilización si el tamaño del contenedor de aplicaciones esperado es adecuado para una instancia diferente. Por ejemplo, con `m5.xlarge`, incluya instancias optimizadas para la computación, como `c5`, o instancias optimizadas para memoria, como `r5`, para obtener flexibilidad de familias de instancias.

- Costo: para optimizar el costo, se recomienda la flexibilidad de las instancias en todas las variantes. Estas instancias tienen la misma proporción de memoria y vCPU que la instancia inicial. La diferencia con la flexibilidad de variantes es que estas instancias tienen grupos de capacidad más pequeños, lo que puede dar como resultado una capacidad adicional limitada o un mayor número de interrupciones de spot. Con `m5.xlarge` como ejemplo, incluya instancias basadas en AMD (`m5a`), instancias basadas en SSD (`m5d`) o instancias optimizadas para la red (`m5n`) para obtener flexibilidad de variantes de instancias.

Prácticas recomendadas para la configuración del clúster

Utilice la orientación de esta sección como ayuda para determinar los tipos de instancias, las opciones de compra y la cantidad de almacenamiento para aprovisionar cada tipo de nodo en un clúster de EMR.

¿Qué tipo de instancia debería utilizar?

Hay varias formas de agregar instancias de Amazon EC2 a su clúster. El método que debe elegir depende de si utiliza la configuración de grupos de instancias o la configuración de flotas de instancias para el clúster.

- Grupos de instancias
  - Añada manualmente instancias del mismo tipo a los grupos de instancias de tareas y secundarias existentes.
  - Añada manualmente un grupo de instancias de tareas, que pueden utilizar un tipo de instancia diferente.
  - Configure el escalado automático en Amazon EMR para un grupo de instancias, agregando y eliminando instancias automáticamente en función del valor de la métrica de Amazon CloudWatch que especifique. Para obtener más información, consulte Usar el escalado de clústeres.
- Flotas de instancias
  - Añadir una única flota de instancias de tarea.
  - Cambiar la capacidad de destino de las instancias bajo demanda y de spot para las flotas de instancias secundarias y de tareas. Para obtener más información, consulte Configurar flotas de instancias.

Una forma de planificar las instancias del clúster consiste en ejecutar un clúster de prueba con un conjunto representativo de datos de ejemplo y monitorizar la utilización de los nodos del clúster. Para obtener más información, consulte Ver y monitorizar un clúster. Otra forma consiste en calcular la capacidad de las instancias que se estén planeando y comparar dicho valor con el tamaño de los datos.

En general, el tipo de nodo principal, que asigna tareas, no requiere una instancia de EC2 con una gran potencia de procesamiento; las instancias de Amazon EC2 para el tipo de nodo básico, que procesan tareas y almacenan datos en HDFS, necesitan tanto potencia de procesamiento como capacidad de almacenamiento; las instancias de Amazon EC2 para el tipo de nodo de tarea, que no almacenan datos, solo necesitan potencia de procesamiento. Para conocer directrices sobre las instancias de Amazon EC2 disponibles y su configuración, consulte Configuración de instancias de Amazon EC2.

Las siguientes directrices se aplican a la mayoría de los clústeres de Amazon EMR.

- Hay un límite de vCPU para el número total de instancias de Amazon EC2 bajo demanda que se ejecutan en una cuenta de AWS por Región de AWS. Para más información sobre el límite de vCPU y cómo solicitar un aumento del límite para su cuenta, consulte Instancias bajo demanda en la Guía del usuario de Amazon EC2 para instancias de Linux.
- Generalmente, el nodo principal no tiene grandes requisitos informáticos. Para los clústeres con una gran cantidad de nodos o para los clústeres con aplicaciones que se implementan específicamente en el nodo principal (JupyterHub, Hue, etc.), es posible que se requiera un nodo principal más grande que ayude a mejorar el rendimiento del clúster. Por ejemplo, considere la posibilidad de utilizar una instancia m5.xlarge para clústeres pequeños (50 nodos o menos) y aumentarla a un tipo de instancia más grande para clústeres más grandes.
- Las necesidades informáticas de los nodos secundarios y de tareas dependen del tipo de procesamiento que realiza la aplicación. Muchos trabajos se pueden ejecutar en tipos de instancia de uso general, que ofrecen un rendimiento equilibrado en términos de CPU, espacio en disco y entrada/salida. Los clústeres con cálculo intensivo podrían beneficiarse de ejecutarse en instancias de uso elevado de CPU, que tienen en proporción más CPU que RAM. Las aplicaciones de base de datos y almacenamiento en memoria caché podrían beneficiarse al ejecutarse en instancias de memoria elevada. Las aplicaciones con uso intensivo de red y uso intensivo de CPU como análisis, NLP y machine learning pueden beneficiarse de las instancias de computación en clúster, que ofrecen proporcionalmente recursos de CPU elevada y mayor rendimiento de red.

- Si distintas fases del clúster tienen diferentes necesidades de capacidad, puede empezar con un pequeño número de nodos secundarios y aumentar o reducir el número de nodos de tareas para satisfacer los requisitos de capacidad variable del flujo de trabajo.
- La cantidad de datos que puede procesar depende de la capacidad de los nodos secundarios y del tamaño de los datos como entrada, durante el procesamiento y como salida. Los conjuntos de datos entrantes, intermedios y salientes residen en el clúster durante el procesamiento.

¿Cuándo se deben utilizar las instancias de spot?

Al lanzar un clúster en Amazon EMR, puede elegir lanzar instancias principales, básicas y de tarea en instancias de spot. Dado que cada tipo de grupo de instancias desempeña un papel diferente en el clúster, hay distintas consecuencias al lanzar cada tipo de nodo en instancias de spot. No se puede cambiar una opción de compra de instancias mientras se ejecuta un clúster. Para cambiar de instancias bajo demanda a instancias de spot o viceversa para el nodo principal y los nodos básicos, debe terminar el clúster y lanzar uno nuevo. Para los nodos de tareas, puede lanzar un nuevo grupo de instancias de tareas o una nueva flota de instancias y eliminar la anterior.

Temas

- Configuración de Amazon EMR para evitar errores en los trabajos debido a la terminación de instancias de spot de los nodos de tarea
- Nodo principal en una instancia de spot
- Nodos básicos en instancias de spot
- Nodos de tarea en instancias de spot
- Configuraciones de instancias para escenarios de aplicaciones

Configuración de Amazon EMR para evitar errores en los trabajos debido a la terminación de instancias de spot de los nodos de tarea

Dado que las instancias de spot se utilizan a menudo para ejecutar nodos de tarea, Amazon EMR tiene una funcionalidad predeterminada para programar trabajos de YARN, de modo que los trabajos en ejecución no presenten errores cuando los nodos de tarea que se ejecutan en las instancias de spot se terminen. Para ello, Amazon EMR permite que los procesos maestros de la aplicación se ejecuten únicamente en los nodos principales. El proceso maestro de la aplicación controla los trabajos en ejecución y debe mantenerse activo durante toda la vida del trabajo.

La versión 5.19.0 y posteriores de Amazon EMR utilizan la característica integrada de etiquetas de nodo YARN para lograrlo. (Las versiones anteriores utilizaban una revisión de código). Las propiedades en las clasificaciones de configuración `yarn-site` y `capacity-scheduler` se ajustan de forma predeterminada para que capacity-scheduler y fair-scheduler de YARN utilicen las etiquetas de nodo. Amazon EMR etiqueta automáticamente los nodos principales con la etiqueta `CORE` y establece las propiedades para que los maestros de la aplicación se programen únicamente en los nodos con la etiqueta CORE. La modificación manual de las propiedades relacionadas en las clasificaciones de configuración yarn-site y capacity-scheduler o directamente en los archivos XML asociados podría interrumpir esta característica o modificar esta funcionalidad.

Amazon EMR configura las siguientes propiedades y valores de forma predeterminada. Actúe con precaución al configurar estas propiedades.

- yarn-site (yarn-site.xml) en todos los nodos
  - `yarn.node-labels.enabled: true`
  - `yarn.node-labels.am.default-node-label-expression: 'CORE'`
  - `yarn.node-labels.fs-store.root-dir: '/apps/yarn/nodelabels'`
  - `yarn.node-labels.configuration-type: 'distributed'`
- yarn-site (yarn-site.xml) en los nodos principal y básicos
  - `yarn.nodemanager.node-labels.provider: 'config'`
  - `yarn.nodemanager.node-labels.provider.configured-node-partition: 'CORE'`
- capacity-scheduler (capacity-scheduler.xml) en todos los nodos
  - `yarn.scheduler.capacity.root.accessible-node-labels: '*'`
  - `yarn.scheduler.capacity.root.accessible-node-labels.CORE.capacity: 100`
  - `yarn.scheduler.capacity.root.default.accessible-node-labels: '*'`
  - `yarn.scheduler.capacity.root.default.accessible-node-labels.CORE.capacity: 100`

> **Note**
>
> A partir de la serie de versiones 6.x de Amazon EMR, la característica de etiquetas de nodo YARN está desactivada de forma predeterminada. De forma predeterminada, los procesos principales de la aplicación se pueden ejecutar tanto en nodos básicos como en nodos de

tarea. Puede habilitar la función de etiquetas de nodo YARN configurando las siguientes propiedades:

- `yarn.node-labels.enabled: true`
- `yarn.node-labels.am.default-node-label-expression: 'CORE'`

Nodo principal en una instancia de spot

El nodo principal controla y dirige el clúster. Cuando se termina, el clúster finaliza, por lo que solo debe lanzar el nodo principal como una instancia de spot si está ejecutando un clúster que se acepta que termine de forma repentina. Podría ser el caso si está probando una aplicación nueva, tiene un clúster que guarda periódicamente datos en un almacén externo como Amazon S3 o está ejecutando un clúster donde el costo es más importante que garantizar que este se complete.

Cuando se lanza el grupo de instancias principales como instancia de spot, el clúster no se inicia hasta que se completa la solicitud de instancia de spot. Esto es algo que debe considerar a la hora de seleccionar el precio de spot máximo.

Solo puede agregar un nodo principal de instancias de spot al lanzar el clúster. No se puede agregar ni eliminar nodos principales de un clúster en ejecución.

Normalmente, solo se ejecutaría el nodo principal como instancia de spot si se ejecuta todo el clúster (todos los grupos de instancia) como instancias de spot.

Nodos básicos en instancias de spot

Los nodos secundarios procesan los datos y almacenan información mediante HDFS. La terminación de una instancia secundaria conlleva el riesgo de pérdida de datos. Por este motivo, solo debe ejecutar los nodos secundarios en instancias de spot cuando sea admisible la pérdida parcial de datos de HDFS.

Cuando se lanza el grupo de instancias básicas como instancias de spot, Amazon EMR espera hasta que se puedan aprovisionar todas las instancias básicas solicitadas antes de lanzar el grupo de instancias. En otras palabras, si solicitas seis instancias de Amazon EC2 y solo hay cinco disponibles al precio de spot máximo o a un precio inferior, el grupo de instancias no se lanzará. Amazon EMR sigue esperando hasta que las seis instancias de Amazon EC2 estén disponibles o hasta que termine el clúster. Puede cambiar el número de instancias de spot de un grupo de instancias secundario para añadir capacidad a un clúster en ejecución. Para obtener más información sobre

cómo trabajar con los grupos de instancias y cómo funcionan las instancias de spot con las flotas de instancias, consulte the section called “Configurar flotas de instancias o grupos de instancias”.

Nodos de tarea en instancias de spot

Los nodos de tareas procesan datos pero no guardan datos persistentes en HDFS. Si se terminan porque el precio de spot ha superado su precio de spot máximo, no se pierden los datos y el efecto sobre el clúster es mínimo.

Al lanzar uno o varios grupos de instancias de tarea como instancias de spot, Amazon EMR aprovisiona tantos nodos de tarea como puede con su precio de spot máximo. Esto significa que, si solicita un grupo de instancias de tarea con seis nodos y solo hay cinco instancias de spot disponibles a su precio de spot máximo o por debajo de este, Amazon EMR lanza el grupo de instancias con cinco nodos y agrega el sexto más tarde, si es posible.

El lanzamiento de grupos de instancias de tareas como instancias de spot es una forma estratégica de ampliar la capacidad del clúster minimizando los costos. Si lanza los grupos de instancias principales y básicas como instancias bajo demanda, su capacidad está garantizada para la ejecución del clúster. Puede añadir instancias de tarea a los grupos de instancias de tarea según sea necesario para gestionar los picos de tráfico o acelerar el procesamiento de datos.

Puede agregar o eliminar nodos de tareas utilizando la consola, la AWS CLI o la API. También puede añadir grupos de tareas adicionales, pero no puede quitar un grupo de tareas después de haberlo creado.

Configuraciones de instancias para escenarios de aplicaciones

La siguiente tabla es una referencia rápida para las opciones de compra de tipos de nodos y las configuraciones que suelen ser adecuadas para los distintos escenarios de aplicaciones. Haga clic en el enlace correspondiente para ver más información sobre cada escenario.

Escenario de aplicaciones	Opción de compra del nodo principal	Opción de compra de los nodos básicos	Opción de compra de nodos de tarea
Clústeres en ejecución prolongada y almacenamientos de datos	Bajo demanda	Bajo demanda o combinación de flotas de instancias	Spot o combinación de flotas de instancias

Escenario de aplicaciones	Opción de compra del nodo principal	Opción de compra de los nodos básicos	Opción de compra de nodos de tarea
Cargas de trabajo dirigidas por costos	Spot	Spot	Spot
Cargas de trabajo críticas para datos	Bajo demanda	Bajo demanda	Spot o combinación de flotas de instancias
Prueba de aplicación	Spot	Spot	Spot

Existen varios escenarios en las que las instancias de spot son útiles para ejecutar un clúster de Amazon EMR.

Clústeres en ejecución prolongada y almacenamientos de datos

Si ejecuta un clúster de Amazon EMR persistente que tiene una variación predecible de capacidad informática, como un almacenamiento de datos, puede gestionar los picos de demanda a menor costo con las instancias de spot. Puede lanzar los grupos de instancias principales y básicas como instancias bajo demanda para gestionar la capacidad normal y lanzar los grupos de instancias de tarea como instancias de spot para gestionar los requisitos de picos de carga.

Cargas de trabajo dirigidas por costos

Si está ejecutando clústeres transitorios para los que un costo inferior es más importante que el tiempo de finalización y es aceptable la pérdida parcial de trabajo, puede ejecutar todo el clúster (grupos de instancias principales, básicas y de tarea) como instancias de spot para beneficiarse de los mayores ahorros de costos.

Cargas de trabajo críticas para datos

Si está ejecutando un clúster para el que un costo inferior es más importante que el tiempo de finalización, pero no es aceptable la pérdida parcial de trabajo, lance los grupos de instancias principales y básicas como instancias bajo demanda y compleméntelas con uno o varios grupos de instancias de tarea de instancias de spot. La ejecución de los grupos de instancias principales y básicas como instancias bajo demanda garantiza que los datos se almacenen en HDFS y que el clúster se proteja de la terminación debido a fluctuaciones en el mercado de spot, además de

proporcionar un ahorro de costos que supone la ejecución de los grupos de instancias de tarea como instancias de spot.

Prueba de aplicación

Cuando se prueba una nueva aplicación para prepararla para su lanzamiento en un entorno de producción, se puede ejecutar todo el clúster (grupos de instancias principales, básicas y de tarea) como instancias de spot para reducir los costos de las pruebas.

Cálculo de la capacidad de HDFS requerida de un clúster

La cantidad de almacenamiento HDFS disponible para su clúster depende de los siguientes factores:

- El número de instancias de Amazon EC2 utilizadas para los nodos básicos.
- La capacidad del almacén de instancias de Amazon EC2 para el tipo de instancia que se utilice. Para más información acerca de los volúmenes de almacén de instancias, consulte Almacén de instancias de Amazon EC2 en la Guía del usuario de Amazon EC2 para instancias de Linux.
- El número y el tamaño de los volúmenes de Amazon EBS asociados a los nodos básicos.
- Un factor de replicación, que tiene en cuenta la cantidad de bloques de datos que se almacenan en HDFS para redundancia similar a RAID. De forma predeterminada, el factor de replicación es tres para un clúster de 10 o más nodos secundarios, dos para un clúster de 4-9 nodos secundarios y uno para un clúster de 3 nodos o menos.

Para calcular la capacidad de HDFS de un clúster, por cada nodo principal, agregue la capacidad del volumen del almacén de instancias a la capacidad de almacenamiento de Amazon EBS (si se utiliza). Multiplique el resultado por el número de nodos secundarios y, a continuación, divida el total por el factor de replicación que depende del número de nodos secundarios. Por ejemplo, un clúster con 10 nodos básicos de tipo i2.xlarge, que tiene 800 GB de almacenamiento de instancias y no tiene ningún volumen de Amazon EBS asociado, tiene un total de aproximadamente 2666 GB disponibles para HDFS (10 nodos x 800 GB ÷ 3 de factor de replicación).

Si el valor de capacidad de HDFS calculado es inferior a sus datos, puede aumentar la cantidad de almacenamiento de HDFS de las siguientes formas:

- Creación de un clúster con volúmenes de Amazon EBS adicionales o adición de grupos de instancias con volúmenes de Amazon EBS asociados a un clúster existente
- Agregar más nodos secundarios
- Selección de un tipo de instancia de Amazon EC2 con una mayor capacidad de almacenamiento

- Uso de la compresión de datos
- Cambio de la configuración de Hadoop para reducir el factor de replicación

La reducción del factor de replicación se debería utilizar con precaución, ya que reduce la redundancia de los datos de HDFS y la capacidad del clúster de recuperarse frente a bloques de HDFS perdidos o dañados.

Configurar el registro y la depuración de un clúster

Una de las cosas que decidir a la hora de planificar el clúster es la cantidad de soporte de depuración que tiene que hacer que esté disponible. Al desarrollar por primera vez la aplicación de procesamiento de datos, le recomendamos probar la aplicación en un clúster que procese un subconjunto pequeño, pero representativo, de sus datos. Al hacerlo, probablemente querrá aprovechar todas las herramientas de depuración que ofrece Amazon EMR, como el archivado de archivos de registro en Amazon S3.

Cuando haya terminado el desarrollo y tenga la aplicación de procesamiento de datos a pleno rendimiento, puede optar por reducir la depuración. Al hacerlo puede ahorrar el costo de almacenamiento del conjunto de los archivos de registro en Amazon S3 y reducir la carga de procesamiento en el clúster, ya que no es necesario escribir el estado en Amazon S3. El inconveniente, por supuesto, es que si algo va mal, tendrá menos herramientas disponibles para investigar el problema.

Archivos de registro predeterminados

De forma predeterminada, cada clúster escribe archivos de registro en el nodo principal. Se escriben en el directorio `/mnt/var/log/`. Puede acceder a ellos a través de SSH para conectarse al nodo principal tal como se describe en Conectarse al nodo principal mediante SSH.

> **Note**
>
> Si utiliza la versión 6.8.0 o anterior de Amazon EMR, los archivos de registro se guardan en Amazon S3 durante la terminación del clúster, por lo que no podrá acceder a los archivos de registro una vez que termine el nodo principal. Las versiones 6.9.0 y posteriores de Amazon EMR archivan los registros en Amazon S3 durante la reducción vertical del clúster, de modo que los archivos de registro generados en él persisten incluso después de terminar el nodo.

No tiene que habilitar nada para que los archivos de registro se escriban en el nodo principal. Este es el comportamiento predeterminado de Amazon EMR y Hadoop.

Un clúster genera varios tipos de archivos de registro, incluidos:

- Registros de paso: estos registros se generan mediante el servicio de Amazon EMR y contienen información acerca del clúster y los resultados de cada paso. Los archivos de registro se almacenan en el directorio `/mnt/var/log/hadoop/steps/` en el nodo principal. Cada paso registra sus resultados en un subdirectorio numerado independiente: `/mnt/var/log/hadoop/steps/s-`*`stepId1`*`/` para el primer paso, `/mnt/var/log/hadoop/steps/s-`*`stepId2`*`/` para el segundo paso, y así sucesivamente. Los identificadores de paso de 13 caracteres (p. ej. stepId1, stepId2) son exclusivos de un clúster.
- Registros de componentes de YARN y Hadoop: los registros de componentes asociados con Apache YARN y MapReduce, por ejemplo, se encuentran en carpetas independientes en `/mnt/var/log`. Las ubicaciones de archivo de registro para los componentes de Hadoop en `/mnt/var/log` son las siguientes: hadoop-hdfs, hadoop-mapreduce, hadoop-httpfs y hadoop-yarn. El directorio hadoop-state-pusher es para la salida del proceso state pusher de Hadoop.
- Registros de acción de arranque: si su trabajo utiliza acciones de arranque, se registran los resultados de estas acciones. Los archivos de registro se almacenan en /mnt/var/log/bootstrap-actions/ en el nodo principal. Cada acción de arranque registra sus resultados en un subdirectorio numerado independiente: `/mnt/var/log/bootstrap-actions/1/` para la primera acción de arranque, `/mnt/var/log/bootstrap-actions/2/` para la segunda acción de arranque y así sucesivamente.
- Registros de estado de instancia: estos registros proporcionan información sobre la CPU, el estado de la memoria y los subprocesos del recolector de elementos no utilizados del nodo. Los archivos de registro se almacenan en `/mnt/var/log/instance-state/` en el nodo principal.

Archivar archivos de registro en Amazon S3

> **Note**
>
> Actualmente, no puede utilizar la agregación de registros para Amazon S3 con la utilidad `yarn logs`.

Las versiones 6.9.0 y posteriores de Amazon EMR archivan los registros en Amazon S3 durante la reducción vertical del clúster, de modo que los archivos de registro generados en él persisten incluso

después de terminar el nodo. Este comportamiento se habilita de forma automática, por lo que no es necesario que haga nada para activarlo. Para las versiones 6.8.0 y anteriores de Amazon EMR, puede configurar un clúster para archivar periódicamente los archivos de registro almacenados en el nodo principal en Amazon S3. Esto garantiza que los archivos de registro estén disponibles después de que el clúster se termine, ya sea a través de un apagado normal o debido a un error. Amazon EMR archiva los archivos de registro en Amazon S3 en intervalos de 5 minutos.

Para que los archivos de registro se archiven en Amazon S3 en el caso de las versiones 6.8.0 y anteriores de Amazon EMR, debe habilitar esta característica al lanzar el clúster. Puede hacerlo utilizando la consola, la CLI o la API. De manera predeterminada, los clústeres lanzados con la consola tienen habilitado el archivado de registros. Para clústeres lanzados con la CLI o la API, el registro en Amazon S3 se debe habilitar manualmente.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para archivar los archivos de registro en Amazon S3 con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Registros de clúster, seleccione la casilla Publicar los registros específicos del clúster en Amazon S3.
4. En el campo Ubicación de Amazon S3, escriba (o busque) una ruta de Amazon S3 para almacenar sus registros. Si escribe el nombre de una carpeta que no existe en el bucket, Amazon S3 la creará.

   Cuando define este valor, Amazon EMR copia los archivos de registro de las instancias de EC2 en el clúster de Amazon S3. Esto evita que se pierdan los archivos de registro cuando el clúster finaliza y EC2 termina las instancias que alojan el clúster. Estos registros resultan útiles en la solución de problemas. Para más información, consulte Ver archivos de registro.

5. De manera opcional, seleccione la casilla Cifrar los registros específicos del clúster. A continuación, seleccione una clave de AWS KMS de la lista, escriba un ARN de clave o cree una clave nueva. Esta opción solo está disponible con la versión 5.30.0 y versiones posteriores de Amazon EMR, excluyendo la versión 6.0.0. Para utilizar esta opción, agregue permiso a AWS KMS para su perfil de instancia de EC2 y su rol de Amazon EMR. Para obtener más información, consulte Para cifrar los archivos de registro almacenados en Amazon S3 con una clave administrada por el cliente de AWS KMS.
6. Elija cualquier otra opción que se aplique a su clúster.
7. Para lanzar el clúster, elija Crear clúster.

Old console

Para archivar los archivos de registro en Amazon S3 con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Go to advanced options (Ir a las opciones avanzadas).
4. En la sección General options (Opciones generales) en el campo Logging (Registro), acepte la opción predeterminada: Enabled (Habilitada).

   Esta opción determina si Amazon EMR captura datos de registro detallados en Amazon S3. Solo puede definirlo cuando se crea el clúster. Para obtener más información, consulte Ver archivos de registro de .
5. En el campo escriba Carpeta de S3, escriba (o busque) una ruta de Amazon S3 para almacenar los registros. También puede permitir que la consola genere una ruta de Amazon S3 para usted. Si escribe el nombre de una carpeta que no exista en el bucket, se creará.

   Cuando se define este valor, Amazon EMR copia los archivos de registro de las instancias de EC2 en el clúster de Amazon S3. Esto evita que se pierdan cuando el clúster finaliza y las instancias EC2 que lo alojan terminan. Estos registros resultan útiles en la solución de problemas.

   Para más información, consulte Ver archivos de registro.

6. En el campo Cifrado de registro seleccione Cifrar registros almacenados en S3 con una clave de AWS KMS administrada por el cliente. A continuación, seleccione una clave AWS KMS de la lista o escriba un ARN de clave. También puede crear una nueva clave de AWS KMS.

   Esta opción solo está disponible con la versión 5.30.0 y versiones posteriores de Amazon EMR, excluyendo la versión 6.0.0. Para utilizar esta opción, agregue permiso a AWS KMS para su perfil de instancia de EC2 y su rol de Amazon EMR. Para obtener más información, consulte Para cifrar los archivos de registro almacenados en Amazon S3 con una clave administrada por el cliente de AWS KMS.

7. Continúe con la creación del clúster tal como se describe en Planificar y configurar clústeres.

CLI

Para archivar los archivos de registro en Amazon S3 con la AWS CLI

Para archivar los archivos de registro en Amazon S3 mediante la AWS CLI, escriba el comando `create-cluster` y especifique la ruta de registros de Amazon S3 mediante el parámetro `--log-uri`.

1. Para registrar archivos en Amazon S3, escriba el siguiente comando y sustituya *myKey* por el nombre del par de claves de EC2.

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 --log-
uri s3://DOC-EXAMPLE-BUCKET/logs --applications Name=Hadoop Name=Hive Name=Pig
 --use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge --
instance-count 3
```

2. Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos básicos. Todos los nodos utilizarán el tipo de instancia especificado en el comando.

> **Note**
>
> Si no ha creado con anterioridad el rol de servicio de Amazon EMR predeterminado y el perfil de instancia de EC2, ingrese `aws emr create-default-roles` para crearlos antes de escribir el subcomando `create-cluster`.

Para cifrar los archivos de registro almacenados en Amazon S3 con una clave administrada por el cliente de AWS KMS

Con la versión 5.30.0 y posterior de Amazon EMR (excepto Amazon EMR 6.0.0), puede cifrar los archivos de registro almacenados en Amazon S3 con una clave administrada por el cliente de AWS KMS. Para habilitar esta opción en la consola, siga los pasos descritos en Archivar archivos de registro en Amazon S3. Su perfil de instancia de Amazon EC2 y su rol de Amazon EMR deben cumplir los siguientes requisitos previos:

- El perfil de instancia de Amazon EC2 utilizado para el clúster debe tener permiso para utilizar `kms:GenerateDataKey`.
- El rol de Amazon EMR utilizado para el clúster debe tener permiso para utilizar `kms:DescribeKey`.
- El perfil de instancia de Amazon EC2 y el rol de Amazon EMR deben agregarse a la lista de usuarios de claves para la clave de AWS KMS administrada por el cliente especificada, como se demuestra en los pasos siguientes:
  1. Abra la consola de AWS Key Management Service (AWS KMS) en https://console.aws.amazon.com/kms.
  2. Para cambiar la región de AWS, utilice el Region selector (Selector de regiones) en la esquina superior derecha de la página.
  3. Seleccione el alias de la clave de KMS que desee modificar.
  4. En la página de detalles de la clave, en Key Users (Usuarios de claves), seleccione Add (Añadir).
  5. En el cuadro de diálogo Agregar usuarios clave, seleccione su perfil de instancia de Amazon EC2 y el rol de Amazon EMR.
  6. Elija Add (Agregar).

Para más información, consulte Roles de servicio de IAM usados por Amazon EMR y Uso de las políticas de claves en la guía para desarrolladores de AWS Key Management Service.

Para agregar registros en Amazon S3 con la AWS CLI

> **Note**
>
> No puede utilizar actualmente la agregación de registros con la utilidad `yarn logs`. Solo puede utilizar la agregación compatible con este procedimiento.

La agregación de registros (Hadoop 2.x) compila los registros desde todos los contenedores para una aplicación individual en un único archivo. Para habilitar la agregación de registros en Amazon S3 mediante la AWS CLI, utilice una acción de arranque al lanzar el clúster para habilitar la agregación de registros y especificar el bucket donde se deben almacenar los registros.

- Para habilitar la agregación de registros, cree el siguiente archivo de configuración llamado `myConfig.json`, que contiene lo siguiente:

```
[
  {
    "Classification": "yarn-site",
    "Properties": {
      "yarn.log-aggregation-enable": "true",
      "yarn.log-aggregation.retain-seconds": "-1",
      "yarn.nodemanager.remote-app-log-dir": "s3:\/\/DOC-EXAMPLE-BUCKET\/logs"
    }
  }
]
```

  Escriba el comando siguiente y sustituya *myKey* por el nombre del par de claves de EC2. Además, puede sustituir cualquiera de los textos en rojo por sus propias configuraciones.

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-5.36.1 \
--applications Name=Hadoop \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-type m5.xlarge \
--instance-count 3 \
--configurations file://./myConfig.json
```

Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos básicos. Todos los nodos utilizarán el tipo de instancia especificado en el comando.

> **Note**
>
> Si no ha creado con anterioridad el rol de servicio de EMR predeterminado y el perfil de instancia de EC2, ejecute `aws emr create-default-roles` para crearlos antes de ejecutar el subcomando `create-cluster`.

Para más información acerca del uso de comandos de Amazon EMR en la AWS CLI, consulte la Referencia de los comandos de la AWS CLI.

Localización de registros

La siguiente lista incluye todos los tipos de registro y sus ubicaciones en Amazon S3. Puede utilizarlos para solucionar problemas de Amazon EMR.

Registros de pasos

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/steps/<step-id>/
```

Logs de aplicaciones

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/containers/
```

Esta ubicación incluye el contenedor `stderr` y los registros `stdout`, `directory.info`, `prelaunch.out` y `launch_container.sh`.

Registros del administrador de recursos

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
applications/hadoop-yarn/
```

HDFS de Hadoop

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/
applications/hadoop-hdfs/
```

Esta ubicación incluye los registros de NameNode, DataNode y YARN TimelineServer.

Registros del administrador de nodos

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/
applications/hadoop-yarn/
```

Registros de estado de la instancia

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/daemons/
instance-state/
```

Registros de aprovisionamiento de Amazon EMR

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
provision-node/*
```

Registros de Hive

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
applications/hive/*
```

- Para encontrar los registros de Hive en su clúster, elimine el asterisco (*) y agregue `/var/log/hive/` al enlace anterior.
- Para encontrar los registros de HiveServer2, elimine el asterisco (*) y agregue `var/log/hive/hiveserver2.log` al enlace anterior.
- Para encontrar los registros de HiveCLI, elimine el asterisco (*) y agregue `/var/log/hive/user/hadoop/hive.log` al enlace anterior.
- Para encontrar los registros de Hive Metastore Server, elimine el asterisco (*) y agregue `/var/log/hive/user/hive/hive.log` al enlace anterior.

Si el error se encuentra en el nodo principal o de tarea de su aplicación de Tez, proporcione los registros del contenedor de Hadoop correspondiente.

Habilitar la herramienta de depuración

La herramienta de depuración permite examinar fácilmente los archivos de registro desde la consola de Amazon EMR. Para obtener más información, consulte Ver archivos de registro en la herramienta de depuración. Cuando se habilita la depuración en un clúster, Amazon EMR archiva los archivos de registro en Amazon S3 y, a continuación, indexa dichos archivos. A continuación, puede utilizar la consola para examinar de manera intuitiva los archivos de registro de paso, trabajo, tarea e intento de tarea del clúster.

Para utilizar la herramienta de depuración en la consola de Amazon EMR, debe habilitar la depuración al lanzar el clúster con la consola, la CLI o la API. Tenga en cuenta que la nueva consola de Amazon EMR no ofrece la herramienta de depuración.

Old console

Para activar la herramienta de depuración con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. Elija Go to advanced options (Ir a las opciones avanzadas).
4. En la sección Cluster Configuration (Configuración del clúster), en el campo Logging (Registro), elija Enabled (Habilitado). No puede habilitar la depuración sin habilitar el registro.
5. En el campo Ubicación de carpeta de registros en S3, escriba una ruta de Amazon S3 para almacenar sus registros.
6. En el campo Debugging (Depuración), elija Enabled (Habilitada). La opción de depuración crea un intercambio de Amazon SQS para publicar mensajes de depuración en el backend de servicio de Amazon EMR. Podrían aplicarse cargos por la publicación de mensajes en el intercambio. Para más información, consulte la página del producto de Amazon SQS.
7. Continúe con la creación del clúster tal como se describe en Planificar y configurar clústeres.

AWS CLI

Para activar la herramienta de depuración con la AWS CLI

Para habilitar la depuración mediante la AWS CLI, escriba el subcomando `create-cluster` con el parámetro `--enable-debugging`. Asimismo, debe especificar el parámetro `--log-uri` a la hora de habilitar la depuración.

- Para habilitar la depuración utilizando la AWS CLI, escriba el siguiente comando y sustituya *myKey* por el nombre del par de claves de EC2.

```
aws emr create-cluster --name "Test cluster" \
--release-label emr-5.36.1 \
--log-uri s3://DOC-EXAMPLE-BUCKET/logs \
--enable-debugging \
```

```
--applications Name=Hadoop Name=Hive Name=Pig \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-type m5.xlarge \
--instance-count 3
```

Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos básicos. Todos los nodos utilizarán el tipo de instancia especificado en el comando.

> **Note**
>
> Si no ha creado con anterioridad el rol de servicio de EMR predeterminado y el perfil de instancia EC2, escriba aws `emr create-default-roles` para crearlos antes de escribir el subcomando `create-cluster`.

API

Para activar la herramienta de depuración con la API de Amazon EMR

- Habilite la depuración mediante la siguiente configuración del SDK para Java.

```
   StepFactory stepFactory = new StepFactory();
StepConfig enabledebugging = new StepConfig()
    .withName("Enable debugging")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newEnableDebuggingStep());
```

En este ejemplo, `new StepFactory()` utiliza `us-east-1` como la región predeterminada. Si el clúster se lanza en otra región, debe especificarla utilizando `new StepFactory("region.elasticmapreduce")`, como, por ejemplo, `new StepFactory("ap-northeast-2.elasticmapreduce")`.

Información sobre la opción de depuración

Las versiones 4.1.0 a 5.27.0 de Amazon EMR admiten la depuración en todas las regiones. Otras versiones de Amazon EMR no admiten la opción de depuración. A partir del 23 de enero de 2023, Amazon EMR suspenderá la herramienta de depuración para todas las versiones.

Amazon EMR crea una cola de Amazon SQS para procesar los datos de depuración. Podrían aplicarse cargos de mensajes. Sin embargo, Amazon SQS tiene capa gratuita de hasta 1 000 000 solicitudes disponibles. Para obtener más información, consulte https://aws.amazon.com/https://aws.amazon.com/sqs.

La depuración requiere el uso de roles; su rol de servicio y perfil de instancia deben permitirle utilizar todas las operaciones de la API de Amazon SQS. Si sus roles se asociar a políticas administradas de Amazon EMR, no tiene que realizar ninguna operación para modificarlos. Si tiene roles personalizados, tendrá que añadir permisos `sqs:*`. Para obtener más información, consulte Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS.

Etiquetado de clústeres

Puede resultar práctico clasificar los recursos de AWS de diversas maneras; por ejemplo, según su finalidad, propietario o entorno. Puede conseguirlo en Amazon EMR asignando metadatos personalizados a los clústeres de Amazon EMR mediante etiquetas. Una etiqueta consta de una clave y un valor, ambos definidos por el usuario. En el caso de Amazon EMR, el clúster es el nivel de recursos que puede etiquetar. Por ejemplo, podría definir un conjunto de etiquetas para los clústeres de su cuenta que le ayude a realizar un seguimiento de cada propietario del clúster o identificar un clúster de producción frente a un clúster de pruebas. Le recomendamos que cree un conjunto de etiquetas coherente para satisfacer los requisitos de su organización.

Cuando se agrega una etiqueta a un clúster de Amazon EMR, la etiqueta también se propaga a cada una de las instancias de Amazon EC2 activas asociadas al clúster. Del mismo modo, si elimina una etiqueta de un clúster de Amazon EMR, dicha etiqueta se elimina de cada una de las instancias de Amazon EC2 activas asociadas.

> **Important**
>
> Utilice la consola o la CLI de Amazon EMR para administrar etiquetas en instancias de Amazon EC2 que forman parte de un clúster en lugar de la CLI o de la consola de Amazon EC2, ya que los cambios que realice en Amazon EC2 no se vuelven a sincronizar con el sistema de etiquetado de Amazon EMR.

Puede identificar una instancia de Amazon EC2 que forme parte de un clúster de Amazon EMR buscando las siguientes etiquetas del sistema. En este ejemplo, *CORE* es el valor del rol de grupo de instancias y *j-12345678* es un ejemplo de valor de identificador de flujo de trabajo (clúster):

- aws:elasticmapreduce:instance-group-role=*CORE*
- aws:elasticmapreduce:job-flow-id=*j-12345678*

> **Note**
>
> Amazon EMR y Amazon EC2 interpretan las etiquetas como una cadena de caracteres sin significado semántico.

Puede trabajar con etiquetas utilizando la AWS Management Console la CLI y la API.

Puede agregar etiquetas al crear un clúster nuevo de Amazon EMR y puede agregar, editar o eliminar etiquetas de un clúster de Amazon EMR en ejecución. Editar una etiqueta es un concepto que se aplica a la consola de Amazon EMR; sin embargo, con la CLI y la API, para editar una etiqueta, se elimina la etiqueta antigua y se agrega una nueva. Puede editar las claves y los valores de las etiquetas y también puede eliminar etiquetas de un recurso en cualquier momento que se ejecuta un clúster. Sin embargo, no puede añadir, editar o suprimir etiquetas desde un clúster terminado o instancias terminadas que se asociaron anteriormente con un clúster que sigue activo. Además, puede establecer el valor de una etiqueta como una cadena vacía, pero no puede establecer el valor de una etiqueta como nulo.

Si está utilizando AWS Identity and Access Management (IAM) con sus instancias de Amazon EC2 para permisos basados en recursos por etiqueta, las políticas de IAM se aplican a las etiquetas que Amazon EMR propaga a las instancias de Amazon EC2 de un clúster. Para que las etiquetas de Amazon EMR se propaguen a las instancias de Amazon EC2, la política de IAM para Amazon EC2 debe conceder permisos para llamar a las API CreateTags y DeleteTags de Amazon EC2. Además, las etiquetas propagadas pueden afectar a sus permisos basados en recursos de Amazon EC2. Las etiquetas propagadas a Amazon EC2 se pueden leer como condiciones en la política de IAM, como las demás etiquetas de Amazon EC2. Tenga en cuenta la política de IAM a la hora de agregar etiquetas a las clústeres de Amazon EMR para evitar que los usuarios tengan permisos incorrectos para un clúster. Para evitar problemas, asegúrese de que sus políticas de IAM no incluyan condiciones de etiquetas que también planee utilizar en sus clústeres de Amazon EMR. Para más información, consulte Control de acceso a recursos de Amazon EC2.

Restricciones de las etiquetas

Se aplican las siguientes restricciones básicas a las etiquetas:

- Las restricciones que se aplican a los recursos de Amazon EC2 se aplican también a Amazon EMR. Para obtener más información, consulte https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Using_Tags.html#tag-restrictions.
- No utilice el prefijo `aws:` en los nombres y valores de las etiquetas, porque está reservado para uso de AWS. Además, no puede editar o eliminar los nombres y valores de etiquetas que tienen este prefijo.
- No puede cambiar o editar etiquetas en un clúster terminado.
- Un valor de etiqueta puede ser una cadena vacía, pero no nulo. Además, una clave de etiqueta no puede ser una cadena vacía.
- Las claves y los valores pueden contener caracteres alfabéticos en cualquier idioma, cualquier carácter numérico, espacios en blanco, separadores invisibles y los siguientes símbolos: _ . : / = + - @

Para más información sobre el etiquetado mediante la AWS Management Console, consulte Uso de etiquetas en la consola en la Guía del usuario de Amazon EC2 para instancias de Linux. Para más información sobre el etiquetado mediante la API de Amazon EC2 o la línea de comando, consulte Información general de la API y la CLI en la Guía del usuario de Amazon EC2 para instancias de Linux.

Recursos de etiquetas para facturación

Puede utilizar etiquetas para organizar su factura de AWS para reflejar su propia estructura de costos. Para ello, inscríbase para obtener una factura de la cuenta de AWS que incluya valores de clave de etiquetas. A continuación, puede organizar su información de facturación por valores clave de etiqueta, para ver el costo de los recursos combinados. Aunque Amazon EMR y Amazon EC2 tienen facturaciones distintas, las etiquetas de cada clúster también se colocan en cada instancia asociada a fin de que pueda utilizar las etiquetas para enlazar los costos relacionados de Amazon EMR y Amazon EC2.

Por ejemplo, puede etiquetar varios recursos con un nombre de aplicación específico y luego organizar su información de facturación para ver el costo total de la aplicación en distintos servicios. Para más información, consulte Etiquetado y asignación de costos en la Guía del usuario de AWS Billing.

Agregar etiquetas a un clúster

También puede agregar etiquetas a un clúster al momento de crearlo.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para agregar etiquetas al crear un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Etiquetas, elija Agregar etiqueta nueva. Especifique una etiqueta en el campo Clave. Si lo desea, especifique una etiqueta en el campo Valor.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para agregar etiquetas al crear un clúster con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Crear clúster e Ir a las opciones avanzadas.
3. En la página Step 3: General Cluster Settings (Paso 3: Configuración general del clúster), en la sección Tags (Etiquetas), escriba un valor en Key (Clave) para la etiqueta.

   Cuando empiece a escribir el valor en Key (Clave), aparece automáticamente una nueva fila para dejar espacio a la etiqueta nueva siguiente.

4. Si lo desea, escriba un Value (Valor) para la etiqueta.
5. Repita los pasos anteriores para cada par clave/valor de etiqueta que añadir al clúster. Cuando se lanza el clúster, las etiquetas que introduzca se asocian automáticamente al clúster.

AWS CLI

Para agregar etiquetas al crear un clúster con la AWS CLI

El siguiente ejemplo muestra cómo añadir una etiqueta a un nuevo clúster con la AWS CLI. Para añadir etiquetas al crear un clúster, escriba el subcomando `create-cluster` con el parámetro `--tags`.

- Para añadir una etiqueta denominada *costCenter* con el valor de clave *marketing* al crear un clúster, escriba el siguiente comando y sustituya *myKey* por el nombre del par de claves de EC2.

```
aws emr create-cluster --name "Test cluster" --release-label emr-4.0.0 --
applications Name=Hadoop Name=Hive Name=Pig --tags "costCenter=marketing" --
use-default-roles --ec2-attributes KeyName=myKey --instance-type m5.xlarge --
instance-count 3
```

Cuando especifica el recuento de instancias sin utilizar el parámetro `--instance-groups`, se lanza un nodo principal único y el resto de las instancias se lanzan como nodos secundarios. Todos los nodos utilizarán el tipo de instancia especificado en el comando.

> **Note**
>
> Si no ha creado con anterioridad el rol de servicio de EMR predeterminado y el perfil de instancia EC2, escriba aws `emr create-default-roles` para crearlos antes de escribir el subcomando `create-cluster`.

Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

También puede añadir etiquetas a un clúster existente.

New console

Para agregar etiquetas a un clúster existente utilizando la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar.
3. En la pestaña Etiquetas de la página de detalles del clúster, seleccione Administrar etiquetas. Especifique una etiqueta en el campo Clave. Si lo desea, especifique una etiqueta en el campo Valor.
4. Seleccione Save changes (Guardar cambios). La pestaña Etiquetas se actualiza con el nuevo número de etiquetas que tiene en el clúster. Por ejemplo, si ahora tiene dos etiquetas, la etiqueta de su pestaña será Etiquetas (2).

Old console

Para agregar etiquetas a un clúster existente con la consola antigua

1. En la consola de Amazon EMR, seleccione la página Lista de clústeres y haga clic en un clúster al que desee agregar etiquetas.
2. En la página Cluster Details (Detalles del clúster), en el campo Tags (Etiquetas), haga clic en View All/Edit (Ver todo/Editar).
3. En la página View All/Edit (Ver todo/Editar), haga clic en Add (Añadir).
4. Haga clic en el campo vacío en la columna Key (Clave) y escriba el nombre de la clave.
5. Opcionalmente, haga clic en el campo vacío en la columna Value (Valor) y escriba el valor.
6. Con cada nueva etiqueta que comience, aparece otra línea de etiqueta vacía en la etiqueta que está editando. Repita los pasos anteriores en la nueva línea de etiqueta para cada etiqueta que desee añadir.

AWS CLI

Para agregar etiquetas a un clúster en ejecución utilizando la AWS CLI

- Escriba el subcomando `add-tags` con el parámetro `--tag` para asignar etiquetas a un ID de clúster. Puede encontrar el ID del clúster mediante la consola o el comando `list-clusters`. El subcomando `add-tags` actualmente solo acepta un ID de recurso.

  Para agregar dos etiquetas a un clúster en ejecución, una con una clave denominada *`costCenter`* con un valor de *`marketing`* y otra llamada *`otra`* con valor de *`contabilidad`*, escriba el siguiente comando y sustituya *`j-KT4XXXXXXXX1NM`* por su ID de clúster.

  ```
  aws emr add-tags --resource-id j-KT4XXXXXXXX1NM --tag "costCenter=marketing" --
  tag "other=accounting"
  ```

  Cuando se agregan etiquetas utilizando la CLI de AWS, no hay ninguna salida del comando. Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Ver etiquetas en un clúster

Si desea ver todas las etiquetas asociadas a un clúster, puede verlas en la consola o en la AWS CLI.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para ver las etiquetas en un clúster mediante la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar.

3. Para ver todas las etiquetas, seleccione la pestaña Etiquetas en la página de detalles del clúster.

Old console

Para ver etiquetas en un clúster mediante la consola antigua

1. En la consola de Amazon EMR, seleccione la página Lista de clústeres y haga clic en un clúster para ver las etiquetas.
2. En la página Cluster Detailsse (Detalles del clúster), en el campo Tags (Etiquetas), se muestran algunas etiquetas. Haga clic en View All/Edit (Ver todo/Editar) para mostrar todas las etiquetas disponibles en el clúster.

AWS CLI

Para ver etiquetas en un clúster mediante la AWS CLI

Para ver las etiquetas en un clúster mediante la AWS CLI, escriba el subcomando `describe-cluster` con el parámetro `--query`.

- Para ver las etiquetas de un clúster, escriba el siguiente comando y sustituya *`j-KT4XXXXXXXX1NM`* por el ID del clúster.

```
aws emr describe-cluster --cluster-id j-KT4XXXXXX1NM --query Cluster.Tags
```

La salida muestra toda la información de etiquetas sobre el clúster similar a la siguiente:

```
Value: accounting     Value: marketing
Key: other            Key: costCenter
```

Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Eliminar etiquetas de un clúster

Si ya no necesita una etiqueta, puede eliminarla del clúster.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para eliminar etiquetas de un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar.
3. En la pestaña Etiquetas de la página de detalles del clúster, seleccione Administrar etiquetas.
4. Seleccione Eliminar en cada par de clave-valor que desee eliminar.
5. Elija Save changes (Guardar cambios).

Old console

Para eliminar etiquetas de un clúster con la consola antigua

1. En la consola de Amazon EMR, seleccione la página Lista de clústeres y haga clic en un clúster del que desee eliminar las etiquetas.
2. En la página Cluster Details (Detalles del clúster), en el campo Tags (Etiquetas), haga clic en View All/Edit (Ver todo/Editar).
3. En el cuadro de diálogo View All/Edit (Ver todo/Editar), haga clic en el icono X situado junto a la etiqueta que desea eliminar y haga clic en Save (Guardar).
4. (Opcional) Repita el paso anterior para cada par clave-valor de etiqueta que desee eliminar del clúster.

AWS CLI

Para eliminar etiquetas en un clúster mediante la AWS CLI

Escriba el subcomando `remove-tags` con el parámetro `--tag-keys`. Al eliminar una etiqueta, solo se necesita el nombre de clave.

- Para eliminar una etiqueta de un clúster, escriba el siguiente comando y sustituya *j-KT4XXXXXXXX1NM* por el ID del clúster.

```
aws emr remove-tags --resource-id j-KT4XXXXXX1NM --tag-keys "costCenter"
```

> **Note**
>
> Actualmente no puede eliminar varias etiquetas con un único comando.

Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Integración de controladores y aplicaciones de terceros

Puede ejecutar varias aplicaciones de macrodatos populares en Amazon EMR con los precios de servicios públicos. Esto significa que paga una tarifa nominal adicional por hora para la aplicación de terceros, mientras el clúster está en ejecución. Le permite usar la aplicación sin tener que adquirir una licencia anual. En las siguientes secciones se describen algunas de las herramientas que puede utilizar con EMR.

Temas

- Utilizar herramientas de inteligencia empresarial con Amazon EMR

Utilizar herramientas de inteligencia empresarial con Amazon EMR

Puede utilizar herramientas de inteligencia empresarial populares como Microsoft Excel, MicroStrategy, QlikView y Tableau con Amazon EMR para explorar y visualizar los datos. Muchas de estas herramientas requieren un controlador ODBC (Open Database Connectivity) o JDBC (Java

Database Connectivity). Para descargar e instalar los controladores más recientes, consulte http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/.

Para buscar versiones anteriores de los controladores, consulte http://awssupportdatasvcs.com/bootstrap-actions/Simba/.

Seguridad en Amazon EMR

La seguridad en la nube de AWS es la mayor prioridad. Como cliente de AWS, se beneficia de una arquitectura de red y un centro de datos que se han diseñado para satisfacer los requisitos de seguridad de las organizaciones más exigentes.

La seguridad es una responsabilidad compartida entre AWS y usted. El modelo de responsabilidad compartida la describe como seguridad de la nube y seguridad en la nube:

- Seguridad de la nube: AWS es responsable de proteger la infraestructura que ejecuta los servicios de AWS en la nube de AWS. AWS también proporciona servicios que puede utilizar de forma segura. Auditores externos prueban y verifican periódicamente la eficacia de nuestra seguridad en el marco de los programas de conformidad de AWS. Para obtener más información sobre los programas de conformidad que se aplican a Amazon EMR, consulte Servicios de AWS en el ámbito del programa de conformidad.
- Seguridad en la nube: su responsabilidad viene determinada por el servicio de AWS que utilice. También es responsable de otros factores, incluida la confidencialidad de los datos, los requisitos de la empresa y la legislación y los reglamentos vigentes.

Esta documentación le permite comprender cómo aplicar el modelo de responsabilidad compartida cuando se utiliza Amazon EMR. Cuando desarrolle soluciones en Amazon EMR, utilice las siguientes tecnologías para poder proteger datos y recursos de clúster de acuerdo con sus requisitos empresariales. Los temas de este capítulo muestran cómo configurar Amazon EMR y utilizar otros productos de AWS para satisfacer sus necesidades de seguridad y objetivos de conformidad.

Configuraciones de seguridad

Las configuraciones de seguridad de Amazon EMR son plantillas que contienen diferentes ajustes de seguridad. Puede crear una configuración de seguridad para volver a utilizar ciertos ajustes de seguridad cada vez que cree un clúster. Para obtener más información, consulte Uso de configuraciones de seguridad para configurar la seguridad del clúster.

Protección de los datos

Puede implementar el cifrado de datos para proteger los datos en reposo en Amazon S3, los datos en reposo en el almacenamiento de las instancias del clúster y los datos en tránsito. Para obtener más información, consulte Cifrado de datos en reposo y en tránsito.

AWS Identity and Access Management con Amazon EMR

AWS Identity and Access Management (IAM) es un servicio de AWS que ayuda al administrador a controlar de forma segura el acceso a los recursos de AWS. Los administradores de IAM controlan quién se puede autenticar (iniciar sesión) y autorizar (tener permisos) para utilizar los recursos de Amazon EMR. IAM es un servicio de AWS que puede utilizar sin cargo adicional.

- Políticas basadas en identidades de IAM: las políticas de IAM conceden o deniegan permisos a los usuarios y grupos para realizar acciones. Las políticas se pueden combinar con las etiquetas para controlar el acceso en un clúster a clúster. Para obtener más información, consulte AWS Identity and Access Management para Amazon EMR.
- Roles de IAM: el rol de servicio, el perfil de instancia y el rol basado en servicio de Amazon EMR controlan cómo Amazon EMR es capaz de tener acceso a otros productos de AWS. Para obtener más información, consulte Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS.
- Roles de IAM para las solicitudes de EMRFS a Amazon S3: cuando Amazon EMR accede a Amazon S3, puede especificar el rol de IAM que se va a utilizar en función del usuario, el grupo o la ubicación de los datos de EMRFS en Amazon S3. Esto le permite controlar con precisión si los usuarios del clúster pueden acceder a los archivos desde dentro de Amazon EMR. Para obtener más información, consulte Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3.

Kerberos

Puede configurar Kerberos para proporcionar una autenticación robusta a través de la criptografía de clave secreta. Para obtener más información, consulte Uso de Kerberos para la autenticación con Amazon EMR.

Lake Formation

Puede utilizar los permisos de Lake Formation junto con Catálogo de datos de AWS Glue para proporcionar un acceso de nivel de columna detallado a las bases de datos y a las tablas en Catálogo de datos de AWS Glue. Lake Formation permite un inicio de sesión único federado en blocs de notas de EMR o Apache Zeppelin desde un sistema de identidades de la compañía. Para obtener más información, consulte Integración de Amazon EMR con AWS Lake Formation.

Secure Socket Shell (SSH)

SSH ayuda a proporcionar un método seguro a los usuarios para conectarse a la línea de comandos de las instancias de clúster. También proporciona tunelización para ver las interfaces web que las aplicaciones alojan en el nodo principal. Los clientes pueden autenticar mediante Kerberos o un par de claves de Amazon EC2. Para obtener más información, consulte Uso de un par de claves de EC2 para credenciales de SSH y Conectarse a un clúster.

Grupos de seguridad de Amazon EC2

Los grupos de seguridad actúan como un firewall virtual para las instancias de clúster de EMR y controlan el tráfico entrante y saliente. Para obtener más información, consulte Control del tráfico de red con grupos de seguridad.

Actualizaciones de la AMI de Amazon Linux predeterminada para Amazon EMR

Important

Los clústeres de Amazon EMR que ejecutan las AMI (imágenes de máquina de Amazon) de Amazon Linux o Amazon Linux 2 utilizan el comportamiento predeterminado de Amazon Linux y no descargan ni instalan automáticamente actualizaciones importantes y críticas del kernel que requieren un reinicio. Este comportamiento es el mismo que el de otras instancias de Amazon EC2 que ejecutan la AMI de Amazon Linux predeterminada. Si aparecen nuevas actualizaciones de software de Amazon Linux que requieren un reinicio (por ejemplo, actualizaciones del kernel, NVIDIA y CUDA) tras el lanzamiento de una versión de Amazon EMR, las instancias de clúster de Amazon EMR que ejecutan la AMI predeterminada no descargan ni instalan automáticamente esas actualizaciones. Para obtener actualizaciones

del kernel, puede personalizar la AMI de Amazon EMR para que utilice la AMI de Amazon Linux más reciente.

En función del nivel de seguridad de la aplicación y del tiempo que lleve ejecutándose un clúster, puede optar por reiniciar el clúster periódicamente para aplicar las actualizaciones de seguridad, o crear una acción de arranque para personalizar la instalación y las actualizaciones de los paquetes. También puede optar por probar y, a continuación, instalar determinadas actualizaciones de seguridad en las instancias del clúster en ejecución. Para obtener más información, consulte Uso de la AMI de Amazon Linux predeterminada para Amazon EMR. Tenga en cuenta que la configuración de red debe permitir la salida de HTTP y HTTPS a los repositorios de Amazon Linux en Amazon S3; de lo contrario, las actualizaciones de seguridad no se realizarán correctamente.

Uso de configuraciones de seguridad para configurar la seguridad del clúster

Con la versión 4.8.0 o posteriores de Amazon EMR, puede utilizar las configuraciones de seguridad para configurar el cifrado de datos, la autenticación de Kerberos (disponible a partir de la versión 5.10.0 y posteriores) y la autorización de Amazon S3 para EMRFS (disponible en la versión 5.10.0 o posteriores).

Después de crear una configuración de seguridad, debe especificarla cuando crea un clúster y puede volver a utilizarla para cualquier número de clústeres.

Puede utilizar la consola, la AWS Command Line Interface (AWS CLI) o los SDK de AWS para crear configuraciones de seguridad. También puede utilizar una plantilla de AWS CloudFormation para crear una configuración de seguridad. Para obtener más información, consulte la Guía del usuario de AWS CloudFormation y la referencia de plantilla de AWS::EMR::SecurityConfiguration.

Temas

- Creación de una configuración de seguridad
- Especificación de una configuración de seguridad para un clúster

Creación de una configuración de seguridad

En este tema se abordan los procedimientos generales para crear una configuración de seguridad utilizando la consola de EMR y la AWS CLI, así como una referencia para los parámetros de cifrado,

autenticación y roles de IAM para EMRFS. Para obtener más información sobre estas características, consulte los siguientes temas:

- Cifrado de datos en reposo y en tránsito
- Uso de Kerberos para la autenticación con Amazon EMR
- Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3

Creación de una configuración de seguridad utilizando la consola

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/emr.
2. En el panel de navegación, elija Security Configurations (Configuraciones de seguridad), Create security configuration (Crear configuración de seguridad).
3. Escriba un nombre para la configuración de seguridad en el campo Name (Nombre).
4. Elija opciones de Cifrado y Autenticación, tal y como se describe en las secciones siguientes y, a continuación, elija Crear.

Creación de una configuración de seguridad utilizando la AWS CLI

- Utilice el comando `create-security-configuration`, tal y como se muestra en el ejemplo siguiente.
  - En *`SecConfigName`*, especifique el nombre de la configuración de seguridad. Este es el nombre que especifica cuando crea un clúster que usa esta configuración de seguridad.
  - En *`SecConfigDef`*, especifique una estructura JSON en línea o la ruta a un archivo JSON local, como por ejemplo, *`file://MySecConfig.json`*. Los parámetros de JSON definen opciones de Cifrado, Roles de IAM para el acceso de EMRFS a Amazon S3 y Autenticación, tal como se describe en las secciones siguientes.

```
aws emr create-security-configuration --name "SecConfigName" --security-
configuration SecConfigDef
```

Configuración del cifrado de datos

Antes de configurar el cifrado en una configuración de seguridad, cree las claves y los certificados que se utilizan para el cifrado. Para obtener más información, consulte Proporcionar claves para

cifrado de datos en reposo con Amazon EMR y Proporcionar certificados para el cifrado de datos en tránsito con el cifrado de Amazon EMR.

Al crear una configuración de seguridad, debe especificar dos conjuntos de opciones de cifrado: cifrado de datos en reposo y cifrado de datos en tránsito. Las opciones para el cifrado de datos en reposo incluyen tanto Amazon S3 con EMRFS y el cifrado de disco local. Las opciones de cifrado en tránsito habilitan las características de cifrado de código abierto para determinadas aplicaciones que admiten Transport Layer Security (TLS). Las opciones en reposo y en tránsito se pueden habilitar juntas o por separado. Para obtener más información, consulte Cifrado de datos en reposo y en tránsito.

> **Note**
>
> Cuando utilice AWS KMS, se aplican cargos por el almacenamiento y el uso de las claves de cifrado. Para obtener más información, consulte Precios de AWS KMS.

Especificación de las opciones de cifrado con la consola

Elija opciones en Encryption (Cifrado) de acuerdo con las siguientes directrices.

- Elija las opciones en At rest encryption (Cifrado en reposo) para cifrar los datos almacenados en el sistema de archivos.

  Puede elegir cifrar datos en Amazon S3, discos locales o ambas opciones.
- En Cifrado de datos de S3, para Modo de cifrado, seleccione un valor para determinar cómo Amazon EMR cifra los datos de Amazon S3 con EMRFS.

  Lo que haga a continuación depende del modo de cifrado que haya elegido:
  - SSE-S3

    Especifica el cifrado del servidor con claves de cifrado administradas por Amazon S3. No necesita hacer nada más, ya que Amazon S3 se encarga de gestionar las claves.
  - SSE-KMS o CSE-KMS

    Especifica el cifrado del lado del servidor con claves administradas por AWS KMS (SSE-KMS) o el cifrado del lado del cliente con claves administradas por AWS KMS (CSE-KMS). En AWS KMS key, seleccione una clave. La clave debe existir en la misma región que su clúster de EMR. Para los requisitos de clave, consulte Uso de AWS KMS keys para el cifrado.

- CSE-Custom

  Especifica el cifrado del cliente utilizando una clave raíz del cliente (CSE-Custom). En Objeto de S3, especifique la ubicación en Amazon S3 o el ARN de Amazon S3 del archivo JAR del proveedor de claves personalizadas. A continuación, en Key provider class (Clase de proveedor de claves), escriba el nombre de clase completo de una clase declarada en la aplicación que implementa la interfaz EncryptionMaterialsProvider.

- En Local disk encryption (Cifrado de disco local), seleccione un valor para Key provider type (Tipo de proveedor de clave).
  - AWS KMS key

    Seleccione esta opción para especificar una AWS KMS key. En AWS KMS key, seleccione una clave. La clave debe existir en la misma región que su clúster de EMR. Para obtener más información sobre los requisitos de claves, consulte Uso de AWS KMS keys para el cifrado.

    Cifrado de EBS

    Cuando especifique AWS KMS como su proveedor de claves, puede habilitar el cifrado de EBS para cifrar los volúmenes de almacenamiento y de dispositivo raíz de EBS. Para habilitar esa opción, debe conceder el rol de servicio de EMR `EMR_DefaultRole` con permisos para utilizar la AWS KMS key que especifique. Para obtener más información sobre los requisitos de claves, consulte Habilitación del cifrado de EBS proporcionando permisos adicionales para las claves de KMS.

  - Custom (Personalizado)

    Seleccione esta opción para especificar un proveedor de claves personalizadas. En Objeto de S3, especifique la ubicación en Amazon S3 o el ARN de Amazon S3 del archivo JAR del proveedor de claves personalizadas. En Key provider class (Clase de proveedor de claves), escriba el nombre de clase completo de una clase declarada en la aplicación que implementa la interfaz EncryptionMaterialsProvider. El nombre de clase que proporcione aquí debe ser distinto al nombre de clase proporcionado para CSE-Custom.

- Seleccione In-transit encryption (Cifrado en tránsito.) para habilitar las características de cifrado TLS de código abierto para los datos en tránsito. Elija un Certificate provider type (Tipo de proveedor de certificados) de acuerdo con las directrices siguientes:
  - PEM

Seleccione esta opción para utilizar los archivos PEM que proporcione dentro de un archivo zip. Se requieren dos artefactos dentro del archivo zip: privateKey.pem y certificateChain.pem. Un tercer archivo, trustedCertificates.pem, es opcional. Para obtener más información, consulte Proporcionar certificados para el cifrado de datos en tránsito con el cifrado de Amazon EMR. En Objeto de S3, especifique la ubicación en Amazon S3 o el ARN de Amazon S3 del campo del archivo ZIP.

- Custom (Personalizado)

  Seleccione esta opción para especificar un proveedor de certificados personalizado y, a continuación, en Objeto de S3, especifique la ubicación en Amazon S3 o el ARN de Amazon S3 del archivo JAR del proveedor de certificados personalizado. En Key provider class (Clase de proveedor de claves), escriba el nombre de clase completo de una clase declarada en la aplicación que implemente la interfaz TLSArtifactsProvider.

Especificación de las opciones de cifrado con AWS CLI

Las secciones que aparecen a continuación utilizan escenarios de ejemplo para ilustrar una estructura JSON --security-configuration con el formato correcto para diferentes configuraciones y proveedores de claves, así como una referencia para los parámetros JSON y sus valores adecuados.

Opciones de cifrado de datos en tránsito de ejemplo

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está habilitado y el cifrado de datos en reposo está deshabilitado.
- Un archivo ZIP con certificados en Amazon S3 se utiliza como proveedor de claves (consulte Proporcionar certificados para el cifrado de datos en tránsito con el cifrado de Amazon EMR para conocer los requisitos de los certificados).

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://MyConfigStore/artifacts/MyCerts.zip"
```

```
    }
   }
  }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está habilitado y el cifrado de datos en reposo está deshabilitado.
- Se utiliza un proveedor de claves personalizadas (consulte Proporcionar certificados para el cifrado de datos en tránsito con el cifrado de Amazon EMR para requisitos de certificado).

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "CertificateProviderClass": "com.mycompany.MyCertProvider"
   }
  }
  }
}'
```

Opciones de cifrado de datos en reposo de ejemplo

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está deshabilitado y el cifrado de datos en reposo está habilitado.
- SSE-S3 se utiliza para el cifrado de Amazon S3.
- El cifrado de disco local utiliza AWS KMS como proveedor de claves.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": false,
  "EnableAtRestEncryption": true,
```

```
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "SSE-S3"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "AwsKms",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   }
  }
  }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está habilitado y hace referencia a un archivo ZIP con certificados PEM en Amazon S3 utilizando el ARN.
- SSE-KMS se utiliza para el cifrado de Amazon S3.
- El cifrado de disco local utiliza AWS KMS como proveedor de claves.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": true,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "arn:aws:s3:::MyConfigStore/artifacts/MyCerts.zip"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "SSE-KMS",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "AwsKms",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
```

```
   }
  }
 }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está habilitado y hace referencia a un archivo ZIP con certificados PEM en Amazon S3.
- CSE-KMS se utiliza para el cifrado de Amazon S3.
- El cifrado en disco local utiliza un proveedor de claves personalizadas al que hace referencia su ARN.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": true,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://MyConfigStore/artifacts/MyCerts.zip"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "CSE-KMS",
    "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
   },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "Custom",
    "S3Object": "arn:aws:s3:::artifacts/MyKeyProvider.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
   }
  }
 }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está habilitado con un proveedor de claves personalizadas.
- CSE-Custom se utiliza para los datos de Amazon S3.
- El cifrado de disco local utiliza un proveedor de claves personalizadas.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": "true",
  "EnableAtRestEncryption": "true",
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "CertificateProviderClass": "com.mycompany.MyCertProvider"
   }
  },
  "AtRestEncryptionConfiguration": {
   "S3EncryptionConfiguration": {
    "EncryptionMode": "CSE-Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
    },
   "LocalDiskEncryptionConfiguration": {
    "EncryptionKeyProviderType": "Custom",
    "S3Object": "s3://MyConfig/artifacts/MyCerts.jar",
    "EncryptionKeyProviderClass": "com.mycompany.MyKeyProvider"
   }
  }
 }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está deshabilitado y el cifrado de datos en reposo está habilitado.
- El cifrado de Amazon S3 está habilitado con SSE-KMS.
- Se utilizan varias claves de AWS KMS, una por cada bucket de S3, y se aplican excepciones de cifrado a estos buckets de S3 individuales.
- El cifrado de disco local se ha deshabilitado.

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
   "EncryptionConfiguration": {
     "AtRestEncryptionConfiguration": {
             "S3EncryptionConfiguration": {
           "EncryptionMode": "SSE-KMS",
           "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012",
           "Overrides": [
              {
                "BucketName": "sse-s3-bucket-name",
                "EncryptionMode": "SSE-S3"
              },
              {
                "BucketName": "cse-kms-bucket-name",
                "EncryptionMode": "CSE-KMS",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
              },
              {
                "BucketName": "sse-kms-bucket-name",
                "EncryptionMode": "SSE-KMS",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
              }
             ]
             }
           },
     "EnableInTransitEncryption": false,
      "EnableAtRestEncryption": true
  }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está deshabilitado y el cifrado de datos en reposo está habilitado.
- El cifrado de Amazon S3 se habilita con SSE-S3 y el cifrado de disco local se deshabilita.

```
aws emr create-security-configuration --name "MyS3EncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
```

```
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "S3EncryptionConfiguration": {
                "EncryptionMode": "SSE-S3"
            }
        }
    }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está deshabilitado y el cifrado de datos en reposo está habilitado.
- El cifrado de disco local se habilita con AWS KMS como el proveedor de claves y el cifrado de Amazon S3 se deshabilita.

```
aws emr create-security-configuration --name "MyLocalDiskEncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "LocalDiskEncryptionConfiguration": {
                "EncryptionKeyProviderType": "AwsKms",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
            }
        }
    }
}'
```

El ejemplo siguiente ilustra el supuesto siguiente:

- El cifrado de datos en tránsito está deshabilitado y el cifrado de datos en reposo está habilitado.
- El cifrado de disco local se habilita con AWS KMS como el proveedor de claves y el cifrado de Amazon S3 se deshabilita.
- El cifrado de la EBS está habilitado.

```
aws emr create-security-configuration --name "MyLocalDiskEncryptionConfig" --security-
configuration '{
```

```
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "LocalDiskEncryptionConfiguration": {
                "EnableEbsEncryption": true,
                "EncryptionKeyProviderType": "AwsKms",
                "AwsKmsKey": "arn:aws:kms:us-
east-1:123456789012:key/12345678-1234-1234-1234-123456789012"
            }
        }
     }
}'
```

Referencia de JSON para la configuración de cifrado

En la siguiente tabla se muestran los parámetros JSON de configuración de cifrado y se ofrece una descripción de valores aceptados para cada parámetro.

Parámetro	Descripción
`"EnableInTransitEncryption": true | falso`	Specify `true` to enable in-transit encryption and `falso` to disable it. If omitted, `falso` is assumed, and in-transit encryption is disabled.
`"EnableAtRestEncryption": true | falso`	Specify `true` to enable at-rest encryption and `falso` to disable it. If omitted, `falso` is assumed and at-rest encryption is disabled.
Parámetros de cifrado en tránsito	
`"InTransitEncryptionConfiguration":`	Specifies a collection of values used to configure in-transit encryption when `EnableInTransitEncryption` is `true`.
`"CertificateProviderType": "PEM" | "Custom"`	Specifies whether to use `PEM` certificates referenced with a zipped file, or a `Personalizado` certificate provider. If `PEM` is specified, `S3Object` must be a reference to the location in Amazon S3 of a zip file containing the certificates. If Custom is specified, `S3Object`

Parámetro	Descripción
	must be a reference to the location in Amazon S3 of a JAR file, followed by a `CertificateProviderClass` entry.
`"S3Object": "` *`ZipLocation`* `" | "` *`JarLocation`* `"`	Provides the location in Amazon S3 to a zip file when `PEM` is specified, or to a JAR file when `Personalizado` is specified. The format can be a path (for example, `s3://MyConfig/artifacts/CertFiles.zip`) or an ARN (for example, `arn:aws:s3:::Code/MyCertProvider.jar`). If a zip file is specified, it must contain files named exactly `privateKey.pem` and `certificateChain.pem`. A file named `trustedCertificates.pem` is optional.
`"CertificateProviderClass": "` *`MyClassID`* `"`	Required only if `Personalizado` is specified for `CertificateProviderType`. *`MyClassID`* specifies a full class name declared in the JAR file, which implements the TLSArtifactsProvider interface. For example, `com.mycompany.MyCertProvider`.
Parámetros de cifrado en reposo	
`"AtRestEncryptionConfiguration":`	Specifies a collection of values for at-rest encryption when `EnableAtRestEncryption` is `true`, including Amazon S3 encryption and local disk encryption.
Parámetros de cifrado de Amazon S3	
`"S3EncryptionConfiguration":`	Specifies a collection of values used for Amazon S3 encryption with the EMR File System (EMRFS).

Parámetro	Descripción
`"EncryptionMode" : "SSE-S3" | "SSE-KMS" | "CSE-KMS" | "CSE-Custom"`	Specifies the type of Amazon S3 encryption to use. If `SSE-S3` is specified, no further Amazon S3 encryption values are required. If either `SSE-KMS` or `CSE-KMS` is specified, an AWS KMS key ARN must be specified as the `AwsKmsKey` value. If `CSE-Custom` is specified, `S3Object` and `EncryptionKeyProviderClass` values must be specified.
`"AwsKmsKey": " MyKeyARN"`	Required only when either `SSE-KMS` or `CSE-KMS` is specified for `EncryptionMode`. *MyKeyARN* must be a fully specified ARN to a key (for example, `arn:aws:kms:us-east-1:123456789012:key/12345678-1234-1234-1234-123456789012`).
`"S3Object": "JarLocation "`	Required only when `CSE-Custom` is specified for `CertificateProviderType`. *JarLocation* provides the location in Amazon S3 to a JAR file. The format can be a path (for example, `s3://MyConfig/artifacts/MyKeyProvider.jar`) or an ARN (for example, `arn:aws:s3:::Code/MyKeyProvider.jar`).
`"EncryptionKeyProviderClass": "MyS3KeyClassID "`	Required only when `CSE-Custom` is specified for `EncryptionMode`. *MyS3KeyClassID* specifies a full class name of a class declared in the application that implements the EncryptionMaterialsProvider interface; for example, *com.mycompany.MyS3KeyProvider*.

Parámetros de cifrado de disco local

Parámetro	Descripción
`"LocalDiskEncryptionConfiguration"`	Specifies the key provider and corresponding values to be used for local disk encryption.
`"EnableEbsEncryption": true | false`	Specify `true` to enable EBS encryption. EBS encryption encrypts the EBS root device volume and attached storage volumes. To use EBS encryption, you must specify `AwsKms` as your `EncryptionKeyProviderType`.
`"EncryptionKeyProviderType": "AwsKms" | "Custom"`	Specifies the key provider. If `AwsKms` is specified, an KMS key ARN must be specified as the `AwsKmsKey` value. If **`Personalizado`** is specified, `S3Object` and `EncryptionKeyProviderClass` values must be specified.
`"AwsKmsKey: " *MyKeyARN*"`	Required only when `AwsKms` is specified for `Tipo`. *`MyKeyARN`* must be a fully specified ARN to a key (for example, `arn:aws:kms:us-east-1:123456789012:key/12345678-1234-1234-1234-456789012123`).
`"S3Object": "*JarLocation* "`	Required only when `CSE-Custom` is specified for `CertificateProviderType`. *`JarLocation`* provides the location in Amazon S3 to a JAR file. The format can be a path (for example, `s3://MyConfig/artifacts/MyKeyProvider.jar`) or an ARN (for example, `arn:aws:s3:::Code/MyKeyProvider.jar)`.

Parámetro	Descripción
`"EncryptionKeyProviderClass" : "MyLocalDiskKeyClassID "`	Required only when `Personalizado` is specified for `Tipo`. *`MyLocalDiskKeyClassID`* specifies a full class name of a class declared in the application that implements the EncryptionMaterialsProvider interface; for example, *`com.mycompany.MyLocalDiskKeyProvider`*.

Configuración de la autenticación de Kerberos

Una configuración de seguridad con la configuración de Kerberos solo se puede utilizar en un clúster que se crea con los atributos de Kerberos o si se produce un error. Para obtener más información, consulte Uso de Kerberos para la autenticación con Amazon EMR. Kerberos solo está disponible en la versión 5.10.0 y posteriores de Amazon EMR.

Especificación de la configuración de Kerberos con la consola

Elija opciones en Kerberos authentication (Autenticación Kerberos) de acuerdo con las siguientes directrices.

Parámetro		Descripción
Kerberos		Especifica que Kerberos está habilitado para los clústeres que utilizan esta configuración de seguridad. Si un clúster usa esta configuración de seguridad, también debe tener la configuración de Kerberos especificada o se producirá un error.
Proveedor	KDC dedicado del clúster	Especifica que Amazon EMR crea un KDC en el nodo principal de cualquier clúster que utilice esta configuración de seguridad. Al crear el clúster, debe especificar el nombre del dominio y la contraseña de administrador del KDC. Si es necesario, puede hacer referencia a este KDC desde otros clústeres. Cree esos clústeres con una

Parámetro		Descripción
		configuración de seguridad diferente, especifique un KDC externo y utilice el nombre de dominio y la contraseña de administrador del KDC que especifique para el KDC dedicado al clúster.
	KDC externo	Solo está disponible en la versión de Amazon EMR 5.20.0 y posteriores. Especifica que los clústeres que utilizan esta configuración de seguridad autentican las entidades principales de Kerberos mediante un servidor de KDC externo al clúster. No se crea un KDC en el clúster. Al crear el clúster, debe especificar el nombre de dominio y la contraseña de administrador del KDC para el KDC externo.
Ciclo de vida del ticket		Opcional. Especifica el período durante el que un ticket de Kerberos emitido por el KDC es válido en los clústeres que utilizan esta configuración de seguridad. El ciclo de vida de los tickets es limitado por motivos de seguridad. Las aplicaciones de clúster y los servicios renuevan automáticamente los tickets después de que expiren. Los usuarios que se conecten al clúster mediante SSH utilizando las credenciales de Kerberos tienen que ejecutar el comando `kinit` desde la línea de comandos del nodo principal para la renovación tras la expiración de un ticket.

Parámetro		Descripción
Relación de confianza entre ámbitos		Especifica una relación de confianza entre ámbitos entre un KDC dedicado a un clúster en los clústeres que utilizan esta configuración de seguridad y un KDC situado en un ámbito de Kerberos diferente. Las entidades principales (normalmente los usuarios) de otro ámbito se autentican en los clústeres que utilizan esta configuración. Se requiere una configuración adicional en el otro ámbito de Kerberos. Para obtener más información, consulte Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory.
Propiedades de la relación de confianza entre ámbitos	Ámbito	Especifica el nombre de ámbito de Kerberos del otro ámbito en la relación de confianza. Por convención, los nombres de ámbito de Kerberos son los mismos que los nombres de dominio, pero en mayúsculas.
	Dominio	Especifica el nombre de dominio del otro ámbito en la relación de confianza.
	Servidor de administración	Especifica el nombre de dominio completo (FQDN) o dirección IP del servidor de administración del otro ámbito de la relación de confianza. El servidor de administración y el servidor de KDC suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos para comunicarse. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 749. También se puede especificar el puerto (por ejemplo, `domain.example.com:749`).

Parámetro		Descripción
	Servidor de KDC	Especifica el nombre de dominio completo (FQDN) o dirección IP del servidor de KDC del otro ámbito de la relación de confianza. El servidor de KDC y el servidor de administración suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 88. También se puede especificar el puerto (por ejemplo, `domain.example.com:88`).
KDC externo		Especifica que el clúster utiliza el KDC externo del clúster.
Propiedades de KDC externo	Servidor de administración	Especifica el nombre de dominio completo (FQDN) o la dirección IP del servidor de administración externo. El servidor de administración y el servidor de KDC suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos para comunicarse. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 749. También se puede especificar el puerto (por ejemplo, `domain.example.com:749`).

Parámetro			Descripción
	Servidor de KDC		Especifica el nombre de dominio completo (FQDN) del servidor de KDC externo. El servidor de KDC y el servidor de administración suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 88. También se puede especificar el puerto (por ejemplo, `domain.example.com :88`).
	Integración de Active Directory		Especifica que la autenticación de la entidad principal de Kerberos está integrada en un dominio de Microsoft Active Directory.
	Propiedades de la integración de Active Directory	Ámbito de Active Directory	Especifica el nombre de ámbito de Kerberos del dominio de Active Directory. Por convención, los nombres de ámbito de Kerberos suelen ser los mismos que los nombres de dominio, pero en mayúsculas.
		Dominio de Active Directory	Especifica el nombre de dominio de Active Directory.
		Servidor de Active Directory	Especifica el nombre de dominio completo (FQDN) del controlador de dominio de Microsoft Active Directory.

Especificación de la configuración de Kerberos con AWS CLI

En la siguiente tabla de referencia, se muestran los parámetros JSON para la configuración de Kerberos en una configuración de seguridad. Para ver configuraciones de ejemplo, consulte Ejemplos de configuraciones.

Parámetro	Descripción
"AuthenticationConfiguration": {	Obligatorio para Kerberos. Especifica que una configuración de autenticación forma parte de esta configuración de seguridad.
"KerberosConfiguration": {	Obligatorio para Kerberos. Especifica las propiedades de configuración de Kerberos.
"Provider": *"ClusterDedicatedKdc",* —o— "Provider": *"ExternalKdc",*	*ClusterDedicatedKdc* especifica que Amazon EMR crea un KDC en el nodo principal de cualquier clúster que utilice esta configuración de seguridad. Al crear el clúster, debe especificar el nombre del dominio y la contraseña de administrador del KDC. Si es necesario, puede hacer referencia a este KDC desde otros clústeres. Cree esos clústeres con una configuración de seguridad diferente, especifique un KDC externo y utilice el nombre de ámbito y la contraseña de administrador del KDC que especificó al crear el clúster con el KDC dedicado a un clúster. *ExternalKdc* especifica que el clúster usa un KDC externo. Amazon EMR no crea un KDC en el nodo principal. Un clúster que utilice esta configuración de seguridad debe especificar el nombre de ámbito y la contraseña de administrador del KDC del KDC externo.

Parámetro	Descripción
"ClusterDedicatedKdcConfiguration": {	Es obligatorio cuando se especifica *ClusterDedicatedKdc* .
"TicketLifetimeInHours": *24*,	Opcional. Especifica el período durante el que un ticket de Kerberos emitido por el KDC es válido en los clústeres que utilizan esta configuración de seguridad. El ciclo de vida de los tickets es limitado por motivos de seguridad . Las aplicaciones de clúster y los servicios renuevan automáticamente los tickets después de que expiren. Los usuarios que se conecten al clúster mediante SSH utilizando las credenciales de Kerberos tienen que ejecutar el comando `kinit` desde la línea de comandos del nodo principal para la renovación tras la expiración de un ticket.

Parámetro	Descripción
`"CrossRealmTrustConfiguration": {`	Especifica una relación de confianza entre ámbitos entre un KDC dedicado a un clúster en los clústeres que utilizan esta configuración de seguridad y un KDC situado en un ámbito de Kerberos diferente. Las entidades principales (normalmente los usuarios) de otro ámbito se autentican en los clústeres que utilizan esta configuración. Se requiere una configuración adicional en el otro ámbito de Kerberos. Para obtener más información, consulte Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory.
`"Realm": "KDC2.COM",`	Especifica el nombre de ámbito de Kerberos del otro ámbito en la relación de confianza. Por convención, los nombres de ámbito de Kerberos son los mismos que los nombres de dominio, pero en mayúsculas.
`"Domain": "kdc2.com",`	Especifica el nombre de dominio del otro ámbito en la relación de confianza.

Parámetro	Descripción
"AdminServer": "*kdc.com:749* ",	Especifica el nombre de dominio completo (FQDN) o dirección IP del servidor de administración del otro ámbito de la relación de confianza . El servidor de administración y el servidor de KDC suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos para comunicarse. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 749. También se puede especificar el puerto (por ejemplo, `domain.example.com` *:749*).
"KdcServer": "*kdc.com:88* "	Especifica el nombre de dominio completo (FQDN) o dirección IP del servidor de KDC del otro ámbito de la relación de confianza. El servidor de KDC y el servidor de administración suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 88. También se puede especificar el puerto (por ejemplo, `domain.example.com` *:88*).
}	
}	

Parámetro	Descripción
"ExternalKdcConfiguration": {	Es obligatorio cuando se especifica *ExternalKdc* .
"TicketLifetimeInHours": *24*,	Opcional. Especifica el período durante el que un ticket de Kerberos emitido por el KDC es válido en los clústeres que utilizan esta configuración de seguridad. El ciclo de vida de los tickets es limitado por motivos de seguridad . Las aplicaciones de clúster y los servicios renuevan automáticamente los tickets después de que expiren. Los usuarios que se conecten al clúster mediante SSH utilizando las credenciales de Kerberos tienen que ejecutar el comando `kinit` desde la línea de comandos del nodo principal para la renovación tras la expiración de un ticket.
"KdcServerType": "Single",	Especifica que se hace referencia a un único servidor de KDC. `Single` es el único valor admitido en este momento.

Parámetro	Descripción
"AdminServer": "*kdc.com:749* ",	Especifica el nombre de dominio completo (FQDN) o la dirección IP del servidor de administración externo. El servidor de administración y el servidor de KDC suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos para comunicarse. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 749. También se puede especificar el puerto (por ejemplo, `domain.example.com` *:749*).
"KdcServer": "*kdc.com:88* ",	Especifica el nombre de dominio completo (FQDN) del servidor de KDC externo. El servidor de KDC y el servidor de administración suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 88. También se puede especificar el puerto (por ejemplo, `domain.example.com` *:88*).
`"AdIntegrationConfiguration": {`	Especifica que la autenticación de la entidad principal de Kerberos está integrada en un dominio de Microsoft Active Directory.

Parámetro	Descripción
"AdRealm": "*AD.DOMAIN.COM* ",	Especifica el nombre de ámbito de Kerberos del dominio de Active Directory. Por convención, los nombres de ámbito de Kerberos suelen ser los mismos que los nombres de dominio, pero en mayúsculas.
"AdDomain": "*ad.domain.com* "	Especifica el nombre de dominio de Active Directory.
"AdServer": "*ad.domain.com* "	Especifica el nombre de dominio completo (FQDN) del controlador de dominio de Microsoft Active Directory .
}	
}	
}	
}	

Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3

Los roles de IAM para EMRFS le permiten proporcionar diferentes permisos a los datos de EMRFS en Amazon S3. Puede crear asignaciones que especifiquen un rol de IAM que se use para obtener permisos cuando una solicitud de acceso contenga un identificador que haya especificado. El identificador puede ser un usuario o un rol de Hadoop o un prefijo de Amazon S3.

Para obtener más información, consulte Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3.

Especificación de roles de IAM para EMRFS mediante AWS CLI

A continuación, se muestra un ejemplo de fragmento de JSON para especificar roles de IAM personalizados para EMRFS dentro de una configuración de seguridad. Muestra las asignaciones de roles para los tres tipos de identificadores diferentes, seguidas de una referencia de parámetros.

```
{
  "AuthorizationConfiguration": {
    "EmrFsConfiguration": {
      "RoleMappings": [{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_user1",
        "IdentifierType": "User",
        "Identifiers": [ "user1" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_to_MyBuckets",
        "IdentifierType": "Prefix",
        "Identifiers": [ "s3://MyBucket/","s3://MyOtherBucket/" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_AdminGroup",
        "IdentifierType": "Group",
        "Identifiers": [ "AdminGroup" ]
      }]
    }
  }
}
```

Parámetro	Descripción
`"AuthorizationConfiguration":`	Obligatorio.
`"EmrFsConfiguration":`	Obligatorio. Contiene asignaciones de roles.
`"RoleMappings":`	Obligatorio. Contiene una o más definiciones de asignación de roles. Las asignaciones de roles se evalúan de forma descendente por orden de aparición. Si una asignación de roles se evalúa como true en una invocación de datos de EMRFS en Amazon S3, no se evalúan más asignaciones de roles y EMRFS utiliza el rol de IAM especificado para la

Parámetro	Descripción
	solicitud. Las asignaciones de roles tienen los siguientes parámetros obligatorios:
`"Role":`	Especifica el identificador de ARN de un rol de IAM en el formato `arn:aws:iam::` *`account-id`* `:role/`*`role-name`*. Este es el rol de IAM que asume Amazon EMR si la solicitud de EMRFS a Amazon S3 coincide con alguno de los `Identifiers` especificados.
`"IdentifierType":`	Puede ser uno de los siguientes: • `"User"` especifica que los identificadores son uno o más usuarios de Hadoop, que pueden ser usuarios de cuentas de Linux o entidades principales de Kerberos. Cuando la solicitud de EMRFS se origina en el usuario o los usuarios especificados, se asume el rol de IAM. • `"Prefix"` especifica que el identificador es una ubicación de Amazon S3. El rol de IAM se asume para las llamadas a la ubicación o ubicaciones con los prefijos especificados. Por ejemplo, el prefijo `s3://mybucket/` coincide con `s3://mybucket/mydir` y `s3://mybucket/yetanotherdir`. • `"Group"` especifica que los identificadores son uno o más grupos de Hadoop. El rol de IAM se asume si la solicitud proviene de un usuario del grupo o grupos especificados.
`"Identifiers":`	Especifica uno o más identificadores del tipo de identificador adecuado. Separe varios identificadores con comas sin espacios.

Configuración de las solicitudes de servicios de metadatos a las instancias de Amazon EC2

Los metadatos de instancia son datos sobre una instancia que se pueden utilizar para configurar o administrar la instancia en ejecución. Para acceder a los metadatos de instancia desde una instancia en ejecución puede utilizar uno de los métodos siguientes:

- Servicio de metadatos de instancia, versión 1 (IMDSv1): un método de solicitud y respuesta
- Servicio de metadatos de instancia, versión 2 (IMDSv2): un método orientado a la sesión

Si bien Amazon EC2 admite IMDSv1 e IMDSv2, Amazon EMR admite IMDSv2 en Amazon EMR 5.23.1, 5.27.1, 5.32 o versiones posteriores y 6.2 o versiones posteriores. En estas versiones, los componentes de Amazon EMR utilizan IMDSv2 para todas las llamadas al IMDS. Para las llamadas al IMDS en el código de la aplicación, puede utilizar IMDSv1 e IMDSv2, o configurar el IMDS para que utilice solo IMDSv2 para mayor seguridad. Si especifica que debe usarse IMDSv2, IMDSv1 dejará de funcionar.

Para obtener más información, consulte Configuración del servicio de metadatos de instancia en la Guía del usuario de Amazon EC2 para instancias de Linux.

> **Note**
>
> En versiones anteriores de Amazon EMR 5.x o 6.x, la desactivación de IMDSv1 provocaba un error en el arranque del clúster, ya que los componentes de Amazon EMR utilizaban IMDSv1 para todas las llamadas al IMDS. Al desactivar IMDSv1, asegúrese de que cualquier software personalizado que utilice IMDSv1 esté actualizado para usar IMDSv2.

Especificación de la configuración del servicio de metadatos de la instancia mediante AWS CLI

A continuación, se muestra un ejemplo de un fragmento de JSON para especificar el servicio de metadatos de instancias (IMDS) de Amazon EC2 dentro de una configuración de seguridad.

```
{
  "InstanceMetadataServiceConfiguration" : {
      "MinimumInstanceMetadataServiceVersion": integer,
      "HttpPutResponseHopLimit": integer
   }
```

```
}
```

Parámetro	Descripción
`"InstanceMetadataServiceConfiguration":`	Obligatorio.
`"MinimumInstanceMetadataServiceVersion":`	Obligatorio. Especifique 1 o 2. Un valor de 1 permite el uso de IMDSv1 e IMDSv2. Un valor de 2 solo permite el uso de IMDSv2.
`"HttpPutResponseHopLimit":`	Obligatorio. Límite de saltos de respuesta HTTP PUT deseado para las solicitudes de metadatos de instancia. Cuanto mayor sea el número, más solicitudes de metadatos de instancia pueden viajar. Valor predeterminado: 1. Puede especificar un número entero del 1 al 64.

Especificación la configuración del servicio de metadatos de la instancia con la consola

Puede configurar el uso del IMDS para un clúster al lanzarlo desde la consola de Amazon EMR. IMDS Security configura los controles en la consola Amazon EMR

Para configurar el uso del IMDS mediante la consola:

1. Al crear una nueva configuración de seguridad en la página Configuraciones de seguridad, seleccione Configurar el servicio de metadatos de instancias de EC2 en la configuración del servicio de metadatos de instancias de EC2. Esta configuración solo se admite en Amazon EMR 5.23.1, 5.27.1, 5.32 o versiones posteriores y 6.2 o versiones posteriores.
2. Para ver la opción Versión mínima del servicio de metadatos de instancias, seleccione una de las siguientes opciones:
   - Desactivar IMDSv1 y permitir solo IMDSv2 si quiere permitir únicamente el uso de IMDSv2 en este clúster. Consulte Transición al uso de servicio de metadatos de instancias, versión 2 en la Guía del usuario de Amazon EC2 para instancias de Linux.
   - Permitir tanto IMDSv1 como IMDSv2 en el clúster si quiere permitir el uso de IMDSv1 e IMDSv2 orientado a la sesión en este clúster.

3. En IMDSv2, también puede configurar el número permitido de saltos de red para el token de metadatos configurando un número entero entre 1 y 64 en Límite de saltos de respuesta de HTTP PUT.

Para obtener más información, consulte Configuración del servicio de metadatos de instancia en la Guía del usuario de Amazon EC2 para instancias de Linux.

Consulte Configuración de los detalles de la instancia y Configuración del servicio de metadatos de instancia en la Guía del usuario de Amazon EC2 para instancias de Linux.

Especificación de una configuración de seguridad para un clúster

Puede especificar la configuración de cifrado al crear un clúster especificando la configuración de seguridad. Puede utilizar la AWS Management Console o la AWS CLI.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Especificación de una configuración de seguridad con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Configuración y permisos de seguridad, busque el campo Configuración de seguridad. Seleccione el menú desplegable o elija Examinar para seleccionar el nombre de una configuración de seguridad que haya creado anteriormente. También puede elegir Crear configuración de seguridad para crear una configuración que pueda usar para su clúster.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Especificación de una configuración de seguridad con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/emr.
2. Elija Crear clúster e Ir a las opciones avanzadas.
3. En la pantalla Paso 1: software y pasos, en la lista Versión, elija emr-4.8.0 o una versión más reciente. Elija la configuración que desee y haga clic en Next (Siguiente).
4. En la pantalla Step 2: Hardware (Paso 2: Hardware), elija la configuración que desea y elija Next (Siguiente). Haga lo mismo para el Step 3: General Cluster Settings (Paso 3: Configuración general del clúster).
5. En la pantalla Step 4: Security (Paso 4: Seguridad), en Encryption Options (Opciones de cifrado), elija un valor para Security configuration (Configuración de seguridad).
6. Configure las demás opciones de seguridad como desee y elija Create cluster (Crear clúster).

CLI

Para especificar una configuración de seguridad con AWS CLI

- Utilice `aws emr create-cluster` para aplicar una configuración de seguridad con `--security-configuration` *`MySecConfig`*, donde *`MySecConfig`* es el nombre de la configuración de seguridad, tal y como se muestra en el siguiente ejemplo. La etiqueta `--release-label` especificada debe ser 4.8.0 o una versión posterior y `--instance-type` puede ser cualquiera disponible.

```
aws emr create-cluster --instance-type m5.xlarge --release-label emr-5.0.0 --
security-configuration mySecConfig
```

Protección de los datos en Amazon EMR

El modelo de responsabilidad compartida de AWS se aplica a la protección de datos en Amazon EMR. Tal como se describe en este modelo, AWS es responsable de proteger la infraestructura global que ejecuta toda la nube de AWS. Usted es responsable de mantener el control sobre el contenido alojado en esta infraestructura. Este contenido incluye la configuración de seguridad y las tareas de administración para el que utiliza AWS. Para obtener más información sobre la privacidad de datos, consulte Data Privacy FAQ (Preguntas frecuentes sobre la privacidad de datos). Para

obtener más información sobre la protección de datos en Europa, consulte la publicación de blog The Amazon Shared Responsibility Model and GDPR en el blog de seguridad de AWS.

Para fines de protección de datos, recomendamos proteger las credenciales de la cuenta de AWS y configurar cuentas individuales con AWS Identity and Access Management. De esta manera, solo se otorgan a cada usuario los permisos necesarios para cumplir con sus obligaciones laborales. También recomendamos proteger sus datos de las siguientes formas:

- Utilice autenticación multifactor (MFA) en cada cuenta.
- Utilice TLS para comunicarse con los recursos de AWS. Se requiere usar TLS 1.2.
- Configure la API y el registro de actividad del usuario con AWS CloudTrail.
- Utilice las soluciones de cifrado de AWS, junto con todos los controles de seguridad predeterminados dentro de los servicios de AWS.
- Utilice avanzados servicios de seguridad administrados, como Amazon Macie, que lo ayuden a detectar y proteger los datos personales almacenados en Amazon S3.
- Si necesita módulos criptográficos validados FIPS 140-2 al acceder a AWS a través de una interfaz de línea de comandos o una API, utilice un punto de enlace de FIPS. Para obtener más información acerca de los puntos de enlace de FIPS disponibles, consulte Estándar de procesamiento de la información federal (FIPS) 140-2.

Le recomendamos encarecidamente que nunca introduzca información de identificación confidencial, como, por ejemplo, números de cuenta de sus clientes, en los campos de formato libre, como el campo Name (Nombre). Esto incluye los casos en los que trabaje con Amazon EMR u otros servicios de AWS mediante la consola, la API, AWS CLI o los SDK de AWS. Es posible que cualquier dato que ingrese en Amazon EMR u otros servicios se incluya en los registros de diagnóstico. Cuando proporcione una URL a un servidor externo, no incluya información de credenciales en la URL para validar la solicitud para ese servidor.

Cifrado de datos en reposo y en tránsito

El cifrado de datos ayuda a impedir que los usuarios no autorizados lean los datos en un clúster y sistemas de almacenamiento de datos asociados. Esto incluye los datos guardados en medios persistentes, conocidos como datos en reposo y datos que pueden ser interceptados cuando recorren la red, conocidos como datos en tránsito.

A partir de la versión 4.8.0 de Amazon EMR, puede utilizar configuraciones de seguridad de Amazon EMR para definir configuraciones de cifrado de datos para clústeres de manera más sencilla. Las

configuraciones de seguridad ofrecen ajustes para habilitar la seguridad de los datos en tránsito y de los datos en reposo en volúmenes de Amazon Elastic Block Store (Amazon EBS) y EMRFS en Amazon S3.

Opcionalmente, a partir de la versión 4.1.0 de Amazon EMR y posteriores, puede elegir configurar el cifrado transparente en el HDFS, que no se configura utilizando las configuraciones de seguridad. Para obtener más información, consulte Cifrado transparente en el HDFS en Amazon EMR en la Guía de lanzamiento de Amazon EMR.

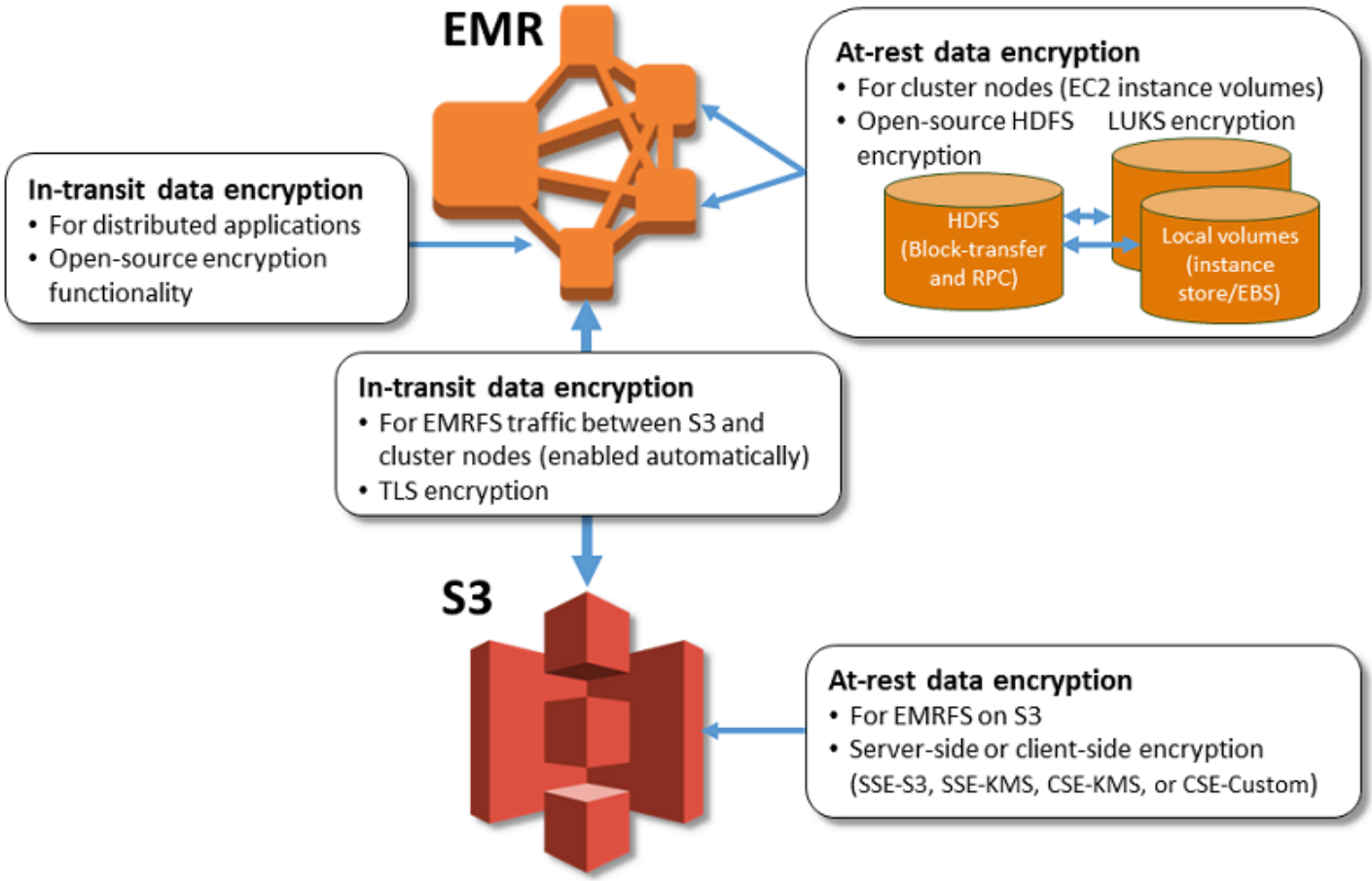
Temas

- Opciones de cifrado
- Creación de claves y certificados para el cifrado de datos

Opciones de cifrado

Con las versiones 4.8.0 y posteriores de Amazon EMR, puede utilizar una configuración de seguridad para especificar las opciones de cifrado de datos en reposo, datos en tránsito o ambos. Al habilitar el cifrado de datos en reposo, puede elegir cifrar datos de EMRFS en Amazon S3, datos en discos locales o ambos. Cada configuración de seguridad que se crea se almacena en Amazon EMR en lugar de en la configuración del clúster, por lo que puede volver a utilizar con facilidad una configuración para especificar ajustes de cifrado de datos cada vez que cree un clúster. Para obtener más información, consulte Creación de una configuración de seguridad.

En el siguiente diagrama se muestran las distintas opciones de cifrado de datos disponibles con las configuraciones de seguridad.



Las siguientes opciones de cifrado también están disponibles y no se configuran utilizando una configuración de seguridad:

- Opcionalmente, con las versiones 4.1.0 y posteriores de Amazon EMR, puede elegir configurar el cifrado transparente en el HDFS. Para obtener más información, consulte Cifrado transparente en el HDFS en Amazon EMR en la Guía de lanzamiento de Amazon EMR.
- Si utiliza una versión de Amazon EMR que no admite configuraciones de seguridad, puede configurar manualmente el cifrado para los datos de EMRFS en Amazon S3. Para obtener más información, consulte Especificación del cifrado de Amazon S3 con propiedades de EMRFS.
- Si utiliza una versión de Amazon EMR anterior a 5.24.0, un volumen de dispositivo raíz cifrado de EBS se admite únicamente cuando se utiliza una AMI personalizada. Para obtener más información, consulte Creación de una AMI personalizada con un volumen de dispositivo raíz de Amazon EBS cifrado en la Guía de administración de Amazon EMR.

> **Note**
>
> A partir de la versión 5.24.0 de Amazon EMR, podrá utilizar una opción de configuración de seguridad para cifrar los volúmenes de almacenamiento y el dispositivo raíz de EBS cuando especifique AWS KMS como su proveedor de claves. Para obtener más información, consulte Cifrado de disco local.

El cifrado de datos requiere las claves y los certificados. Una configuración de seguridad le ofrece la flexibilidad para elegir entre varias opciones, incluidas claves administradas por AWS Key Management Service, claves administradas por Amazon S3, así como claves y certificados de proveedores personalizados que proporcione. Cuando se utiliza AWS KMS como proveedor de claves, se aplican cargos por el almacenamiento y el uso de las claves de cifrado. Para obtener más información, consulte Precios de AWS KMS.

Antes de especificar las opciones de cifrado, decida los sistemas de administración de clave y certificado que desee utilizar, para poder crear primero las claves y los certificados o los proveedores personalizados que especifique como parte de la configuración de cifrado.

Cifrado en reposo para datos de EMRFS en Amazon S3

El cifrado de Amazon S3 funciona con objetos del sistema de archivos de EMR (EMRFS) que se leen y se escriben en Amazon S3. Se especifica el cifrado del servidor (SSE) o el cifrado del cliente (CSE) de Amazon S3 como Modo de cifrado predeterminado al habilitar el cifrado en reposo. También puede especificar métodos de cifrado diferentes para buckets individuales utilizando Per bucket encryption overrides (Reemplazos de cifrado por bucket). Independientemente de si el cifrado de Amazon S3 está habilitado, la seguridad de la capa de transporte (TLS) cifra los objetos de EMRFS en tránsito entre los nodos del clúster de EMR y Amazon S3. Para obtener más información acerca del cifrado de Amazon S3, consulte Protección de datos mediante cifrado en la Guía del usuario de Amazon Simple Storage Service.

> **Note**
>
> Cuando utilice AWS KMS, se aplican cargos por el almacenamiento y el uso de las claves de cifrado. Para obtener más información, consulte Precios de AWS KMS.

Cifrado del servidor de Amazon S3

Cuando configura el cifrado del servidor de Amazon S3, Amazon S3 cifra los datos del objeto a medida que escribe los datos en el disco y descifra los datos cuando se accede. Para obtener más información sobre SSE, consulte Protección de los datos con el cifrado del servidor en la Guía del usuario de Amazon Simple Storage Service.

Puede elegir entre dos sistemas de administración de claves distintos al especificar SSE en Amazon EMR:

- SSE-S3: Amazon S3 administra las claves en su nombre.
- SSE-KMS: utiliza una AWS KMS key para configurar políticas adecuadas para Amazon EMR. Para obtener más información sobre los requisitos de las claves de Amazon EMR, consulte Uso de AWS KMS keys para el cifrado.

SSE con claves proporcionadas por el cliente (SSE-C) no está disponible para su uso con Amazon EMR.

Cifrado del cliente de Amazon S3

Con el cifrado del cliente de Amazon S3, el proceso de cifrado y descifrado de Amazon S3 se produce en el cliente de EMRFS en su clúster. Los objetos se cifran antes de cargarlos en Amazon S3 y se descifran después de que se descarguen. El proveedor que especifique proporciona la clave de cifrado que utiliza el cliente. El cliente puede usar claves proporcionadas por AWS KMS (CSE-KMS) o una clase de Java personalizada que proporciona la clave raíz del cliente (CSE-C). Los detalles de cifrado son ligeramente diferentes entre CSE-KMS y CSE-C, en función del proveedor especificado y de los metadatos del objeto que se descifra o se cifra. Para obtener más información sobre estas diferencias, consulte Protección de los datos con el cifrado del cliente en la Guía del usuario de Amazon Simple Storage Service.

> **Note**
>
> El CSE de Amazon S3 solo garantiza que los datos de EMRFS intercambiados con Amazon S3 se cifren; no se cifran todos los datos en volúmenes de instancias de clúster. Además, ya que Hue no utiliza EMRFS, los objetos que Hue S3 File Browser escribe en Amazon S3 no se cifran.

Cifrado de disco local

Los siguientes mecanismos funcionan juntos para cifrar discos locales cuando habilita el cifrado de discos locales utilizando una configuración de seguridad de Amazon EMR.

Cifrado HDFS de código abierto

HDFS intercambia datos entre las instancias de clúster durante el procesamiento distribuido. También lee y escribe datos a volúmenes de almacenes de instancias y a los volúmenes de EBS asociado a las instancias. Las siguientes opciones de cifrado de Hadoop de código abierto se activan cuando se habilita el cifrado de disco local:

- Secure Hadoop RPC se define en `Privacy`, que utiliza nivel de seguridad y autenticación simples (SASL).
- Data encryption on HDFS block data transfer se define como `true` y se configura para utilizar el cifrado AES 256.

> **Note**
>
> Puede activar el cifrado de Apache Hadoop adicional habilitando el cifrado en tránsito. Para obtener más información, consulte Cifrado en tránsito. Estos ajustes de cifrado no activan el cifrado transparente de HDFS, que puede configurar manualmente. Para obtener más información, consulte Cifrado transparente en el HDFS en Amazon EMR en la Guía de lanzamiento de Amazon EMR.

Cifrado del almacén de instancias

En los tipos de instancias de EC2 que utilizan SSD basados en NVMe como el volumen de almacén de instancias, el cifrado NVMe se utiliza independientemente de la configuración de cifrado de Amazon EMR. Para obtener más información, consulte Volúmenes SSD de NVMe en la Guía del usuario de Amazon EC2 para instancias de Linux. En otros volúmenes de almacén de instancias, Amazon EMR utiliza LUKS para cifrar el volumen de almacén de instancias cuando se ha habilitado el cifrado de disco local independientemente de si los volúmenes de EBS se han cifrado utilizando el cifrado de EBS o LUKS.

Cifrado de volumen de EBS

Si crea un clúster en una región donde el cifrado de Amazon EC2 de volúmenes de EBS se ha habilitado por defecto para su cuenta, los volúmenes de EBS se cifran incluso si el cifrado de disco local no está habilitado. Para obtener más información, consulte Cifrado de forma predeterminada en la Guía del usuario de Amazon EC2 para instancias de Linux. Con el cifrado de disco local habilitado en una configuración de seguridad, la configuración de Amazon EMR tiene preferencia sobre la configuración de cifrado predeterminada de Amazon EC2 para las instancias de EC2 de clústeres.

Las siguientes opciones están disponibles para volúmenes de cifrado de EBS que utilizan una configuración de seguridad:

- Cifrado EBS: a partir de la versión 5.24.0 de Amazon EMR, puede optar por habilitar el cifrado EBS. La opción de cifrado de EBS cifra el volumen de dispositivo raíz de EBS y los volúmenes de almacenamiento adjuntos. La opción de cifrado de EBS está disponible solo cuando especifica AWS Key Management Service como su proveedor principal. Recomendamos el uso del cifrado de EBS.
- Cifrado LUKS: si decide utilizar el cifrado LUKS para los volúmenes de Amazon EBS, el cifrado LUKS se aplica únicamente a los volúmenes de almacenamiento adjuntos, no al volumen del dispositivo raíz. Para obtener más información sobre el cifrado de LUKS, consulte la especificación de LUKS en disco.

  Puede configurar una AWS KMS key con políticas adecuadas para Amazon EMR o una clase de Java personalizada que proporcione los artefactos de cifrado para su proveedor de claves. Cuando utilice AWS KMS, se aplican cargos por el almacenamiento y el uso de las claves de cifrado. Para obtener más información, consulte Precios de AWS KMS.

> **Note**
>
> Para comprobar si el cifrado de EBS está habilitado en el clúster, se recomienda utilizar una llamada a la API `DescribeVolumes`. Para obtener más información, consulte DescribeVolumes. La ejecución de `lsblk` en el clúster solo comprobará el estado del cifrado LUKS, en lugar del cifrado de EBS.

Cifrado en tránsito

Hay habilitados diversos mecanismos de cifrado con el cifrado en tránsito. Se trata de características de código abierto, específicas de la aplicación y pueden variar según la versión de Amazon EMR. Se pueden utilizar configuraciones de aplicación de Apache para habilitar las siguientes características de cifrado específicas de la aplicación: Para obtener más información, consulte Configurar aplicaciones.

Hadoop

- Hadoop MapReduce Encrypted Shuffle usa TLS.
- Secure Hadoop RPC se ha definido en “Privacy” y utiliza SASL (activado en Amazon EMR cuando se habilita el cifrado en reposo).
- El cifrado de datos en la transferencia de datos en bloques de HDFS usa AES 256 (activado en Amazon EMR cuando está habilitado el cifrado en reposo en la configuración de seguridad).
- Para obtener más información, consulte Hadoop in Secure Mode en la documentación de Apache Hadoop:

HBase

- Al habilitar Kerberos, la propiedad `hbase.rpc.protection` se configura en `privacy` para la comunicación privada.
- Para obtener más información, consulte Client-side configuration for secure operation en la documentación de Apache HBase.
- Para obtener más información sobre Kerberos con Amazon EMR, consulte Uso de Kerberos para la autenticación con Amazon EMR.

Hive

- La comunicación del cliente JDBC/ODBC con HiveServer2 (HS2) se cifra mediante configuraciones SSL en las versiones 6.9.0 y posteriores de Amazon EMR.
- Para obtener más información, consulte la sección SSL encryption de la documentación de Apache Hive.

Spark

- Las comunicaciones RPC internas entre componentes Spark, como el servicio de transferencia de bloques y el servicio de reorganización externo, se cifran mediante el cifrado AES-256 en las versiones 5.9.0 y posteriores de Amazon EMR. En versiones anteriores, las comunicaciones RPC internas se cifran mediante SASL con DIGEST-MD5 como cifrado.
- Las comunicaciones del protocolo HTTP con interfaces de usuario como Spark History Server y servidores de archivos compatibles con HTTPS se cifran mediante la configuración SSL de Spark. Para obtener más información, consulte SSL configuration en la documentación de Spark.
- Para obtener más información, consulte Spark security settings en la documentación de Apache Spark.

Tez

- Tez shuffle handler usa TLS (`tez.runtime.ssl.enable`).

Presto

- La comunicación interna entre nodos de Presto utiliza SSL/TLS (solo la versión 5.6.0 y posteriores de Amazon EMR).

Tiene que especificar los artefactos de cifrado utilizados para el cifrado en tránsito de una de estas dos maneras: facilitando un archivo comprimido con los certificados que carga en Amazon S3, o bien, haciendo referencia a una clase de Java personalizada que proporcione artefactos de cifrado. Para obtener más información, consulte Proporcionar certificados para el cifrado de datos en tránsito con el cifrado de Amazon EMR.

Creación de claves y certificados para el cifrado de datos

Antes de especificar las opciones de cifrado mediante una configuración de seguridad, decida el proveedor que desea usar para las claves y los artefactos de cifrado. Por ejemplo, puede utilizar AWS KMS o un proveedor personalizado que cree. A continuación, cree las claves o el proveedor tal y como se describe en esta sección.

Proporcionar claves para cifrado de datos en reposo con Amazon EMR

Puede utilizar AWS Key Management Service (AWS KMS) o un proveedor de claves personalizadas para el cifrado de datos en reposo en Amazon EMR. Cuando utilice AWS KMS, se aplican cargos por el almacenamiento y el uso de las claves de cifrado. Para obtener más información, consulte Precios de AWS KMS.

En este tema se ofrecen detalles sobre las políticas de claves para una clave de KMS que se vaya a usar con Amazon EMR, así como directrices y ejemplos de código para escribir una clase de proveedor de claves personalizadas para el cifrado de Amazon S3. Para más información sobre la creación de claves, consulte Creación de claves en la Guía para desarrolladores de AWS Key Management Service.

Uso de AWS KMS keys para el cifrado

La clave de cifrado de AWS KMS debe estar creada en la misma región que su instancia de clúster de Amazon EMR y los buckets de Amazon S3 utilizados con EMRFS. Si la clave que especifica está en una cuenta diferente de la que usa para configurar un clúster, debe especificar la clave mediante su ARN.

El rol del perfil de instancia de Amazon EC2 debe tener permisos para usar la clave de KMS que especifique. El rol predeterminado del perfil de instancia en Amazon EMR es `EMR_EC2_DefaultRole`. Si usa un rol diferente para el perfil de instancia o usa roles de IAM para las solicitudes de EMRFS a Amazon S3, asegúrese de agregar cada rol como usuario clave, según corresponda. Esto proporciona al rol los permisos para utilizar la clave de KMS. Para obtener más información, consulte Uso de políticas de claves en la Guía para desarrolladores de AWS Key Management Service y Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3.

Puede utilizar la AWS Management Console para agregar su perfil de instancia o el perfil de instancia de EC2 a la lista de usuarios de la clave de KMS especificada. También puede usar la AWS CLI o un AWS SDK para adjuntar una política de claves adecuada.

Tenga en cuenta que Amazon EMR solo admite claves de KMS simétricas. No se puede utilizar una clave KMS asimétrica para cifrar datos en reposo en un clúster de Amazon EMR. Para obtener ayuda para determinar si una clave de KMS es simétrica o asimétrica, consulte Identificación de claves de KMS simétricas y asimétricas.

El siguiente procedimiento describe cómo agregar el perfil de instancia de Amazon EMR predeterminado, `EMR_EC2_DefaultRole` como un usuario clave con la AWS Management

Console. Se supone que ya ha creado una clave de KMS. Para crear una nueva clave de KMS, consulte Creación de claves en la Guía para desarrolladores de AWS Key Management Service.

Para agregar el perfil de instancia de EC2 para Amazon EMR a la lista de usuarios de claves de cifrado

1. Inicie sesión en la AWS Management Console y abra la consola AWS Key Management Service (AWS KMS) en https://console.aws.amazon.com/kms.
2. Para cambiar la Región de AWS, utilice el Selector de regiones ubicado en la esquina superior derecha de la página.
3. Seleccione el alias de la clave de KMS que desee modificar.
4. En la página de detalles de la clave, en Key Users (Usuarios de claves), seleccione Add (Añadir).
5. En el cuadro de diálogo Add key users (Añadir usuarios clave), seleccione el rol adecuado. El nombre del rol predeterminado es `EMR_EC2_DefaultRole`.
6. Elija Add (Agregar).

Habilitación del cifrado de EBS proporcionando permisos adicionales para las claves de KMS

A partir de la versión 5.24.0 de Amazon EMR, puede cifrar el dispositivo raíz y los volúmenes de almacenamiento de EBS utilizando una opción de configuración de seguridad. Para habilitar esa opción, debe especificar a AWS KMS como su proveedor de claves. Además, debe conceder el rol de servicio de EMR `EMR_DefaultRole` con permisos para utilizar la AWS KMS key que especifique.

Puede utilizar la AWS Management Console para agregar el rol de servicio de EMR a la lista de usuarios clave de la clave de KMS especificada, o bien puede usar la AWS o un AWS CLI SDK para adjuntar una política de claves adecuada.

El siguiente procedimiento describe cómo añadir el rol de servicio de EMR predeterminado, `EMR_DefaultRole`, como un usuario de claves con la AWS Management Console. Se supone que ya ha creado una clave de KMS. Para crear una nueva clave de KMS, consulte Creación de claves en la Guía para desarrolladores de AWS Key Management Service.

Para añadir el rol de servicio de EMR a la lista de usuarios de claves de cifrado

1. Inicie sesión en la AWS Management Console y abra la consola AWS Key Management Service (AWS KMS) en https://console.aws.amazon.com/kms.

2. Para cambiar la Región de AWS, utilice el Selector de regiones ubicado en la esquina superior derecha de la página.
3. Elija Claves administradas por el cliente en la barra lateral izquierda.
4. Seleccione el alias de la clave de KMS que desee modificar.
5. En la página de detalles de la clave, en Key Users (Usuarios de claves), seleccione Add (Añadir).
6. En el cuadro de diálogo Add key users (Añadir usuarios clave), seleccione el rol adecuado. El nombre del rol de servicio de EMR es `EMR_DefaultRole`.
7. Elija Add (Agregar).

Creación de un proveedor de claves personalizadas

Cuando se utiliza una configuración de seguridad, es necesario especificar un nombre de clase de proveedor distinto para el cifrado de disco local y el cifrado de Amazon S3.

Cuando se crea un proveedor de claves personalizadas, se espera que la aplicación implemente la interfaz EncryptionMaterialsProvider, que está disponible en la versión 1.11.0 y posteriores de AWS SDK for Java. La aplicación puede utilizar cualquier estrategia para proporcionar materiales de cifrado. Puede, por ejemplo, elegir proporcionar materiales de cifrado estáticos o integrar con un sistema de administración de claves más complejo.

El algoritmo de cifrado utilizado para los materiales de cifrado personalizados debe ser AES/GCM/NoPadding.

La clase EncryptionMaterialsProvider obtiene los materiales de cifrado por contexto de cifrado. Amazon EMR rellena el contexto de cifrado en tiempo de ejecución para ayudar al intermediario a determinar qué materiales de cifrado debe devolver.

Example Ejemplo: uso de un proveedor de claves de cifrado personalizadas para el cifrado de Amazon S3 con EMRFS

Cuando Amazon EMR recupera los materiales de cifrado desde la clase EncryptionMaterialsProvider para realizar el cifrado, EMRFS rellena opcionalmente el argumento materialsDescription con dos campos: el URI de Amazon S3 para el objeto y JobFlowId del clúster, que puede utilizar la clase EncryptionMaterialsProvider para devolver materiales de cifrado de manera selectiva.

Por ejemplo, el proveedor podría devolver claves distintas para diferentes prefijos URI de Amazon S3. Se trata de la descripción de los materiales de cifrado devuelta que se almacena finalmente con

el objeto de Amazon S3 en lugar del valor materialsDescription que genera EMRFS y se transfiere al proveedor. Al descifrar un objeto de Amazon S3, la descripción de materiales de cifrado se transfiere a la clase EncryptionMaterialsProvider, de modo que puede, nuevamente, devolver de forma selectiva la clave coincidente para descifrar el objeto.

A continuación, se proporciona una implementación de referencia de EncryptionMaterialsProvider. Otro proveedor personalizado, EMRFSRSAEncryptionMaterialsProvider, está disponible desde 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;
  }
}
```

Proporcionar certificados para el cifrado de datos en tránsito con el cifrado de Amazon EMR

Con la versión 4.8.0 o posterior de Amazon EMR, dispone de dos opciones para especificar artefactos para el cifrado de datos en tránsito utilizando una configuración de seguridad:

- Puede crear manualmente certificados PEM, incluirlos en un archivo .zip y, a continuación, hacer referencia al archivo .zip en Amazon S3.
- Puede implementar un proveedor de certificados personalizado como una clase Java. Deberá especificar el archivo JAR de la aplicación en Amazon S3 y, a continuación, proporcionar el nombre de la clase completa del proveedor tal como se declara en la aplicación. La clase debe implementar la interfaz TLSArtifactsProvider disponible a partir del AWS SDK for Java versión 1.11.0.

Amazon EMR descarga automáticamente artefactos en cada nodo del clúster y, posteriormente, los utiliza para implementar las características de cifrado en tránsito de código abierto. Para obtener más información sobre las opciones disponibles, consulte Cifrado en tránsito.

Uso de certificados PEM

Cuando especifique un archivo zip para el cifrado en tránsito, la configuración de seguridad espera que los archivos PEM dentro del archivo zip se nombren exactamente tal y como aparecen a continuación:

Certificados de cifrado en tránsito

Nombre de archivo	Obligatorio/opcional	Detalles
privateKey.pem	Obligatorio	Clave privada
certificateChain.pem	Obligatorio	Cadena de certificados
trustedCertificates.pem	Opcional	Obligatorio si el certificado proporcionado no está firmado por la autoridad de certificación (CA) raíz de confianza predeterminada de Java o una CA intermedia que enlace a la CA de raíz de confianza predeterminada de Java. Las CA raíz de confianza predeterminadas de Java pueden encontrarse en `jre/lib/security/cacerts` .

Es probable que desee configurar el archivo PEM de clave privada como un certificado comodín que permita el acceso al dominio de Amazon VPC en el que residen las instancias del clúster. Por ejemplo, si el clúster se encuentran en la región us-east-1 (Norte de Virginia), puede elegir especificar un nombre común en la configuración de certificado que permita el acceso al clúster especificando `CN=*.ec2.internal` en la definición del sujeto del certificado. Si el clúster reside en us-west-2 (Oregón), puede especificar `CN=*.us-west-2.compute.internal`.

Si el archivo PEM proporcionado en el artefacto de cifrado no tiene un carácter comodín en el nombre común (CN) del dominio, debe cambiar el valor de `hadoop.ssl.hostname.verifier` a `ALLOW_ALL`. Esto se hace con la clasificación `core-site` al enviar las configuraciones a un clúster o al agregar este valor en el archivo `core-site.xml`. Este cambio es necesario porque el verificador de nombres de host predeterminado no aceptará un nombre de host sin el comodín, lo que provocará un error. Para obtener más información sobre la configuración del clúster de EMR en una instancia de Amazon VPC, consulte Configurar redes.

En el siguiente ejemplo, se muestra cómo utilizar OpenSSL para generar un certificado X.509 autofirmado con una clave privada RSA de 1024 bits. La clave permite el acceso a las instancias de clúster de Amazon EMR en la región `us-west-2` (Oregón), tal y como se especifica mediante el nombre de dominio *`*.us-west-2.compute.internal`* como nombre común.

Se especifican otros elementos del sujeto opcionales como país (C), estado (S) y configuración regional (L). Dado que se genera un certificado autofirmado, el segundo comando del ejemplo copia el archivo `certificateChain.pem` en el archivo `trustedCertificates.pem`. El tercer comando utiliza `zip` para crear el archivo `my-certs.zip` que contiene los certificados.

> **Important**
>
> Este ejemplo es solo una demostración de la prueba de concepto. El uso de certificados autofirmados no se recomienda y presenta un posible riesgo para la seguridad. En el caso de los sistemas de producción, utilice una autoridad de certificación (CA) de confianza para emitir certificados.

```
$ openssl req -x509 -newkey rsa:1024 -keyout privateKey.pem -out certificateChain.pem
 -days 365 -nodes -subj '/C=US/ST=Washington/L=Seattle/O=MyOrg/OU=MyDept/CN=*.us-
west-2.compute.internal'
$ cp certificateChain.pem trustedCertificates.pem
$ zip -r -X my-certs.zip certificateChain.pem privateKey.pem trustedCertificates.pem
```

AWS Identity and Access Management para Amazon EMR

AWS Identity and Access Management (IAM) es un Servicio de AWS que ayuda a los administradores a controlar de forma segura el acceso a los recursos de AWS. Los administradores de IAM controlan quién se puede autenticar (iniciar sesión) y autorizar (tener permisos) para utilizar los recursos de Amazon EMR. IAM es un Servicio de AWS que se puede utilizar sin cargo adicional.

Temas

- Público
- Autenticación con identidades
- Administración de acceso mediante políticas
- Cómo funciona Amazon EMR con IAM

- Roles en tiempo de ejecución para los pasos de Amazon EMR
- Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS
- Ejemplos de políticas de Amazon EMR basadas en identidades

Público

La forma en que utilice AWS Identity and Access Management (IAM) varía en función del trabajo que realice en Amazon EMR.

Usuario de servicio: si utiliza el servicio Amazon EMR para realizar el trabajo, el administrador proporciona las credenciales y los permisos que necesita. A medida que utilice más características de Amazon EMR para realizar el trabajo, es posible que necesite permisos adicionales. Entender cómo se administra el acceso puede ayudarlo a solicitar los permisos correctos a su administrador. Si no puede acceder a una característica de Amazon EMR, consulte Solución de problemas de identidad y acceso de Amazon EMR.

Administrador de servicio: si está a cargo de los recursos de Amazon EMR de la empresa, probablemente tenga acceso completo a Amazon EMR. El trabajo consiste en determinar a qué características y recursos de Amazon EMR deben acceder los usuarios del servicio. Luego, debe enviar solicitudes a su administrador de IAM para cambiar los permisos de los usuarios de su servicio. Revise la información de esta página para conocer los conceptos básicos de IAM. Para obtener más información acerca de cómo la empresa puede utilizar IAM con Amazon EMR, consulte Cómo funciona Amazon EMR con IAM.

Administrador de IAM: si es un administrador de IAM, es posible que desee obtener información sobre cómo escribir políticas para administrar el acceso a Amazon EMR. Para consultar ejemplos de políticas de Amazon EMR basadas en identidades que puede utilizar en IAM, consulte Ejemplos de políticas de Amazon EMR basadas en identidades.

Autenticación con identidades

La autenticación es la manera de iniciar sesión en AWS mediante credenciales de identidad. Debe estar autenticado (haber iniciado sesión en AWS) como el Usuario raíz de la cuenta de AWS, como un usuario de IAM o asumiendo un rol de IAM.

Puede iniciar sesión en AWS como una identidad federada mediante las credenciales proporcionadas a través de una fuente de identidad. AWS IAM Identity Center Los usuarios (IAM

Identity Center), la autenticación de inicio de sesión único de su empresa y sus credenciales de Google o Facebook son ejemplos de identidades federadas. Al iniciar sesión como identidad federada, su administrador habrá configurado previamente la federación de identidades mediante roles de IAM. Cuando accede a AWS mediante la federación, está asumiendo un rol de forma indirecta.

Según el tipo de usuario que sea, puede iniciar sesión en la AWS Management Console o en el portal de acceso AWS. Para obtener más información sobre el inicio de sesión en AWS, consulte Cómo iniciar sesión en su Cuenta de AWS en la Guía del usuario de AWS Sign-In.

Si accede a AWS mediante programación, AWS proporciona un kit de desarrollo de software (SDK) y una interfaz de línea de comandos (CLI) para firmar criptográficamente las solicitudes mediante el uso de las credenciales. Si no usa las herramientas de AWS, debe firmar usted mismo las solicitudes. Para obtener más información sobre la firma de solicitudes, consulte Firma de solicitudes API de AWS en la Guía del usuario de IAM.

Independientemente del método de autenticación que use, es posible que deba proporcionar información de seguridad adicional. Por ejemplo, AWS le recomienda el uso de la autenticación multifactor (MFA) para aumentar la seguridad de su cuenta. Para obtener más información, consulte Autenticación multifactor en la Guía del usuario de AWS IAM Identity Center y Uso de la autenticación multifactor (MFA) en AWS en la Guía del usuario de IAM.

Usuario raíz de Cuenta de AWS

Cuando se crea una Cuenta de AWS, se comienza con una identidad de inicio de sesión que tiene acceso completo a todos los recursos y Servicios de AWS de la cuenta. Esta identidad recibe el nombre de usuario raíz de la Cuenta de AWS y se accede a ella iniciando sesión con el email y la contraseña que utilizó para crear la cuenta. Recomendamos que no utilice el usuario raíz para las tareas cotidianas. Proteja las credenciales del usuario raíz y utilícelas solo para las tareas que este pueda realizar. Para ver la lista completa de las tareas que requieren que inicie sesión como usuario raíz, consulte Tareas que requieren credenciales de usuario raíz en la Guía de usuario de IAM.

Identidad federada

Como práctica recomendada, solicite que los usuarios humanos, incluidos los que requieren acceso de administrador, utilicen la federación con un proveedor de identidades para acceder a Servicios de AWS utilizando credenciales temporales.

Una identidad federada es un usuario del directorio de usuarios de su empresa, un proveedor de identidad web, el AWS Directory Service, el directorio del Identity Center, o cualquier usuario que

acceda a Servicios de AWS utilizando credenciales proporcionadas a través de una fuente de identidad. Cuando identidades federadas acceden a Cuentas de AWS, asumen roles y los roles proporcionan credenciales temporales.

Para una administración de acceso centralizada, le recomendamos que utilice AWS IAM Identity Center. Puede crear usuarios y grupos en el IAM Identity Center o puede conectarse y sincronizar con un conjunto de usuarios y grupos de su propia fuente de identidad para usarlos en todas su sus aplicaciones y Cuentas de AWS. Para obtener más información, consulte ¿Qué es IAM Identity Center? en la Guía del usuario de AWS IAM Identity Center.

Usuarios y grupos de IAM

Un usuario de IAM es una identidad de la Cuenta de AWS que dispone de permisos específicos para una sola persona o aplicación. Siempre que sea posible, recomendamos emplear credenciales temporales, en lugar de crear usuarios de IAM que tengan credenciales de larga duración como contraseñas y claves de acceso. No obstante, si tiene casos de uso específicos que requieran credenciales de larga duración con usuarios de IAM, recomendamos rotar las claves de acceso. Para obtener más información, consulte Rotar las claves de acceso periódicamente para casos de uso que requieran credenciales de larga duración en la Guía del usuario de IAM.

Un grupo de IAM es una identidad que especifica un conjunto de usuarios de IAM. No puede iniciar sesión como grupo. Puede usar los grupos para especificar permisos para varios usuarios a la vez. Los grupos facilitan la administración de los permisos de grandes conjuntos de usuarios. Por ejemplo, podría tener un grupo cuyo nombre fuese IAMAdmins y conceder permisos a dicho grupo para administrar los recursos de IAM.

Los usuarios son diferentes de los roles. Un usuario se asocia exclusivamente a una persona o aplicación, pero la intención es que cualquier usuario pueda asumir un rol que necesite. Los usuarios tienen credenciales permanentes a largo plazo y los roles proporcionan credenciales temporales. Para obtener más información, consulte Cuándo crear un usuario de IAM (en lugar de un rol) en la Guía del usuario de IAM.

IAM roles

Un rol de IAM es una identidad de la Cuenta de AWS que dispone de permisos específicos. Es similar a un usuario de IAM, pero no está asociado a una determinada persona. Puede asumir temporalmente un rol de IAM en la AWS Management Consolecambiando de roles. Puede asumir un rol llamando a una operación de la AWS CLI o de la API de AWS, o utilizando una URL

personalizada. Para obtener más información acerca de los métodos para el uso de roles, consulte Uso de roles de IAM en la Guía del usuario de IAM.

Los roles de IAM con credenciales temporales son útiles en las siguientes situaciones:

- Acceso de usuario federado: para asignar permisos a una identidad federada, puede crear un rol y definir permisos para este. Cuando se autentica una identidad federada, se asocia la identidad al rol y se le conceden los permisos que están definidos en este. Para obtener información acerca de roles para federación, consulte Creación de un rol para un proveedor de identidades de terceros en la Guía del usuario de IAM. Si utiliza el Centro de identidades de IAM, debe configurar un conjunto de permisos. El Centro de identidades de IAM correlaciona el conjunto de permisos con un rol en IAM para controlar a qué pueden acceder las identidades después de autenticarse. Para obtener información acerca de los conjuntos de permisos, consulte Conjuntos de permisos en la Guía del usuario de AWS IAM Identity Center.
- Permisos de usuario de IAM temporales: un usuario de IAM puede asumir un rol de IAM para recibir temporalmente permisos distintos que le permitan realizar una tarea concreta.
- Acceso entre cuentas: puede utilizar un rol de IAM para permitir que alguien (una entidad principal de confianza) de otra cuenta acceda a los recursos de la cuenta. Los roles son la forma principal de conceder acceso entre cuentas. No obstante, con algunos Servicios de AWS se puede adjuntar una política directamente a un recurso (en lugar de utilizar un rol como representante). Para obtener información acerca de la diferencia entre los roles y las políticas basadas en recursos para el acceso entre cuentas, consulte Cómo los roles de IAM difieren de las políticas basadas en recursos en la Guía del usuario de IAM.
- Acceso entre servicios: algunos Servicios de AWS utilizan características de otros Servicios de AWS. Por ejemplo, cuando realiza una llamada en un servicio, es común que ese servicio ejecute aplicaciones en Amazon EC2 o almacene objetos en Amazon S3. Es posible que un servicio haga esto usando los permisos de la entidad principal, usando un rol de servicio o usando un rol vinculado a servicios.
  - Permisos principales: cuando utiliza un usuario o un rol de IAM para llevar a cabo acciones en AWS, se lo considera una entidad principal. Las políticas conceden permisos a una entidad principal. Cuando utiliza algunos servicios, es posible que realice una acción que desencadene otra acción en un servicio diferente. En este caso, debe tener permisos para realizar ambas acciones. Para ver si una acción requiere acciones dependientes adicionales en una política, consulte Acciones, recursos y claves de condición para Amazon EMR en la Referencia de autorizaciones de servicio.

  - Rol de servicio: un rol de servicio es un rol de IAM que adopta un servicio para realizar acciones en su nombre. Un administrador de IAM puede crear, modificar y eliminar un rol de servicio desde IAM. Para obtener más información, consulte Creación de un rol para delegar permisos a un Servicio de AWS en la Guía del usuario de IAM.
  - Rol vinculado a servicio: un rol vinculado a servicio es un tipo de rol de servicio que está vinculado a un Servicio de AWS. El servicio puede asumir el rol para realizar una acción en su nombre. Los roles vinculados a servicios aparecen en la Cuenta de AWS y son propiedad del servicio. Un administrador de IAM puede ver, pero no editar, los permisos de los roles vinculados a servicios.
- Aplicaciones que se ejecutan en Amazon EC2: puede utilizar un rol de IAM que le permita administrar credenciales temporales para las aplicaciones que se ejecutan en una instancia de EC2 y realizan solicitudes a la AWS CLI o a la API de AWS. Es preferible hacerlo de este modo a almacenar claves de acceso en la instancia de EC2. Para asignar un rol de AWS a una instancia de EC2 y ponerla a disposición de todas las aplicaciones, cree un perfil de instancia adjuntado a la instancia. Un perfil de instancia contiene el rol y permite a los programas que se ejecutan en la instancia de EC2 obtener credenciales temporales. Para obtener más información, consulte Uso de un rol de IAM para conceder permisos a aplicaciones que se ejecutan en instancias Amazon EC2 en la Guía del usuario de IAM.

Para obtener información sobre el uso de los roles de IAM, consulte Cuándo crear un rol de IAM (en lugar de un usuario) en la Guía del usuario de IAM.

Administración de acceso mediante políticas

Para controlar el acceso en AWS, se crean políticas y se asocian a identidades o recursos de AWS. Una política es un objeto de AWS que, cuando se asocia a una identidad o un recurso, define sus permisos. AWS evalúa estas políticas cuando una entidad principal (sesión de rol, usuario o usuario raíz) realiza una solicitud. Los permisos en las políticas determinan si la solicitud se permite o se deniega. Las mayoría de las políticas se almacenan en AWS como documentos JSON. Para obtener más información sobre la estructura y el contenido de los documentos de política JSON, consulte Información general de políticas JSON en la Guía del usuario de IAM.

Los administradores pueden utilizar las políticas JSON de AWS para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y bajo qué condiciones.

De forma predeterminada, los usuarios y los roles no tienen permisos. Un administrador de IAM puede crear políticas de IAM para conceder permisos a los usuarios para realizar acciones en los

recursos que necesitan. A continuación, el administrador puede agregar las políticas de IAM a los roles y los usuarios pueden asumirlos.

Las políticas de IAM definen permisos para una acción independientemente del método que se utilice para realizar la operación. Por ejemplo, suponga que dispone de una política que permite la acción `iam:GetRole`. Un usuario con dicha política puede obtener información del usuario de la AWS Management Console, la AWS CLI o la API de AWS.

Políticas basadas en identidad

Las políticas basadas en identidad son documentos de políticas de permisos JSON que puede adjuntar a una identidad, como un usuario de IAM, un grupo de usuarios o un rol. Estas políticas controlan qué acciones pueden realizar los usuarios y los roles, en qué recursos y bajo qué condiciones. Para obtener más información sobre cómo crear una política basada en identidad, consulte Creación de políticas de IAM en la Guía del usuario de IAM.

Las políticas basadas en identidad pueden clasificarse además como políticas insertadas o políticas administradas. Las políticas insertadas se integran directamente en un único usuario, grupo o rol. Las políticas administradas son políticas independientes que puede adjuntar a varios usuarios, grupos y roles de su Cuenta de AWS. Las políticas administradas incluyen las políticas administradas por AWS y las políticas administradas por el cliente. Para obtener más información acerca de cómo elegir una política administrada o una política insertada, consulte Elegir entre políticas administradas y políticas insertadas en la Guía del usuario de IAM.

Políticas basadas en recursos

Las políticas basadas en recursos son documentos de política JSON que se adjuntan a un recurso. Ejemplos de políticas basadas en recursos son las políticas de confianza de roles de IAM y las políticas de bucket de Amazon S3. En los servicios que admiten políticas basadas en recursos, los administradores de servicios pueden utilizarlos para controlar el acceso a un recurso específico. Para el recurso al que se adjunta la política, la política define qué acciones puede realizar una entidad principal especificada en ese recurso y en qué condiciones. Debe especificar una entidad principal en una política basada en recursos. Las entidades principales pueden incluir cuentas, usuarios, roles, usuarios federados o Servicios de AWS.

Las políticas basadas en recursos son políticas insertadas que se encuentran en ese servicio. No se puede utilizar políticas de IAM administradas por AWS en una política basada en recursos.

Listas de control de acceso (ACL)

Las listas de control de acceso (ACL) controlan qué entidades principales (miembros de cuentas, usuarios o roles) tienen permisos para acceder a un recurso. Las ACL son similares a las políticas basadas en recursos, aunque no utilizan el formato de documento de política JSON.

Amazon S3, AWS WAF y Amazon VPC son ejemplos de servicios que admiten las ACL. Para obtener más información sobre las ACL, consulte Información general de Lista de control de acceso (ACL) en la Guía para desarrolladores de Amazon Simple Storage Service.

Otros tipos de políticas

AWS admite otros tipos de políticas adicionales menos frecuentes. Estos tipos de políticas pueden establecer el máximo de permisos que los tipos de políticas más frecuentes le otorgan.

- Límites de permisos: un límite de permisos es una característica avanzada que le permite establecer los permisos máximos que una política basada en identidad puede conceder a una entidad de IAM (usuario o rol de IAM). Puede establecer un límite de permisos para una identidad. Los permisos resultantes son la intersección de las políticas basadas en identidad de la entidad y los límites de permisos. Las políticas basadas en recursos que especifiquen el usuario o rol en el campo `Principal` no estarán restringidas por el límite de permisos. Una denegación explícita en cualquiera de estas políticas anulará el permiso. Para obtener más información sobre los límites de los permisos, consulte Límites de permisos para las entidades de IAM en la Guía del usuario de IAM.
- Políticas de control de servicio (SCP): las SCP son políticas de JSON que especifican los permisos máximos de una organización o una unidad organizativa en AWS Organizations. AWS Organizations es un servicio que le permite agrupar y administrar de manera centralizada varias Cuentas de AWS que posea su empresa. Si habilita todas las características en una organización, entonces podrá aplicar políticas de control de servicio (SCP) a una o todas sus cuentas. Una SCP limita los permisos para las entidades de las cuentas de miembros, incluido cada Usuario raíz de la cuenta de AWS. Para obtener más información acerca de Organizations y las SCP, consulte Funcionamiento de las SCP en la Guía del usuario de AWS Organizations.
- Políticas de sesión: las políticas de sesión son políticas avanzadas que se pasan como parámetro cuando se crea una sesión temporal mediante programación para un rol o un usuario federado. Los permisos de la sesión resultantes son la intersección de las políticas basadas en identidad del rol y las políticas de la sesión. Los permisos también pueden proceder de una política basada en recursos. Una denegación explícita en cualquiera de estas políticas anulará el permiso. Para obtener más información, consulte Políticas de sesión en la Guía del usuario de IAM.

Varios tipos de políticas

Cuando se aplican varios tipos de políticas a una solicitud, los permisos resultantes son más complicados de entender. Para obtener información acerca de cómo AWS decide si permitir o no una solicitud cuando hay varios tipos de políticas implicados, consulte Lógica de evaluación de políticas en la Guía del usuario de IAM.

Cómo funciona Amazon EMR con IAM

Antes de utilizar IAM para administrar el acceso a Amazon EMR, obtenga información sobre qué características de IAM se encuentran disponibles con Amazon EMR.

Características de IAM que puede utilizar con Amazon EMR

Características de IAM	Compatibilidad con Amazon EMR
Políticas basadas en identidad	Sí
Políticas basadas en recursos	Sí
Acciones de política	Sí
Recursos de políticas	Sí
Claves de condiciones de políticas	Sí
ACL	No
ABAC (etiquetas en políticas)	Sí
Credenciales temporales	Sí
Permisos de entidades principales	Sí
Roles de servicio	No
Roles vinculados a servicios	Sí

Para obtener una perspectiva general sobre cómo funcionan Amazon EMR y otros productos de AWS con la mayoría de las características de IAM, consulte Servicios de AWS que funcionan con IAM en la Guía del usuario de IAM.

Políticas de Amazon EMR basadas en identidades

Compatibilidad con las políticas basadas en identidad	Sí

Las políticas basadas en identidad son documentos de políticas de permisos JSON que puede adjuntar a una identidad, como un usuario de IAM, un grupo de usuarios o un rol. Estas políticas controlan qué acciones pueden realizar los usuarios y los roles, en qué recursos y bajo qué condiciones. Para obtener más información sobre cómo crear una política basada en identidad, consulte Creación de políticas de IAM en la Guía del usuario de IAM.

Con las políticas basadas en identidad de IAM, puede especificar las acciones y recursos permitidos o denegados, así como las condiciones en las que se permiten o deniegan las acciones. No es posible especificar la entidad principal en una política basada en identidad porque se aplica al usuario o rol al que está asociado. Para obtener más información acerca de los elementos que puede utilizar en una política de JSON, consulte Referencia de los elementos de las políticas de JSON de IAM en la Guía del usuario de IAM.

Ejemplos de políticas basadas en identidades para Amazon EMR

Para ver ejemplos de políticas basadas en identidades de Amazon EMR, consulte Ejemplos de políticas de Amazon EMR basadas en identidades.

Políticas basadas en recursos de Amazon EMR

Compatibilidad con las políticas basadas en recursos	Sí

Las políticas basadas en recursos son documentos de política JSON que se adjuntan a un recurso. Ejemplos de políticas basadas en recursos son las políticas de confianza de roles de IAM y las políticas de bucket de Amazon S3. En los servicios que admiten políticas basadas en recursos, los administradores de servicios pueden utilizarlos para controlar el acceso a un recurso específico. Para el recurso al que se adjunta la política, la política define qué acciones puede realizar una entidad principal especificada en ese recurso y en qué condiciones. Debe especificar una entidad principal en una política basada en recursos. Las entidades principales pueden incluir cuentas, usuarios, roles, usuarios federados o Servicios de AWS.

Para habilitar el acceso entre cuentas, puede especificar toda una cuenta o entidades de IAM de otra cuenta como la entidad principal de una política basada en recursos. Añadir a una política basada en recursos una entidad principal entre cuentas es solo una parte del establecimiento de una relación de confianza. Cuando la entidad principal y el recurso se encuentran en Cuentas de AWS diferentes, un administrador de IAM de la cuenta de confianza también debe conceder a la entidad principal (usuario o rol) permiso para acceder al recurso. Para conceder el permiso, asocie la entidad a una política basada en identidad. Sin embargo, si la política basada en recursos concede el acceso a una entidad principal de la misma cuenta, no es necesaria una política basada en identidad adicional. Para obtener más información, consulte Cómo los roles de IAM difieren de las políticas basadas en recursos en la Guía del usuario de IAM.

Acciones de políticas para Amazon EMR

Admite acciones de política	Sí

Los administradores pueden utilizar las políticas JSON de AWS para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y bajo qué condiciones.

El elemento `Action` de una política JSON describe las acciones que puede utilizar para permitir o denegar el acceso en una política. Las acciones de la política generalmente tienen el mismo nombre que la operación de API de AWS asociada. Hay algunas excepciones, como acciones de solo permiso que no tienen una operación de API coincidente. También hay algunas operaciones que requieren varias acciones en una política. Estas acciones adicionales se denominan acciones dependientes.

Incluya acciones en una política para conceder permisos y así llevar a cabo la operación asociada.

A fin de conocer una lista completa de acciones de políticas para Amazon EMR, consulte Acciones, recursos y claves de condición para Amazon EMR en la Referencia de autorizaciones de servicio.

Las acciones de políticas de Amazon EMR utilizan el siguiente prefijo antes de la acción:

```
EMR
```

Para especificar varias acciones en una única instrucción, sepárelas con comas.

```
"Action": [
```

```
      "EMR:action1",
      "EMR:action2"
         ]
```

Para ver ejemplos de políticas basadas en identidades de Amazon EMR, consulte Ejemplos de políticas de Amazon EMR basadas en identidades.

Recursos de políticas para Amazon EMR

Admite recursos de políticas	Sí

Los administradores pueden utilizar las políticas JSON de AWS para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y bajo qué condiciones.

El elemento `Resource` de la política JSON especifica el objeto u objetos a los que se aplica la acción. Las instrucciones deben contener un elemento `Resource` o `NotResource`. Como práctica recomendada, especifique un recurso utilizando el Nombre de recurso de Amazon (ARN). Puede hacerlo para acciones que admitan un tipo de recurso específico, conocido como permisos de nivel de recurso.

Para las acciones que no admiten permisos de nivel de recurso, como las operaciones de descripción, utilice un carácter comodín (*) para indicar que la instrucción se aplica a todos los recursos.

```
"Resource": "*"
```

Para ver una lista de tipos de recursos de Amazon EMR y sus ARN, consulte Tipos de recurso definidos por Amazon EMR en la Referencia de autorizaciones de servicio. Para saber con qué acciones puede especificar el ARN de cada recurso, consulte Acciones, recursos y claves de condición para Amazon EMR.

Para ver ejemplos de políticas basadas en identidades de Amazon EMR, consulte Ejemplos de políticas de Amazon EMR basadas en identidades.

Claves de condición de políticas para Amazon EMR

Admite claves de condición de política específicas del servicio	Sí

Los administradores pueden utilizar las políticas JSON de AWS para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y bajo qué condiciones.

El elemento `Condition` (o bloque de `Condition`) permite especificar condiciones en las que entra en vigor una instrucción. El elemento `Condition` es opcional. Puede crear expresiones condicionales que utilicen operadores de condición, tales como igual o menor que, para que la condición de la política coincida con los valores de la solicitud.

Si especifica varios elementos de `Condition` en una instrucción o varias claves en un único elemento de `Condition`, AWS las evalúa mediante una operación `AND` lógica. Si especifica varios valores para una única clave de condición, AWS evalúa la condición con una operación lógica `OR`. Se deben cumplir todas las condiciones antes de que se concedan los permisos de la instrucción.

También puede utilizar variables de marcador de posición al especificar condiciones. Por ejemplo, puede conceder un permiso de usuario de IAM para acceder a un recurso solo si está etiquetado con su nombre de usuario de IAM. Para obtener más información, consulte Elementos de la política de IAM: variables y etiquetas en la Guía del usuario de IAM.

AWS admite claves de condición globales y claves de condición específicas del servicio. Para ver todas las claves de condición globales de AWS, consulte Claves de contexto de condición globales de AWS en la Guía del usuario de IAM.

Para obtener una lista de las claves de condición de Amazon EMR y para más información sobre qué acciones y recursos admiten el uso de una clave de condición, consulte Acciones, recursos y claves de condición para Amazon EMR en la Referencia de autorizaciones de servicio.

Para ver ejemplos de políticas basadas en identidades de Amazon EMR, consulte Ejemplos de políticas de Amazon EMR basadas en identidades.

Listas de control de acceso (ACL) de Amazon EMR

Admite las ACL	No

Las listas de control de acceso (ACL) controlan qué entidades principales (miembros de cuentas, usuarios o roles) tienen permisos para acceder a un recurso. Las ACL son similares a las políticas basadas en recursos, aunque no utilizan el formato de documento de política JSON.

Control de acceso basado en atributos (ABAC) con Amazon EMR

Admite ABAC (etiquetas en las políticas)	Sí

El control de acceso basado en atributos (ABAC) es una estrategia de autorización que define permisos basados en atributos. En AWS, estos atributos se denominan etiquetas. Puede asociar etiquetas a entidades de IAM (usuarios o roles) y a muchos recursos de AWS. El etiquetado de entidades y recursos es el primer paso de ABAC. A continuación, designa las políticas de ABAC para permitir operaciones cuando la etiqueta de la entidad principal coincida con la etiqueta del recurso al que se intenta acceder.

ABAC es útil en entornos que crecen con rapidez y ayuda en situaciones en las que la administración de las políticas resulta engorrosa.

Para controlar el acceso en función de etiquetas, debe proporcionar información de las etiquetas en el elemento de condición de una política utilizando las claves de condición `aws:ResourceTag/key-name`, `aws:RequestTag/key-name` o `aws:TagKeys`.

Si un servicio admite las tres claves de condición para cada tipo de recurso, el valor es Sí para el servicio. Si un servicio admite las tres claves de condición solo para algunos tipos de recursos, el valor es Parcial.

Para obtener más información sobre ABAC, consulte ¿Qué es ABAC? en la Guía del usuario de IAM. Para ver un tutorial con los pasos para configurar ABAC, consulte Uso del control de acceso basado en atributos (ABAC) en la Guía del usuario de IAM.

Uso de credenciales temporales con Amazon EMR

Compatible con el uso de credenciales temporales.	Sí

Algunos servicios de Servicios de AWS no funcionan cuando inicia sesión con credenciales temporales. Para obtener información adicional, incluida la información sobre qué servicios de

Servicios de AWS funcionan con credenciales temporales, consulte Servicios de Servicios de AWS que funcionan con IAM en la Guía del usuario de IAM.

Utiliza credenciales temporales si inicia sesión en la AWS Management Console con cualquier método excepto un nombre de usuario y una contraseña. Por ejemplo, cuando accede a AWS utilizando el enlace de inicio de sesión único (SSO) de la empresa, ese proceso crea automáticamente credenciales temporales. También crea automáticamente credenciales temporales cuando inicia sesión en la consola como usuario y luego cambia de rol. Para obtener más información acerca del cambio de roles, consulte Cambio a un rol (consola) en la Guía del usuario de IAM.

Puede crear credenciales temporales de forma manual mediante la AWS CLI o la API de AWS. A continuación, puede usar esas credenciales temporales para acceder a AWS. AWS recomienda generar credenciales temporales de forma dinámica en lugar de usar claves de acceso a largo plazo. Para obtener más información, consulte Credenciales de seguridad temporales en IAM.

Permisos de entidades principales entre servicios de Amazon EMR

Admite permisos de entidades principales	Sí

Cuando utiliza un usuario o un rol de IAM para llevar a cabo acciones en AWS, se lo considera una entidad principal. Las políticas conceden permisos a una entidad principal. Cuando utiliza algunos servicios, es posible que realice una acción que desencadene otra acción en un servicio diferente. En este caso, debe tener permisos para realizar ambas acciones. Para ver si una acción requiere acciones dependientes adicionales en una política, consulte Acciones, recursos y claves de condición para Amazon EMR en la Referencia de autorizaciones de servicio.

Roles de servicio para Amazon EMR

Compatible con funciones del servicio	No

Roles vinculado a servicios para Amazon EMR

Compatible con roles vinculados a servicios	Sí

Para obtener más información acerca de cómo crear o administrar roles vinculados a servicios, consulte Servicios de AWS que funcionan con IAM. Busque un servicio en la tabla que incluya `Yes` en la columna Service-linked role (Rol vinculado al servicio). Elija el vínculo Sí para ver la documentación acerca del rol vinculado al servicio en cuestión.

Uso de etiquetas de clúster y cuaderno con políticas de IAM para el control de acceso

Es posible ajustar el permiso de las acciones de Amazon EMR asociadas con Cuadernos de Amazon EMR y clústeres de EMR mediante el control de acceso basado en etiquetas con políticas de IAM basadas en identidades. Puede utilizar claves de condición en un `Condition`elemento (también denominado bloque`Condition`) para permitir ciertas acciones solo cuando un bloc de notas, un clúster o ambos tengan una clave de etiqueta o una combinación clave-valor determinadas. También puede limitar la acción `CreateEditor` (que crea un cuaderno de EMR) y la acción `RunJobFlow` (que crea un clúster), de modo que una solicitud se envíe al crear dicho recurso.

En Amazon EMR, las claves de condición que se pueden utilizar en un elemento `Condition` se aplican únicamente a las acciones de la API de Amazon EMR donde `ClusterID` o `NotebookID` sean un parámetro de solicitud necesario. Por ejemplo, la acción ModifyInstanceGroups no admite las claves de contexto porque `ClusterID` es un parámetro opcional.

Al crear un cuaderno de EMR, se aplica una etiqueta predeterminada con una cadena de claves de `creatorUserId` que se convierte en el valor del ID de usuario de IAM que ha creado el cuaderno. Esto resulta útil para limitar las acciones permitidas en el bloc de notas únicamente al creador.

Las siguientes claves de condición están disponibles en Amazon EMR:

- Utilice la clave de contexto de condición `elasticmapreduce:ResourceTag/`*`TagKeyString`* para permitir o denegar las acciones de los usuarios en clústeres o blocs de notas con etiquetas que tienen la cadena *`TagKeyString`* que especifique. Si una acción pasa `ClusterID` y `NotebookID`, la condición se aplica al clúster y al bloc de notas. Esto significa que ambos recursos deben tener la cadena de clave de etiqueta o la combinación clave-valor que especifique. Puede utilizar el elemento `Resource` para limitar la instrucción de modo que se aplique únicamente a los clústeres o los blocs de notas que lo necesiten. Para obtener más información, consulte Ejemplos de políticas de Amazon EMR basadas en identidades.
- Utilice la clave de contexto de condición `elasticmapreduce:RequestTag/`*`TagKeyString`* para requerir una etiqueta específica con acciones/llamadas al API. Por ejemplo, puede utilizar la clave de contexto de condición junto con la acción `CreateEditor` para requerir que se aplique una clave con la cadena *`TagKeyString`* a un bloc de notas cuando se crea este.

Ejemplos

Para ver una lista de las acciones de Amazon EMR, consulte Acciones definidas por Amazon EMR en la Guía del usuario de IAM.

Roles en tiempo de ejecución para los pasos de Amazon EMR

Un rol en tiempo de ejecución es un rol de AWS Identity and Access Management (IAM) que puede especificar al enviar un trabajo o una consulta a un clúster de Amazon EMR. El trabajo o la consulta que envíe a su clúster de Amazon EMR utiliza el rol en tiempo de ejecución para acceder a los recursos de AWS, como los objetos de Amazon S3. Puede especificar roles en tiempo de ejecución con Amazon EMR para los trabajos de Spark y Hive.

También puede especificar los roles en tiempo de ejecución al conectarse a los clústeres de Amazon EMR en Amazon SageMaker y al adjuntar un espacio de trabajo de Amazon EMR Studio a un clúster de EMR. Para obtener más información, consulte Conectarse a un clúster de Amazon EMR desde SageMaker Studio y Ejecutar un espacio de trabajo de EMR Studio con un rol de tiempo de ejecución.

Anteriormente, los clústeres de Amazon EMR ejecutaban trabajos o consultas de Amazon EMR con permisos basados en la política de IAM adjunta al perfil de instancia que utilizaba para lanzar el clúster. Esto significaba que las políticas tenían que contener la unión de todos los permisos para todos los trabajos y consultas que se ejecutaban en un clúster de Amazon EMR. Con los roles en tiempo de ejecución, ahora puede administrar el control de acceso de cada trabajo o consulta de forma individual, en lugar de compartir el perfil de instancia de Amazon EMR del clúster.

En los clústeres de Amazon EMR con roles en tiempo de ejecución, también puede aplicar un control de acceso basado en AWS Lake Formation a los trabajos y consultas de Spark, Hive y Presto en sus lagos de datos. Para obtener más información sobre cómo realizar la integración conAWS Lake Formation, consulte Integración de Amazon EMR con AWS Lake Formation.

> **Note**
>
> Cuando especifica un rol en tiempo de ejecución para un paso de Amazon EMR, los trabajos o consultas que envíe solo pueden acceder a los recursos de AWS que las políticas asociadas al rol en tiempo de ejecución permitan. Estos trabajos y consultas no pueden acceder al servicio de metadatos de instancias en las instancias de EC2 del clúster ni utilizar el perfil de instancia de EC2 del clúster para acceder a ningún recurso de AWS.

Requisitos previos para lanzar un clúster de Amazon EMR con un rol en tiempo de ejecución

Temas

- Paso 1: configurar los controles de seguridad en Amazon EMR
- Paso 2: configurar un perfil de instancia de EC2 para el clúster de Amazon EMR
- Paso 3: configurar una política de confianza

Paso 1: configurar los controles de seguridad en Amazon EMR

Utilice la siguiente estructura de JSON para crear una configuración de seguridad en AWS Command Line Interface (AWS CLI) y configure `EnableApplicationScopedIAMRole` en `true`. Para obtener más información acerca de las configuraciones de seguridad, consulte Uso de configuraciones de seguridad para configurar la seguridad del clúster.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true
        }
    }
}
```

Le recomendamos que habilite siempre las opciones de cifrado en tránsito en la configuración de seguridad, de modo que los datos que se transfieran a través de Internet estén cifrados y no sean texto sin formato. Puede omitir estas opciones si no quiere conectarse a clústeres de Amazon EMR con roles en tiempo de ejecución de SageMaker Studio o EMR Studio. Para configurar el cifrado de datos, consulte Configuración del cifrado de datos.

Como alternativa, puede crear una configuración de seguridad con ajustes personalizados con la AWS Management Console.

Paso 2: configurar un perfil de instancia de EC2 para el clúster de Amazon EMR

Los clústeres de Amazon EMR utilizan el rol de perfil de instancia de Amazon EC2 para asumir los roles en tiempo de ejecución. Para usar roles en tiempo de ejecución con los pasos de Amazon EMR, agregue las siguientes políticas al rol de IAM que planea usar como rol de perfil de instancia. Para agregar políticas a un rol de IAM o editar una política integrada o administrada existente, consulte Adición y eliminación de permisos de identidad de IAM.

```
{
    "Version":"2012-10-17",
    "Statement":[
        {
            "Sid":"AllowRuntimeRoleUsage",
            "Effect":"Allow",
            "Action":[
                "sts:AssumeRole",
                "sts:TagSession"
            ],
            "Resource":[
                <runtime-role-ARN>
            ]
        }
    ]
}
```

Paso 3: configurar una política de confianza

Para cada rol de IAM que vaya a utilizar como rol en tiempo de ejecución, defina la siguiente política de confianza y sustituya `EMR_EC2_DefaultRole` por el rol de perfil de instancia. Para modificar la política de confianza de un rol de IAM, consulte Modificación de una política de confianza de rol.

```
{
    "Sid":"AllowAssumeRole",
    "Effect":"Allow",
    "Principal":{
        "AWS":"arn:aws:iam::<AWS_ACCOUNT_ID>:role/EMR_EC2_DefaultRole"
    },
    "Action":"sts:AssumeRole"
}
```

Lanzamiento de un clúster de Amazon EMR con un control de acceso basado en roles

Tras realizar las configuraciones, puede lanzar un clúster de Amazon EMR con la configuración de seguridad de Paso 1: configurar los controles de seguridad en Amazon EMR. Para usar los roles en tiempo de ejecución con los pasos de Amazon EMR, utilice la etiqueta de lanzamiento `emr-6.7.0` o una versión posterior y seleccione Hive, Spark o ambas como aplicación de clúster. Para conectarse desde SageMaker Studio, utilice la versión `emr-6.9.0` o posterior y seleccione Livy, Spark, Hive o Presto como aplicación de clúster. Para obtener instrucciones sobre cómo lanzar el clúster, consulte Especificación de una configuración de seguridad para un clúster.

Envío de trabajos de Spark siguiendo los pasos de Amazon EMR

El siguiente es un ejemplo de cómo ejecutar el ejemplo de HdfsTest incluido con Apache Spark. Esta llamada a la API solo se realiza correctamente si el rol en tiempo de ejecución de Amazon EMR proporcionado puede acceder a `S3_LOCATION`.

```
RUNTIME_ROLE_ARN=<runtime-role-arn>
S3_LOCATION=<s3-path>
REGION=<aws-region>
CLUSTER_ID=<cluster-id>

aws emr add-steps --cluster-id $CLUSTER_ID \
--steps '[{ "Name": "Spark Example", "ActionOnFailure": "CONTINUE","HadoopJarStep":
 { "Jar":"command-runner.jar","Args" : ["spark-example","HdfsTest",
 "$S3_LOCATION"] } }]' \
--execution-role-arn $RUNTIME_ROLE_ARN \
--region $REGION
```

> **Note**
>
> Le recomendamos que desactive el acceso SSH al clúster de Amazon EMR y que solo permita que la API de Amazon EMR `AddJobFlowSteps` acceda al clúster.

Envío de trabajos de Hive siguiendo los pasos de Amazon EMR

En el siguiente ejemplo, se utilizan los pasos de Apache Hive con Amazon EMR para enviar un trabajo y ejecutar el archivo `QUERY_FILE.hql`. Esta consulta solo se realiza correctamente si el rol en tiempo de ejecución proporcionado puede acceder a la ruta de Amazon S3 del archivo de consulta.

```
RUNTIME_ROLE_ARN=<runtime-role-arn>
REGION=<aws-region>
CLUSTER_ID=<cluster-id>

aws emr add-steps --cluster-id $CLUSTER_ID \
--steps '[{ "Name": "Run hive query using command-runner.jar - simple
 select","ActionOnFailure":"CONTINUE","HadoopJarStep": { "Jar": "command-
runner.jar","Args" :["hive -
f","s3://DOC_EXAMPLE_BUCKET/QUERY_FILE.hql"] } }]' \
--execution-role-arn $RUNTIME_ROLE_ARN \
```

```
--region $REGION
```

Conexión a clústeres de Amazon EMR con roles en tiempo de ejecución desde un cuaderno de SageMaker Studio

Puede aplicar roles en tiempo de ejecución de Amazon EMR a las consultas que ejecute en los clústeres de Amazon EMR desde SageMaker Studio. Para hacerlo, siga estos pasos:

1. Siga las instrucciones de Iniciar Amazon SageMaker Studio para crear un SageMaker Studio.
2. En la interfaz de usuario de SageMaker Studio, inicie un cuaderno con kernels compatibles. Por ejemplo, inicie una imagen de SparkMagic con un kernel de PySpark.
3. Seleccione un clúster de Amazon EMR en SageMaker Studio y, a continuación, elija Conectar.
4. Seleccione un rol en tiempo de ejecución y, a continuación, elija Conectar.

Esto creará una celda de cuaderno de SageMaker con comandos mágicos para conectarse a su clúster de Amazon EMR con el rol en tiempo de ejecución de Amazon EMR elegido. En la celda del cuaderno, puede introducir y ejecutar consultas con el rol en tiempo de ejecución y el control de acceso basado en Lake Formation. Para ver un ejemplo más detallado, consulte Apply fine-grained data access controls with AWS Lake Formation and Amazon EMR from Amazon SageMaker Studio.

Control del acceso al rol en tiempo de ejecución de Amazon EMR

Puede controlar el acceso al rol en tiempo de ejecución con la clave de condición `elasticmapreduce:ExecutionRoleArn`. La siguiente política permite a una entidad principal de IAM utilizar un rol de IAM denominado `Caller`, o cualquier rol de IAM que comience por la cadena `CallerTeamRole`, como rol en tiempo de ejecución.

> **Important**
>
> Debe crear una condición basada en la clave de contexto `elasticmapreduce:ExecutionRoleArn` al conceder acceso a la persona que llama para llamar a las API `GetClusterSessionCredentials` o `AddJobFlowSteps`, como se muestra en el siguiente ejemplo.

```
{
    "Sid":"AddStepsWithSpecificExecRoleArn",
```

```
    "Effect":"Allow",
    "Action":[
        "elasticmapreduce:AddJobFlowSteps"
    ],
    "Resource":"*",
    "Condition":{
        "StringEquals":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::<AWS_ACCOUNT_ID>:role/Caller"
            ]
        },
        "StringLike":{
            "elasticmapreduce:ExecutionRoleArn":[
                "arn:aws:iam::<AWS_ACCOUNT_ID>:role/CallerTeamRole*"
            ]
        }
    }
}
```

Establecimiento de confianza entre los roles en tiempo de ejecución y los clústeres de Amazon EMR

Amazon EMR genera un identificador único `ExternalId` para cada configuración de seguridad con la autorización del rol en tiempo de ejecución activada. Esta autorización permite a cada usuario ser propietario de un conjunto de roles en tiempo de ejecución para utilizarlos en los clústeres que les pertenecen. Por ejemplo, en una empresa, cada departamento puede usar su identificador externo para actualizar la política de confianza en su propio conjunto de roles en tiempo de ejecución.

Puede obtener el identificador externo con la API `DescribeSecurityConfiguration` de Amazon EMR de Amazon, tal y como se muestra en el ejemplo siguiente.

```
aws emr describe-security-configuration --name 'iamconfig-with-lf'{"Name": "iamconfig-
with-lf",
    "SecurityConfiguration":
        "{\"AuthorizationConfiguration\":{\"IAMConfiguration\":
{\"EnableApplicationScopedIAMRole\
        ":true,\"ApplicationScopedIAMRoleConfiguration\":{\"PropagateSourceIdentity
\":true,\"Exter
        nalId\":\"FXH5TSACFDWUCDSR3YQE2O7ETPUSM4OBCGLYWODSCUZDNZ4Y\"}},\"Lake
        FormationConfiguration\":{\"AuthorizedSessionTagValue\":\"Amazon EMR\"}}}",
    "CreationDateTime": "2022-06-03T12:52:35.308000-07:00"
}
```

Para obtener más información, consulte Cómo utilizar un ID externo al otorgar acceso a los recursos de AWS a terceros.

Auditoría

Para supervisar y controlar las acciones que realizan los usuarios finales con los roles de IAM, puede activar la característica de identidad de origen. Para obtener más información sobre la identidad de origen, consulte Monitoreo y control de las acciones realizadas con roles asumidos.

Para hacer un seguimiento de la identidad de origen, establezca `ApplicationScopedIAMRoleConfiguration/PropagateSourceIdentity` en `true` en la configuración de seguridad de la siguiente manera.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true,
            "ApplicationScopedIAMRoleConfiguration":{
                "PropagateSourceIdentity":true
            }
        }
    }
}
```

Cuando establece `PropagateSourceIdentity` en `true`, Amazon EMR aplica la identidad de origen de las credenciales de llamada a un trabajo o sesión de consulta que cree con el rol en tiempo de ejecución. Si no hay identidad de origen en las credenciales de llamada, Amazon EMR no establece la identidad de origen.

Para usar esta propiedad, proporcione permisos `sts:SetSourceIdentity` a su perfil de instancia de la siguiente manera.

```
{ // PropagateSourceIdentity statement
    "Sid":"PropagateSourceIdentity",
    "Effect":"Allow",
    "Action":"sts:SetSourceIdentity",
    "Resource":[
        <runtime-role-ARN>
    ],
    "Condition":{
        "StringEquals":{
```

```
            "sts:SourceIdentity":<source-identity>
        }
    }
}
```

También debe agregar la instrucción `AllowSetSourceIdentity` a la política de confianza de sus roles en tiempo de ejecución.

```
{ // AllowSetSourceIdentity statement
    "Sid":"AllowSetSourceIdentity",
    "Effect":"Allow",
    "Principal":{
        "AWS":"arn:aws:iam::<AWS_ACCOUNT_ID>:role/EMR_EC2_DefaultRole"
    },
    "Action":[
        "sts:SetSourceIdentity",
        "sts:AssumeRole"
    ],
    "Condition":{
        "StringEquals":{
            "sts:SourceIdentity":<source-identity>
        }
    }
}
```

Consideraciones adicionales

> **Note**
>
> Con la versión `emr-6.9.0` de Amazon EMR, es posible que se produzcan errores intermitentes al conectarse a clústeres de Amazon EMR desde SageMaker Studio. Si desea solucionar este problema, puede instalar el parche con una acción de arranque al lanzar el clúster. Para obtener más información sobre el parche, consulte Problemas conocidos de la versión 6.9.0 de Amazon EMR.

Además, tenga en cuenta lo siguiente al configurar los roles en tiempo de ejecución para Amazon EMR.

- Amazon EMR admite roles en tiempo de ejecución en todas las Regiones de AWS comerciales.

- Los pasos de Amazon EMR admiten los trabajos de Apache Spark y Apache Hive con roles en tiempo de ejecución cuando se utiliza la versión `emr-6.7.0` o versiones posteriores.
- SageMaker Studio admite consultas de Spark, Hive y Presto con roles en tiempo de ejecución cuando se utiliza la versión `emr-6.9.0` o versiones posteriores.
- Los siguientes kernels de cuaderno de SageMaker admiten roles en tiempo de ejecución:
  - DataScience: kernel de Python 3
  - DataScience 2.0: kernel de Python 3
  - DataScience 3.0: kernel de Python 3
  - SparkAnalytics 1.0: kernels de SparkMagic y PySpark
  - SparkAnalytics 2.0: kernels de SparkMagic y PySpark
  - SparkMagic: kernel de PySpark
- Amazon EMR admite pasos que utilizan `RunJobFlow` únicamente en el momento de la creación del clúster. Esta API no admite roles en tiempo de ejecución.
- Amazon EMR no admite roles en tiempo de ejecución en clústeres que usted configure para que tengan una alta disponibilidad.
- Los roles en tiempo de ejecución no permiten controlar el acceso a los recursos del clúster, como HDFS y HMS.

Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS

Amazon EMR y las aplicaciones como Hadoop necesitan permisos para obtener acceso a otros recursos de AWS y realizar acciones cuando se ejecutan. Cada clúster de Amazon EMR debe tener un rol de servicio y un rol para el perfil de instancia de Amazon EC2. Para obtener más información, consulte Roles de IAM y Uso de perfiles de instancia en la Guía del usuario de IAM. Las políticas de IAM asociadas a estos roles proporcionan permisos que permiten al clúster interoperar con otros servicios de AWS en nombre de un usuario.

Es necesario otro rol adicional, el rol de escalado automático, si el clúster utiliza el escalado automático en Amazon EMR. El rol de servicio de AWS de Cuadernos de Amazon EMR es obligatorio si utiliza Cuadernos de Amazon EMR.

Amazon EMR proporciona roles predeterminados y políticas administradas predeterminadas que determinan permisos los permisos de cada rol. Las políticas administradas las crea y las mantiene

AWS, por lo que se actualizan automáticamente si cambian los requisitos de servicio. Para obtener más información, consulte Políticas administradas por AWS en la Guía del usuario de IAM.

Si está creando un clúster o un cuaderno por primera vez en una cuenta, los roles de Cuadernos de Amazon EMR aún no existen. Una vez creados, es posible ver los roles, las políticas asociadas a ellos y los permisos concedidos o denegados por estas políticas en la consola de IAM (https://console.aws.amazon.com/iam/). Puede especificar los roles predeterminados para Amazon EMR para crear y utilizar, puede crear sus propios roles y especificarlos de manera individual al crear un clúster para personalizar los permisos, y puede especificar roles predeterminados que se utilizarán al crear un clúster mediante la AWS CLI. Para obtener más información, consulte Personalización de roles de IAM.

Modificación de políticas basadas en identidades para obtener permisos y así transferir roles de servicio para Amazon EMR

Las políticas administradas de forma predeterminada con todos los permisos de Amazon EMR incorporan configuraciones de seguridad `iam:PassRole`, entre las que se incluyen las siguientes:

- Permisos `iam:PassRole` solo para roles específicos de Amazon EMR predeterminados.
- Condiciones `iam:PassedToService` que le permiten usar la política solo con los productos de AWS especificados, como `elasticmapreduce.amazonaws.com` y `ec2.amazonaws.com`.

Puede ver la versión JSON de las políticas AmazonEMRFullAccessPolicy_v2 y AmazonEMRServicePolicy_v2 en la consola de IAM. Le recomendamos crear nuevos clústeres con las políticas administradas de la versión 2.

Resumen de rol de servicio

En la siguiente tabla se muestran los roles de servicio de IAM asociados con Amazon EMR para una consulta rápida.

Función	Rol predeterminado	Descripción	Política administrada predeterminada
Rol de servicio para Amazon EMR (rol de EMR)	`EMR_DefaultRole_V2`	Permite que Amazon EMR llame a otros productos de AWS en su nombre al	`AmazonEMRServicePolicy_v2`

Función	Rol predeterminado	Descripción	Política administrada predeterminada
		aprovisionar recursos y llevar a cabo acciones de servicio. Este rol es necesario para todos los clústeres.	**Important** La solicitud de instancias de spot requiere un rol vinculado a un servicio. Si este rol no existe, el rol de servicio de Amazon EMR debe tener permisos para crearlo; en caso contrario, se producirá un error de permisos. Si planea solicitar instancias de spot, debe actualizar esta política para incluir una instrucción que permita la creación de este rol vinculado a un servicio. Para obtener más información,

Función	Rol predeterminado	Descripción	Política administrada predeterminada
			consulte Rol de servicio para Amazon EMR (rol de EMR) y Rol vinculado al servicio para solicitudes de instancias de spot en la Guía del usuario de Amazon EC2 para instancias de Linux.

Función	Rol predeterminado	Descripción	Política administrada predeterminada
Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2)	`EMR_EC2_DefaultRole`	Los procesos de aplicación que se ejecutan sobre el ecosistema de Hadoop, en las instancias del clúster, utilizan este rol cuando hacen llamadas a otros servicios de AWS. Para obtener acceso a los datos en Amazon S3 usando EMRFS, puede especificar diferentes roles que se asumirán en función de la ubicación de los datos en Amazon S3. Por ejemplo, varios equipos pueden acceder a una única “cuenta de almacenamiento” de datos de Amazon S3. Para obtener más información, consulte Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3. Este rol es necesario para todos los clústeres.	`AmazonElasticMapReduceforEC2Role` . Para obtener más información, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2).

Función	Rol predeterminado	Descripción	Política administrada predeterminada
Rol de servicio para el escalado automático en Amazon EMR (rol de Auto Scaling)	`EMR_AutoScaling_DefaultRole`	Permite acciones adicionales para entornos de escalado dinámico. Solo es obligatorio para los clústeres que utilizan el escalado automático en Amazon EMR. Para obtener más información, consulte Uso del escalado automático con una política personalizada para grupos de instancias .	`AmazonElasticMapReduceforAutoScalingRole` . Para obtener más información, consulte Rol de servicio para el escalado automático en Amazon EMR (rol de Auto Scaling).

Función	Rol predeterminado	Descripción	Política administrada predeterminada
Rol de servicio para Cuadernos de Amazon EMR	`EMR_Notebooks_DefaultRole`	Proporciona permisos que un cuaderno de EMR necesita para obtener acceso a otros recursos de AWS y realizar acciones. Necesario solo si se utiliza Cuadernos de Amazon EMR.	`AmazonElasticMapReduceEditorsRole` . Para obtener más información, consulte Rol de servicio para Cuadernos de Amazon EMR. También se asocia `S3FullAccessPolicy` de forma predeterminada. El contenido de esta política se muestra a continuación. `{ "Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Action": "s3:*", "Resource": "*" } ] }`

Función	Rol predeterminado	Descripción	Política administrada predeterminada
Rol vinculado a servicio	`AWSServiceRoleForEMRCleanup`	Amazon EMR crea automáticamente un rol vinculado a un servicio. Si el servicio de Amazon EMR ha perdido la capacidad de limpiar los recursos de Amazon EC2, Amazon EMR puede utilizar este rol para limpiar. Si un clúster utiliza instancias de spot, la política de permisos vinculada al Rol de servicio para Amazon EMR (rol de EMR) debe permitir la creación de un rol vinculado al servicio. Para obtener más información, consulte Permisos de roles vinculados a un servicio para Amazon EMR.	`AmazonEMRCleanupPolicy`

Temas

- Roles de servicio de IAM utilizados por Amazon EMR
- Personalización de roles de IAM
- Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3

- Uso de políticas basadas en recursos para el acceso de Amazon EMR a Catálogo de datos de AWS Glue
- Uso de roles de IAM con las aplicaciones que llaman directamente a los servicios de AWS
- Cómo permitir a los usuarios y grupos crear y modificar roles

Roles de servicio de IAM utilizados por Amazon EMR

Amazon EMR utiliza roles de servicio de IAM para llevar a cabo acciones en su nombre al aprovisionar los recursos del clúster, ejecutar aplicaciones, escalar recursos de forma dinámica y crear y ejecutar Cuadernos de Amazon EMR. Amazon EMR utiliza los siguientes roles cuando interactúa con otros servicios de AWS. Cada rol tiene una función exclusiva en Amazon EMR. Los temas de esta sección describen la función del rol y facilitan los roles y la política de permisos predeterminados para cada rol.

Si tiene código de aplicación en el clúster que llama directamente a los servicios de AWS, es posible que tenga que utilizar el SDK para especificar las funciones. Para obtener más información, consulte Uso de roles de IAM con las aplicaciones que llaman directamente a los servicios de AWS.

Temas

- Rol de servicio para Amazon EMR (rol de EMR)
- Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2)
- Rol de servicio para el escalado automático en Amazon EMR (rol de Auto Scaling)
- Rol de servicio para Cuadernos de Amazon EMR
- Uso de roles vinculados a un servicio para Amazon EMR

Rol de servicio para Amazon EMR (rol de EMR)

El rol de Amazon EMR define las acciones permitidas para Amazon EMR al aprovisionar recursos y realizar tareas de nivel de servicio que no se llevan a cabo en el contexto de una instancia de Amazon EC2 que se ejecuta dentro de un clúster. Por ejemplo, el rol de servicio se utiliza para aprovisionar instancias EC2 cuando se lanza un clúster.

- El nombre del rol predeterminado es `EMR_DefaultRole_V2`.
- La política administrada predeterminada con ámbito de aplicación de Amazon EMR y asociada a `EMR_DefaultRole_V2` es `AmazonEMRServicePolicy_v2`. Esta política, versión 2, sustituye a la política administrada predeterminada, `AmazonElasticMapReduceRole`, ya obsoleta.

`AmazonEMRServicePolicy_v2` depende del acceso limitado a los recursos que Amazon EMR aprovisiona o utiliza. Cuando utilice esta política, tendrá que pasar la etiqueta de usuario `for-use-with-amazon-emr-managed-policies = true` al aprovisionar el clúster. Amazon EMR propagará automáticamente esas etiquetas. Además, es posible que tenga que agregar manualmente una etiqueta de usuario a tipos específicos de recursos, como grupos de seguridad de EC2 que no fueron creados por Amazon EMR. Consulte Etiquetado de recursos para usar políticas administradas.

> **Important**
>
> Amazon EMR utiliza este rol de servicio de Amazon EMR y el rol `AWSServiceRoleForEMRCleanup` para limpiar los recursos del clúster de su cuenta que ya no utiliza, como las instancias de Amazon EC2. Debe incluir acciones para que las políticas de rol eliminen o terminen los recursos. De lo contrario, Amazon EMR no podrá realizar estas acciones de limpieza y podría incurrir en costos por los recursos no utilizados que permanecen en el clúster.

A continuación se muestra el contenido de la política `AmazonEMRServicePolicy_v2`. También puede ver el contenido actual de la política `AmazonEMRServicePolicy_v2`administrada en la consola de IAM.

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Sid": "CreateInTaggedNetwork",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateNetworkInterface",
    "ec2:RunInstances",
    "ec2:CreateFleet",
    "ec2:CreateLaunchTemplate",
    "ec2:CreateLaunchTemplateVersion"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:subnet/*",
    "arn:aws:ec2:*:*:security-group/*"
   ],
   "Condition": {
```

```
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "CreateWithEMRTaggedLaunchTemplate",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateFleet",
    "ec2:RunInstances",
    "ec2:CreateLaunchTemplateVersion"
   ],
   "Resource": "arn:aws:ec2:*:*:launch-template/*",
   "Condition": {
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "CreateEMRTaggedLaunchTemplate",
   "Effect": "Allow",
   "Action": "ec2:CreateLaunchTemplate",
   "Resource": "arn:aws:ec2:*:*:launch-template/*",
   "Condition": {
    "StringEquals": {
     "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "CreateEMRTaggedInstancesAndVolumes",
   "Effect": "Allow",
   "Action": [
    "ec2:RunInstances",
    "ec2:CreateFleet"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:instance/*",
    "arn:aws:ec2:*:*:volume/*"
   ],
   "Condition": {
    "StringEquals": {
```

```
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "ResourcesToLaunchEC2",
    "Effect": "Allow",
    "Action": [
     "ec2:RunInstances",
     "ec2:CreateFleet",
     "ec2:CreateLaunchTemplate",
     "ec2:CreateLaunchTemplateVersion"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:network-interface/*",
     "arn:aws:ec2:*::image/ami-*",
     "arn:aws:ec2:*:*:key-pair/*",
     "arn:aws:ec2:*:*:capacity-reservation/*",
     "arn:aws:ec2:*:*:placement-group/pg-*",
     "arn:aws:ec2:*:*:fleet/*",
     "arn:aws:ec2:*:*:dedicated-host/*",
     "arn:aws:resource-groups:*:*:group/*"
    ]
   },
   {
    "Sid": "ManageEMRTaggedResources",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateLaunchTemplateVersion",
     "ec2:DeleteLaunchTemplate",
     "ec2:DeleteNetworkInterface",
     "ec2:ModifyInstanceAttribute",
     "ec2:TerminateInstances"
    ],
    "Resource": "*",
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "ManageTagsOnEMRTaggedResources",
    "Effect": "Allow",
```

```
    "Action": [
     "ec2:CreateTags",
     "ec2:DeleteTags"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:instance/*",
     "arn:aws:ec2:*:*:volume/*",
     "arn:aws:ec2:*:*:network-interface/*",
     "arn:aws:ec2:*:*:launch-template/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "CreateNetworkInterfaceNeededForPrivateSubnet",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateNetworkInterface"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:network-interface/*"
    ],
    "Condition": {
     "StringEquals": {
      "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
     }
    }
   },
   {
    "Sid": "TagOnCreateTaggedEMRResources",
    "Effect": "Allow",
    "Action": [
     "ec2:CreateTags"
    ],
    "Resource": [
     "arn:aws:ec2:*:*:network-interface/*",
     "arn:aws:ec2:*:*:instance/*",
     "arn:aws:ec2:*:*:volume/*",
     "arn:aws:ec2:*:*:launch-template/*"
    ],
    "Condition": {
```

```
    "StringEquals": {
     "ec2:CreateAction": [
      "RunInstances",
      "CreateFleet",
      "CreateLaunchTemplate",
      "CreateNetworkInterface"
     ]
    }
   }
  },
  {
   "Sid": "TagPlacementGroups",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateTags",
    "ec2:DeleteTags"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:placement-group/pg-*"
   ]
  },
  {
   "Sid": "ListActionsForEC2Resources",
   "Effect": "Allow",
   "Action": [
    "ec2:DescribeAccountAttributes",
    "ec2:DescribeCapacityReservations",
    "ec2:DescribeDhcpOptions",
    "ec2:DescribeImages",
    "ec2:DescribeInstances",
    "ec2:DescribeLaunchTemplates",
    "ec2:DescribeNetworkAcls",
    "ec2:DescribeNetworkInterfaces",
    "ec2:DescribePlacementGroups",
    "ec2:DescribeRouteTables",
    "ec2:DescribeSecurityGroups",
    "ec2:DescribeSubnets",
    "ec2:DescribeVolumes",
    "ec2:DescribeVolumeStatus",
    "ec2:DescribeVpcAttribute",
    "ec2:DescribeVpcEndpoints",
    "ec2:DescribeVpcs"
   ],
   "Resource": "*"
```

```
  },
  {
   "Sid": "CreateDefaultSecurityGroupWithEMRTags",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateSecurityGroup"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:security-group/*"
   ],
   "Condition": {
    "StringEquals": {
     "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "CreateDefaultSecurityGroupInVPCWithEMRTags",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateSecurityGroup"
   ],
   "Resource": [
    "arn:aws:ec2:*:*:vpc/*"
   ],
   "Condition": {
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "TagOnCreateDefaultSecurityGroupWithEMRTags",
   "Effect": "Allow",
   "Action": [
    "ec2:CreateTags"
   ],
   "Resource": "arn:aws:ec2:*:*:security-group/*",
   "Condition": {
    "StringEquals": {
     "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true",
     "ec2:CreateAction": "CreateSecurityGroup"
    }
   }
```

```
  },
  {
   "Sid": "ManageSecurityGroups",
   "Effect": "Allow",
   "Action": [
    "ec2:AuthorizeSecurityGroupEgress",
    "ec2:AuthorizeSecurityGroupIngress",
    "ec2:RevokeSecurityGroupEgress",
    "ec2:RevokeSecurityGroupIngress"
   ],
   "Resource": "*",
   "Condition": {
    "StringEquals": {
     "aws:ResourceTag/for-use-with-amazon-emr-managed-policies": "true"
    }
   }
  },
  {
   "Sid": "CreateEMRPlacementGroups",
   "Effect": "Allow",
   "Action": [
    "ec2:CreatePlacementGroup"
   ],
   "Resource": "arn:aws:ec2:*:*:placement-group/pg-*"
  },
  {
   "Sid": "DeletePlacementGroups",
   "Effect": "Allow",
   "Action": [
    "ec2:DeletePlacementGroup"
   ],
   "Resource": "*"
  },
  {
   "Sid": "AutoScaling",
   "Effect": "Allow",
   "Action": [
    "application-autoscaling:DeleteScalingPolicy",
    "application-autoscaling:DeregisterScalableTarget",
    "application-autoscaling:DescribeScalableTargets",
    "application-autoscaling:DescribeScalingPolicies",
    "application-autoscaling:PutScalingPolicy",
    "application-autoscaling:RegisterScalableTarget"
   ],
```

```
    "Resource": "*"
   },
   {
    "Sid": "ResourceGroupsForCapacityReservations",
    "Effect": "Allow",
    "Action": [
     "resource-groups:ListGroupResources"
    ],
    "Resource": "*"
   },
   {
    "Sid": "AutoScalingCloudWatch",
    "Effect": "Allow",
    "Action": [
     "cloudwatch:PutMetricAlarm",
     "cloudwatch:DeleteAlarms",
     "cloudwatch:DescribeAlarms"
    ],
    "Resource": "arn:aws:cloudwatch:*:*:alarm:*_EMR_Auto_Scaling"
   },
   {
    "Sid": "PassRoleForAutoScaling",
    "Effect": "Allow",
    "Action": "iam:PassRole",
    "Resource": "arn:aws:iam::*:role/EMR_AutoScaling_DefaultRole",
    "Condition": {
     "StringLike": {
      "iam:PassedToService": "application-autoscaling.amazonaws.com*"
     }
    }
   },
   {
    "Sid": "PassRoleForEC2",
    "Effect": "Allow",
    "Action": "iam:PassRole",
    "Resource": "arn:aws:iam::*:role/EMR_EC2_DefaultRole",
    "Condition": {
     "StringLike": {
      "iam:PassedToService": "ec2.amazonaws.com*"
     }
    }
   }
  ]
```

```
}
```

Su rol de servicio debe usar la siguiente política de confianza.

> **Important**
>
> La siguiente política de confianza incluye las claves de condición globales `aws:SourceArn` y `aws:SourceAccount` para limitar los permisos que concede a Amazon EMR a determinados recursos de su cuenta. Con estas, podrá protegerse contra el problema del suplente confuso.

```
{
    "Version": "2008-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "elasticmapreduce.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
                }
            }
        }
    ]
}
```

Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2)

El rol de servicio para instancias de EC2 de clúster (también conocido como el perfil de instancia de EC2 para Amazon EMR) es un tipo especial de rol de servicio que está asignado a cada instancia de EC2 de un clúster de Amazon EMR cuando se lanza la instancia. Los procesos de aplicación que se ejecutan sobre el ecosistema de Hadoop asumen este rol para los permisos, para interactuar así con otros productos de AWS.

Para obtener más información sobre los roles de servicio para instancias de EC2, consulte Uso de un rol de IAM para conceder permisos a aplicaciones que se ejecutan en instancias de Amazon EC2 en la Guía del usuario de IAM.

> **Important**
>
> El rol de servicio predeterminado para las instancias de EC2 del clúster y su política administrada predeterminada asociada de AWS, `AmazonElasticMapReduceforEC2Role`, están en vías de quedar obsoletos y no se proporcionan políticas administradas de AWS de reemplazo. Tendrá que crear y especificar un perfil de instancia para reemplazar la política predeterminada y el rol obsoletos.

Política administrada y rol predeterminados

- El nombre del rol predeterminado es `EMR_EC2_DefaultRole`.
- La política administrada `EMR_EC2_DefaultRole` predeterminada, `AmazonElasticMapReduceforEC2Role`, está a punto de finalizar su soporte. En lugar de utilizar una política administrada predeterminada para el perfil de instancia de EC2, aplique políticas basadas en recursos a los buckets de S3 y otros recursos que Amazon EMR necesite, o utilice su propia política administrada por el cliente con un rol de IAM como perfil de instancia. Para obtener más información, consulte Creación de un rol de servicio para las instancias de EC2 del clúster con permisos de privilegios mínimos.

Lo siguiente muestra el contenido de la versión 3 de `AmazonElasticMapReduceforEC2Role`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Resource": "*",
            "Action": [
                "cloudwatch:*",
                "dynamodb:*",
                "ec2:Describe*",
                "elasticmapreduce:Describe*",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
```

```
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSteps",
                "kinesis:CreateStream",
                "kinesis:DeleteStream",
                "kinesis:DescribeStream",
                "kinesis:GetRecords",
                "kinesis:GetShardIterator",
                "kinesis:MergeShards",
                "kinesis:PutRecord",
                "kinesis:SplitShard",
                "rds:Describe*",
                "s3:*",
                "sdb:*",
                "sns:*",
                "sqs:*",
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTable",
                "glue:GetTables",
                "glue:GetTableVersions",
                "glue:CreatePartition",
                "glue:BatchCreatePartition",
                "glue:UpdatePartition",
                "glue:DeletePartition",
                "glue:BatchDeletePartition",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:BatchGetPartition",
                "glue:CreateUserDefinedFunction",
                "glue:UpdateUserDefinedFunction",
                "glue:DeleteUserDefinedFunction",
                "glue:GetUserDefinedFunction",
                "glue:GetUserDefinedFunctions"
            ]
        }
    ]
```

```
}
```

Su rol de servicio debe usar la siguiente política de confianza.

```
{
    "Version": "2008-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "ec2.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

Creación de un rol de servicio para las instancias de EC2 del clúster con permisos de privilegios mínimos

Como práctica recomendada, le recomendamos encarecidamente que cree un rol de servicio para las instancias de EC2 del clúster y una política de permisos que tenga los permisos mínimos para otros productos de AWS que necesite la aplicación.

La política administrada predeterminada, `AmazonElasticMapReduceforEC2Role`, proporciona los permisos que facilitan el lanzar un clúster inicial. Sin embargo, `AmazonElasticMapReduceforEC2Role` está en vías de quedar obsoleta y Amazon EMR no proporcionará una política predeterminada administrada de AWS que reemplace el rol obsoleto. Para lanzar un clúster inicial, debe proporcionar una política basada en los recursos o en la identificación administrada por el cliente.

Las siguientes instrucciones de política facilitan ejemplos de permisos necesarios para las distintas características de Amazon EMR. Le recomendamos que utilice estos permisos para crear una política de permisos que restrinja el acceso tan solo a aquellas funciones y recursos que necesite el clúster. Todas las instrucciones de política de ejemplo utilizan la región *`us-west-2`* y el ID de cuenta de AWS ficticio *`123456789012`*. Sustituya estos según corresponda para el clúster.

Para obtener más información sobre la creación y la especificación de roles personalizados, consulte Personalización de roles de IAM.

> **Note**
>
> Si crea un rol de EMR personalizado para EC2, siga el flujo de trabajo básico, que crea automáticamente un perfil de instancia con el mismo nombre. Amazon EC2 le permite crear roles y perfiles de instancia con nombres diferentes, pero Amazon EMR no admite esta configuración y se produce un error “Perfil de instancia no válido” al crear el clúster.

Lectura y escritura de datos en Amazon S3 con EMRFS

Cuando una aplicación que se ejecuta en un clúster de Amazon EMR hace referencia a los datos con el formato `s3://mydata`, Amazon EMR utiliza el perfil de instancia de EC2 para realizar la solicitud. Por lo general, los clústeres leen y escriben datos en Amazon S3 de esta forma, y Amazon EMR utiliza los permisos asociados al rol de servicio para instancias de EC2 del clúster de forma predeterminada. Para obtener más información, consulte Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3.

Dado que los roles de IAM para EMRFS seguirán usando los permisos asociados al rol de servicio para las instancias de EC2 del clúster, como práctica recomendada, le recomendamos que utilice los roles de IAM para EMRFS y limite los permisos de Amazon S3 y EMRFS asociados al rol de servicio para las instancias de EC2 del clúster.

La siguiente instrucción de ejemplo señala los permisos que EMRFS necesita para hacer solicitudes a Amazon S3.

- *my-data-bucket-in-s3-for-emrfs-reads-and-writes* especifica el bucket de Amazon S3 en el que el clúster lee y escribe los datos y todas las subcarpetas con */**. Añada solo los buckets y carpetas que necesita su aplicación.
- La instrucción de política que permite realizar acciones de dynamodb solo es necesaria si la vista coherente de EMRFS está habilitada. *EmrFSMetadata* especifica la carpeta predeterminada para la vista coherente de EMRFS.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
```

```
                "s3:AbortMultipartUpload",
                "s3:CreateBucket",
                "s3:DeleteObject",
                "s3:GetBucketVersioning",
                "s3:GetObject",
                "s3:GetObjectTagging",
                "s3:GetObjectVersion",
                "s3:ListBucket",
                "s3:ListBucketMultipartUploads",
                "s3:ListBucketVersions",
                "s3:ListMultipartUploadParts",
                "s3:PutBucketVersioning",
                "s3:PutObject",
                "s3:PutObjectTagging"
            ],
            "Resource": [
                "arn:aws:s3:::my-data-bucket-in-s3-for-emrfs-reads-and-writes",
                "arn:aws:s3:::my-data-bucket-in-s3-for-emrfs-reads-and-writes/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "dynamodb:CreateTable",
                "dynamodb:BatchGetItem",
                "dynamodb:BatchWriteItem",
                "dynamodb:PutItem",
                "dynamodb:DescribeTable",
                "dynamodb:DeleteItem",
                "dynamodb:GetItem",
                "dynamodb:Scan",
                "dynamodb:Query",
                "dynamodb:UpdateItem",
                "dynamodb:DeleteTable",
                "dynamodb:UpdateTable"
            ],
            "Resource": "arn:aws:dynamodb:us-west-2:123456789012:table/EmrFSMetadata"
        },
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData",
                "dynamodb:ListTables",
                "s3:ListBucket"
```

```
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "sqs:GetQueueUrl",
                "sqs:ReceiveMessage",
                "sqs:DeleteQueue",
                "sqs:SendMessage",
                "sqs:CreateQueue"
            ],
            "Resource": "arn:aws:sqs:us-west-2:123456789012:EMRFS-Inconsistency-*"
        }
    ]
}
```

Almacenamiento de archivos de registro en Amazon S3

La siguiente instrucción de política permite al clúster de Amazon EMR almacenar los archivos de registro en la ubicación de Amazon S3 indicada. En el siguiente ejemplo, cuando se creó el clúster, *s3://MyLoggingBucket/MyEMRClusterLogs* se especificó mediante Ubicación de la carpeta de registros en S3 en la consola, utilizando la opción `--log-uri` de la AWS CLI o el parámetro `LogUri` en el comando `RunJobFlow`. Para obtener más información, consulte Archivar archivos de registro en Amazon S3.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:PutObject",
            "Resource": "arn:aws:s3:::MyLoggingBucket/MyEMRClusterLogs/*"
        }
    ]
}
```

Uso de herramientas de depuración

La siguiente instrucción de política permite las acciones que se requieren si habilita la herramienta de depuración de Amazon EMR. Para la depuración, se requiere el almacenamiento de los archivos de

registro en Amazon S3 y los permisos asociados que aparecen en el ejemplo anterior. Para obtener más información, consulte Habilitar la herramienta de depuración.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "sqs:GetQueueUrl",
                "sqs:SendMessage"
            ],
            "Resource": "arn:aws:sqs:us-west-2:123456789012:AWS-ElasticMapReduce-*"
        }
    ]
}
```

Uso de Catálogo de datos de AWS Glue

La siguiente instrucción de política permite las acciones que se requieren si utiliza Catálogo de datos de AWS Glue como almacén de metadatos para las aplicaciones. Para obtener más información, consulte Uso del Catálogo de datos de AWS Glue como metaalmacén para Spark SQL, Uso del Catálogo de datos de AWS Glue como metaalmacén de Hive y Uso de Presto con el Catálogo de datos de AWS Glue en la Guía de versiones de Amazon EMR.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "glue:CreateDatabase",
                "glue:UpdateDatabase",
                "glue:DeleteDatabase",
                "glue:GetDatabase",
                "glue:GetDatabases",
                "glue:CreateTable",
                "glue:UpdateTable",
                "glue:DeleteTable",
                "glue:GetTable",
                "glue:GetTables",
                "glue:GetTableVersions",
```

```
                "glue:CreatePartition",
                "glue:BatchCreatePartition",
                "glue:UpdatePartition",
                "glue:DeletePartition",
                "glue:BatchDeletePartition",
                "glue:GetPartition",
                "glue:GetPartitions",
                "glue:BatchGetPartition",
                "glue:CreateUserDefinedFunction",
                "glue:UpdateUserDefinedFunction",
                "glue:DeleteUserDefinedFunction",
                "glue:GetUserDefinedFunction",
                "glue:GetUserDefinedFunctions"
            ],
            "Resource": "*",
        }
    ]
}
```

Rol de servicio para el escalado automático en Amazon EMR (rol de Auto Scaling)

El rol de Auto Scaling de Amazon EMR sirve básicamente para lo mismo que el rol de servicio, pero permite acciones adicionales para entornos de escalado dinámico.

- El nombre del rol predeterminado es `EMR_AutoScaling_DefaultRole`.
- La política administrada predeterminada y asociada a `EMR_AutoScaling_DefaultRole` es `AmazonElasticMapReduceforAutoScalingRole`.

El contenido de la versión 1 de `AmazonElasticMapReduceforAutoScalingRole` se muestra a continuación.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "cloudwatch:DescribeAlarms",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
```

```
        }
    ]
}
```

Su rol de servicio debe usar la siguiente política de confianza.

> **Important**
>
> La siguiente política de confianza incluye las claves de condición globales `aws:SourceArn` y `aws:SourceAccount` para limitar los permisos que concede a Amazon EMR a determinados recursos de su cuenta. Con estas, podrá protegerse contra el problema del suplente confuso.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service":  [
                    "elasticmapreduce.amazonaws.com",
                    "application-autoscaling.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
                }
            }
        }
    ]
}
```

Rol de servicio para Cuadernos de Amazon EMR

Cada cuaderno de EMR requiere permisos para obtener acceso a otros recursos de AWS, así como para realizar acciones. Las políticas de IAM asociadas a este rol de servicio proporcionan permisos que permiten al cuaderno interoperar con otros servicios de AWS. Cuando crea un cuaderno mediante la AWS Management Console, tiene que especificar un rol de servicio de AWS. Puede utilizar el rol predeterminado, `EMR_Notebooks_DefaultRole` o especificar un rol que haya creado. Si no se ha creado un bloc de notas anteriormente, puede elegir crear el rol predeterminado.

- El nombre del rol predeterminado es `EMR_Notebooks_DefaultRole`.
- Las políticas administradas predeterminadas asociadas a `EMR_Notebooks_DefaultRole` son `AmazonElasticMapReduceEditorsRole` y `S3FullAccessPolicy`.

Su rol de servicio debe usar la siguiente política de confianza.

> **Important**
>
> La siguiente política de confianza incluye las claves de condición globales `aws:SourceArn` y `aws:SourceAccount` para limitar los permisos que concede a Amazon EMR a determinados recursos de su cuenta. Con estas, podrá protegerse contra el problema del suplente confuso.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "elasticmapreduce.amazonaws.com"
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "<account-id>"
                },
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:elasticmapreduce:<region>:<account-id>:*"
                }
            }
```

```
        }
    ]
}
```

El contenido de la versión 1 de `AmazonElasticMapReduceEditorsRole` es el siguiente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:AuthorizeSecurityGroupEgress",
                "ec2:AuthorizeSecurityGroupIngress",
                "ec2:CreateSecurityGroup",
                "ec2:DescribeSecurityGroups",
                "ec2:RevokeSecurityGroupEgress",
                "ec2:CreateNetworkInterface",
                "ec2:CreateNetworkInterfacePermission",
                "ec2:DeleteNetworkInterface",
                "ec2:DeleteNetworkInterfacePermission",
                "ec2:DescribeNetworkInterfaces",
                "ec2:ModifyNetworkInterfaceAttribute",
                "ec2:DescribeTags",
                "ec2:DescribeInstances",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:ListSteps"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateTags",
            "Resource": "arn:aws:ec2:*:*:network-interface/*",
            "Condition": {
                "ForAllValues:StringEquals": {
                    "aws:TagKeys": [
                        "aws:elasticmapreduce:editor-id",
                        "aws:elasticmapreduce:job-flow-id"
                    ]
```

```
                }
            }
        }
    ]
}
```

A continuación, se muestra el contenido de `S3FullAccessPolicy`. La `S3FullAccessPolicy` permite que su rol de servicio para Cuadernos de Amazon EMR realice todas las acciones de Amazon S3 en los objetos de su Cuenta de AWS. Al crear un rol de servicio personalizado para Cuadernos de Amazon EMR, debe otorgarle permisos a Amazon S3 para ese rol de servicio.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:*",
            "Resource": "*"
        }
    ]
}
```

Puede limitar el acceso de lectura y escritura para su rol de servicio a la ubicación de Amazon S3 en la que desee guardar los archivos de su cuaderno. Utilice el siguiente conjunto mínimo de permisos de Amazon S3.

```
"s3:PutObject",
"s3:GetObject",
"s3:GetEncryptionConfiguration",
"s3:ListBucket",
"s3:DeleteObject"
```

Si su bucket de Amazon S3 está cifrado, debe incluir los siguientes permisos para AWS Key Management Service.

```
"kms:Decrypt",
"kms:GenerateDataKey",
"kms:ReEncryptFrom",
"kms:ReEncryptTo",
"kms:DescribeKey"
```

Cuando vincula repositorios de Git a su cuaderno y necesita crear un secreto para el repositorio, debe agregar el permiso `secretsmanager:GetSecretValue` a la política de IAM asociada al rol de servicio de Cuadernos de Amazon EMR. A continuación se muestra una política de ejemplo:

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": "secretsmanager:GetSecretValue",
            "Resource": "*"
        }
    ]
}
```

Permisos del rol de servicio de Cuadernos de Amazon EMR

En esta tabla se enumeran las acciones que Cuadernos de Amazon EMR lleva a cabo con el rol de servicio, junto con los permisos necesarios para cada acción.

Acción	Permisos

Establezca un canal de red seguro entre un cuaderno y un clúster de Amazon EMR y lleve a cabo las acciones de limpieza necesarias.	`"ec2:CreateNetworkInterface",`
	`"ec2:CreateNetworkInterfacePermission",`
	`"ec2:DeleteNetworkInterface",`
	`"ec2:DeleteNetworkInterfacePermission",`
	`"ec2:DescribeNetworkInterfaces",`
	`"ec2:ModifyNetworkInterfaceAttribute",`
	`"ec2:AuthorizeSecurityGroupEgress",`
	`"ec2:AuthorizeSecurityGroupIngress",`
	`"ec2:CreateSecurityGroup",`
	`"ec2:DescribeSecurityGroups",`
	`"ec2:RevokeSecurityGroupEgress",`
	`"ec2:DescribeTags",`
	`"ec2:DescribeInstances",`
	`"ec2:DescribeSubnets",`
	`"ec2:DescribeVpcs",`
	`"elasticmapreduce:ListInstances",`
	`"elasticmapreduce:DescribeCluster",`
	`"elasticmapreduce:ListSteps"`

Acción	Permisos
Usa las credenciales de Git almacenadas en AWS Secrets Manager para vincular los repositorios de Git a un cuaderno.	`"secretsmanager:GetSecretValue"`
Aplique etiquetas de AWS a la interfaz de red y a los grupos de seguridad predeterminados que Cuadernos de Amazon EMR crea al configurar el canal de red seguro. Para obtener más información, consulte Tagging AWS resources (Etiquetado de los recursos de).	`"ec2:CreateTags"`

Acceda a los archivos y metadatos de los cuadernos en Amazon S3 o cárguelos.	`"s3:PutObject",`
	`"s3:GetObject",`
	`"s3:GetEncryptionConfiguration",`
	`"s3:ListBucket",`
	`"s3:DeleteObject"`
	
	Los siguientes permisos solo son necesarios si utiliza un bucket de Amazon S3 cifrado.
	
	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Actualizaciones de Cuadernos de Amazon EMR a las políticas administradas por AWS

Consulte los detalles sobre las actualizaciones de las políticas administradas por AWS para Cuadernos de Amazon EMR desde el 1 de marzo de 2021.

Cambio	Descripción	Fecha
`AmazonElasticMapReduceEditorsRole - Added permissions`	Cuadernos de Amazon EMR agregó los permisos `ec2:describeVPCs` y `elastmicmapreduce:ListSteps` a `AmazonElasticMapReduceEditorsRole` .	8 de febrero de 2023
Cuadernos de Amazon EMR comenzó a registrar cambios	Cuadernos de Amazon EMR comenzó a registrar los cambios de las políticas administradas por AWS.	8 de febrero de 2023

Uso de roles vinculados a un servicio para Amazon EMR

Amazon EMR utiliza roles vinculados a un servicio de AWS Identity and Access Management (IAM). Un rol vinculado a un servicio es un tipo único de rol de IAM que se encuentra vinculado directamente a Amazon EMR. El rol vinculado a un servicio está predefinido por Amazon EMR e incluye los permisos que Amazon EMR requiere para llamar a Amazon EC2 en su nombre para limpiar los recursos del clúster una vez que ya no estén en uso. El rol vinculado a un servicio funciona conjuntamente con el rol de servicio de Amazon EMR y el perfil de instancia de Amazon EC2 para Amazon EMR. Para obtener más información acerca del rol de servicio y del perfil de instancia, consulte Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS.

Amazon EMR define los permisos de este rol vinculado a un servicio y, a menos que esté definido de otra manera, solo Amazon EMR puede asumir el rol. Los permisos definidos incluyen las políticas de confianza y de permisos y que la política de permisos no se pueda adjuntar a ninguna otra entidad de IAM. Puede eliminar el rol únicamente después de terminar todos los clústeres de EMR de la cuenta.

Para obtener información sobre otros servicios que admiten roles vinculados al servicio, consulte Servicios de AWS que funcionan con IAM y busque los servicios que muestran Sí en la columna Rol vinculado al servicio. Seleccione una opción Sí con un enlace para ver la documentación acerca del rol vinculado al servicio en cuestión.

Permisos de roles vinculados a un servicio para Amazon EMR

Amazon EMR utiliza el rol AWSServiceRoleForEMRCleanup, que es un rol basado en un servicio que permite a Amazon EMR cancelar y eliminar los recursos de Amazon EC2 en su nombre si el rol de servicio de Amazon EMR ha perdido esa capacidad. Si aún no existe, Amazon EMR crea el rol automáticamente durante la creación del clúster.

El rol vinculado a un servicio AWSServiceRoleForEMRCleanup depende de los siguientes servicios para asumir el rol:

- `elasticmapreduce.amazonaws.com`

La política de permisos del rol se denomina AWSServiceRoleForEMRCleanup y permite a Amazon EMR realizar las siguientes acciones en los recursos especificados:

- Acción: `DescribeInstances` en `ec2`
- Acción: `DescribeSpotInstanceRequests` en `ec2`
- Acción: `ModifyInstanceAttribute` en `ec2`
- Acción: `TerminateInstances` en `ec2`
- Acción: `CancelSpotInstanceRequests` en `ec2`
- Acción: `DeleteNetworkInterface` en `ec2`
- Acción: `DescribeInstanceAttribute` en `ec2`
- Acción: `DescribeVolumeStatus` en `ec2`
- Acción: `DescribeVolumes` en `ec2`
- Acción: `DetachVolume` en `ec2`
- Acción: `DeleteVolume` en `ec2`

Debe configurar permisos para permitir a una entidad de IAM (como un usuario, grupo o rol) crear, editar o eliminar un rol vinculado a servicios.

Para permitir a una entidad de IAM crear el rol vinculado a un servicio AWSServiceRoleForEMRCleanup

Agregue la siguiente instrucción a la política de permisos de la entidad de IAM que tiene que crear el rol vinculado al servicio:

```
{
```

```
    "Effect": "Allow",
    "Action": [
        "iam:CreateServiceLinkedRole",
        "iam:PutRolePolicy"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/elasticmapreduce.amazonaws.com*/
AWSServiceRoleForEMRCleanup*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "elasticmapreduce.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

Para permitir que una entidad de IAM pueda editar la descripción del rol vinculado a un servicio AWSServiceRoleForEMRCleanup

Agregue la siguiente instrucción a la política de permisos de la entidad de IAM que tiene que editar la descripción del rol vinculado al servicio:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:UpdateRoleDescription"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/elasticmapreduce.amazonaws.com*/
AWSServiceRoleForEMRCleanup*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "elasticmapreduce.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

Para permitir que una entidad de IAM pueda eliminar el rol vinculado a un servicio AWSServiceRoleForEMRCleanup

Agregue la siguiente instrucción a la política de permisos de la entidad de IAM que tiene que eliminar un rol vinculado al servicio:

```
{
    "Effect": "Allow",
    "Action": [
        "iam:DeleteServiceLinkedRole",
        "iam:GetServiceLinkedRoleDeletionStatus"
    ],
    "Resource": "arn:aws:iam::*:role/aws-service-role/elasticmapreduce.amazonaws.com*/
AWSServiceRoleForEMRCleanup*",
    "Condition": {
        "StringLike": {
            "iam:AWSServiceName": [
                "elasticmapreduce.amazonaws.com",
                "elasticmapreduce.amazonaws.com.cn"
            ]
        }
    }
}
```

Creación de un rol vinculado a un servicio para Amazon EMR

No es necesario crear manualmente el rol AWSServiceRoleForEMRCleanup. Cuando se lanza un clúster, ya sea por primera vez o porque no hay ningún rol vinculado a un servicio, Amazon EMR crea el rol vinculado a un servicio en su nombre. Debe tener permisos de para crear el rol vinculado a servicios. Para obtener una instrucción de ejemplo que agregue esta capacidad a la política de permisos de una entidad de IAM (como un usuario, grupo o rol), consulte Permisos de roles vinculados a un servicio para Amazon EMR.

> **Important**
>
> Si utilizaba el servicio de Amazon EMR antes del 24 de octubre de 2017, fecha en que comenzó a admitir los roles vinculados a un servicio, Amazon EMR creó el rol AWSServiceRoleForEMRCleanup en su cuenta. Para obtener más información, consulte Un nuevo rol ha aparecido en la cuenta de IAM.

Edición de un rol vinculado a un servicio para Amazon EMR

Amazon EMR no permite editar el rol vinculado a un servicio AWSServiceRoleForEMRCleanup. Después de crear un rol vinculado a servicios, no puede cambiarle el nombre, ya que varias entidades pueden hacer referencia al mismo. Sin embargo, puede editar la descripción del rol mediante IAM.

Edición de la descripción de un rol vinculado a un servicio (consola de IAM)

Puede utilizar la consola de IAM para editar la descripción de un rol vinculado a un servicio.

Para editar la descripción de un rol vinculado a un servicio (consola)

1. En el panel de navegación de la consola de IAM, elija Roles (Roles).
2. Seleccione el nombre del rol que desea modificar.
3. En el extremo derecho de Descripción del rol, seleccione Editar.
4. Ingrese una descripción nueva en el cuadro y elija Save changes (Guardar cambios).

Edición de la descripción de un rol vinculado a un servicio (CLI de IAM)

Puede utilizar comandos de IAM desde la AWS Command Line Interface para editar la descripción de un rol vinculado a un servicio.

Para cambiar la descripción de un rol vinculado a un servicio (CLI)

1. (Opcional) Para ver la descripción actual de un rol, ejecute uno de los siguientes comandos:

   ```
   $ aws iam get-role --role-name role-name
   ```

   Utilice el nombre del rol, no el ARN, para hacer referencia a los roles con los comandos de CLI. Por ejemplo, si un rol tiene el ARN `arn:aws:iam::123456789012:role/myrole`, debe referirse a él como **`myrole`**.
2. Para actualizar la descripción de un rol vinculado a un servicio, ejecute uno de los siguientes comandos:

   ```
   $ aws iam update-role-description --role-name role-name --description description
   ```

Edición de la descripción de un rol vinculado a un servicio (API de IAM)

Puede utilizar la API de IAM para editar la descripción de un rol vinculado a un servicio.

Para cambiar la descripción de un rol vinculado a un servicio (API)

1. (Opcional) Para ver la descripción actual de una función, ejecute el siguiente comando:

   API de IAM: GetRole
2. Para actualizar la descripción de una función, use el siguiente comando:

   API de IAM: UpdateRoleDescription

Eliminación de un rol vinculado a un servicio para Amazon EMR

Si ya no necesita utilizar una característica o servicio que requiere un rol vinculado a servicios, recomendamos que elimine dicho rol. De esta forma, no tendrá una entidad no utilizada cuya supervisión o mantenimiento no se realizan de forma activa. Sin embargo, debe limpiar el rol vinculado al servicio antes de eliminarlo.

Limpiar un rol vinculado a servicios

Antes de poder utilizar IAM para eliminar un rol vinculado a un servicio, primero debe confirmar que dicho rol no tiene sesiones activas y eliminar los recursos que utiliza.

Para comprobar si el rol vinculado a un servicio tiene una sesión activa en la consola de IAM

1. Abra la consola de IAM en https://console.aws.amazon.com/iam/.
2. Seleccione Roles (Roles) en el panel de navegación. Seleccione el nombre (no la casilla de verificación) del rol AWSServiceRoleForEMRCleanup.
3. En la página Summary (Resumen) del rol seleccionado, seleccione Access Advisor.
4. En la pestaña Access Advisor, revise la actividad reciente del rol vinculado al servicio.

> **Note**
>
> Si no sabe si Amazon EMR utiliza el rol AWSServiceRoleForEMRCleanup, puede intentar eliminar el rol para comprobarlo. Si el servicio está utilizando el rol, este no podrá eliminarse y podrá ver las regiones de en las que se está utilizando. Si el rol se

está utilizando, debe esperar que la sesión finalice para poder eliminarlo. No se puede revocar la sesión de un rol vinculado a servicios.

Para eliminar los recursos de Amazon EMR que utiliza el rol vinculado a un servicio AWSServiceRoleForEMRCleanup

- Termine todos los clústeres de su cuenta. Para obtener más información, consulte Terminar un clúster.

Eliminación de un rol vinculado a un servicio (consola de IAM)

Puede utilizar la consola de IAM para eliminar un rol vinculado a un servicio.

Para eliminar un rol vinculado a un servicio (consola)

1. Abra la consola de IAM en https://console.aws.amazon.com/iam/.
2. Seleccione Roles (Roles) en el panel de navegación. A continuación, seleccione la casilla de verificación situada junto a AWSServiceRoleForEMRCleanup, no el nombre ni la propia fila.
3. En Role actions (Acciones de rol) en la parte superior de la página, elija Delete role (Eliminar rol).
4. En el cuadro de diálogo de confirmación, revise los datos del último acceso al servicio, que muestra cuándo cada uno de los roles seleccionados tuvo acceso a un servicio de AWS por última vez. Esto lo ayuda a confirmar si el rol está actualmente activo. Para continuar, elija Yes, Delete.
5. Consulte las notificaciones de la consola de IAM para monitorear el progreso de la eliminación del rol vinculado al servicio. Como el proceso de eliminación del rol vinculado al servicio de IAM es asíncrono, dicha tarea puede realizarse correctamente o fallar después de que envía la solicitud de eliminación. Si la tarea no se realiza correctamente, puede seleccionar View details (Ver detalles) o View Resources (Ver recursos) desde las notificaciones para obtener información sobre el motivo por el que no se pudo eliminar el rol. Si la eliminación no pudo producirse porque hay recursos en el servicio que está utilizando el rol, entonces el motivo del error incluye una lista de recursos.

Eliminación de un rol vinculado a un servicio (CLI de IAM)

Puede utilizar los comandos de IAM desde la AWS Command Line Interface para eliminar un rol vinculado a un servicio. Como los roles vinculados a servicios no se puede eliminar si están en uso o tienen recursos asociados, debe enviar una solicitud de eliminación. Si estas condiciones no se cumplen, dicha solicitud se puede denegar.

Para eliminar un rol vinculado a un servicio (CLI)

1. Para comprobar el estado de la tarea de eliminación, debe apuntar el valor de `deletion-task-id` de la respuesta. Escriba el siguiente comando para enviar una solicitud de eliminación de un rol vinculado a un servicio:

   ```
   $ aws iam delete-service-linked-role --role-name AWSServiceRoleForEMRCleanup
   ```

2. Escriba el siguiente comando para comprobar el estado de la tarea de eliminación:

   ```
   $ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
   ```

   El estado de la tarea de eliminación puede ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED` o `FAILED`. Si ocurre un error durante la eliminación, la llamada devuelve el motivo del error para que pueda resolver el problema.

Eliminación de un rol vinculado a un servicio (API de IAM)

Puede utilizar la API de IAM para eliminar un rol vinculado a un servicio. Como los roles vinculados a servicios no se puede eliminar si están en uso o tienen recursos asociados, debe enviar una solicitud de eliminación. Si estas condiciones no se cumplen, dicha solicitud se puede denegar.

Para eliminar un rol vinculado a un servicio (API)

1. Para enviar una solicitud de eliminación de un rol vinculado a un servicio, realice una llamada a DeleteServiceLinkedRole. En la solicitud, especifique el nombre de rol AWSServiceRoleForEMRCleanup.

   Para comprobar el estado de la tarea de eliminación, debe apuntar el valor de `DeletionTaskId` de la respuesta.

2. Para comprobar el estado de la tarea de eliminación, realice una llamada a GetServiceLinkedRoleDeletionStatus. En la solicitud, especifique el valor de `DeletionTaskId`.

   El estado de la tarea de eliminación puede ser `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED` o `FAILED`. Si ocurre un error durante la eliminación, la llamada devuelve el motivo del error para que pueda resolver el problema.

Regiones admitidas para los roles vinculados a un servicio de Amazon EMR

Amazon EMR admite el uso de roles vinculados a un servicio en las siguientes regiones:

Nombre de la región de	Identidad de la región	Compatibilidad en Amazon EMR
US East (N. Virginia)	us-east-1	Sí
US East (Ohio)	us-east-2	Sí
EE. UU Oeste (Norte de California)	us-west-1	Sí
EE. UU. Oeste (Oregón)	us-west-2	Sí
Asia Pacífico (Mumbai)	ap-south-1	Sí
Asia Pacífico (Osaka)	ap-northeast-3	Sí
Asia Pacífico (Seúl)	ap-northeast-2	Sí
Asia Pacífico (Singapur)	ap-southeast-1	Sí
Asia Pacífico (Sídney)	ap-southeast-2	Sí
Asia Pacífico (Tokio)	ap-northeast-1	Sí
Canada (Central)	ca-central-1	Sí
Europa (Fráncfort)	eu-central-1	Sí
Europe (Ireland)	eu-west-1	Sí
Europa (Londres)	eu-west-2	Sí

Nombre de la región de	Identidad de la región	Compatibilidad en Amazon EMR
Europa (París)	eu-west-3	Sí
South America (São Paulo)	sa-east-1	Sí

Personalización de roles de IAM

Es posible que quiera personalizar los permisos y los roles de servicio de IAM para limitar los privilegios de acuerdo con los requisitos de seguridad. Para personalizar los permisos, le recomendamos que cree nuevos roles y políticas. Comience con los permisos de las políticas administradas para los roles predeterminados (por ejemplo, `AmazonElasticMapReduceforEC2Role` y `AmazonElasticMapReduceRole`). A continuación, copie y pegue el contenido de las nuevas instrucciones de política, modifique los permisos según corresponda y asocie las políticas de permisos modificadas a los roles que cree. Debe disponer de los permisos de IAM adecuados para trabajar con los roles y las políticas. Para obtener más información, consulte Cómo permitir a los usuarios y grupos crear y modificar roles.

Si crea un rol de EMR personalizado para EC2, siga el flujo de trabajo básico, que crea automáticamente un perfil de instancia con el mismo nombre. Amazon EC2 le permite crear roles y perfiles de instancia con nombres diferentes, pero Amazon EMR no admite esta configuración y se produce un error “Perfil de instancia no válido” al crear el clúster.

> **Important**
>
> Las políticas insertadas no se actualizan automáticamente cuando cambian los requisitos de servicio. Si crea y adjunta políticas insertadas, tenga en cuenta que se pueden producir actualizaciones de servicio que provoquen errores de permisos de forma repentina. Para obtener más información, consulte Políticas administradas y políticas insertadas en la Guía del usuario de IAM y Cómo especificar roles de IAM personalizados al crear un clúster.

Para obtener más información sobre cómo trabajar con los roles de IAM, consulte los siguientes temas en la Guía del usuario de IAM:

- Creación de un rol para delegar permisos a un servicio de AWS
- Modificación de un rol

- Eliminación de un rol

Cómo especificar roles de IAM personalizados al crear un clúster

El rol de servicio para Amazon EMR y el rol para el perfil de instancia de Amazon EC2 se especifican al crear un clúster. El usuario que crea los clústeres necesita permisos para recuperar y asignar roles a Amazon EMR y a las instancias de EC2. De lo contrario, se produce el error La cuenta no tiene autorización para llamar a EC2. Para obtener más información, consulte Cómo permitir a los usuarios y grupos crear y modificar roles.

Uso de la consola para especificar roles personalizados

Cuando cree un clúster, puede especificar un rol de servicio personalizado para Amazon EMR, un rol personalizado para el perfil de instancia de EC2 y un rol de Auto Scaling personalizado mediante Opciones avanzadas. Si utiliza las Quick options (Opciones rápidas), se especifican el rol de servicio y el rol para el perfil de instancia EC2 predeterminados. Para obtener más información, consulte Roles de servicio de IAM utilizados por Amazon EMR.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para especificar roles de IAM personalizados mediante la nueva consola

Al crear un clúster con la nueva consola, debe especificar un rol de servicio personalizado para Amazon EMR y un rol personalizado para el perfil de instancia de EC2. Para obtener más información, consulte Roles de servicio de IAM utilizados por Amazon EMR.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Configuración y permisos de seguridad, busque los campos Rol de IAM para el perfil de instancia y Rol de servicio para Amazon EMR. Seleccione un rol en la lista para cada tipo de

rol. Solo aparecerán los roles de la cuenta que tengan la política de confianza adecuada para ese tipo de rol.

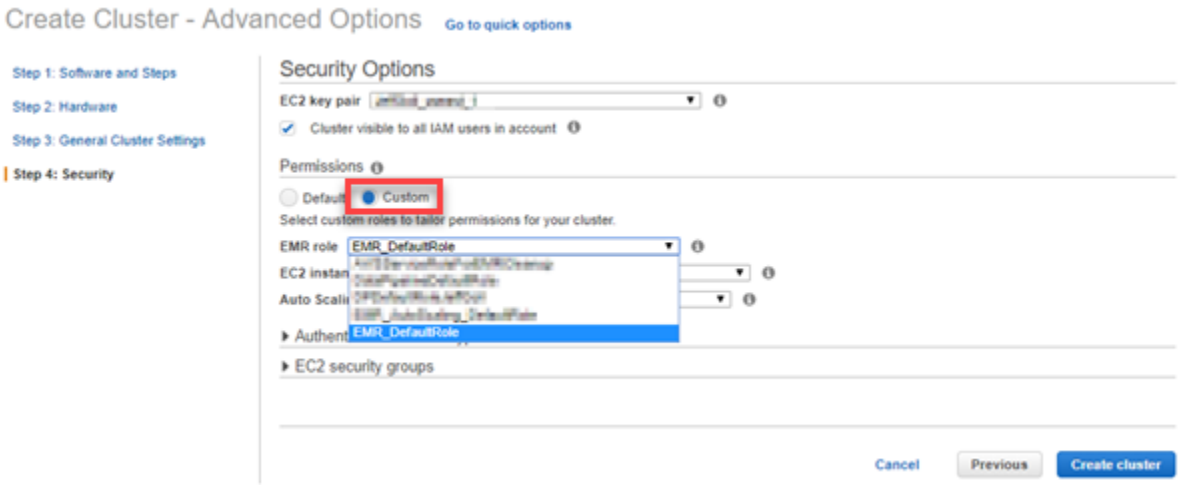
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para especificar roles de IAM personalizados mediante la consola antigua

Cuando cree un clúster con la consola antigua, puede especificar un rol de servicio personalizado para Amazon EMR, un rol personalizado para el perfil de instancia de EC2 y un rol de Auto Scaling personalizado mediante Opciones avanzadas. Si utiliza las Quick options (Opciones rápidas), se especifican el rol de servicio y el rol para el perfil de instancia EC2 predeterminados. Para obtener más información, consulte Roles de servicio de IAM utilizados por Amazon EMR.

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Crear clúster e Ir a las opciones avanzadas.
3. Elija la configuración del clúster apropiada para la aplicación hasta llegar a Security Options (Opciones de seguridad). En Permisos, están seleccionados los roles Predeterminados para Amazon EMR.
4. Elija Custom (Personalizados).
5. Seleccione un rol en la lista para cada tipo de rol. Solo aparecerán los roles de la cuenta que tengan la política de confianza adecuada para ese tipo de rol.



6. Elija las demás opciones que necesite para el clúster y, a continuación, elija Create Cluster (Crear clúster).

Uso de la AWS CLI para especificar roles personalizados

Puede especificar un rol de servicio para Amazon EMR y un rol de servicio para las instancias de EC2 del clúster de forma explícita mediante las opciones con el comando `create-cluster` desde la AWS CLI. Utilice la opción `--service-role` para especificar el rol de servicio. Utilice el argumento `InstanceProfile` de la opción `--ec2-attributes` para especificar el rol para el perfil de instancia EC2.

El rol de Auto Scaling se especifica mediante la opción `--auto-scaling-role`. Para obtener más información, consulte Uso del escalado automático con una política personalizada para grupos de instancias .

Para especificar roles de IAM personalizados mediante la AWS CLI

- El siguiente comando especifica el rol de servicio personalizado *MyCustomServiceRoleForEMR* y el rol personalizado para el perfil de instancia EC2 *MyCustomServiceRoleForClusterEC2Instances* al lanzar un clúster. Este ejemplo utiliza el rol de Amazon EMR predeterminado.

  > **Note**
  >
  > Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name "Test cluster" --release-label emr-5.36.1 \
--applications Name=Hive Name=Pig --service-role MyCustomServiceRoleForEMR \
--ec2-attributes InstanceProfile=MyCustomServiceRoleForClusterEC2Instances,\
KeyName=myKey --instance-type m5.xlarge --instance-count 3
```

Puede utilizar estas opciones para especificar roles predeterminados de forma explícita en lugar de usar la opción `--use-default-roles`. La opción `--use-default-roles` especifica el rol de

servicio y el rol para el perfil de la instancia EC2 que se ha definido en el archivo `config` de la AWS CLI.

El siguiente ejemplo muestra el contenido de un archivo `config` para el que AWS CLI especifica los roles personalizados de Amazon EMR. Con este archivo de configuración, cuando se especifica la opción `--use-default-roles`, el clúster se crea con *`MyCustomServiceRoleForEMR`* y *`MyCustomServiceRoleForClusterEC2Instances`*. De forma predeterminada, el archivo `config` especifica el valor predeterminado `service_role` como `AmazonElasticMapReduceRole` y el valor predeterminado `instance_profile` como `EMR_EC2_DefaultRole`.

```
[default]
output = json
region = us-west-1
aws_access_key_id = myAccessKeyID
aws_secret_access_key = mySecretAccessKey
emr =
     service_role = MyCustomServiceRoleForEMR
     instance_profile = MyCustomServiceRoleForClusterEC2Instances
```

Configuración de roles de IAM para solicitudes de EMRFS a Amazon S3

Cuando una aplicación que se ejecuta en un clúster hace referencia a los datos con el formato `s3://mydata`, Amazon EMR utiliza EMRFS para realizar la solicitud. Para interactuar con Amazon S3, EMRFS asume las políticas de permisos que se adjuntan a su perfil de instancia de Amazon EC2. El mismo perfil de instancia de Amazon EC2 se utiliza independientemente del usuario o grupo mediante la aplicación o la ubicación de los datos en Amazon S3.

Si tiene clústeres con varios usuarios que necesitan diferentes niveles de acceso a los datos en Amazon S3 a través de EMRFS, puede establecer una configuración de seguridad con los roles de IAM para EMRFS. EMRFS puede asumir un rol de servicio diferente para las instancias de EC2 del clúster en función del usuario o grupo que realiza la solicitud o en función de la ubicación de los datos en Amazon S3. Cada rol de IAM para EMRFS puede tener diferentes permisos para obtener acceso a los datos en Amazon S3. Para obtener más información acerca del rol de servicio para las instancias de EC2 del clúster, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2).

Las versiones 5.10.0 y posteriores de Amazon EMR admiten el uso de roles de IAM personalizados para EMRFS. Si utiliza una versión anterior o tiene requisitos de autorización que los roles de

IAM para EMRFS no son capaces de proporcionar, puede crear un proveedor de credenciales personalizado en su lugar. Para obtener más información, consulte Autorización de acceso a los datos de EMRFS en Amazon S3.

Cuando se usa una configuración de seguridad para especificar roles de IAM para EMRFS, debe definir asignaciones de roles. Cada asignación de roles especifica un rol de IAM que corresponde a los identificadores. Dichos identificadores determinan la base para el acceso a Amazon S3 a través de EMRFS. Los identificadores pueden ser usuarios, grupos o prefijos de Amazon S3 que indican una ubicación de datos. Cuando EMRFS realiza una solicitud a Amazon S3, y la solicitud coincide con la base para el acceso, EMRFS tiene las instancias de EC2 del clúster que asumen el rol de IAM correspondiente para la solicitud. Los permisos de IAM asociados a ese rol se aplican en lugar de los permisos de IAM adjuntos al rol de servicio para las instancias de EC2 del clúster.

Los usuarios y grupos de un mapeo de roles son usuarios y grupos de Hadoop que se definen en el clúster. Los usuarios y grupos se transfieren a EMRFS en el contexto de la aplicación que lo utiliza (por ejemplo, una suplantación de usuario de YARN). El prefijo de Amazon S3 puede ser un especificador de bucket de cualquier profundidad (por ejemplo, `s3://mybucket` o `s3://mybucket/myproject/mydata`). Puede especificar varios identificadores dentro de un único mapeo de roles, pero todos deben ser del mismo tipo.

> **Important**
>
> Los roles de IAM para EMRFS proporcionan aislamiento de la aplicación entre los usuarios de la aplicación. No proporcionan aislamiento de nivel del host entre los usuarios del host. Cualquier usuario con acceso a el clúster puede omitir el aislamiento para asumir cualquiera de los roles.

Cuando una aplicación del clúster realiza una solicitud a Amazon S3 a través de EMRFS, EMRFS evalúa las asignaciones de roles en el orden descendente en el que aparecen en la configuración de seguridad. Si una solicitud realizada a través de EMRFS no coincide con ningún identificador, EMRFS vuelve a usar el rol de servicio de las instancias de EC2 del clúster. Por este motivo, le recomendamos que las políticas asociadas a este rol limiten los permisos a Amazon S3. Para obtener más información, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2).

Configurar roles de

Antes de definir una configuración de seguridad con roles de IAM para EMRFS, planifique y cree los roles y las políticas de permisos que va a adjuntar a los roles. Para obtener más información, consulte ¿Cómo funcionan los roles de instancias de Amazon EC2? en la Guía del usuario de IAM. Al crear políticas de permisos, le recomendamos que empiece con la política administrada asociada al rol de Amazon EMR predeterminado para EC2 y que luego edite esta política de acuerdo con sus necesidades. El nombre del rol predeterminado es `EMR_EC2_DefaultRole` y la política administrada predeterminada para editar es `AmazonElasticMapReduceforEC2Role`. Para obtener más información, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2).

Actualización de políticas de confianza para asumir los permisos del rol

Cada rol que utiliza EMRFS debe tener una política de confianza que permita que el rol de Amazon EMR del clúster de EC2 lo asuma. Igualmente, el rol de Amazon EMR del clúster para EC2 debe tener una política de confianza que permita que los roles de EMRFS lo asuman.

En el ejemplo siguiente, la política de confianza se asocia a los roles para EMRFS. La instrucción permite que el rol de Amazon EMR predeterminado de EC2 asuma el rol. Por ejemplo, si tiene dos roles de EMRFS ficticios, `EMRFSRole_First` y `EMRFSRole_Second`, esta instrucción de la política se añade a las políticas de confianza para cada uno de ellos.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "AWS":"arn:aws:iam::AWSAcctID:role/EMR_EC2_DefaultRole"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

Además, en el ejemplo siguiente, se añade la instrucción de la política de confianza a `EMR_EC2_DefaultRole` para permitir que los dos roles de EMRFS ficticios la asuman.

```
{
```

```
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "AWS": ["arn:aws:iam::AWSAcctID:role/EMRFSRole_First",
 "arn:aws:iam::AWSAcctID:role/EMRFSRole_Second"]
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

Para actualizar la política de confianza de un rol de IAM

Abra la consola de IAM en https://console.aws.amazon.com/iam/.

1. Elija Roles, escriba el nombre del rol en Search (Buscar) y, a continuación, seleccione su Role name (Nombre de rol).
2. Seleccione Trust relationships (Relaciones de confianza), Edit trust relationship (Editar relaciones de confianza).
3. Agregue una instrucción de confianza de acuerdo con Documento de política siguiendo las directrices anteriores y, a continuación, seleccione Actualizar la política de confianza.

Especificación de un rol como un usuario de claves

Si un rol permite el acceso a una ubicación en Amazon S3 que está cifrada mediante una AWS KMS key, asegúrese de que el rol se especifica como un usuario de claves. Esto proporciona al rol permiso para utilizar la clave de KMS. Para obtener más información, consulte Políticas de claves en AWS KMS en la Guía para desarrolladores de AWS Key Management Service.

Definición de una configuración de seguridad con roles de IAM para EMRFS

> **Important**
>
> Si no se aplica ninguno de los roles de IAM para EMRFS especificados, EMRFS vuelve a usar el rol de Amazon EMR para EC2. Considere la posibilidad de personalizar este rol para restringir los permisos en Amazon S3 según sea necesario para su aplicación y de especificar después este rol personalizado en lugar de `EMR_EC2_DefaultRole` cuando

cree un clúster. Para obtener más información, consulte Personalización de roles de IAM y Cómo especificar roles de IAM personalizados al crear un clúster.

Para especificar roles de IAM para solicitudes de EMRFS a Amazon S3 con la consola

1. Cree una configuración de seguridad que especifique los mapeos de roles:
   a. En la consola de Amazon EMR, seleccione Configuraciones de seguridad, Crear.
   b. Escriba un nombre para la configuración de seguridad en el campo Name (Nombre). Este nombre se utiliza para especificar la configuración de seguridad cuando crea un clúster.
   c. Elija Usar roles de IAM para solicitudes de EMRFS a Amazon S3.
   d. Seleccione el rol de IAM que desea aplicar, elija en la lista Base para el acceso un identificador de tipo (Usuarios, Grupos o Prefijos de S3) e ingrese los identificadores correspondientes. Si utiliza varios identificadores, sepárelos con una coma y sin espacios. Para obtener más información acerca de cada tipo de identificador, consulte JSON configuration reference a continuación.
   e. Elija Add role (Añadir rol) para configurar mapeos de roles adicionales, tal y como se describe en el paso anterior.
   f. Defina otras opciones de configuración de seguridad según corresponda y elija Create (Crear). Para obtener más información, consulte Creación de una configuración de seguridad.
2. Especifique la configuración de seguridad que creó anteriormente cuando cree un clúster. Para obtener más información, consulte Especificación de una configuración de seguridad para un clúster.

Para especificar roles de IAM para las solicitudes de EMRFS a Amazon S3 con la AWS CLI

1. Utilice el comando `aws emr create-security-configuration`, especificando un nombre para la configuración de seguridad y los detalles de configuración de seguridad en formato JSON.

   El comando de ejemplo que se muestra a continuación crea una configuración de seguridad con el nombre `EMRFS_Roles_Security_Configuration`. Se basa en una estructura JSON del archivo `MyEmrfsSecConfig.json`, que se guarda en el mismo directorio en el que se ejecuta el comando.

```
aws emr create-security-configuration --name EMRFS_Roles_Security_Configuration --
security-configuration file://MyEmrFsSecConfig.json.
```

Utilice las siguientes directrices para la estructura del archivo `MyEmrFsSecConfig.json`. Puede especificar esta estructura junto con estructuras de otras opciones de configuración de seguridad. Para obtener más información, consulte Creación de una configuración de seguridad.

A continuación, se muestra un ejemplo de fragmento de JSON para especificar roles de IAM personalizados para EMRFS dentro de una configuración de seguridad. Muestra las asignaciones de roles para los tres tipos de identificadores diferentes, seguidas de una referencia de parámetros.

```
{
  "AuthorizationConfiguration": {
    "EmrFsConfiguration": {
      "RoleMappings": [{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_user1",
        "IdentifierType": "User",
        "Identifiers": [ "user1" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_to_MyBuckets",
        "IdentifierType": "Prefix",
        "Identifiers": [ "s3://MyBucket/","s3://MyOtherBucket/" ]
      },{
        "Role": "arn:aws:iam::123456789101:role/allow_EMRFS_access_for_AdminGroup",
        "IdentifierType": "Group",
        "Identifiers": [ "AdminGroup" ]
      }]
    }
  }
}
```

Parámetro	Descripción
`"AuthorizationConfiguration":`	Obligatorio.
`"EmrFsConfiguration":`	Obligatorio. Contiene asignaciones de roles.

Parámetro	Descripción
`"RoleMappings":`	Obligatorio. Contiene una o más definiciones de asignación de roles. Las asignaciones de roles se evalúan de forma descendente por orden de aparición. Si una asignación de roles se evalúa como true en una invocación de datos de EMRFS en Amazon S3, no se evalúan más asignaciones de roles y EMRFS utiliza el rol de IAM especificado para la solicitud. Las asignaciones de roles tienen los siguientes parámetros obligatorios:
`"Role":`	Especifica el identificador de ARN de un rol de IAM en el formato `arn:aws:iam:: account-id :role/role-name`. Este es el rol de IAM que asume Amazon EMR si la solicitud de EMRFS a Amazon S3 coincide con alguno de los `Identifiers` especificados.

Parámetro	Descripción
`"IdentifierType":`	Puede ser uno de los siguientes: • `"User"` especifica que los identificadores son uno o más usuarios de Hadoop, que pueden ser usuarios de cuentas de Linux o entidades principales de Kerberos. Cuando la solicitud de EMRFS se origina en el usuario o los usuarios especificados, se asume el rol de IAM. • `"Prefix"` especifica que el identificador es una ubicación de Amazon S3. El rol de IAM se asume para las llamadas a la ubicación o ubicaciones con los prefijos especificados. Por ejemplo, el prefijo `s3://mybucket/` coincide con `s3://mybucket/mydir` y `s3://mybucket/yetanotherdir` . • `"Group"` especifica que los identificadores son uno o más grupos de Hadoop. El rol de IAM se asume si la solicitud proviene de un usuario del grupo o grupos especificados.
`"Identifiers":`	Especifica uno o más identificadores del tipo de identificador adecuado. Separe varios identificadores con comas sin espacios.

2. Utilice el comando `aws emr create-cluster` para crear un clúster y especificar la configuración de seguridad que creó en el paso anterior.

   En el siguiente ejemplo, se crea un clúster con las aplicaciones de Hadoop básicas predeterminadas instaladas. El clúster utiliza la configuración de seguridad creada anteriormente como `EMRFS_Roles_Security_Configuration` y también utiliza un rol de Amazon EMR personalizado para EC2, `EC2_Role_EMR_Restrict_S3`, que se especifica mediante el argumento `InstanceProfile` del parámetro `--ec2-attributes`.

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name MyEmrFsS3RolesCluster \
--release-label emr-5.36.1 --ec2-attributes
 InstanceProfile=EC2_Role_EMR_Restrict_S3,KeyName=MyKey \
--instance-type m5.xlarge --instance-count 3 \
--security-configuration EMRFS_Roles_Security_Configuration
```

Uso de políticas basadas en recursos para el acceso de Amazon EMR a Catálogo de datos de AWS Glue

Si utiliza AWS Glue junto con Hive, Spark o Presto en Amazon EMR, AWS Glue admite políticas basadas en recursos para controlar el acceso a los recursos del Catálogo de datos. Estos recursos incluyen bases de datos, tablas, conexiones y funciones definidas por el usuario. Para obtener más información, consulte Políticas de recursos de AWS Glue en la Guía para desarrolladores de AWS Glue.

Cuando utilice políticas basadas en recursos para limitar el acceso a AWS Glue desde Amazon EMR, la entidad principal que especifique en la política de permisos debe ser el ARN del rol asociado al perfil de instancia de EC2 que se especifica al crear un clúster. Por ejemplo, para una política basada en recursos adjunta a un catálogo, puede especificar el ARN del rol como rol de servicio predeterminado para las instancias de EC2 del clúster, *EMR_EC2_DefaultRole* como `Principal`, con el formato que se muestra en el siguiente ejemplo:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

El *acct-id* puede ser diferente del ID de la cuenta de AWS Glue. Esto permite el acceso desde clústeres de EMR en diferentes cuentas. Puede especificar varias entidades principales, cada una de ellas desde una cuenta diferente.

Uso de roles de IAM con las aplicaciones que llaman directamente a los servicios de AWS

Las aplicaciones que se ejecutan en las instancias EC2 de un clúster pueden usar el perfil de instancia de EC2 para obtener credenciales de seguridad temporales al llamar a los servicios de AWS.

Las versiones de Hadoop disponibles con la versión 2.3.0 y posteriores de Amazon EMR ya se han actualizado para utilizar roles de IAM. Si la aplicación se ejecuta estrictamente sobre la arquitectura de Hadoop y no llama directamente a ningún servicio en AWS, debería funcionar con roles de IAM, sin ningún tipo de modificación.

Si la aplicación llama a los servicios de AWS directamente, debe actualizarla para que utilice los roles de IAM. Esto significa que, en lugar de obtener credenciales de cuenta de `/etc/hadoop/conf/core-site.xml` en las instancias de EC2 del clúster, la aplicación utiliza un SDK para acceder a los recursos que utilizan roles de IAM o llama a los metadatos de la instancia de EC2 para obtener las credenciales temporales.

Para tener acceso a los recursos de AWS con roles de IAM mediante un SDK

- Los siguientes temas muestran cómo utilizar varios de los SDK de AWS para tener acceso a credenciales temporales utilizando roles de IAM. Cada tema comienza con una versión de una aplicación que no utiliza roles de IAM y, a continuación, realiza un recorrido por el proceso de conversión de dicha aplicación para utilizar roles de IAM.
  - Uso de roles de IAM para instancias de EC2 con el SDK para Java en la Guía para desarrolladores de AWS SDK for Java
  - Uso de roles de IAM para instancias de EC2 con el SDK para .NET en la Guía para desarrolladores de AWS SDK for .NET
  - Uso de roles de IAM para instancias de EC2 con el SDK para PHP en la Guía para desarrolladores de AWS SDK for PHP;
  - Uso de roles de IAM para instancias de EC2 con el SDK para Ruby en la Guía para desarrolladores de AWS SDK for Ruby

Para obtener credenciales temporales de metadatos de instancias EC2

- Llame a la siguiente URL desde una instancia EC2 que se ejecuta con el rol de IAM especificado, que devuelve las credenciales de seguridad temporales asociadas (AccessKeyId,

SecretAccessKey, SessionToken y Expiration). En el ejemplo que aparece a continuación se utiliza el perfil de instancia predeterminado para Amazon EMR, `EMR_EC2_DefaultRole`.

```
GET http://169.254.169.254/latest/meta-data/iam/security-
credentials/EMR_EC2_DefaultRole
```

Para obtener más información sobre la escritura de aplicaciones que utilizan roles de IAM, consulte Permitir el acceso a los recursos de AWS a aplicaciones que se ejecutan en instancias de Amazon EC2.

Para obtener más información acerca de las credenciales de seguridad temporales, consulte Uso de credenciales de seguridad temporales en la guía Uso de credenciales de seguridad temporales.

Cómo permitir a los usuarios y grupos crear y modificar roles

Las entidades principales de IAM (usuarios y grupos) que crean, modifican y especifican roles para un clúster, incluidos los roles predeterminados, deben tener permisos para realizar las acciones siguientes. Para obtener información sobre cada una de las acciones, consulte Acciones en la Referencia de la API de IAM.

- `iam:CreateRole`
- `iam:PutRolePolicy`
- `iam:CreateInstanceProfile`
- `iam:AddRoleToInstanceProfile`
- `iam:ListRoles`
- `iam:GetPolicy`
- `iam:GetInstanceProfile`
- `iam:GetPolicyVersion`
- `iam:AttachRolePolicy`
- `iam:PassRole`

El permiso `iam:PassRole` permite la creación del clúster. Los permisos restantes permiten la creación de las funciones predeterminadas.

Para obtener más información acerca de la actualización de permisos en IAM, consulte Cambio de los permisos de un usuario en la Guía del usuario de IAM.

Ejemplos de políticas de Amazon EMR basadas en identidades

De forma predeterminada, los usuarios y roles no tienen permiso para crear ni modificar los recursos de Amazon EMR. Tampoco pueden realizar tareas mediante la AWS Management Console, la AWS CLI, o la API de AWS. Un administrador de IAM debe crear políticas de IAM que concedan permisos a los usuarios y a los roles para realizar operaciones de la API concretas en los recursos especificados que necesiten. El administrador debe adjuntar esas políticas a los usuarios o grupos que necesiten esos permisos.

Para obtener más información acerca de cómo crear una política basada en identidad de IAM con estos documentos de políticas de JSON de ejemplo, consulte Creación de políticas en la pestaña JSON en la Guía del usuario de IAM.

Temas

- Prácticas recomendadas en materia de políticas para Amazon EMR
- Permitir a los usuarios consultar sus propios permisos
- Políticas administradas por Amazon EMR
- Políticas de IAM para el acceso basado en etiquetas para clústeres y cuadernos de EMR
- Denegación de la acción ModifyInstanceGroup
- Solución de problemas de identidad y acceso de Amazon EMR

Prácticas recomendadas en materia de políticas para Amazon EMR

Las políticas basadas en identidad son muy eficaces. Determinan si alguien puede crear o eliminar los recursos de Amazon EMR de su cuenta o acceder a ellos. Estas acciones pueden generar costes adicionales para su cuenta de AWS. Siga estas directrices y recomendaciones al crear o editar políticas basadas en identidad:

- Comenzar a utilizar políticas administradas por AWS: para comenzar a utilizar Amazon EMR rápidamente, utilice las políticas administradas por AWS a fin de proporcionar los permisos necesarios a los empleados. Estas políticas ya están disponibles en su cuenta, y las mantiene y actualiza AWS. Para obtener más información, consulte Introducción al uso de permisos con políticas administradas de AWS en la Guía del usuario de IAM y Políticas administradas por Amazon EMR.

- Conceder privilegios mínimos: al crear políticas personalizadas, conceda solo los permisos necesarios para llevar a cabo una tarea. Comience con un conjunto mínimo de permisos y conceda permisos adicionales según sea necesario. Por lo general, es más seguro que comenzar con permisos que son demasiado tolerantes e intentar hacerlos más estrictos más adelante. Para obtener más información, consulte Conceder privilegios mínimos en la Guía del usuario de IAM.
- Habilitar MFA para operaciones confidenciales: para mayor seguridad, obligue a los usuarios de que utilicen la autenticación multifactor (MFA) para acceder a recursos u operaciones de API confidenciales. Para obtener más información, consulte Uso de la autenticación multifactor (MFA) en AWS en la Guía del usuario de IAM.
- Utilizar condiciones de política para mayor seguridad: en la medida en que sea práctico, defina las condiciones en las que sus políticas basadas en identidad permitan el acceso a un recurso. Por ejemplo, puede escribir condiciones para especificar un rango de direcciones IP permitidas desde el que debe proceder una solicitud. También puede escribir condiciones para permitir solicitudes solo en un intervalo de hora o fecha especificado o para solicitar el uso de SSL o MFA. Para obtener más información, consulte Elementos de la política JSON de IAM: condición en la Guía del usuario de IAM.

Permitir a los usuarios consultar sus propios permisos

En este ejemplo, se muestra cómo podría crear una política que permita a los usuarios de IAM ver las políticas administradas e insertadas que se adjuntan a la identidad de sus usuarios. Esta política incluye permisos para llevar a cabo esta acción en la consola o mediante programación con la AWS CLI o la API de AWS.

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Sid":"ViewOwnUserInfo",
         "Effect":"Allow",
         "Action":[
            "iam:GetUser",
            "iam:GetUserPolicy",
            "iam:ListAttachedUserPolicies",
            "iam:ListGroupsForUser",
            "iam:ListUserPolicies"
         ],
         "Resource":[
            "arn:aws:iam:::user/${aws:username}"
```

```
         ]
      },
      {
         "Sid":"NavigateInConsole",
         "Effect":"Allow",
         "Action":[
            "iam:GetGroupPolicy",
            "iam:GetPolicy",
            "iam:GetPolicyVersion",
            "iam:ListAttachedGroupPolicies",
            "iam:ListGroupPolicies",
            "iam:ListGroups",
            "iam:ListPolicies",
            "iam:ListPolicyVersions",
            "iam:ListUsers"
         ],
         "Resource":""
      }
   ]
}
```

Políticas administradas por Amazon EMR

La forma más sencilla de conceder acceso completo o acceso de solo lectura a las acciones de Amazon EMR requeridas consiste en utilizar las políticas administradas de IAM para Amazon EMR. Las políticas administradas ofrecen el beneficio de que se actualizan automáticamente si cambian los requisitos de permisos. Si utiliza políticas insertadas, pueden producirse cambios en los servicios que provoquen la aparición de errores de permisos.

Amazon EMR dejará de utilizar las políticas administradas existentes (políticas de la versión 1) en favor de las nuevas políticas administradas (políticas de la versión 2). Se ha reducido el ámbito de aplicación de las nuevas políticas administradas para que estén en consonancia con las prácticas recomendadas de AWS. Una vez que las políticas administradas de la versión 1 estén obsoletas, no podrá adjuntarlas a ningún rol o usuario de IAM nuevo. Los roles y usuarios existentes que usan políticas obsoletas pueden seguir usándolas. Las políticas administradas de la versión 2 restringen el acceso mediante etiquetas. Solo permiten acciones específicas de Amazon EMR y requieren recursos de clúster etiquetados con una clave específica de EMR. Le recomendamos que revise detenidamente la documentación antes de utilizar las nuevas políticas de la versión 2.

Las políticas de la versión 1 se marcarán como obsoletas con un icono de advertencia a su lado en la lista Políticas de la consola de IAM. Las políticas obsoletas tienen las siguientes características:

- Siguen funcionando para todos los usuarios, grupos y roles asociados en ese momento. Ningún elemento deja de funcionar.
- No pueden asociarse a ningún usuario, grupo o rol nuevo. Si separa una política de una entidad actual, no puede volver a asociarla.
- Después de separar una política de la versión 1 de todas las entidades actuales, la política dejará de estar visible y ya no podrá utilizarse.

La siguiente tabla resume los cambios entre las políticas actuales (versión 1) y las políticas de la versión 2.

Cambios en las políticas administradas por EMR

Tipo de política	Nombres de las políticas	Propósito de la política	Cambios en la política de la versión 2
Política administrada por IAM para un acceso integral a EMR por parte del usuario, rol o grupo asociado	Política de la versión 1 (quedará obsoleta): **AmazonElasticMapReduceFullAccess** Nombre de la política de la versión 2 (con ámbito de aplicación): **AmazonEMRFullAccessPolicy_v2**	Concede a los usuarios permisos completos para realizar acciones de EMR. Incluye los permisos iam:PassRole para los recursos.	La política agrega un requisito previo según el cual los usuarios deben agregar etiquetas de usuario a los recursos antes de poder usar esta política. Consulte Etiquetado de recursos para usar políticas administradas. La acción iam:PassRole requiere que la condición iam:PassedToService esté configurada en un servicio específico. El acceso a Amazon EC2, Amazon S3 y otros servicios no está

Tipo de política	Nombres de las políticas	Propósito de la política	Cambios en la política de la versión 2
			permitido de forma predeterminada. Consulte Política administrada por IAM para un acceso total (política predeterminada administrada de la versión 2).
Política administrada por IAM para el acceso de solo lectura por parte del usuario, rol o grupo asociado	Política de la versión 1 (quedará obsoleta): **AmazonElasticMapReduceReadOnlyAccess** Nombre de la política de la versión 2 (con ámbito de aplicación): **AmazonEMRReadOnlyAccessPolicy_v2**	Concede a los usuarios permisos de solo lectura para realizar acciones de Amazon EMR.	Los permisos permiten únicamente acciones específicas de solo lectura de elasticmapreduce. El acceso a Amazon S3 no está permitido de forma predeterminada. Consulte Política administrada por IAM para un acceso de solo lectura (política predeterminada administrada de la versión 2).

Tipo de política	Nombres de las políticas	Propósito de la política	Cambios en la política de la versión 2
Rol de servicio de EMR predeterminado y política administrada adjunta	Nombre del rol: EMR_DefaultRole Política de la versión 1 (quedará obsoleta): AmazonElasticMapReduceRole (rol de servicio de EMR) Nombre de la política de la versión 2 (con ámbito de aplicación reducido): **AmazonEMRServicePolicy_v2**	Permite que Amazon EMR llame a otros productos de AWS en su nombre al aprovisionar recursos y llevar a cabo acciones de servicio. Este rol es necesario para todos los clústeres.	El rol de servicio de la versión 2 y la política predeterminada de la versión 2 sustituyen a la política y al rol obsoletos. La política agrega un requisito previo según el cual los usuarios deben agregar etiquetas de usuario a los recursos antes de poder usar esta política. Consulte Etiquetado de recursos para usar políticas administradas. Consulte Rol de servicio para Amazon EMR (rol de EMR).

Tipo de política	Nombres de las políticas	Propósito de la política	Cambios en la política de la versión 2
Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2)	Política de la versión 1 (quedará obsoleta): EMR_EC2_DefaultRole (perfil de instancia) Nombre de la política obsoleta: AmazonElasticMapReduceforEC2Role	Permite que las aplicaciones que se ejecutan en un clúster de EMR accedan a otros recursos de AWS, como Amazon S3. Por ejemplo, si ejecuta trabajos de Apache Spark que procesan datos de Amazon S3, la política debe permitir el acceso a dichos recursos.	Tanto el rol predeterminado como la política predeterminada están en vías de quedar obsoletos. No hay ninguna política o rol administrado por AWS predeterminado de reemplazo. Debe proporcionar una política basada en los recursos o en la identidad. Esto significa que, de forma predeterminada, las aplicaciones que se ejecuten en un clúster de EMR no tienen acceso a Amazon S3 ni a ningún otro recurso, a menos que las agregue manualmente a la política. Consulte Política administrada y rol predeterminados.

Tipo de política	Nombres de las políticas	Propósito de la política	Cambios en la política de la versión 2
Otras políticas de rol de servicio de EC2	Nombres de políticas actuales: AmazonElasticMapReduceforAutoScalingRole, AmazonElasticMapReduceEditorsRole, AmazonEMRCleanupPolicy	Proporciona los permisos que EMR necesita para acceder a otros recursos de AWS y realizar acciones si utiliza el escalado automático, y los cuadernos, o bien para limpiar los recursos de EC2.	No hay cambios para la versión 2.

Protección de iam:PassRole

Las políticas administradas de forma predeterminada con todos los permisos de Amazon EMR incorporan configuraciones de seguridad `iam:PassRole`, entre las que se incluyen las siguientes:

- Permisos `iam:PassRole` solo para roles específicos de Amazon EMR predeterminados.
- Condiciones `iam:PassedToService` que le permiten usar la política solo con los productos de AWS especificados, como `elasticmapreduce.amazonaws.com` y `ec2.amazonaws.com`.

Puede ver la versión JSON de las políticas AmazonEMRFullAccessPolicy_v2 y AmazonEMRServicePolicy_v2 en la consola de IAM. Le recomendamos crear nuevos clústeres con las políticas administradas de la versión 2.

Para crear políticas personalizadas, le recomendamos que comience a partir de las políticas administradas y las edite de acuerdo con sus requisitos.

Para obtener información sobre cómo asociar políticas a los usuarios (entidades principales), consulte Uso de políticas administradas con la AWS Management Console en la Guía del usuario de IAM.

Etiquetado de recursos para usar políticas administradas

AmazonEMRServicePolicy_v2 and AmazonEMRFullAccessPolicy_v2 dependen del acceso limitado a los recursos que Amazon EMR aprovisiona o utiliza. La reducción del alcance se logra restringiendo el acceso únicamente a los recursos que tienen una etiqueta de usuario predefinida asociada a ellos. Si utiliza cualquiera de estas dos políticas, debe pasar la etiqueta de usuario predefinida `for-use-with-amazon-emr-managed-policies = true` al aprovisionar el clúster. A continuación, Amazon EMR propagará automáticamente esa etiqueta. Además, debe agregar una etiqueta de usuario a los recursos que se enumeran en la siguiente sección. Si utiliza la consola de Amazon EMR para lanzar el clúster, consulte Consideraciones sobre el uso de la consola de Amazon EMR para lanzar clústeres con políticas administradas de la versión 2.

Para usar políticas administradas, pase la etiqueta de usuario `for-use-with-amazon-emr-managed-policies = true` al aprovisionar un clúster con la CLI, el SDK u otro método.

Al pasar la etiqueta, Amazon EMR la propaga a los volúmenes de EBS, las instancias de EC2 y las ENI de subred privada que cree. Amazon EMR también etiqueta automáticamente los grupos de seguridad que cree. Sin embargo, si desea que Amazon EMR se lance con un grupo de seguridad determinado, debe etiquetarlo. En el caso de los recursos que no haya creado Amazon EMR, debe agregar etiquetas a esos recursos. Por ejemplo, debe etiquetar las subredes de Amazon EC2, los grupos de seguridad de EC2 (si no los creó Amazon EMR) y las VPC (si desea que Amazon EMR cree grupos de seguridad). Para lanzar clústeres con políticas administradas desde la versión 2 en las VPC, debe etiquetar esas VPC con la etiqueta de usuario predefinida. Consulte, Consideraciones sobre el uso de la consola de Amazon EMR para lanzar clústeres con políticas administradas de la versión 2.

Etiquetado propagado especificado por el usuario

Amazon EMR etiqueta los recursos que crea mediante las etiquetas de Amazon EMR que especifique al crear un clúster. Amazon EMR aplica etiquetas a los recursos que cree durante la vida útil del clúster.

Amazon EMR propaga las etiquetas de usuario de los siguientes recursos:

- ENI de Subred privada (interfaces de red elásticas de acceso a servicios)
- Instancias EC2
- Volúmenes de EBS
- Plantillas de lanzamiento de EC2

Grupos de seguridad etiquetados automáticamente

Amazon EMR etiqueta los grupos de seguridad de EC2 que cree con la etiqueta necesaria para las políticas administradas de la versión 2 de Amazon EMR, `for-use-with-amazon-emr-managed-policies`, independientemente de las etiquetas que especifique en el comando create cluster. En el caso de un grupo de seguridad que se creó antes de la introducción de las políticas administradas de la versión 2, Amazon EMR no etiqueta automáticamente el grupo de seguridad. Si desea utilizar políticas administradas de la versión 2 con los grupos de seguridad predeterminados que ya existen en la cuenta, debe etiquetar los grupos de seguridad manualmente con `for-use-with-amazon-emr-managed-policies = true`.

Recursos de clúster etiquetados manualmente

Debe etiquetar manualmente algunos recursos del clúster para que los roles predeterminados de Amazon EMR puedan acceder a ellos.

- Debe etiquetar manualmente los grupos de seguridad de EC2 y las subredes de EC2 con la etiqueta de política administrada de Amazon EMR `for-use-with-amazon-emr-managed-policies`.
- Debe etiquetar manualmente una VPC si quiere que Amazon EMR cree grupos de seguridad predeterminados. EMR intentará crear un grupo de seguridad con la etiqueta específica si el grupo de seguridad predeterminado aún no existe.

Amazon EMR etiqueta automáticamente los siguientes recursos:

- Grupos de seguridad de EC2 creados por EMR

Debe etiquetar de forma manual los siguientes recursos:

- Subred de EC2
- Grupos de seguridad de la EC2

De forma opcional, puede etiquetar de forma manual los siguientes recursos:

- VPC: solo cuando desee que Amazon EMR cree grupos de seguridad

Consideraciones sobre el uso de la consola de Amazon EMR para lanzar clústeres con políticas administradas de la versión 2

Puede aprovisionar clústeres con políticas administradas de la versión 2 mediante la consola de Amazon EMR. Estas son algunas consideraciones que debe tener en cuenta al utilizar la consola para lanzar clústeres de Amazon EMR.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR. La capacidad de etiquetado automático aún no está disponible en la nueva consola, y la nueva consola tampoco muestra qué recursos (VPC/subredes) deben etiquetarse. Consulte Novedades de la consola para obtener más información sobre las diferencias entre la consola antigua y la nueva.

- No es necesario pasar la etiqueta predefinida. Amazon EMR agrega automáticamente la etiqueta y la propaga a los componentes correspondientes.
- En el caso de los componentes que deben etiquetarse manualmente, la antigua consola de Amazon EMR intenta etiquetarlos automáticamente si tiene los permisos necesarios para etiquetar los recursos. Si no tiene los permisos para etiquetar los recursos o si desea utilizar la nueva consola, pida al administrador que etiquete esos recursos.
- No puede lanzar clústeres con políticas administradas de la versión 2, a menos que se cumplan todos los requisitos previos.
- La consola antigua de Amazon EMR le muestra qué recursos (VPC/subredes) deben etiquetarse.

Política administrada por IAM para un acceso total (política predeterminada administrada de la versión 2)

Las políticas administradas por defecto de EMR con el ámbito de aplicación de la versión 2 otorgan privilegios de acceso específicos a los usuarios. Requieren una etiqueta de recurso de Amazon EMR predefinida y claves de condición de `iam:PassRole` para los recursos que utiliza Amazon EMR, como la `Subnet` y el `SecurityGroup` que utiliza para lanzar el clúster.

Para conceder todas las acciones necesarias dentro del ámbito de aplicación de Amazon EMR, adjunte la política administrada `AmazonEMRFullAccessPolicy_v2`. Esta política administrada predeterminada actualizada sustituye a la política administrada `AmazonElasticMapReduceFullAccess`.

`AmazonEMRFullAccessPolicy_v2` depende del acceso limitado a los recursos que Amazon EMR aprovisiona o utiliza. Cuando utilice esta política, tendrá que pasar la etiqueta de usuario `for-use-with-amazon-emr-managed-policies = true` al aprovisionar el clúster. Amazon EMR propagará automáticamente la etiqueta. Además, es posible que tenga que agregar manualmente una etiqueta de usuario a tipos específicos de recursos, como grupos de seguridad de EC2 que no fueron creados por Amazon EMR. Para obtener más información, consulte Etiquetado de recursos para usar políticas administradas.

La política `AmazonEMRFullAccessPolicy_v2` protege los recursos de la siguiente manera:

- Requiere que los recursos se etiqueten con la etiqueta de políticas administradas de Amazon EMR predefinida `for-use-with-amazon-emr-managed-policies` para la creación de clústeres y el acceso a Amazon EMR.
- Restringe la acción `iam:PassRole` a roles predeterminados específicos y el acceso `iam:PassedToService` a servicios específicos.
- Ya no proporciona acceso a Amazon EC2, Amazon S3 ni otros servicios de forma predeterminada.

El contenido de esta política se muestra a continuación.

> **Note**
>
> También puede utilizar el enlace de la consola `AmazonEMRFullAccessPolicy_v2` para ver esta política.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "RunJobFlowExplicitlyWithEMRManagedTag",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:RunJobFlow"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:RequestTag/for-use-with-amazon-emr-managed-policies": "true"
                }
```

```
            }
        },
        {
            "Sid": "ElasticMapReduceActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:AddInstanceFleet",
                "elasticmapreduce:AddInstanceGroups",
                "elasticmapreduce:AddJobFlowSteps",
                "elasticmapreduce:AddTags",
                "elasticmapreduce:CancelSteps",
                "elasticmapreduce:CreateEditor",
                "elasticmapreduce:CreateSecurityConfiguration",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:DeleteSecurityConfiguration",
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:DescribeJobFlows",
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:DescribeReleaseLabel",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:GetManagedScalingPolicy",
                "elasticmapreduce:GetAutoTerminationPolicy",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListSupportedInstanceTypes",
                "elasticmapreduce:ModifyCluster",
                "elasticmapreduce:ModifyInstanceFleet",
                "elasticmapreduce:ModifyInstanceGroups",
                "elasticmapreduce:OpenEditorInConsole",
                "elasticmapreduce:PutAutoScalingPolicy",
                "elasticmapreduce:PutBlockPublicAccessConfiguration",
                "elasticmapreduce:PutManagedScalingPolicy",
                "elasticmapreduce:RemoveAutoScalingPolicy",
                "elasticmapreduce:RemoveManagedScalingPolicy",
                "elasticmapreduce:RemoveTags",
                "elasticmapreduce:SetTerminationProtection",
```

```
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:TerminateJobFlows",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ViewMetricsInEMRConsole",
            "Effect": "Allow",
            "Action": [
                "cloudwatch:GetMetricStatistics"
            ],
            "Resource": "*"
        },
        {
            "Sid": "PassRoleForElasticMapReduce",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": [
                "arn:aws:iam::*:role/EMR_DefaultRole",
                "arn:aws:iam::*:role/EMR_DefaultRole_V2"
            ],
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "PassRoleForEC2",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::*:role/EMR_EC2_DefaultRole",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "ec2.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "PassRoleForAutoScaling",
            "Effect": "Allow",
            "Action": "iam:PassRole",
```

```
            "Resource": "arn:aws:iam::*:role/EMR_AutoScaling_DefaultRole",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "application-autoscaling.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "ElasticMapReduceServiceLinkedRole",
            "Effect": "Allow",
            "Action": "iam:CreateServiceLinkedRole",
            "Resource": "arn:aws:iam::*:role/aws-service-role/
elasticmapreduce.amazonaws.com*/AWSServiceRoleForEMRCleanup*",
            "Condition": {
                "StringEquals": {
                    "iam:AWSServiceName": [
                        "elasticmapreduce.amazonaws.com",
                        "elasticmapreduce.amazonaws.com.cn"
                    ]
                }
            }
        },
        {
            "Sid": "ConsoleUIActions",
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeAccountAttributes",
                "ec2:DescribeAvailabilityZones",
                "ec2:DescribeImages",
                "ec2:DescribeKeyPairs",
                "ec2:DescribeNatGateways",
                "ec2:DescribeRouteTables",
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "ec2:DescribeVpcEndpoints",
                "s3:ListAllMyBuckets",
                "iam:ListRoles"
            ],
            "Resource": "*"
        }
    ]
}
```

Política administrada por IAM para obtener acceso completo (en vías de quedar obsoleta)

Las políticas administradas `AmazonElasticMapReduceFullAccess` y `AmazonEMRFullAccessPolicy_v2` de AWS Identity and Access Management (IAM) otorgan todas las acciones necesarias para Amazon EMR y otros servicios.

> **Important**
>
> La política administrada `AmazonElasticMapReduceFullAccess` está en vías de quedar obsoleta y ya no se recomienda su uso con Amazon EMR. En su lugar, utilice `AmazonEMRFullAccessPolicy_v2`. Cuando el servicio de IAM haga que la política de la versión 1 quede obsoleta, ya no podrá asociarla a ningún rol. Sin embargo, puedes adjuntar un rol existente a un clúster incluso si ese rol usa la política obsoleta.

Las políticas administradas de forma predeterminada con todos los permisos de Amazon EMR incorporan configuraciones de seguridad `iam:PassRole`, entre las que se incluyen las siguientes:

- Permisos `iam:PassRole` solo para roles específicos de Amazon EMR predeterminados.
- Condiciones `iam:PassedToService` que le permiten usar la política solo con los productos de AWS especificados, como `elasticmapreduce.amazonaws.com` y `ec2.amazonaws.com`.

Puede ver la versión JSON de las políticas AmazonEMRFullAccessPolicy_v2 y AmazonEMRServicePolicy_v2 en la consola de IAM. Le recomendamos crear nuevos clústeres con las políticas administradas de la versión 2.

Puede ver el contenido de la política de la versión 1 obsoleta en la AWS Management Console en `AmazonElasticMapReduceFullAccess`. La acción `ec2:TerminateInstances` de la política otorga permiso a un usuario o rol para cancelar cualquiera de las instancias de Amazon EC2 asociadas a la cuenta de IAM. Esto incluye instancias que no forman parte de un clúster de EMR.

Política administrada por IAM para el acceso de solo lectura (política predeterminada administrada de la versión 2)

Para conceder privilegios de solo lectura a Amazon EMR, adjunte la política administrada AmazonEMRReadOnlyAccessPolicy_v2. Esta política administrada predeterminada reemplaza a la política administrada `AmazonElasticMapReduceReadOnlyAccess`.
El contenido de esta instrucción de la política se muestra en el siguiente fragmento: En comparación con la política `AmazonElasticMapReduceReadOnlyAccess`, la política

`AmazonEMRReadOnlyAccessPolicy_v2` no utiliza caracteres comodín para el elemento `elasticmapreduce`. En su lugar, la política de la versión 2 predeterminada determina el ámbito de aplicación de las acciones `elasticmapreduce` permitidas.

> **Note**
>
> También puede usar el enlace de la AWS Management Console a `AmazonEMRReadOnlyAccessPolicy_v2` para ver la política.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ElasticMapReduceActions",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:DescribeCluster",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:DescribeJobFlows",
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:DescribeReleaseLabel",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:GetManagedScalingPolicy",
                "elasticmapreduce:GetAutoTerminationPolicy",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListSupportedInstanceTypes",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole"
            ],
            "Resource": "*"
        },
        {
            "Sid": "ViewMetricsInEMRConsole",
            "Effect": "Allow",
```

```
            "Action": [
                "cloudwatch:GetMetricStatistics"
            ],
            "Resource": "*"
        }
    ]
}
```

Política administrada por IAM para obtener acceso de solo lecturas (en vías de quedar obsoleta)

La política administrada `AmazonElasticMapReduceReadOnlyAccess` está en vías de quedar obsoleta. No puede adjuntar esta política al lanzar nuevos clústeres. `AmazonElasticMapReduceReadOnlyAccess` se ha sustituido por `AmazonEMRReadOnlyAccessPolicy_v2`, que ahora es la política administrada predeterminada de Amazon EMR. El contenido de esta instrucción de la política se muestra en el siguiente fragmento: Los caracteres comodín para el elemento `elasticmapreduce` especifican que solo se permiten las acciones que empiezan por las cadenas especificadas. Tenga en cuenta que, dado que esta política no deniega acciones explícitamente, se puede seguir utilizando una instrucción de política distinta para otorgar acceso a acciones especificadas.

> **Note**
>
> También puede usar la AWS Management Console para ver la política.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:Describe*",
        "elasticmapreduce:List*",
        "elasticmapreduce:ViewEventsFromAllClustersInConsole",
        "s3:GetObject",
        "s3:ListAllMyBuckets",
        "s3:ListBucket",
        "sdb:Select",
        "cloudwatch:GetMetricStatistics"
      ],
      "Resource": "*"
```

```
    }
  ]
}
```

Políticas administradas por AWS para Amazon EMR

Para agregar permisos a usuarios, grupos y roles, es más fácil utilizar las políticas administradas de AWS que escribirlas uno mismo. Se necesita tiempo y experiencia para crear políticas de IAM administradas por el cliente que proporcionen a su equipo solo los permisos necesarios. Para comenzar a hacerlo con rapidez, puede utilizar nuestras políticas administradas por AWS. Estas políticas cubren casos de uso comunes y están disponibles en su cuenta de AWS. Para obtener más información sobre las políticas administradas por AWS, consulte Políticas administradas por AWS en la Guía del usuario de IAM.

Los servicios de AWS mantienen y actualizan las políticas administradas por AWS. No puede cambiar los permisos en las políticas administradas por AWS. En ocasiones, los servicios agregan permisos adicionales a una política administrada por AWS para admitir características nuevas. Este tipo de actualización afecta a todas las identidades (usuarios, grupos y roles) donde se asocia la política. Es más probable que los servicios actualicen una política administrada por AWS cuando se lanza una nueva característica o cuando se ponen a disposición nuevas operaciones. Los servicios no quitan permisos de una política administrada por AWS, por lo que las actualizaciones de políticas no deteriorarán los permisos existentes.

Además, AWS admite políticas administradas para funciones de trabajo que abarcan varios servicios. Por ejemplo, la política administrada por ReadOnlyAccessAWS proporciona acceso de solo lectura a todos los servicios y los recursos de AWS. Cuando un servicio lanza una nueva característica, AWS agrega permisos de solo lectura para las operaciones y los recursos nuevos. Para obtener una lista y descripciones de las políticas de funciones de trabajo, consulte Políticas administradas de AWS para funciones de trabajo en la Guía del usuario de IAM.

Actualizaciones de Amazon EMR a políticas administradas por AWS

Es posible consultar los detalles sobre las actualizaciones de las políticas administradas por AWS para Amazon EMR desde que este servicio comenzó a hacer un seguimiento de estos cambios. Para

obtener alertas automáticas sobre cambios en esta página, suscríbase a la fuente RSS en la página de historial de documentos de Amazon EMR.

Cambio	Descripción	Fecha
AmazonEMRFullAccessPolicy_v2 y **AmazonEMRReadOnlyAccessPolicy_v2**: actualización de una política existente	`elasticmapreduce:ListSupportedInstanceTypes` añadido.	13 de julio de 2023
AmazonEMRFullAccessPolicy_v2 y **AmazonEMRReadOnlyAccessPolicy_v2**: actualización de una política existente	Se añadieron `elasticmapreduce:DescribeReleaseLabel` y `elasticmapreduce:GetAutoTerminationPolicy` .	21 de abril de 2022
AmazonEMRFullAccessPolicy_v2: actualización de una política actual	Se agregó `ec2:DescribeImages` para Uso de una AMI personalizada.	15 de febrero de 2022
Políticas administradas por Amazon EMR	Se actualizó para aclarar el uso de etiquetas de usuario predefinidas. Se agregó una sección sobre el uso de la consola de AWS para lanzar clústeres con políticas administradas en la versión 2.	29 de septiembre de 2021
AmazonEMRFullAccessPolicy_v2: actualización de una política actual	Se modificaron las acciones `PassRoleForEC2` y `PassRoleForAutoScaling` para que usen el operador de condición `StringLike` y que así coincidan con `"iam:Pass`	20 de mayo de 2021

Cambio	Descripción	Fecha
	`edToService":"appl ication-autoscalin g.amazonaws.com*"` y `"iam:PassedToServi ce":"ec2.amazonaws .com*"` , respectivamente.	
`AmazonEMRFullAcces sPolicy_v2`: actualización de una política actual	Se ha eliminado la acción no válida `s3:ListBuckets` y se ha sustituido por la acción `s3:ListAllMyBuckets` . Se actualizó la creación de roles vinculados a un servicio (SLR) para que se limite explícitamente al único rol que tiene Amazon EMR con principios de servicio explícito s. Los SLR que se pueden crear son exactamente los mismos que antes de este cambio.	23 de marzo de 2021

Cambio	Descripción	Fecha
AmazonEMRFullAccessPolicy_v2: política nueva	Amazon EMR ha agregado nuevos permisos para limitar el acceso a los recursos y para agregar un requisito previo según el cual los usuarios deben agregar una etiqueta de usuario predefinida a los recursos antes de poder utilizar las políticas administradas de Amazon EMR. La acción `iam:PassRole` requiere que se establezca una condición `iam:PassedToService` en un servicio específico. El acceso a Amazon EC2, Amazon S3 y otros servicios no está permitido de forma predeterminada.	11 de marzo de 2021
AmazonEMRServicePolicy_v2: política nueva	Agrega un requisito previo según el cual los usuarios deben agregar etiquetas de usuario a los recursos antes de poder utilizar esta política.	11 de marzo de 2021
AmazonEMRReadOnlyAccessPolicy_v2: política nueva	Los permisos permiten únicamente acciones específicas de solo lectura de elasticmapreduce. El acceso a Amazon S3 no está permitido de forma predeterminada.	11 de marzo de 2021

Cambio	Descripción	Fecha
Amazon EMR ha comenzado a hacer un seguimiento de los cambios	Amazon EMR comenzó a hacer un seguimiento de los cambios en sus políticas administradas por AWS.	11 de marzo de 2021

Políticas de IAM para el acceso basado en etiquetas para clústeres y cuadernos de EMR

Puede utilizar condiciones en su política basada en identidades para controlar el acceso a clústeres y blocs de notas de EMR basándose en etiquetas.

Para obtener más información sobre la adición de etiquetas a los clústeres de EMR, consulte Etiquetado de clústeres de Amazon EMR.

Los siguientes ejemplos muestran distintos supuestos y formas de utilizar los operadores de condición con las claves de condición de Amazon EMR. Estas instrucciones de política de IAM tienen fines demostrativos y no deben utilizarse en entornos de producción. Existen varias maneras de combinar las instrucciones de políticas para conceder y denegar permisos de acuerdo con sus requisitos. Para obtener más información sobre la planificación y las pruebas de políticas de IAM, consulte la Guía del usuario de IAM.

> **Important**
>
> La denegación de permisos explícita para acciones de etiquetado de acciones es un factor importante. Esto impide que los usuarios etiqueten un recurso y, de esta forma, se concedan a sí mismos permisos que usted no tenía previsto conceder. Si no deniega las acciones de etiquetado de un recurso, el usuario puede modificar las etiquetas y eludir la intención de las políticas basadas en etiquetas.

Ejemplo de instrucciones de políticas basadas en identidades para clústeres

Los ejemplos que se muestran a continuación muestran las políticas de permisos basadas en identidades que se utilizan para controlar las acciones que se permiten con clústeres de EMR.

> **Important**
>
> La acción `ModifyInstanceGroup` en Amazon EMR no requiere que especifique un ID de clúster. Por ese motivo, denegar esta acción en función de las etiquetas de clúster requiere una consideración adicional. Para obtener más información, consulte Denegación de la acción ModifyInstanceGroup.

Temas

- Permitir acciones solo en clústeres con valores de etiqueta específicos
- Etiquetado obligatorio de un clúster al crear un clúster
- Permitir acciones en clústeres con una etiqueta específica, con independencia del valor de su etiqueta

Permitir acciones solo en clústeres con valores de etiqueta específicos

Los ejemplos que aparecen a continuación muestran una política que permite a un usuario realizar acciones en función de la etiqueta de clúster *`department`* con el valor *`dev`* y también permite a un usuario etiquetar clústeres con la misma etiqueta. El ejemplo de política muestra cómo denegar los privilegios para etiquetar clústeres de EMR con cualquier otra cosa excepto la misma etiqueta.

En el siguiente ejemplo de política, la condición `StringEquals` intenta hacer coincidir *`dev`* con el valor de la etiqueta *`department`*. Si la etiqueta *`department`* no se ha añadido al clúster o no contiene el valor *`dev`*, la política no se aplica y esta política no permite las acciones. Si no hay otras instrucciones de política que permitan las acciones, el usuario solo puede trabajar con clústeres que tenga esta etiqueta con este valor.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt12345678901234",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:DescribeCluster",
        "elasticmapreduce:ListSteps",
        "elasticmapreduce:TerminateJobFlows",
        "elasticmapreduce:SetTerminationProtection",
        "elasticmapreduce:ListInstances",
```

```
        "elasticmapreduce:ListInstanceGroups",
        "elasticmapreduce:ListBootstrapActions",
        "elasticmapreduce:DescribeStep"
      ],
      "Resource": [
        "*"
      ],
      "Condition": {
        "StringEquals": {
          "elasticmapreduce:ResourceTag/department": "dev"
        }
      }
    }
  ]
}
```

También puede especificar varios valores de etiqueta utilizando un operador de condición. Por ejemplo, a fin de permitir que todas las acciones en clústeres donde la etiqueta *department* contiene el valor *dev* o *test*, podría sustituir el bloque de condición en el ejemplo anterior por los siguientes.

```
            "Condition": {
              "StringEquals": {
                "elasticmapreduce:ResourceTag/department":["dev", "test"]
              }
            }
```

Etiquetado obligatorio de un clúster al crear un clúster

Como en el ejemplo anterior, la siguiente política de ejemplo busca la misma etiqueta coincidente: el valor *dev* para la etiqueta *department*. Sin embargo, en este ejemplo, la clave de condición `RequestTag` especifica que la política se aplica durante la creación de la etiqueta. Por lo tanto, debe crear un clúster con una etiqueta que coincida con el valor especificado.

Para crear un clúster con una etiqueta, también debe tener permiso para realizar la acción `elasticmapredue:AddTags`. En esta instrucción, la clave de condición `elasticmapreduce:ResourceTag` garantiza que IAM solo conceda acceso a los recursos de etiquetas con el valor *dev* en la etiqueta *department*. El elemento `Resource` se utiliza para limitar este permiso a los recursos del clúster.

En el caso de los recursos de `PassRole`, debe proporcionar el ID o alias de la cuenta de AWS, el nombre del rol de servicio en la instrucción `PassRoleForEMR` y el nombre del perfil de instancia en la instrucción `PassRoleForEC2`. Para obtener más información sobre el formato de los ARN en IAM, consulte ARN de IAM en la Guía del usuario de IAM.

Para obtener más información sobre cómo hacer coincidir los valores de las claves de etiqueta, consulte `aws:RequestTag/tag-key` en la Guía del usuario de IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Sid": "RunJobFlowExplicitlyWithTag",
        "Effect": "Allow",
        "Action": [
          "elasticmapreduce:RunJobFlow"
        ],
        "Resource": "*",
        "Condition": {
          "StringEquals": {
            "aws:RequestTag/department": "dev"
          }
        }
      },
      {
        "Sid": "AddTagsForDevClusters",
        "Effect": "Allow",
        "Action": "elasticmapreduce:AddTags",
        "Resource": "arn:aws:elasticmapreduce:*:*:cluster/*",
        "Condition": {
          "StringEquals": {
            "elasticmapreduce:ResourceTag/department": "dev"
          }
        }
      },
      {
        "Sid": "PassRoleForEMR",
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
        "Condition": {
          "StringLike": {
            "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
```

```
          }
        }
      },
      {
        "Sid": "PassRoleForEC2",
        "Effect": "Allow",
        "Action": "iam:PassRole",
        "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
        "Condition": {
          "StringLike": {
            "iam:PassedToService": "ec2.amazonaws.com*"
          }
        }
      }
    ]
}
```

Permitir acciones en clústeres con una etiqueta específica, con independencia del valor de su etiqueta

También puede permitir acciones solo en clústeres que tengan una etiqueta particular, con independencia del valor de la etiqueta. Para ello, puede utilizar el operador `Null`. Para obtener más información, consulte Operador de condición para comprobar la existencia de claves de condición en la Guía del usuario de IAM. Por ejemplo, para permitir acciones solo en clústeres de EMR que tengan la etiqueta *`department`*, con independencia del valor que contenga, podría sustituir los bloques Condition del ejemplo anterior por el siguiente. El operador `Null` buscará la presencia de la etiqueta *`department`* en un clúster de EMR. Si la etiqueta existe, la instrucción `Null` se evalúa como falsa, ajustándose a la condición especificada en esta instrucción de política y se permiten las acciones adecuadas.

```
"Condition": {
  "Null": {
    "elasticmapreduce:ResourceTag/department":"false"
  }
}
```

La siguiente instrucción de política permite a un usuario crear un clúster de EMR solo si el clúster tendrá una etiqueta *`department`*, que puede contener cualquier valor. En el caso del recurso `PassRole`, debe proporcionar el ID o alias de la cuenta de AWS y el nombre del rol de servicio. Para

obtener más información sobre el formato de los ARN en IAM, consulte ARN de IAM en la Guía del usuario de IAM.

Para obtener más información sobre cómo especificar el operador de condición null (“false”), consulte Operador de condición para comprobar la existencia de claves de condición en la Guía del usuario de IAM.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "CreateClusterTagNullCondition",
            "Effect": "Allow",
            "Action": [
                "elasticmapreduce:RunJobFlow"
            ],
            "Resource": [
                "*"
            ],
            "Condition": {
                "Null": {
                    "aws:RequestTag/department": "false"
                }
            }
        },
        {
            "Sid": "AddTagsNullCondition",
            "Effect": "Allow",
            "Action": "elasticmapreduce:AddTags",
            "Resource": "arn:aws:elasticmapreduce:*:*:cluster/*",
            "Condition": {
                "Null": {
                    "elasticmapreduce:ResourceTag/department": "false"
                }
            }
        },
        {
            "Sid": "PassRoleForElasticMapReduce",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
            "Condition": {
                "StringLike": {
```

```
                    "iam:PassedToService": "elasticmapreduce.amazonaws.com*"
                }
            }
        },
        {
            "Sid": "PassRoleForEC2",
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "arn:aws:iam::AccountId:role/Role-Name-With-Path",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "ec2.amazonaws.com*"
                }
            }
        }
    ]
}
```

Ejemplo de instrucciones de políticas basadas en identidades para Cuadernos de Amazon EMR

Las instrucciones de políticas de IAM de ejemplo de esta sección muestran escenarios comunes del uso de claves para limitar las acciones permitidas mediante Cuadernos de Amazon EMR. Siempre que no haya ninguna otra política asociada a la entidad principal (usuario) que permita las acciones, las claves de contexto de condición limitan las acciones permitidas tal como se indica.

Example : permitir el acceso únicamente a Cuadernos de Amazon EMR que un usuario cree en función del etiquetado

La instrucción de política de ejemplo mostrada a continuación, cuando se asocia a un rol o a un usuario, permite al usuario trabajar solamente con los cuadernos que haya creado. Esta instrucción de política utiliza la etiqueta predeterminada que se aplica al crear un bloc de notas.

En el ejemplo, el operador de condición `StringEquals` intenta emparejar una variable que representa el ID de usuario del usuario actual (`{aws:userId}`) con el valor de la etiqueta `creatorUserID`. Si la etiqueta `creatorUserID` no se ha añadido al bloc de notas o no contiene el valor del ID del usuario actual, la política no se aplica y esta política no permite las acciones. Si no hay otras instrucciones de política que permitan las acciones, el usuario solo puede trabajar con blocs de notas que tengan este valor para esta etiqueta.

```
{
    "Version": "2012-10-17",
    "Statement": [
```

```
        {
            "Action": [
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:StartEditor",
                "elasticmapreduce:StopEditor",
                "elasticmapreduce:DeleteEditor",
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userId}"
                }
            }
        }
    ]
}
```

Example — Etiquetado de cuadernos obligatorio al crearlos

En este ejemplo, se utiliza la clave de contexto `RequestTag`. La acción `CreateEditor` solo se permite si el usuario no cambia ni elimina la etiqueta `creatorUserID` que se añade de forma predeterminada. La variable ${aws:userId}, especifica el ID de usuario del usuario activo actualmente, que es el valor predeterminado de la etiqueta.

La instrucción de política se puede utilizar para ayudar a garantizar que los usuarios no eliminan la etiqueta `createUserId` ni cambian su valor.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/creatorUserId": "${aws:userid}"
                }
            }
```

```
        }
    ]
}
```

Este ejemplo requiere que el usuario cree el clúster con una etiqueta que tenga la cadena de clave `dept` establecida en uno de los siguientes valores: `datascience`, `analytics`, `operations`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/dept": [
                        "datascience",
                        "analytics",
                        "operations"
                    ]
                }
            }
        }
    ]
}
```

Example — Limitación de la creación de cuadernos a clústeres etiquetados y etiquetas de cuadernos obligatorias

Este ejemplo permite la creación de blocs de notas solo si el bloc de notas se crea con una etiqueta que tiene la cadena de clave `owner` establecida en uno de los valores especificados. Además, el bloc de notas solo se puede crear si el clúster tiene una etiqueta con la cadena de clave `department` establecida en uno de los valores especificados.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
```

```
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/owner": [
                        "owner1",
                        "owner2",
                        "owner3"
                    ],
                    "elasticmapreduce:ResourceTag/department": [
                        "dep1",
                        "dep3"
                    ]
                }
            }
        }
    ]
}
```

Example — Limitación de la capacidad de iniciar un cuaderno en función de las etiquetas

Este ejemplo limita la capacidad de iniciar blocs de notas únicamente a aquellos blocs de notas que tienen una etiqueta con la cadena de clave `owner` establecida en uno de los valores especificados. Debido a que el elemento `Resource` se utiliza solo para especificar el valor `editor`, la condición no se aplica al clúster y no es necesario etiquetarlo.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "owner1",
                        "owner2"
                    ]
```

```
                }
            }
        }
    ]
}
```

Este ejemplo es similar a uno anterior. Sin embargo, el límite solo se aplica a los clústeres etiquetados, no a los blocs de notas.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "dep1",
                        "dep3"
                    ]
                }
            }
        }
    ]
}
```

En este ejemplo, se utiliza un conjunto diferente de etiquetas de bloc de notas y de clúster. Permite el inicio de un bloc de notas solo si:

- El bloc de notas tiene una etiqueta con la cadena de clave `owner` establecida en cualquiera de los valores especificados

  —y—
- El clúster tiene una etiqueta con la cadena de clave `department` establecida en cualquiera de los valores especificados

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "user1",
                        "user2"
                    ]
                }
            }
        },
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "datascience",
                        "analytics"
                    ]
                }
            }
        }
    ]
}
```

Example — Limitación de la capacidad de abrir el editor de cuadernos en función de las etiquetas

En este ejemplo, se permite la apertura del editor de blocs de notas solo si:

- El bloc de notas tiene una etiqueta con la cadena de clave `owner` establecida en cualquiera de los valores especificados.

  —y—

- El clúster tiene una etiqueta con la cadena de clave `department` establecida en cualquiera de los valores especificados.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "user1",
                        "user2"
                    ]
                }
            }
        },
        {
            "Action": [
                "elasticmapreduce:OpenEditorInConsole"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "datascience",
                        "analytics"
                    ]
                }
            }
        }
    ]
}
```

Denegación de la acción ModifyInstanceGroup

La acción ModifyInstanceGroups de Amazon EMR no requiere que proporcione un ID de clúster con la acción. En su lugar, puede especificar solo un ID de grupo de instancias. Por este motivo, es posible que una política de denegación aparentemente simple para esta acción basada en el ID de clúster o en una etiqueta de clúster no tenga el efecto deseado. Considere la política de ejemplo siguiente.

```
{
 "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-east-1:123456789012:cluster/
j-12345ABCDEFG67"
        }
    ]

}
```

Si un usuario con esta política adjunta realiza una acción `ModifyInstanceGroup` y especifica solo el ID del grupo de instancias, la política no se aplica. Como la acción está permitida en todos los demás recursos, se realiza correctamente.

Una solución a este problema consiste en adjuntar una instrucción de política a la identidad que utilice un elemento NotResource para denegar cualquier acción `ModifyInstanceGroup` realizada sin un ID de clúster. El siguiente ejemplo de política agrega una declaración de denegación de este tipo para que cualquier solicitud `ModifyInstanceGroups` falle, a menos que se especifique un ID de clúster. Como una identidad debe especificar un ID de clúster con la acción, las instrucciones de denegación basadas en el ID de clúster son, por lo tanto, eficaces.

```
{
 "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-east-1:123456789012:cluster/
j-12345ABCDEFG67"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Deny",
            "NotResource": "arn:*:elasticmapreduce:*:*:cluster/*"
        }
    ]

}
```

Se produce un problema similar cuando se quiere denegar la acción `ModifyInstanceGroups` en función del valor asociado a una etiqueta de clúster. La solución es similar. Además de una sentencia de denegación que especifique el valor de la etiqueta, puede agregar una declaración de política que deniegue la acción `ModifyInstanceGroup` si la etiqueta que ha especificado no está presente, independientemente del valor.

El siguiente ejemplo muestra una política que, cuando se adjunta a una identidad, niega la identidad de la acción `ModifyInstanceGroups` a cualquier clúster con la etiqueta `department` establecida en `dev`. Esta declaración solo es efectiva debido a la declaración de denegación, que utiliza la condición `StringNotLike` para denegar la acción, a menos que la etiqueta `department` esté presente.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Condition": {
                "StringEquals": {
                    "aws:ResourceTag/department": "dev"
                }
            },
            "Effect": "Deny",
            "Resource": "*"
        },
        {
            "Action": [
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Condition": {
                "StringNotLike": {
                    "aws:ResourceTag/department": "?*"
                }
            },
            "Effect": "Deny",
            "Resource": "*"
        }
    ],
}
```

Solución de problemas de identidad y acceso de Amazon EMR

Utilice la siguiente información para diagnosticar y solucionar los problemas habituales que pueden surgir cuando se trabaja con Amazon EMR e IAM.

Temas

- No tengo autorización para realizar una acción en Amazon EMR
- No tengo autorización para realizar la operación iam:PassRole
- Quiero permitir que personas ajenas a mi cuenta de AWS accedan a mis recursos de Amazon EMR.

No tengo autorización para realizar una acción en Amazon EMR

Si la AWS Management Console le indica que no está autorizado para llevar a cabo una acción, debe ponerse en contacto con su administrador para recibir ayuda. Su administrador es la persona que le facilitó su nombre de usuario y contraseña.

En el siguiente ejemplo, el error se produce cuando el usuario `mateojackson` intenta utilizar la consola para consultar los detalles acerca de un recurso ficticio *`my-example-widget`*, pero no tiene los permisos ficticios `EMR:`*`GetWidget`*.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 EMR:GetWidget on resource: my-example-widget
```

En este caso, Mateo pide a su administrador que actualice sus políticas de forma que pueda obtener acceso al recurso *`my-example-widget`* mediante la acción `EMR:`*`GetWidget`*.

No tengo autorización para realizar la operación iam:PassRole

Si recibe un error que indica que no tiene autorización para llevar a cabo la acción `iam:PassRole`, las políticas se deben actualizar para permitirle pasar un rol a Amazon EMR.

Algunos servicios de Servicios de AWS le permiten transferir un rol existente a dicho servicio en lugar de crear un nuevo rol de servicio o uno vinculado al servicio. Para ello, debe tener permisos para transferir el rol al servicio.

En el siguiente ejemplo, el error se produce cuando un usuario de IAM denominado `marymajor` intenta utilizar la consola para realizar una acción en Amazon EMR. Sin embargo, la acción requiere que el servicio cuente con permisos que otorga un rol de servicio. Mary no tiene permisos para transferir el rol al servicio.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

En este caso, las políticas de Mary se deben actualizar para permitirle realizar la acción `iam:PassRole`.

Si necesita ayuda, póngase en contacto con su administrador de AWS. El administrador es la persona que le proporcionó las credenciales de inicio de sesión.

Quiero permitir que personas ajenas a mi cuenta de AWS accedan a mis recursos de Amazon EMR.

Puede crear un rol que los usuarios de otras cuentas o las personas externas a la organización puedan utilizar para acceder a sus recursos. Puede especificar una persona de confianza para que asuma el rol. En el caso de los servicios que admitan las políticas basadas en recursos o las listas de control de acceso (ACL), puede utilizar dichas políticas para conceder a las personas acceso a sus recursos.

Para obtener más información, consulte lo siguiente:

- Para saber si Amazon EMR admite estas características, consulte Cómo funciona Amazon EMR con IAM.
- Para obtener información acerca de cómo proporcionar acceso a los recursos de las Cuentas de AWS de su propiedad, consulte Proporcionar acceso a un usuario de IAM a otra Cuenta de AWS de la que es propietario en la Guía del usuario de IAM.
- Para obtener información acerca de cómo proporcionar acceso a los recursos a Cuentas de AWS de terceros, consulte Proporcionar acceso a Cuentas de AWS que son propiedad de terceros en la Guía del usuario de IAM.
- Para obtener información sobre cómo proporcionar acceso mediante una identidad federada, consulte Proporcionar acceso a usuarios autenticados externamente (identidad federada) en la Guía del usuario de IAM.
- Para obtener información sobre la diferencia entre los roles y las políticas basadas en recursos para el acceso entre cuentas, consulte Cómo los roles de IAM difieren de las políticas basadas en recursos en la Guía del usuario de IAM.

Autenticación en nodos de clúster de Amazon EMR

Lo clientes SSH pueden utilizar un par de claves de Amazon EC2 para autenticarse en instancias de clúster. De forma alternativa, con las versiones 5.10.0 y posteriores de Amazon EMR, puede configurar Kerberos para autenticar a los usuarios y las conexiones SSH con el nodo principal. Además, con las versiones 5.12.0 y posteriores de Amazon EMR, puede autenticarse con LDAP.

Temas

- Uso de un par de claves de EC2 para credenciales de SSH
- Uso de Kerberos para la autenticación con Amazon EMR
- Uso de servidores de Active Directory o LDAP para la autenticación con Amazon EMR

Uso de un par de claves de EC2 para credenciales de SSH

Los nodos de clúster de Amazon EMR se ejecutan en instancias de Amazon EC2. Puede conectarse a nodos del clúster de la misma forma que puede conectarse a instancias de Amazon EC2. Puede utilizar Amazon EC2 para crear un par de claves, o bien puede importar un par de claves. Al crear un clúster, puede especificar el par de claves de Amazon EC2 que se utilizará para las conexiones SSH en todas las instancias del clúster. También puede crear un clúster sin un par de claves. Esto se hace con clústeres transitorios que se inician, ejecutan pasos, y luego se terminan de forma automática.

El cliente SSH que se utiliza para conectar al clúster necesita utilizar el archivo de clave privada asociado a este par de claves. Se trata de un archivo.pem para clientes SSH que utilizan Linux, Unix y MacOS. Debe establecer permisos para que solo el propietario de la clave tenga permiso para acceder al archivo. Para los clientes SSH que utilizan Windows, es un archivo .ppk, que normalmente se crea a partir del archivo .pem.

- Para obtener más información sobre la creación de un par de claves de Amazon EC2, consulte Pares de claves de Amazon EC2 en la Guía del usuario de Amazon EC2 para instancias de Linux.
- Para obtener instrucciones sobre el uso de PuTTYgen para crear un archivo .ppk a partir de un archivo .pem, consulte Convierta su clave privada utilizando PuTTYgen en la Guía del usuario de Amazon EC2 para instancias de Linux.
- Para obtener más información acerca de cómo establecer los permisos para el archivo .pem y cómo conectarse a un nodo principal de un clúster de EMR con distintos métodos (incluido `ssh` desde Linux o macOS, PuTTY desde Windows, o la AWS CLI desde cualquier sistema operativo compatible), consulte Conectarse al nodo principal mediante SSH.

Uso de Kerberos para la autenticación con Amazon EMR

Las versiones 5.10.0 y posteriores de Amazon EMR son compatibles con Kerberos. Kerberos es un protocolo de autenticación de redes que utiliza la criptografía de clave secreta para proporcionar una

autenticación sólida, de forma que las contraseñas u otras credenciales no se envíen a través de la red en un formato no cifrado.

En Kerberos, los servicios y los usuarios que necesitan autenticarse se conocen como entidades principales. Las entidades principales existen dentro de un ámbito de Kerberos. En el ámbito, un servidor de Kerberos conocido como centro de distribución de claves (KDC) proporciona los medios para que se autentiquen las entidades principales. El KDC lo hace mediante la emisión de tickets para la autenticación. El KDC mantiene una base de datos de los principales dentro de su ámbito, sus contraseñas y otros datos administrativos sobre cada uno de los principales. Un KDC también puede aceptar credenciales de autenticación de entidades principales de otros ámbitos, lo que se conoce como una confianza entre ámbitos. Además, un clúster de EMR puede utilizar un KDC externo para autenticar principales.

Una forma común de establecer una relación de confianza entre ámbitos o de utilizar un KDC externo es autenticar a los usuarios desde un dominio de Active Directory. Esto permite a los usuarios obtener acceso a un clúster de EMR con su cuenta de dominio cuando utilizan SSH para conectarse a un clúster o trabajan con aplicaciones de macrodatos.

Si utiliza autenticación Kerberos, Amazon EMR configura Kerberos para las aplicaciones, componentes y subsistemas que instala en el clúster, de forma que se autentiquen entre sí.

> **Important**
>
> Amazon EMR no admite AWS Directory Service for Microsoft Active Directory en una relación de confianza entre ámbitos o como un KDC externo.

Antes de configurar Kerberos con Amazon EMR, le recomendamos que se familiarice con los conceptos de Kerberos, los servicios que se ejecutan en un KDC y las herramientas de administración de servicios de Kerberos. Para obtener más información, consulte la documentación de Kerberos del MIT, que publica el consorcio Kerberos.

Temas

- Aplicaciones compatibles
- Opciones de la arquitectura Kerberos
- Configuración de Kerberos en Amazon EMR
- Uso de SSH para conectarse a clústeres que utilizan Kerberos
- Tutorial: Configuración de un KDC dedicado del clúster

- Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory

Aplicaciones compatibles

Dentro de un clúster de EMR, las entidades principales de Kerberos son los servicios de aplicaciones de macrodatos y los subsistemas que se ejecutan en todos los nodos del clúster. Amazon EMR puede configurar las aplicaciones y componentes que se indican a continuación para utilizar Kerberos. Cada aplicación tiene una entidad principal de Kerberos asociada.

Amazon EMR no admite las relaciones de confianza entre ámbitos con AWS Directory Service for Microsoft Active Directory.

Amazon EMR solo configura las características de autenticación de Kerberos de código abierto para las aplicaciones y los componentes indicados a continuación. Cualquier otra aplicación instalada no utiliza Kerberos, lo que puede provocar la incapacidad para comunicarse con los componentes que utilizan Kerberos y generar errores en la aplicación. Las aplicaciones y componentes que no utilizan Kerberos no tienen habilitada la autenticación. Las aplicaciones y los componentes compatibles pueden variar según las distintas versiones de Amazon EMR.

La interfaz de usuario de Livy es la única interfaz de usuario web alojada en el clúster que está kerberizado.

- Hadoop MapReduce
- HBase
- HCatalog
- HDFS
- Hive
  - No active Hive con la autenticación LDAP. Esto puede provocar problemas al comunicarse con el YARN que utiliza Kerberos.
- Hue
  - Hue la autenticación de usuarios no se establece automáticamente y se pueden configurar mediante la configuración de la API.
  - El servidor de Hue utiliza Kerberos. El front-end (la UI) de Hue no está configurado para la autenticación. La autenticación LDAP se puede configurar para la UI de Hue.
- Livy

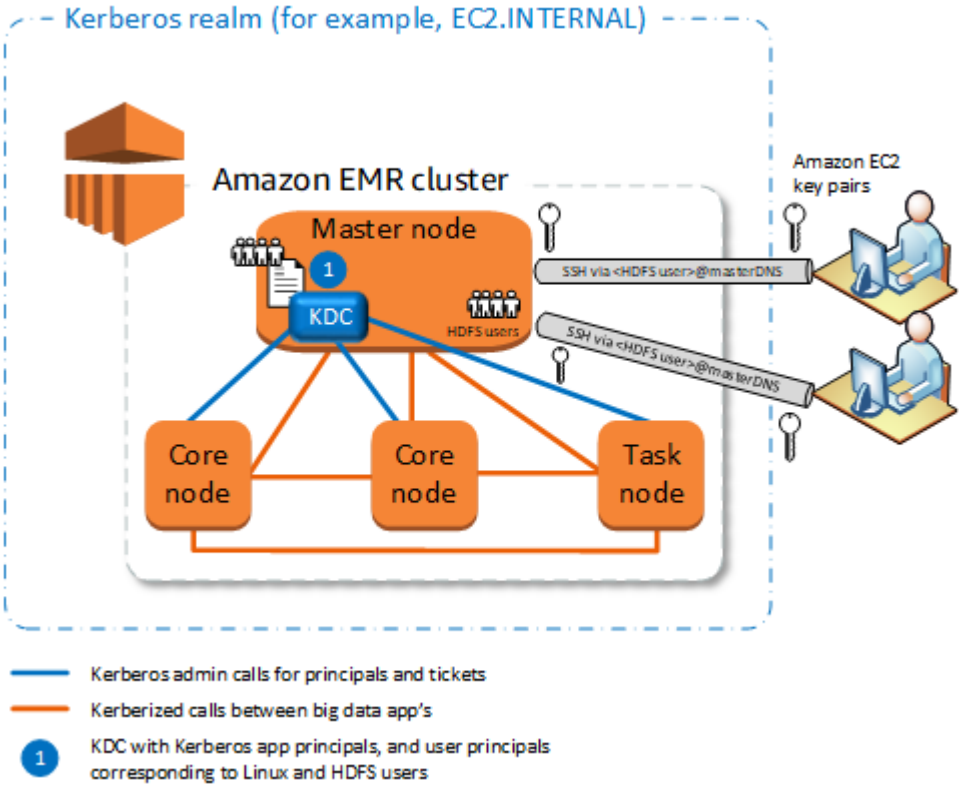
  - La suplantación de Livy con clústeres kerberizados es compatible con las versiones 5.22.0 y posteriores de Amazon EMR.
- Oozie
- Phoenix
- Presto
  - Presto admite la autenticación Kerberos en las versiones 6.9.0 y posteriores de Amazon EMR.
  - Si quiere usar la autenticación Kerberos para Presto, debe habilitar el cifrado en tránsito.
- Spark
- Tez
- Trino
  - Trino admite la autenticación Kerberos en las versiones 6.11.0 y posteriores de Amazon EMR.
  - Si quiere usar la autenticación Kerberos para Trino, debe habilitar el cifrado en tránsito.
- YARN
- Zeppelin
  - Zeppelin solo está configurado para utilizar Kerberos con el intérprete Spark. No está configurado para otros intérpretes.
  - Los intérpretes de Zeppelin kerberizado que no sean Spark no admiten la suplantación de identidad de usuario.
- ZooKeeper
  - Zookeeper cliente no es compatible.

Opciones de la arquitectura Kerberos

Cuando se utiliza Kerberos con Amazon EMR, puede elegir entre las arquitecturas que se indican en esta sección. Independientemente de la arquitectura que elija, para configurar Kerberos hay que seguir los mismos pasos. Debe crear una configuración de seguridad, especificar la configuración de seguridad y las opciones de Kerberos específicas del clúster compatibles al crear el clúster y crear directorios de HDFS para los usuarios de Linux en el clúster que coincidan con las entidades principales de usuarios en el KDC. Para leer una explicación sobre las opciones de configuración y configuraciones de ejemplo de cada arquitectura, consulte Configuración de Kerberos en Amazon EMR.

KDC dedicado del clúster (KDC en nodo principal)

Esta configuración está disponible con las versiones 5.10.0 y posteriores de Amazon EMR.



Ventajas

- Amazon EMR tiene la plena propiedad del KDC.
- El KDC en el clúster de EMR es independiente de las implementaciones de KDC centralizadas, como Microsoft Active Directory o AWS Managed Microsoft AD.
- El impacto en el rendimiento es mínimo, ya que el KDC administra la autenticación solo para los nodos locales del clúster.
- Opcionalmente, otros clústeres que utilizan Kerberos pueden hacer referencia al KDC como un KDC externo. Para obtener más información, consulte KDC externo: nodo principal en un clúster diferente.

Condiciones y limitaciones

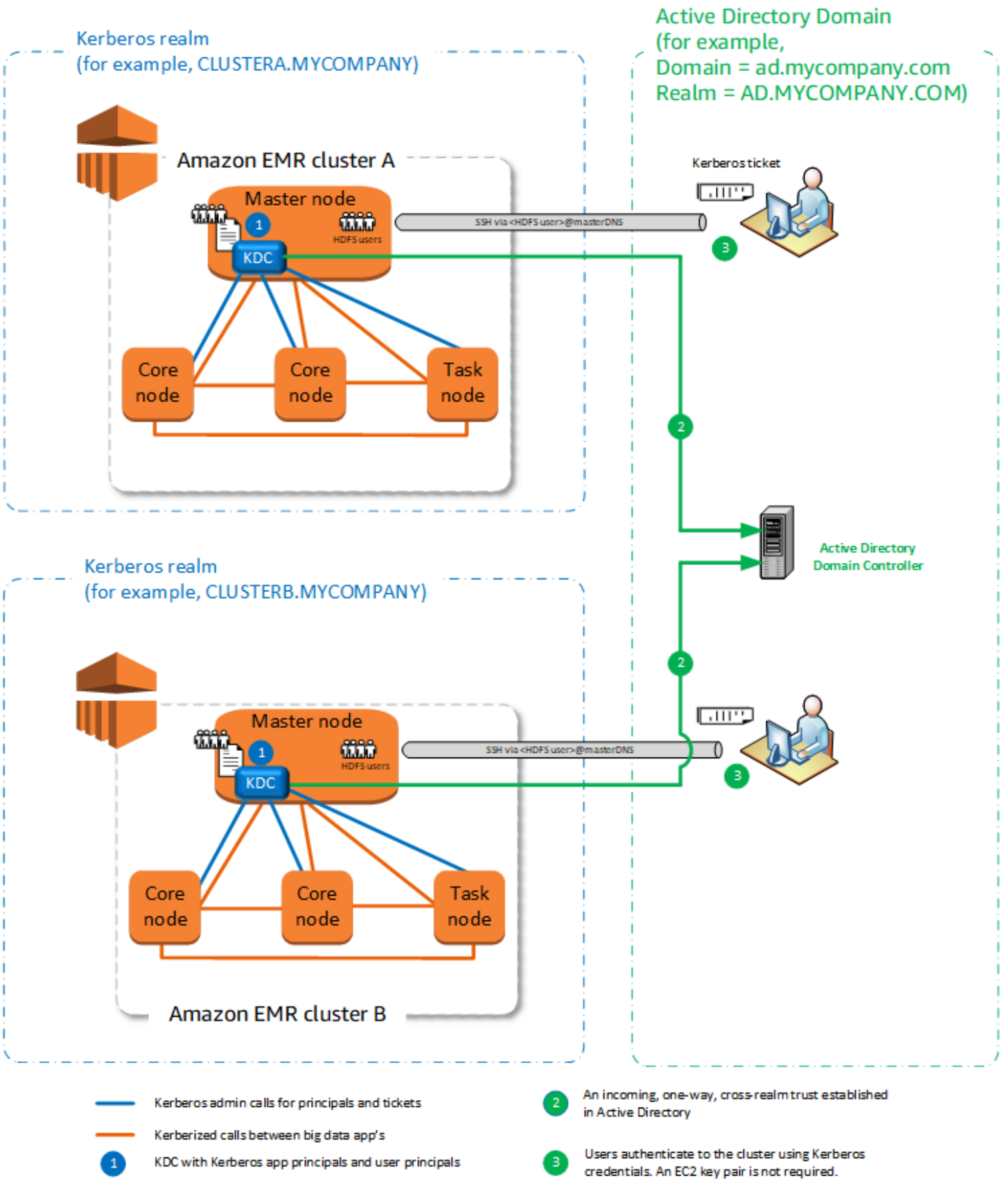
- Los clústeres que utilizan Kerberos no se pueden autenticarse entre sí, por lo que las aplicaciones no pueden interactuar. Si las aplicaciones de clústeres tienen que interactuar, debe establecer una relación de confianza entre ámbitos entre los clústeres o configurar un clúster como el KDC

externo para otros clústeres. Si se establece una relación de confianza entre ámbitos, los KDC deberán tener diferentes ámbitos de Kerberos.

- Debe crear usuarios de Linux en la instancia de EC2 del nodo principal que se correspondan con las entidades principales de usuarios de KDC, junto con los directorios de HDFS de cada usuario.
- Las entidades principales de usuario deben utilizar un archivo de clave privada de EC2 y credenciales de `kinit` para conectarse al clúster mediante SSH.

Relación de confianza entre ámbitos

En esta configuración, las entidades principales (normalmente usuarios) de otro ámbito de Kerberos se autentican para los componentes de la aplicación en un clúster de EMR que utiliza Kerberos y que tiene su propio KDC. El KDC del nodo principal establece una relación de confianza con otro KDC utilizando una entidad principal entre ámbitos que existe en los dos KDC. El nombre y la contraseña de la entidad principal deben coincidir exactamente en cada KDC. Las relaciones de confianza entre ámbitos son más comunes con las implementaciones de Active Directory, tal y como se muestra en el siguiente diagrama. También se admiten relaciones de confianza entre ámbitos con un KDC de MIT externo o un KDC en otro clúster de Amazon EMR.


Active Directory Domain
(for example,
Domain = ad.mycompany.com
Realm = AD.MYCOMPANY.COM)
Kerberos realm
(for example, CLUSTERA.MYCOMPANY)
Amazon EMR cluster A
Master node
1
KDC
HDFS users
SSH via <HDFS user>@masterDNS
Kerberos ticket
3
Core node
Core node
Task node
2
Active Directory Domain Controller
Kerberos realm
(for example, CLUSTERB.MYCOMPANY)
Master node
1
KDC
HDFS users
SSH via <HDFS user>@masterDNS
2
3
Core node
Core node
Task node
Amazon EMR cluster B
Kerberos admin calls for principals and tickets
Kerberized calls between big data app's
1
KDC with Kerberos app principals and user principals
2
An incoming, one-way, cross-realm trust established in Active Directory
3
Users authenticate to the cluster using Kerberos credentials. An EC2 key pair is not required.

Ventajas

- El clúster de EMR en el que está instalado el KDC mantiene la plena propiedad del KDC.
- Con Active Directory, Amazon EMR crea automáticamente usuarios de Linux que se corresponden con las entidades principales de usuario del KDC. Aun así, debe crear directorios de HDFS para cada usuario. Además, las entidades principales de usuario del dominio de Active Directory pueden obtener acceso a los clústeres que utilizan Kerberos mediante credenciales de `kinit`, sin el archivo de clave privada de EC2. Así se elimina la necesidad de compartir el archivo de clave privada entre usuarios de clústeres.
- Dado que cada KDC del clúster administra la autenticación de los nodos del clúster, se minimizan los efectos de la latencia de la red y la sobrecarga de procesamiento de un gran número de nodos entre clústeres.

Condiciones y limitaciones

- Si va a establecer una relación de confianza con un ámbito de Active Directory, debe proporcionar un nombre de usuario y contraseña de Active Directory con permisos para unir entidades principales al dominio al crear el clúster.
- Las relaciones de confianza entre ámbitos no se pueden establecer entre ámbitos de Kerberos con el mismo nombre.
- Las relaciones de confianza entre ámbitos deben establecerse de forma explícita. Por ejemplo, si el Clúster A y B establecen una relación de confianza entre ámbitos con un KDC, no confían intrínsecamente uno en el otro y sus aplicaciones no pueden autenticarse entre sí para interaccionar.
- Los KDC deben mantenerse de forma independiente y deben coordinarse para que las credenciales de las entidades principales de usuario coincidan exactamente.

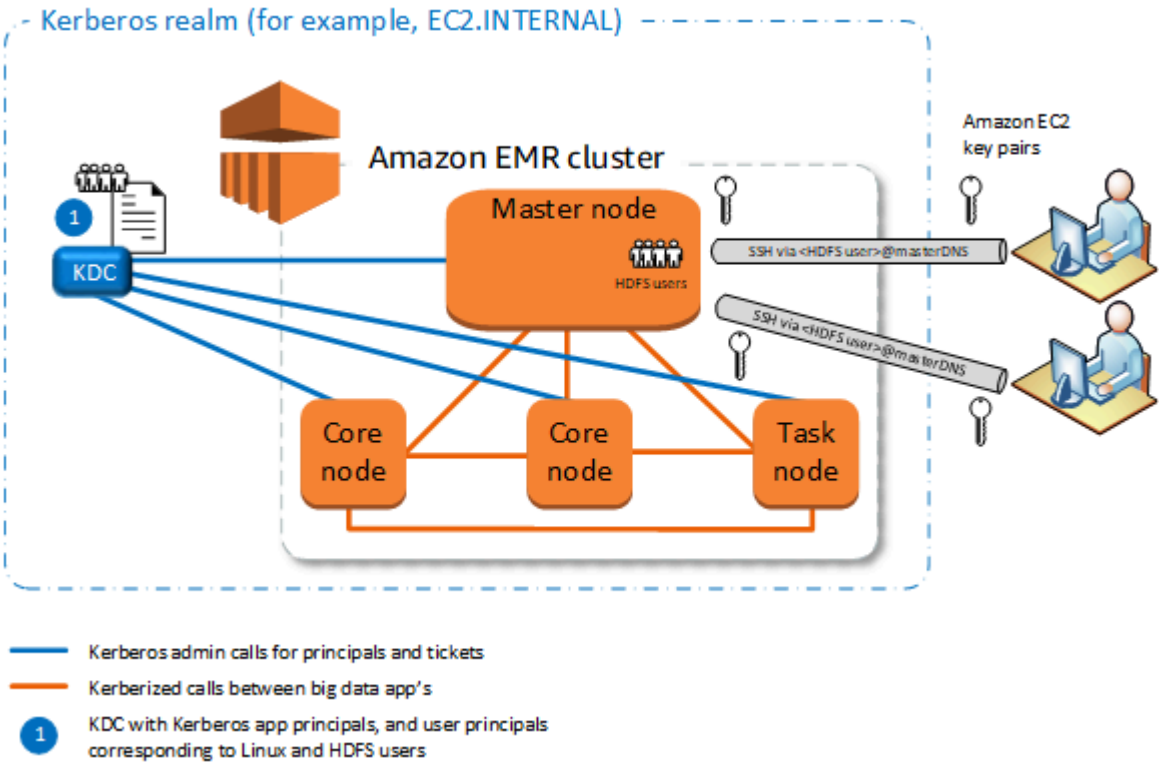
KDC externo

Las configuraciones con un KDC externo son compatibles con las versiones 5.20.0 y posteriores de Amazon EMR.

- KDC externo: KDC de MIT
- KDC externo: nodo principal en un clúster diferente
- KDC externo: KDC de clúster en un clúster diferente con una relación de confianza entre ámbitos de Active Directory

KDC externo: KDC de MIT

Esta configuración permite que uno o varios clústeres de EMR utilicen entidades principales definidas y mantenidos en un servidor de KDC de MIT.



Ventajas

- La administración de principales se consolida en un solo KDC.
- Varios clústeres pueden utilizar el mismo KDC en el mismo ámbito de Kerberos. Para obtener más información, consulte Requisitos para utilizar varios clústeres con el mismo KDC.
- El nodo principal en un clúster que utiliza Kerberos no tiene la carga de rendimiento asociada con el mantenimiento del KDC.

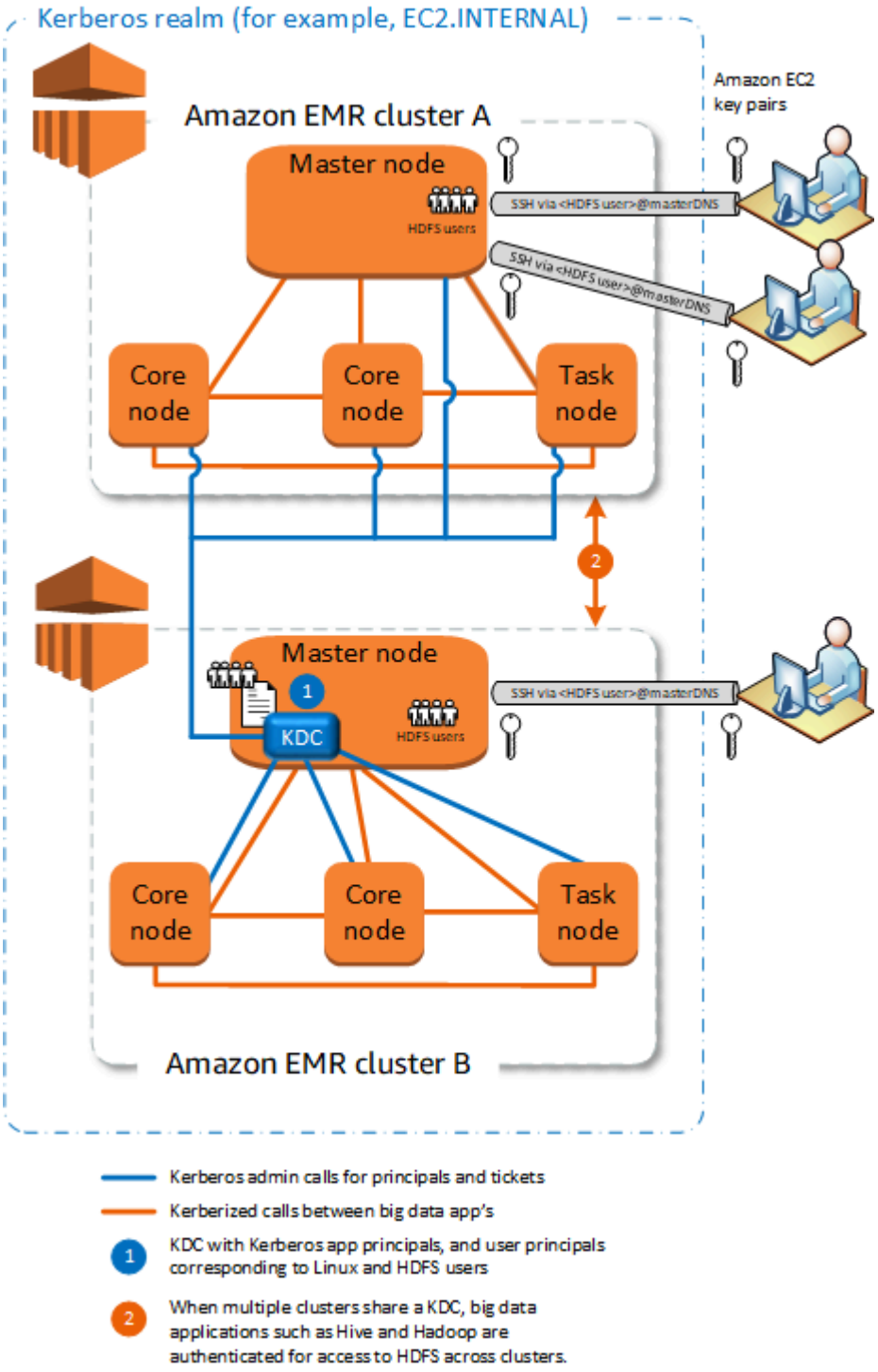
Condiciones y limitaciones

- Debe crear usuarios de Linux en la instancia EC2 del nodo principal del clúster que tiene Kerberos que se correspondan con las entidades principales de usuarios de KDC, junto con los directorios de HDFS de cada usuario.
- Las entidades principales de usuarios deben utilizar un archivo de clave privada de EC2 y credenciales de `kinit` para conectarse a los clústeres que utilizan Kerberos mediante SSH.

- Cada nodo de los clústeres de EMR que utilizan Kerberos debe tener una ruta de red al KDC.
- Cada nodo en los clústeres que utilizan Kerberos impone un carga de autenticación en el KDC externo, por lo que la configuración del KDC afecta al rendimiento del clúster. Cuando configure el hardware del servidor de KDC, tenga en cuenta el número máximo de nodos de Amazon EMR que se pueden admitir de forma simultánea.
- El rendimiento del clúster depende de la latencia de red entre los nodos en los clústeres que utilizan Kerberos y el KDC.
- La solución de problemas puede ser más difícil debido a las interdependencias.

KDC externo: nodo principal en un clúster diferente

Esta configuración es casi idéntica a la implementación de KDC de MIT externo anterior, salvo que el KDC está en el nodo principal de un clúster de EMR. Para obtener más información, consulte KDC dedicado del clúster (KDC en nodo principal) y Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory.



Ventajas

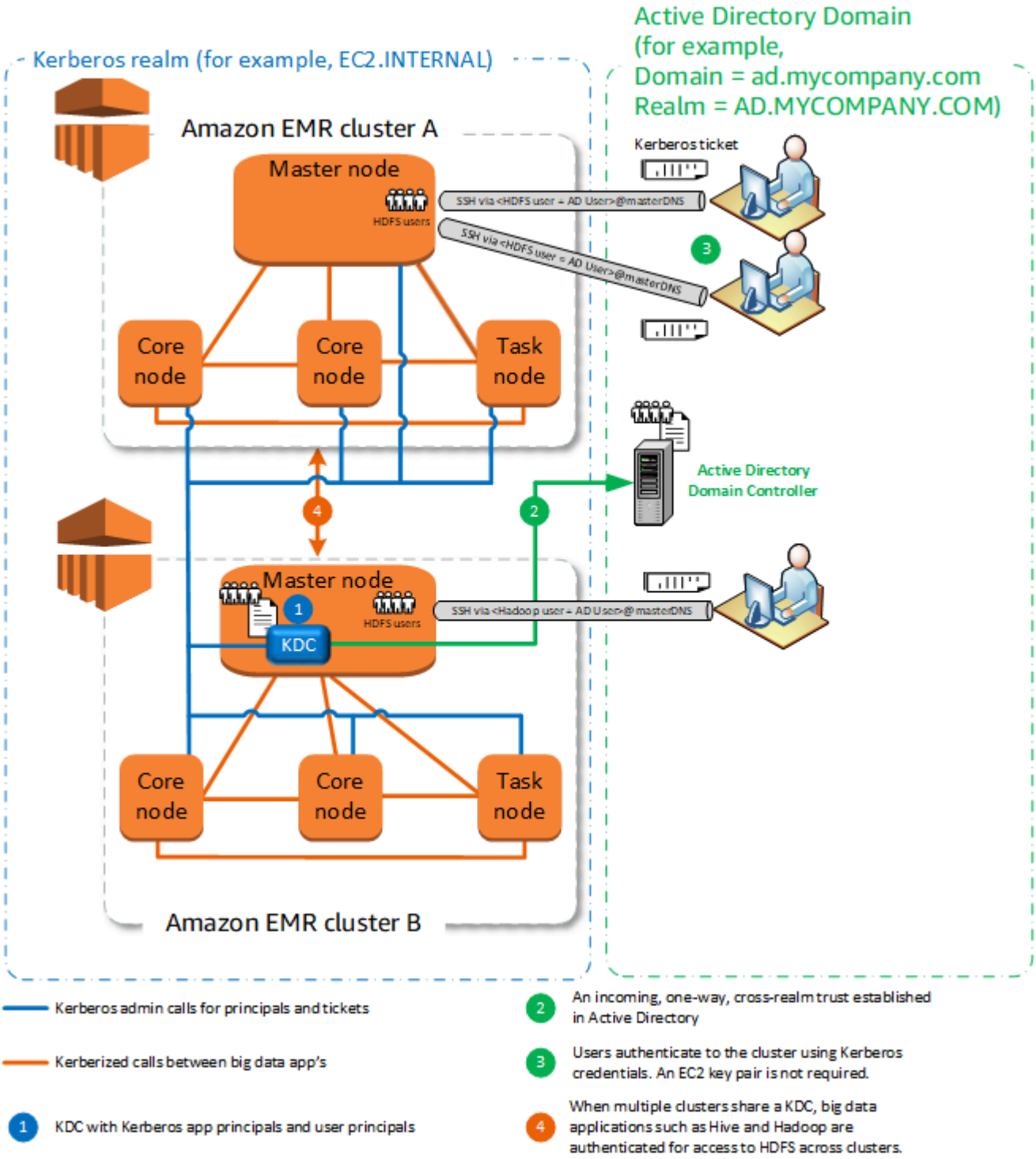
- La administración de principales se consolida en un solo KDC.
- Varios clústeres pueden utilizar el mismo KDC en el mismo ámbito de Kerberos. Para obtener más información, consulte Requisitos para utilizar varios clústeres con el mismo KDC.

Condiciones y limitaciones

- Debe crear usuarios de Linux en la instancia EC2 del nodo principal del clúster que tiene Kerberos que se correspondan con las entidades principales de usuarios de KDC, junto con los directorios de HDFS de cada usuario.
- Las entidades principales de usuarios deben utilizar un archivo de clave privada de EC2 y credenciales de `kinit` para conectarse a los clústeres que utilizan Kerberos mediante SSH.
- Cada nodo de cada clúster de EMR debe tener una ruta de red al KDC.
- Cada nodo de Amazon EMR de los clústeres que utilizan Kerberos impone un carga de autenticación en el KDC externo, por lo que la configuración del KDC afecta al rendimiento del clúster. Cuando configure el hardware del servidor de KDC, tenga en cuenta el número máximo de nodos de Amazon EMR que se pueden admitir de forma simultánea.
- El rendimiento del clúster depende de la latencia de red entre los nodos de clústeres y el KDC.
- La solución de problemas puede ser más difícil debido a las interdependencias.

KDC externo: KDC de clúster en un clúster diferente con una relación de confianza entre ámbitos de Active Directory

En esta configuración, primero se crea un clúster con un KDC dedicado del clúster que tiene una relación de confianza entre ámbitos unidireccional con Active Directory. Para ver un tutorial detallado, consulte Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory. A continuación, lance clústeres adicionales, haciendo referencia al KDC del clúster que tiene la relación de confianza con un KDC externo. Para ver un ejemplo, consulte KDC externo del clúster con una relación de confianza entre ámbitos de Active Directory. De este modo, cada clúster de Amazon EMR puede utilizar el KDC externo para autenticar entidades principales definidas y mantenidas en un dominio de Microsoft Active Directory.



Ventajas

- La administración de entidades principales se consolida en el dominio de Active Directory.

- Amazon EMR se incorpora al ámbito de Active Directory, lo que elimina la necesidad de crear usuarios de Linux que se correspondan con usuarios de Active Directory. Aun así, debe crear directorios de HDFS para cada usuario.
- Varios clústeres pueden utilizar el mismo KDC en el mismo ámbito de Kerberos. Para obtener más información, consulte Requisitos para utilizar varios clústeres con el mismo KDC.
- Las entidades principales de usuarios del dominio de Active Directory pueden acceder a los clústeres que utilizan Kerberos mediante credenciales de `kinit`, sin el archivo de clave privada de EC2. Así se elimina la necesidad de compartir el archivo de clave privada entre usuarios de clústeres.
- Solo un nodo principal de Amazon EMR tiene la carga del mantenimiento del KDC y solo ese clúster debe crearse con credenciales de Active Directory para la relación de confianza entre ámbitos entre el KDC y Active Directory.

Condiciones y limitaciones

- Cada nodo de cada clúster de EMR debe tener una ruta de red al KDC y el controlador de dominio de Active Directory.
- Cada nodo de Amazon EMR impone un carga de autenticación en el KDC externo, por lo que la configuración del KDC afecta al rendimiento del clúster. Cuando configure el hardware del servidor de KDC, tenga en cuenta el número máximo de nodos de Amazon EMR que se pueden admitir de forma simultánea.
- El rendimiento del clúster depende de la latencia de red entre los nodos de los clústeres y el servidor de KDC.
- La solución de problemas puede ser más difícil debido a las interdependencias.

Requisitos para utilizar varios clústeres con el mismo KDC

Varios clústeres pueden utilizar el mismo KDC en el mismo ámbito de Kerberos. Sin embargo, si los clústeres se ejecutan simultáneamente, es posible que se produzcan errores si utilizan nombres de ServicePrincipal de Kerberos que entren en conflicto.

Si tiene varios clústeres simultáneos con el mismo KDC externo, asegúrese de que los clústeres utilicen distintos ámbitos de Kerberos. Si los clústeres deben usar el mismo dominio de Kerberos, asegúrese de que estén en subredes diferentes y de que sus rangos de CIDR no se superpongan.

Configuración de Kerberos en Amazon EMR

En esta sección, se proporcionan detalles y ejemplos de la configuración de Kerberos con arquitecturas comunes. Independientemente de la arquitectura que elija, los aspectos básicos de la configuración son los mismos y se realizan en tres pasos. Si utiliza un KDC externo o configura una relación de confianza entre ámbitos, debe asegurarse de que todos los nodos de un clúster tengan una ruta de red al KDC externo, incluida la configuración de los grupos de seguridad aplicables para permitir el tráfico de entrada y salida de Kerberos.

Paso 1: Cree una configuración de seguridad con propiedades de Kerberos

La configuración de seguridad especifica detalles sobre el KDC de Kerberos y permite reutilizar la configuración de Kerberos cada vez que se crea un clúster. Puede crear una configuración de seguridad utilizando la consola de Amazon EMR, la AWS CLI o la API de EMR. La configuración de seguridad también pueden contener otras opciones de seguridad, como, por ejemplo, el cifrado. Para obtener más información acerca de la creación y especificación de las configuraciones de seguridad cuando se crea un clúster, consulte Uso de configuraciones de seguridad para configurar la seguridad del clúster. Para obtener más información acerca de las propiedades de Kerberos en una configuración de seguridad, consulte Configuración de Kerberos para las configuraciones de seguridad.

Paso 2: Cree un clúster y especifique los atributos de Kerberos específicos del clúster

Al crear un clúster, especifique la configuración de seguridad de Kerberos junto con las opciones de Kerberos específicas del clúster. Cuando utilice la consola de Amazon EMR, solo están disponibles las opciones de Kerberos compatibles con la configuración de seguridad especificada. Cuando utilice la AWS CLI o la API de Amazon EMR, asegúrese de especificar las opciones de Kerberos compatibles con la configuración de seguridad especificada. Por ejemplo, si especifica la contraseña de la entidad principal para una relación de confianza entre ámbitos al crear un clúster mediante la CLI y la configuración de seguridad especificada no tiene parámetros de relación de confianza entre ámbitos, se produce un error. Para obtener más información, consulte Configuración de Kerberos para clústeres.

Paso 3: configure el nodo principal del clúster

En función de los requisitos de su arquitectura e implementación, es posible que sea necesario realizar una configuración adicional en el clúster. Puede hacer esto después de crearla o mediante pasos o acciones de arranque durante el proceso de creación.

Para cada usuario autenticado de Kerberos que se conecte al clúster utilizando SSH, debe asegurarse de que se crean cuentas de Linux que se corresponden con el usuario de Kerberos. Si el controlador de dominio de Active Directory proporciona las entidades principales de usuario, ya sea como el KDC externo o a través de una relación de confianza entre ámbitos, Amazon EMR crea cuentas de usuario de Linux de forma automática. Si no se utiliza Active Directory, debe crear entidades principales para cada usuario que se correspondan con su usuario de Linux. Para obtener más información, consulte Configuración de un clúster para conexiones SSH y usuarios de HDFS autenticados en Kerberos.

Cada usuario también debe tener un directorio de usuarios de HDFS propio que debe crear usted. Además, SSH debe estar configurado con GSSAPI habilitado para permitir conexiones de los usuarios autenticados por Kerberos. GSSAPI debe estar habilitado en el nodo principal y se debe configurar la aplicación SSH del cliente para que utilice GSSAPI. Para obtener más información, consulte Configuración de un clúster para conexiones SSH y usuarios de HDFS autenticados en Kerberos.

Configuración de seguridad y configuración de clústeres para Kerberos en Amazon EMR

Al crear un clúster que utiliza Kerberos, debe especificar la configuración de seguridad junto con los atributos de Kerberos específicos para el clúster. No puede especificar un conjunto sin el otro, o se producirá un error.

En este tema, se ofrece información general de los parámetros de configuración disponibles para Kerberos al crear una configuración de seguridad y un clúster. Además, se proporcionan ejemplos de CLI para crear configuraciones de seguridad compatibles y clústeres para arquitecturas comunes.

Configuración de Kerberos para las configuraciones de seguridad

Puede crear una configuración de seguridad que especifica atributos de Kerberos utilizando la consola de Amazon EMR, la AWS CLI o la API de EMR. La configuración de seguridad también pueden contener otras opciones de seguridad, como, por ejemplo, el cifrado. Para obtener más información, consulte Creación de una configuración de seguridad.

Utilice las siguientes referencias para conocer las opciones de configuración de seguridad disponibles para la arquitectura Kerberos que elija. Se muestra la configuración de la consola de Amazon EMR. Para ver las opciones de la CLI correspondientes, consulte Especificación de la configuración de Kerberos con AWS CLI o Ejemplos de configuraciones.

Parámetro		Descripción
Kerberos		Especifica que Kerberos está habilitado para los clústeres que utilizan esta configuración de seguridad . Si un clúster usa esta configuración de seguridad , también debe tener la configuración de Kerberos especificada o se producirá un error.
Proveedor	KDC dedicado del clúster	Especifica que Amazon EMR crea un KDC en el nodo principal de cualquier clúster que utilice esta configura ción de seguridad. Al crear el clúster, debe especific ar el nombre del dominio y la contraseña de administr ador del KDC. Si es necesario, puede hacer referencia a este KDC desde otros clústeres. Cree esos clústeres con una configuración de seguridad diferente, especifique un KDC externo y utilice el nombre de dominio y la contraseña de administrador del KDC que especifique para el KDC dedicado al clúster.
	KDC externo	Solo está disponible en la versión de Amazon EMR 5.20.0 y posteriores. Especifica que los clústeres que utilizan esta configuración de seguridad autentican las entidades principales de Kerberos mediante un servidor de KDC externo al clúster. No se crea un KDC en el clúster. Al crear el clúster, debe especificar el nombre de dominio y la contraseña de administrador del KDC para el KDC externo.
Ciclo de vida del ticket		Opcional. Especifica el período durante el que un ticket de Kerberos emitido por el KDC es válido en los clústeres que utilizan esta configuración de seguridad. El ciclo de vida de los tickets es limitado por motivos de seguridad. Las aplicaciones de clúster y los servicios renuevan automáticamente los tickets después de que expiren. Los usuarios que se

Parámetro		Descripción
		conecten al clúster mediante SSH utilizando las credenciales de Kerberos tienen que ejecutar el comando `kinit` desde la línea de comandos del nodo principal para la renovación tras la expiración de un ticket.
Relación de confianza entre ámbitos		Especifica una relación de confianza entre ámbitos entre un KDC dedicado a un clúster en los clústeres que utilizan esta configuración de seguridad y un KDC situado en un ámbito de Kerberos diferente. Las entidades principales (normalmente los usuarios) de otro ámbito se autentican en los clústeres que utilizan esta configuración. Se requiere una configura ción adicional en el otro ámbito de Kerberos. Para obtener más información, consulte Tutorial: Configura ción de una relación de confianza entre ámbitos con un dominio de Active Directory.
Propiedad es de la relación de confianza entre ámbitos	Ámbito	Especifica el nombre de ámbito de Kerberos del otro ámbito en la relación de confianza. Por convención, los nombres de ámbito de Kerberos son los mismos que los nombres de dominio, pero en mayúsculas.
	Dominio	Especifica el nombre de dominio del otro ámbito en la relación de confianza.

Parámetro		Descripción
	Servidor de administración	Especifica el nombre de dominio completo (FQDN) o dirección IP del servidor de administración del otro ámbito de la relación de confianza. El servidor de administración y el servidor de KDC suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos para comunicarse. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 749. También se puede especificar el puerto (por ejemplo, `domain.example.com:749`).
	Servidor de KDC	Especifica el nombre de dominio completo (FQDN) o dirección IP del servidor de KDC del otro ámbito de la relación de confianza. El servidor de KDC y el servidor de administración suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 88. También se puede especificar el puerto (por ejemplo, `domain.example.com:88`).
KDC externo		Especifica que el clúster utiliza el KDC externo del clúster.

Parámetro			Descripción
Propiedades de KDC externo	Servidor de administración		Especifica el nombre de dominio completo (FQDN) o la dirección IP del servidor de administración externo. El servidor de administración y el servidor de KDC suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos para comunicarse. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 749. También se puede especificar el puerto (por ejemplo, `domain.example.com:749`).
	Servidor de KDC		Especifica el nombre de dominio completo (FQDN) del servidor de KDC externo. El servidor de KDC y el servidor de administración suelen ejecutarse en el mismo equipo con el mismo FQDN, pero utilizan puertos distintos. Si no se especifica ningún puerto, se usa el puerto predeterminado de Kerberos: el 88. También se puede especificar el puerto (por ejemplo, `domain.example.com:88`).
	Integración de Active Directory		Especifica que la autenticación de la entidad principal de Kerberos está integrada en un dominio de Microsoft Active Directory.
	Propiedades de la integración de Active Directory	Ámbito de Active Directory	Especifica el nombre de ámbito de Kerberos del dominio de Active Directory. Por convención, los nombres de ámbito de Kerberos suelen ser los mismos que los nombres de dominio, pero en mayúsculas.
		Dominio de Active Directory	Especifica el nombre de dominio de Active Directory.

Parámetro			Descripción
		Servidor de Active Directory	Especifica el nombre de dominio completo (FQDN) del controlador de dominio de Microsoft Active Directory.

Configuración de Kerberos para clústeres

Puede especificar la configuración de Kerberos al crear un clúster con la consola de Amazon EMR, la AWS CLI o la API de EMR.

Utilice las siguientes referencias para conocer las opciones de configuración de clústeres disponibles para la arquitectura Kerberos que elija. Se muestra la configuración de la consola de Amazon EMR. Para ver las opciones de la CLI correspondientes, consulte Ejemplos de configuraciones.

Parámetro	Descripción
Ámbito	El nombre del ámbito de Kerberos para el clúster. La convención de Kerberos consiste en establecerlo igual que el nombre del dominio, pero en mayúsculas. Por ejemplo, para el dominio `ec2.internal` , se utiliza `EC2.INTERNAL` como nombre de ámbito.
Contraseña de administración de KDC	La contraseña utilizada en el clúster `kadmin` o `kadmin.local` . Son interfaces de línea de comandos para el sistema de administración de Kerberos V5, que mantiene las entidades principales, las políticas de contraseñas y las tablas de claves de Kerberos para el clúster.
Contraseña de la entidad principal de confianza entre ámbitos (opcional)	Es necesaria para establecer una confianza entre ámbitos. La contraseña de la entidad principal de confianza entre ámbitos, que debe

Parámetro	Descripción
	ser idéntica en los distintos ámbitos. Use una contraseña segura.
Usuario de incorporación al dominio de Active Directory (opcional)	Obligatorio cuando se utiliza Active Directory en una relación de confianza entre ámbitos. Se trata de un nombre de inicio de sesión de usuario de una cuenta de Active Directory con permisos para incorporar equipos al dominio. Amazon EMR utiliza esta identidad para incorporar el clúster al dominio. Para obtener más información, consulte the section called “Paso 3: adición de cuentas al dominio para el clúster de EMR”.
Contraseña de incorporación al dominio de Active Directory (opcional)	La contraseña del usuario de incorporación a un dominio de Active Directory. Para obtener más información, consulte the section called “Paso 3: adición de cuentas al dominio para el clúster de EMR”.

Ejemplos de configuraciones

Los siguientes ejemplos muestran configuraciones de seguridad y configuraciones de clúster para situaciones comunes. Se muestran comandos de AWS CLI por razones de brevedad.

KDC local

Los siguientes comandos crean un clúster con un KDC dedicado del clúster que se ejecuta en el nodo principal. Es necesario realizar una configuración adicional en el clúster. Para obtener más información, consulte Configuración de un clúster para conexiones SSH y usuarios de HDFS autenticados en Kerberos.

Crear configuración de seguridad

```
aws emr create-security-configuration --name LocalKDCSecurityConfig \
```

```
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ClusterDedicatedKdc",\
"ClusterDedicatedKdcConfiguration": {"TicketLifetimeInHours": 24 }}}}'
```

Crear un clúster

```
aws emr create-cluster --release-label emr-5.36.1 \
--instance-count 3 --instance-type m5.xlarge \
--applications Name=Hadoop Name=Hive --ec2-attributes
 InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole \
--security-configuration LocalKDCSecurityConfig \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=MyPassword
```

KDC dedicado del clúster con una relación de confianza entre ámbitos de Active Directory

Los siguientes comandos crean un clúster con un KDC dedicado del clúster que se ejecuta en el nodo principal con una relación de confianza entre ámbitos con un dominio de Active Directory. Se necesita realizar configuración adicional en el clúster y en Active Directory. Para obtener más información, consulte Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory.

Crear configuración de seguridad

```
aws emr create-security-configuration --name LocalKDCWithADTrustSecurityConfig \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ClusterDedicatedKdc", \
"ClusterDedicatedKdcConfiguration": {"TicketLifetimeInHours": 24, \
"CrossRealmTrustConfiguration": {"Realm":"AD.DOMAIN.COM", \
"Domain":"ad.domain.com", "AdminServer":"ad.domain.com", \
"KdcServer":"ad.domain.com"}}}}}'
```

Crear un clúster

```
aws emr create-cluster --release-label emr-5.36.1 \
--instance-count 3 --instance-type m5.xlarge --applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration KDCWithADTrustSecurityConfig \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=MyClusterKDCAdminPassword,\
ADDomainJoinUser=ADUserLogonName,ADDomainJoinPassword=ADUserPassword,\
CrossRealmTrustPrincipalPassword=MatchADTrustPassword
```

KDC externo en un clúster diferente

Los siguientes comandos crean un clúster que hace referencia a un KDC dedicado del clúster en el nodo principal de un clúster diferente para autenticar entidades principales. Es necesario realizar una configuración adicional en el clúster. Para obtener más información, consulte Configuración de un clúster para conexiones SSH y usuarios de HDFS autenticados en Kerberos.

Crear configuración de seguridad

```
aws emr create-security-configuration --name ExtKDCOnDifferentCluster \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ExternalKdc", \
"ExternalKdcConfiguration": {"KdcServerType": "Single", \
"AdminServer": "MasterDNSOfKDCMaster:749", \
"KdcServer": "MasterDNSOfKDCMaster:88"}}}}'
```

Crear un clúster

```
aws emr create-cluster --release-label emr-5.36.1 \
--instance-count 3 --instance-type m5.xlarge \
--applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration ExtKDCOnDifferentCluster \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=KDCOnMasterPassword
```

KDC externo del clúster con una relación de confianza entre ámbitos de Active Directory

Los siguientes comandos crean un clúster sin KDC. El clúster hace referencia a un KDC dedicado del clúster que se ejecuta en el nodo principal de otro clúster para autenticar entidades principales. Esa KDC posee una relación de confianza entre ámbitos con un controlador de dominio de Active Directory. Es necesario realizar una configuración adicional en el nodo principal con el KDC. Para obtener más información, consulte Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory.

Crear configuración de seguridad

```
aws emr create-security-configuration --name ExtKDCWithADIntegration \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ExternalKdc", \
"ExternalKdcConfiguration": {"KdcServerType": "Single", \
```

```
"AdminServer": "MasterDNSofClusterKDC:749", \
"KdcServer": "MasterDNSofClusterKDC.com:88", \
"AdIntegrationConfiguration": {"AdRealm":"AD.DOMAIN.COM", \
"AdDomain":"ad.domain.com", \
"AdServer":"ad.domain.com"}}}}}'
```

Crear un clúster

```
aws emr create-cluster --release-label emr-5.36.1 \
--instance-count 3 --instance-type m5.xlarge --applications Name=Hadoop Name=Hive \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2Key \
--service-role EMR_DefaultRole --security-configuration ExtKDCWithADIntegration \
--kerberos-attributes Realm=EC2.INTERNAL,KdcAdminPassword=KDCOnMasterPassword,\
ADDomainJoinUser=MyPrivilegedADUserName,ADDomainJoinPassword=PasswordForADDomainJoinUser
```

Configuración de un clúster para conexiones SSH y usuarios de HDFS autenticados en Kerberos

Amazon EMR crea clientes de usuario autenticados por Kerberos para las aplicaciones que se ejecutan en el clúster; por ejemplo, el usuario `hadoop`, el usuario `spark` y otros. También puede añadir usuarios autenticados a los procesos del clúster mediante Kerberos. Luego, los usuarios autenticados pueden conectarse al clúster con sus credenciales de Kerberos y trabajar con aplicaciones. Para que un usuario pueda autenticarse en el clúster, es necesario realizar las siguientes configuraciones:

- En el clúster debe existir una cuenta de Linux que coincida con la entidad principal de Kerberos en el KDC. Amazon EMR lo hace automáticamente en arquitecturas que se integran con Active Directory.
- Debe crear un directorio de usuarios de HDFS en el nodo principal para cada usuario y proporcionar permisos de usuario al directorio.
- Debe configurar el servicio SSH para que GSSAPI esté habilitado en el nodo principal. Además, los usuarios deben tener un cliente SSH con GSSAPI habilitado.

Adición de usuarios de Linux y entidades principales de Kerberos al nodo principal

Si no utiliza Active Directory, debe crear las cuentas de Linux en el nodo principal del clúster y agregar entidades principales para estos usuarios de Linux al KDC. Esto incluye una entidad principal en el KDC para el nodo principal. Además de las entidades principales de usuarios, el KDC que se ejecuta en el nodo principal necesita una entidad principal para el host local.

Cuando la arquitectura incluye integración con Active Directory, los usuarios y entidades principales de Linux en el KDC local, si procede, se crean automáticamente. Puede omitir este paso. Para obtener más información, consulte Relación de confianza entre ámbitos y KDC externo: KDC de clúster en un clúster diferente con una relación de confianza entre ámbitos de Active Directory.

> **Important**
>
> El KDC, junto con la base de datos de entidades principales, se pierde cuando el nodo principal termina porque el nodo principal utiliza almacenamiento efímero. Si crea usuarios para las conexiones SSH, le recomendamos que establezca una relación de confianza entre dominios con un KDC externo configurado para lograr una alta disponibilidad. Como alternativa, si crea usuarios para las conexiones SSH mediante cuentas de Linux, automatice el proceso de creación de cuentas mediante acciones y scripts de arranque para que pueda repetirse al crear un clúster nuevo.

La manera más sencilla de añadir usuarios y entidades principales de KDC es enviar un paso al clúster después de crearlo o al crear el clúster. Si lo prefiere, puede conectarse al nodo principal utilizando un par de claves de EC2 como el usuario predeterminado `hadoop` para ejecutar los comandos. Para obtener más información, consulte Conectarse al nodo principal mediante SSH.

En el siguiente ejemplo, se envía un script bash `configureCluster.sh` a un clúster que ya existe, especificando su ID de clúster. El script se almacena en Amazon S3.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,\
Args=["s3://DOC-EXAMPLE-BUCKET/configureCluster.sh"]
```

El siguiente ejemplo muestra el contenido del script `configureCluster.sh`. El script también se encarga de la creación de los directorios de usuario de HDFS y de habilitar GSSAPI para SSH, lo que se explica en las siguientes secciones.

```
#!/bin/bash
#Add a principal to the KDC for the primary node, using the primary node's returned
 host name
sudo kadmin.local -q "ktadd -k /etc/krb5.keytab host/`hostname -f`"
#Declare an associative array of user names and passwords to add
declare -A arr
```

```
arr=([lijuan]=pwd1 [marymajor]=pwd2 [richardroe]=pwd3)
for i in ${!arr[@]}; do
    #Assign plain language variables for clarity
     name=${i}
     password=${arr[${i}]}

     # Create a principal for each user in the primary node and require a new password
 on first logon
     sudo kadmin.local -q "addprinc -pw $password +needchange $name"

     #Add hdfs directory for each user
     hdfs dfs -mkdir /user/$name

     #Change owner of each user's hdfs directory to that user
     hdfs dfs -chown $name:$name /user/$name
done

# Enable GSSAPI authentication for SSH and restart SSH service
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Adición de directorios de HDFS de usuarios

Para permitir a los usuarios iniciar sesión en el clúster para ejecutar los trabajos de Hadoop, debe agregar directorios de usuario de HDFS para sus cuentas de Linux y conceder a cada usuario la propiedad de su directorio.

La manera más sencilla de crear directorios de HDFS es enviar un paso al clúster después de crearlo o al crear el clúster. Si lo prefiere, podría conectarse al nodo principal utilizando un par de claves de EC2 como el usuario predeterminado `hadoop` para ejecutar los comandos. Para obtener más información, consulte Conectarse al nodo principal mediante SSH.

En el siguiente ejemplo, se envía un script bash `AddHDFSUsers.sh` a un clúster que ya existe, especificando su ID de clúster. El script se almacena en Amazon S3.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-
EXAMPLE-BUCKET/AddHDFSUsers.sh"]
```

El siguiente ejemplo muestra el contenido del script `AddHDFSUsers.sh`.

```
#!/bin/bash
# AddHDFSUsers.sh script

# Initialize an array of user names from AD, or Linux users created manually on the
 cluster
ADUSERS=("lijuan" "marymajor" "richardroe" "myusername")

# For each user listed, create an HDFS user directory
# and change ownership to the user

for username in ${ADUSERS[@]}; do
     hdfs dfs -mkdir /user/$username
     hdfs dfs -chown $username:$username /user/$username
done
```

Habilitación de GSSAPI para SSH

Para que los usuarios autenticados por Kerberos se conecten al nodo principal mediante SSH, el servicio SSH debe tener habilitada la autenticación de GSSAPI. Para habilitar GSSAPI, ejecute los siguientes comandos desde la línea de comandos del nodo principal o utilice un paso para ejecutarlo como un script. Después de volver a configurar SSH, reinicie el servicio.

```
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Uso de SSH para conectarse a clústeres que utilizan Kerberos

En esta sección, se muestran los pasos para que un usuario autenticado por Kerberos se conecte al nodo principal de un clúster de EMR.

Cada equipo que se utiliza para una conexión SSH debe tener instalados un cliente SSH y aplicaciones de cliente de Kerberos. Lo más probable es que los equipos Linux los incluyan de forma predeterminada. Por ejemplo, OpenSSH está instalado en la mayoría de los sistemas operativos Linux, Unix y macOS. Puede comprobar si tiene un cliente SSH escribiendo ssh en la línea de comandos. Si su equipo no reconoce el comando, instale un cliente SSH para conectarse al nodo

principal. El proyecto OpenSSH ofrece una implementación gratuita de toda la suite de herramientas de SSH. Para obtener más información, consulte el sitio web OpenSSH. Los usuarios de Windows pueden utilizar aplicaciones como PuTTY como cliente SSH.

Para obtener más información sobre las conexiones SSH persistentes, use Conectarse a un clúster.

SSH utiliza GSSAPI para autenticar a los clientes de Kerberos y debe habilitar la autenticación de GSSAPI para el servicio SSH en el nodo principal del clúster. Para obtener más información, consulte Habilitación de GSSAPI para SSH. Los clientes SSH también deben utilizar GSSAPI.

En los siguientes ejemplos, en *MasterPublicDNS*, utilice el valor que aparece para DNS público principal en la pestaña Resumen del panel de detalles del clúster, por ejemplo, *ec2-11-222-33-44.compute-1.amazonaws.com*.

Requisito previo para krb5.conf (sin Active Directory)

Cuando se utiliza una configuración sin integración con Active Directory, además del cliente SSH y las aplicaciones de cliente de Kerberos, cada equipo cliente debe tener una copia del archivo `/etc/krb5.conf` que coincida con el archivo `/etc/krb5.conf` en el nodo principal del clúster.

Para copiar el archivo krb5.conf

1. Utilice SSH para conectarse al nodo principal mediante un par de claves de EC2 y el usuario `hadoop`predeterminado, por ejemplo, `hadoop@`*MasterPublicDNS*. Para obtener instrucciones detalladas, consulte Conectarse a un clúster.
2. Desde el nodo principal, copie el contenido del archivo `/etc/krb5.conf`. Para obtener más información, consulte Conectarse a un clúster.
3. En cada equipo cliente que se utilice para conectarse al clúster, cree un archivo `/etc/krb5.conf` idéntico a partir de la copia que hizo en el paso anterior.

Uso de kinit y SSH

Cada vez que un usuario se conecta desde un equipo cliente con credenciales de Kerberos, el usuario debe renovar primero los tickets de Kerberos para su usuario en el equipo cliente. Además, se debe configurar el cliente SSH para utilizar la autenticación GSSAPI.

Para utilizar SSH para conectarse a un clúster de EMR que utilice Kerberos

1. Utilice `kinit` para renovar sus tickets de Kerberos, como se muestra en el siguiente ejemplo

```
kinit user1
```

2. Utilice un cliente `ssh` junto con la entidad principal que creó en el KDC dedicado del clúster o el nombre de usuario de Active Directory. Asegúrese de que la autenticación GSSAPI está habilitada, tal y como se muestra en los siguientes ejemplos.

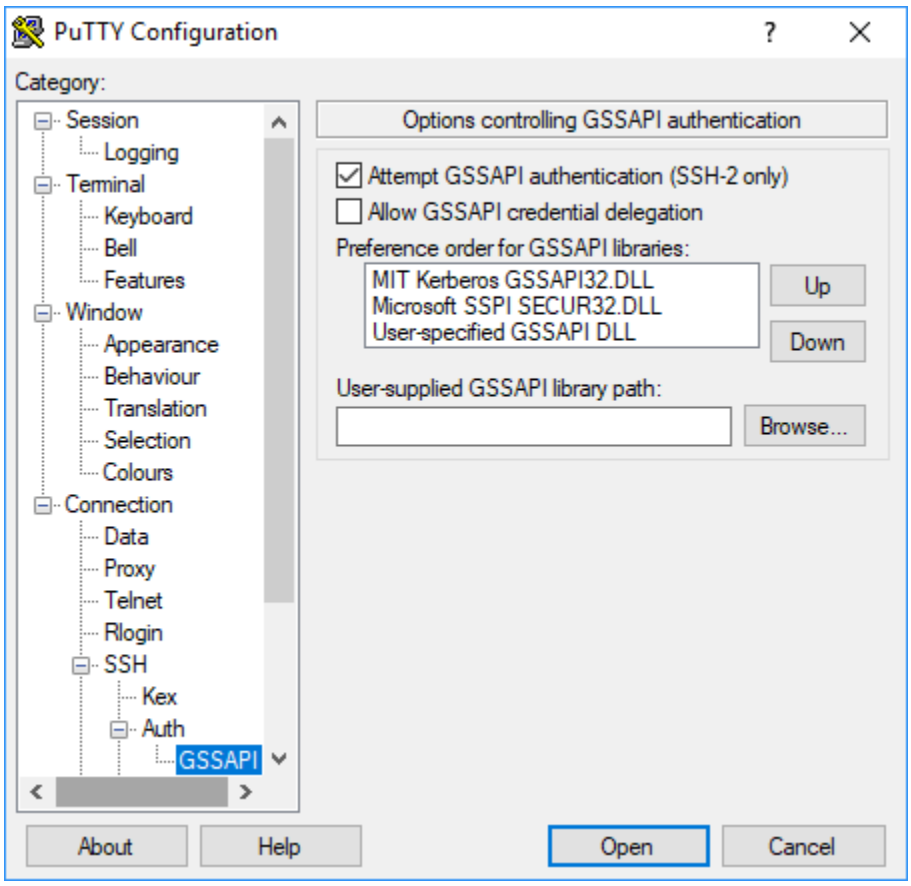
   Ejemplo: usuarios de Linux

   La opción `-K` especifica la autenticación GSSAPI.

   ```
   ssh -K user1@MasterPublicDNS
   ```

   Ejemplo: usuarios de Windows (PuTTY)

   Asegúrese de que la opción de autenticación GSSAPI de la sesión está habilitada, tal y como se muestra:



Tutorial: Configuración de un KDC dedicado del clúster

En este tema, le enseñaremos cómo crear un clúster con un centro de distribución de claves (KDC) dedicado, cómo agregar manualmente cuentas de usuario de Linux a todos los nodos del clúster, cómo agregar entidades principales de Kerberos al KDC en el nodo principal y cómo asegurarse de que los equipos cliente tengan instalado un cliente de Kerberos.

Para obtener más información sobre la compatibilidad de Amazon EMR con Kerberos y KDC, así como enlaces a la documentación de Kerberos del MIT, consulte Uso de Kerberos para la autenticación con Amazon EMR.

Paso 1: creación del clúster que utiliza Kerberos

1. Cree una configuración de seguridad que habilite Kerberos. El siguiente ejemplo muestra el uso de un comando `create-security-configuration` con la AWS CLI que especifica la configuración de seguridad como una estructura JSON en línea. También puede hacer referencia a un archivo guardado de forma local.

```
aws emr create-security-configuration --name MyKerberosConfig \
--security-configuration '{"AuthenticationConfiguration": {"KerberosConfiguration":
{"Provider": "ClusterDedicatedKdc", "ClusterDedicatedKdcConfiguration":
 {"TicketLifetimeInHours": 24}}}}'
```

2. Cree un clúster que haga referencia a la configuración de seguridad, establezca los atributos de Kerberos para el clúster y añada las cuentas de Linux mediante una acción de arranque. El siguiente ejemplo muestra el uso de un comando `create-cluster` en la AWS CLI. El comando hace referencia a la configuración de seguridad que se ha creado anteriormente, `MyKerberosConfig`. También hace referencia a un script sencillo, `createlinuxusers.sh`, como acción de arranque, que debe crear y cargar en Amazon S3 antes de crear el clúster.

```
aws emr create-cluster --name "MyKerberosCluster" \
--release-label emr-5.36.1 \
--instance-type m5.xlarge \
--instance-count 3 \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2KeyPair \
--service-role EMR_DefaultRole \
--security-configuration MyKerberosConfig \
--applications Name=Hadoop Name=Hive Name=Oozie Name=Hue Name=HCatalog Name=Spark \
--kerberos-attributes Realm=EC2.INTERNAL,\
KdcAdminPassword=MyClusterKDCAdminPwd \
```

```
--bootstrap-actions Path=s3://DOC-EXAMPLE-BUCKET/createlinuxusers.sh
```

El siguiente código ilustra el contenido del script `createlinuxusers.sh`, que agrega user1, user2 y user3 en cada nodo del clúster. En el siguiente paso, añadirá estos usuarios como entidades principales de KDC.

```
#!/bin/bash
sudo adduser user1
sudo adduser user2
sudo adduser user3
```

Paso 2: adición de entidades principales al KDC, creación directorios de usuarios de HDFS y configuración de SSH

En el KDC que se ejecuta en el nodo principal, se debe agregar una entidad principal para el host local y para cada usuario que se cree en el clúster. También puede crear directorios de HDFS para cada usuario en caso de que necesitan conectarse al clúster y ejecutar trabajos de Hadoop. Del mismo modo, configurae el servicio SSH para activar la autenticación GSSAPI, que es necesaria para Kerberos. Después de habilitar GSSAPI, reinicie el servicio SSH.

La forma más sencilla de realizar estas tareas es enviar un paso para el clúster. El siguiente ejemplo envía un script bash `configurekdc.sh` al clúster que ha creado en el paso anterior, haciendo referencia a su ID de clúster. El script se almacena en Amazon S3. De forma alternativa, puede conectarse al nodo principal utilizando un par de claves de EC2 para ejecutar los comandos o enviar el paso durante la creación del clúster.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> --steps
 Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,Jar=s3://
myregion.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-
BUCKET/configurekdc.sh"]
```

El siguiente código muestra el contenido del script `configurekdc.sh`.

```
#!/bin/bash
#Add a principal to the KDC for the primary node, using the primary node's returned
 host name
sudo kadmin.local -q "ktadd -k /etc/krb5.keytab host/`hostname -f`"
#Declare an associative array of user names and passwords to add
declare -A arr
```

```
arr=([user1]=pwd1 [user2]=pwd2 [user3]=pwd3)
for i in ${!arr[@]}; do
    #Assign plain language variables for clarity
     name=${i}
     password=${arr[${i}]}

     # Create principal for sshuser in the primary node and require a new password on
 first logon
     sudo kadmin.local -q "addprinc -pw $password +needchange $name"

     #Add user hdfs directory
     hdfs dfs -mkdir /user/$name

     #Change owner of user's hdfs directory to user
     hdfs dfs -chown $name:$name /user/$name
done

# Enable GSSAPI authentication for SSH and restart SSH service
sudo sed -i 's/^.*GSSAPIAuthentication.*$/GSSAPIAuthentication yes/' /etc/ssh/
sshd_config
sudo sed -i 's/^.*GSSAPICleanupCredentials.*$/GSSAPICleanupCredentials yes/' /etc/ssh/
sshd_config
sudo systemctl restart sshd
```

Los usuarios que ha añadido ahora deberían ser capaces de conectarse al clúster mediante SSH. Para obtener más información, consulte Uso de SSH para conectarse a clústeres que utilizan Kerberos.

Tutorial: Configuración de una relación de confianza entre ámbitos con un dominio de Active Directory

Al crear una relación de confianza entre ámbitos, se permite a las entidades principales (normalmente usuarios) de otro ámbito de Kerberos autenticar a los componentes de la aplicación en el clúster de EMR. El centro de distribución de claves (KDC) dedicado del clúster establece una relación de confianza con otro KDC utilizando una entidad principal de confianza entre ámbitos que existe en los dos KDC. El nombre y la contraseña de la entidad principal deben coincidir exactamente.

Una relación de confianza entre ámbitos requiere que los KDC puedan tener acceso entre sí a través de la red y resolver sus nombres de dominio. Pasos para establecer una relación de confianza con un ámbito Microsoft AD controlador de dominio se ejecutan como una instancia de EC2 se

suministran a continuación, junto con un ejemplo de configuración de red que proporciona la conectividad y la resolución de nombres de dominio. Es aceptable cualquier configuración de red que permita el tráfico de red necesario entre KDC.

Opcionalmente, después de establecer una relación de confianza entre ámbitos con Active Directory utilizando un KDC en un clúster, puede crear otro clúster con una configuración de seguridad diferente para hacer referencia al KDC en el primer clúster como un KDC externo. Para ver un ejemplo de la configuración de seguridad y la configuración del clúster, consulte KDC externo del clúster con una relación de confianza entre ámbitos de Active Directory.

Para obtener más información sobre la compatibilidad de Amazon EMR con Kerberos y KDC, así como enlaces a la documentación de Kerberos del MIT, consulte Uso de Kerberos para la autenticación con Amazon EMR.

> **Important**
>
> Amazon EMR no admite las relaciones de confianza entre ámbitos con AWS Directory Service for Microsoft Active Directory.

Paso 1: Configuración de la VPC y la subred

Paso 2: lanzamiento e instalación del controlador de dominio de Active Directory

Paso 3: adición de cuentas al dominio para el clúster de EMR

Paso 4: configuración de una relación de confianza entrante en el controlador de dominio de Active Directory

Paso 5: uso de un conjunto de opciones de DHCP para especificar el controlador de dominio de Active Directory como un servidor DNS de la VPC

Paso 6: Lanzar un clúster de EMR que utiliza Kerberos

Paso 7: creación de usuarios de HDFS y configuración de permisos en el clúster para las cuentas de usuario de Active Directory

Paso 1: Configuración de la VPC y la subred

Los pasos siguientes muestran cómo crear una VPC y una subred para que el KDC dedicado del clúster pueda tener acceso al controlador de dominio de Active Directory y resolver su nombre de

dominio. En estos pasos, la resolución de nombres de dominio se proporciona haciendo referencia al controlador de dominio de Active Directory como el servidor de nombres de dominio en el conjunto de opciones de DHCP. Para obtener más información, consulte Paso 5: uso de un conjunto de opciones de DHCP para especificar el controlador de dominio de Active Directory como un servidor DNS de la VPC.

El KDC y el controlador de dominio de Active Directory deben ser capaces de resolver sus respectivos nombres de dominio. Esto permite a Amazon EMR incorporar equipos al dominio y configurar automáticamente las cuentas de Linux y los parámetros de SSH correspondientes en las instancias de clúster.

Si Amazon EMR no puede resolver el nombre de dominio, se puede hacer referencia a la relación de confianza mediante la dirección IP del controlador de dominio de Active Directory. Sin embargo, debe agregar manualmente las cuentas de usuario de Linux, agregar las correspondientes entidades principales al KDC dedicado del clúster y configurar SSH.

Para configurar la VPC y la subred

1. Cree una instancia de Amazon VPC con una única subred pública. Para obtener más información, consulte Paso 1: creación de la VPC en la Guía de introducción a Amazon VPC.

   > **Important**
   >
   > Cuando utilice un controlador de dominio de Microsoft Active Directory, elija un bloque de CIDR para el clúster de EMR, de forma que todas las direcciones IPv4 tengan menos de nueve caracteres (por ejemplo, 10.0.0.0/16). Esto se debe a que los nombres de DNS de los equipos del clúster se utilizan cuando los equipos se incorporan al directorio de Active Directory. AWS asigna nombres de host de DNS en función de la dirección IPv4, de forma que las direcciones IP más largas pueden dar como resultado nombres de DNS de más de 15 caracteres. Active Directory tiene un límite de 15 caracteres para registrar los nombres de los equipos que se incorporan, y trunca los nombres que son más largos, lo que puede provocar errores impredecibles.

2. Elimine el conjunto de opciones de DHCP predeterminado asignado a la VPC. Para obtener más información, consulte Cambiar una VPC para que no utilice ninguna opción de DHCP. Posteriormente, puede añadir uno nuevo que especifique el controlador de dominio de Active Directory como servidor DNS.

3. Confirme que se ha activado el soporte de DNS para la VPC, es decir, que se han activado los nombres de host DNS y la resolución de DNS. De forma predeterminada, están habilitadas. Para obtener más información, consulte Actualización de la compatibilidad de DNS para su VPC.
4. Confirme que la VPC tiene un asociada una gateway de Internet, que es la opción predeterminada. Para obtener más información, consulte Creación y asociación de una gateway de Internet.

   > **Note**
   >
   > En este ejemplo, se utiliza una gateway de Internet porque se está estableciendo un nuevo controlador de dominio para la VPC. Puede que no sea necesaria una gateway de Internet para su aplicación. El único requisito es que el KDC dedicado del clúster pueda tener acceso al controlador de dominio de Active Directory.

5. Cree una tabla de ruteo personalizada, añada una ruta que se dirija a la gateway de Internet y, a continuación, asóciela a la subred. Para obtener más información, consulte Creación de una tabla de enrutamiento personalizada.
6. Al lanzar la instancia de EC2 para el dominio, debe disponer de un controlador estático de direcciones IPv4 públicas para que pueda conectarse a ella con RDP. La forma más sencilla de hacerlo es configurar la subred pública automáticamente asignar direcciones IPv4. Este no es el valor predeterminado cuando se crea una subred. Para obtener más información, consulte Modificación del atributo de direcciones IPv4 públicas de su subred. Si lo prefiere, puede asignar la dirección al lanzar la instancia. Para obtener más información, consulte Asignación de una dirección IPv4 pública durante el lanzamiento de la instancia.
7. Cuando termine, tome nota de los ID de la VPC y de la subred. Los utilizará posteriormente al lanzar el controlador de dominio de Active Directory y el clúster.

Paso 2: lanzamiento e instalación del controlador de dominio de Active Directory

1. Lance una instancia de EC2 basada en la AMI base de Microsoft Windows Server 2016. Le recomendamos un tipo de instancia m4.xlarge o mejor. Para obtener más información, consulte Lanzamiento de una instancia de AWS Marketplace en la Guía del usuario de Amazon EC2 para instancias de Linux.
2. Anote el Group ID (ID de grupo) del grupo de seguridad asociado a la instancia EC2. Lo necesitará para Paso 6: Lanzar un clúster de EMR que utiliza Kerberos. Utilizamos *sg-012xrlmdomain345*. También puede especificar distintos grupos de seguridad para el

clúster de EMR y esta instancia que permita el tráfico entre ellos. Para obtener más información, consulte Grupos de seguridad de Amazon EC2 para instancias Linux en la Guía del usuario de Amazon EC2 para instancias de Linux.

3. Conéctese a la instancia de EC2 a través de RDP. Para obtener más información, consulte Conexión con la instancia de Windows en la Guía del usuario de Amazon EC2 para instancias de Windows.
4. Inicie Administrador del servidor para instalar y configurar el rol de los servicios de dominio de Active Directory en el servidor. Promocione el servidor a controlador de dominio y asígnele un nombre de dominio (el ejemplo que utilizamos aquí es *ad.domain.com*). Anote el nombre de dominio porque lo necesitará más adelante al crear el clúster y la configuración de seguridad de EMR. Si es la primera vez que configura Active Directory, puede seguir las instrucciones de How to Set Up Active Directory (AD) in Windows Server 2016.

   La instancia se reiniciará cuando termine.

Paso 3: adición de cuentas al dominio para el clúster de EMR

Establezca una conexión RDP con el controlador de dominio de Active Directory para crear cuentas de usuario en Usuarios y equipos de Active Directory para cada usuario del clúster. Para obtener más información, consulte Create a User Account in Active Directory Users and Computers en el sitio Microsoft Learn. Anote el User logon name (Nombre de inicio de sesión de usuario) de cada usuario. Necesitas estas versiones posteriores al configurar el clúster.

Cree también una cuenta con privilegios suficientes para incorporar ordenadores al dominio. Tiene que especificar esta cuenta al crear un clúster. Amazon EMR la utiliza para incorporar las instancias del clúster al dominio Solo debe especificar esta cuenta y su contraseña Paso 6: Lanzar un clúster de EMR que utiliza Kerberos. Para delegar los privilegios de incorporación de equipos a la cuenta, le recomendamos que cree un grupo con privilegios de incorporación y, a continuación, asigne el usuario al grupo. Para obtener instrucciones, consulte Delegación de privilegios de vinculación a directorios en la Guía de administración de AWS Directory Service.

Paso 4: configuración de una relación de confianza entrante en el controlador de dominio de Active Directory

Los comandos del ejemplo siguiente crean una relación de confianza en Active Directory, que es una confianza de ámbito, entrante, unidireccional y no transitiva con el KDC dedicado del clúster. El ejemplo que utilizamos para el ámbito del clúster es *EC2.INTERNAL*. Sustituya *KDC-FQDN* por el nombre de DNS público que aparece para el nodo principal de Amazon EMR que aloja el KDC.

El parámetro `passwordt` especifica la cross-realm principal password (contraseña de la entidad principal de confianza entre ámbitos), que se especifica junto con el realm (ámbito) del clúster al crear un clúster. El nombre del ámbito se deriva del nombre de dominio predeterminado en `us-east-1` para el clúster. El `Domain` es el dominio de Active Directory en el que va a crear la confianza, que es en minúsculas por convención. El ejemplo utiliza *`ad.domain.com`*

Abra el símbolo del sistema de Windows con privilegios de administrador y escriba los siguientes comandos para crear la relación de confianza en el controlador de dominio de Active Directory:

```
C:\Users\Administrator> ksetup /addkdc EC2.INTERNAL KDC-FQDN
C:\Users\Administrator> netdom trust EC2.INTERNAL /Domain:ad.domain.com /add /realm /
passwordt:MyVeryStrongPassword
C:\Users\Administrator> ksetup /SetEncTypeAttr EC2.INTERNAL AES256-CTS-HMAC-SHA1-96
```

Paso 5: uso de un conjunto de opciones de DHCP para especificar el controlador de dominio de Active Directory como un servidor DNS de la VPC

Ahora que está configurado el controlador de dominio de Active Directory, debe configurar la VPC para utilizarla como servidor de nombres de dominio para la resolución de nombres en la VPC. Para ello, asocie un conjunto de opciones de DHCP. En Nombre de dominio, especifique el nombre de dominio del clúster; por ejemplo, `ec2.internal` si el clúster se encuentra en la región us-east-1 o *`region`*`.compute.internal` para las demás regiones. En Domain name servers (Servidores de nombres de dominio), debe especificar la dirección IP del controlador de dominio de Active Directory (al que se debe tener acceso desde el clúster) como la primera entrada, seguida de AmazonProvidedDNS (por ejemplo, ***`xx.xx.xx.xx`***,AmazonProvidedDNS). Para obtener más información, consulte Cambio de los conjuntos de opciones de DHCP.

Paso 6: Lanzar un clúster de EMR que utiliza Kerberos

1. En Amazon EMR, cree una configuración de seguridad que especifique el controlador de dominio de Active Directory que ha creado en los pasos anteriores. A continuación se muestra un ejemplo. Sustituya el dominio, *`ad.domain.com`*por el nombre del dominio especificado en el Paso 2: lanzamiento e instalación del controlador de dominio de Active Directory.

   ```
   aws emr create-security-configuration --name MyKerberosConfig \
   --security-configuration '{
     "AuthenticationConfiguration": {
       "KerberosConfiguration": {
         "Provider": "ClusterDedicatedKdc",
   ```

```
      "ClusterDedicatedKdcConfiguration": {
        "TicketLifetimeInHours": 24,
        "CrossRealmTrustConfiguration": {
          "Realm": "AD.DOMAIN.COM",
          "Domain": "ad.domain.com",
          "AdminServer": "ad.domain.com",
          "KdcServer": "ad.domain.com"
        }
      }
    }
  }
}'
```

2. Cree el clúster con los siguientes atributos:

   - Utilice la opción `--security-configuration` para especificar la configuración de seguridad que ha creado. Utilizamos *`MyKerberosConfig`* en el ejemplo.
   - Utilice la propiedad `SubnetId` de `--ec2-attributes option` para especificar la subred que ha creado en Paso 1: Configuración de la VPC y la subred. Utilizamos *`step1-subnet`* en el ejemplo.
   - Utilice `AdditionalMasterSecurityGroups` y `AdditionalSlaveSecurityGroups` de la opción `--ec2-attributes` para especificar que el grupo de seguridad asociado al controlador de dominio de AD desde Paso 2: lanzamiento e instalación del controlador de dominio de Active Directory está asociado al nodo principal del clúster, así como a los nodos secundarios y de tareas. Utilizamos *`sg-012xrlmdomain345`* en el ejemplo.

   Utilice `--kerberos-attributes` para especificar los siguientes atributos de Kerberos específicos del clúster:

   - El ámbito para el clúster que especificó al configurar el controlador de dominio de Active Directory.
   - La contraseña de la entidad principal de confianza entre ámbitos que especificó como `passwordt` en el Paso 4: configuración de una relación de confianza entrante en el controlador de dominio de Active Directory.
   - Una `KdcAdminPassword`, que se puede utilizar para administrar el KDC dedicado del clúster.
   - El nombre de inicio de sesión de usuario y la contraseña de la cuenta de Active Directory con privilegios para incorporar equipos que ha creado en el Paso 3: adición de cuentas al dominio para el clúster de EMR.

   El siguiente ejemplo lanza un clúster que utiliza Kerberos.

```
aws emr create-cluster --name "MyKerberosCluster" \
--release-label emr-5.10.0 \
--instance-type m5.xlarge \
--instance-count 3 \
--ec2-attributes InstanceProfile=EMR_EC2_DefaultRole,KeyName=MyEC2KeyPair,\
SubnetId=step1-subnet, AdditionalMasterSecurityGroups=sg-012xrlmdomain345,
AdditionalSlaveSecurityGroups=sg-012xrlmdomain345\
--service-role EMR_DefaultRole \
--security-configuration MyKerberosConfig \
--applications Name=Hadoop Name=Hive Name=Oozie Name=Hue Name=HCatalog Name=Spark \
--kerberos-attributes Realm=EC2.INTERNAL,\
KdcAdminPassword=MyClusterKDCAdminPwd,\
ADDomainJoinUser=ADUserLogonName,ADDomainJoinPassword=ADUserPassword,\
CrossRealmTrustPrincipalPassword=MatchADTrustPwd
```

Paso 7: creación de usuarios de HDFS y configuración de permisos en el clúster para las cuentas de usuario de Active Directory

Cuando se configura una relación de confianza con Active Directory, Amazon EMR crea usuarios de Linux en el clúster para cada cuenta de usuario de Active Directory. Por ejemplo, el nombre de inicio de sesión de usuario `LiJuan` en Active Directory se corresponde con la cuenta de Linux de `lijuan`. Los nombres de usuario de Active Directory pueden contener letras mayúsculas, pero Linux no distingue entre mayúsculas y minúsculas para dichos nombres.

Para permitir a los usuarios iniciar sesión en el clúster para ejecutar los trabajos de Hadoop, debe agregar directorios de usuario de HDFS para sus cuentas de Linux y conceder a cada usuario la propiedad de su directorio. Para ello, le recomendamos que ejecute un script almacenado en Amazon S3 como un paso de clúster. De forma alternativa, puede ejecutar los comandos en el script siguiente desde la línea de comandos en el nodo principal. Utilice el par de claves de EC2 especificado al crear el clúster para conectarse al nodo principal a través de SSH como usuario de Hadoop. Para obtener más información, consulte Uso de un par de claves de EC2 para credenciales de SSH.

Ejecute el siguiente comando para añadir un paso al clúster que ejecuta un script, *AddHDFSUsers.sh*.

```
aws emr add-steps --cluster-id <j-2AL4XXXXXX5T9> \
--steps Type=CUSTOM_JAR,Name=CustomJAR,ActionOnFailure=CONTINUE,\
```

```
Jar=s3://region.elasticmapreduce/libs/script-runner/script-runner.jar,Args=["s3://DOC-EXAMPLE-BUCKET/AddHDFSUsers.sh"]
```

El contenido del archivo *AddHDFSUsers.sh* es el siguiente.

```
#!/bin/bash
# AddHDFSUsers.sh script

# Initialize an array of user names from AD or Linux users and KDC principals created
 manually on the cluster
ADUSERS=("lijuan" "marymajor" "richardroe" "myusername")

# For each user listed, create an HDFS user directory
# and change ownership to the user

for username in ${ADUSERS[@]}; do
     hdfs dfs -mkdir /user/$username
     hdfs dfs -chown $username:$username /user/$username
done
```

Grupos de Active Directory asignados a grupos de Hadoop

Amazon EMR utiliza System Security Services Daemon (SSD) para asignar grupos de Active Directory a grupos de Hadoop. Para confirmar las asignaciones de grupos, después de iniciar sesión en el nodo principal tal y como se describe en Uso de SSH para conectarse a clústeres que utilizan Kerberos, puede utilizar el comando `hdfs groups` para confirmar que los grupos de Active Directory a los que pertenece su cuenta de Active Directory se han asignado a los grupos de Hadoop del usuario de Hadoop correspondiente en el clúster. También puede comprobar los mapeos de grupos de otros usuarios especificando uno o varios nombres de usuario con el comando, por ejemplo `hdfs groups lijuan`. Para obtener más información, consulte groups en la Apache HDFS Commands Guide.

Uso de servidores de Active Directory o LDAP para la autenticación con Amazon EMR

Con las versiones 6.12.0 y posteriores de Amazon EMR, puede utilizar el protocolo LDAP sobre SSL (LDAPS) para lanzar un clúster que se integre de forma nativa con su servidor de identidad corporativa. El LDAP (protocolo ligero de acceso a directorios) es un protocolo de aplicaciones abierto e independiente del proveedor que accede a los datos y los mantiene. El LDAP se utiliza

habitualmente para la autenticación de usuarios en servidores de identidad corporativa alojados en aplicaciones como Active Directory (AD) y OpenLDAP. Con esta integración nativa, puede usar su servidor LDAP para autenticar a los usuarios en Amazon EMR.

Entre los aspectos destacados de la integración del LDAP de Amazon EMR se encuentran los siguientes:

- Amazon EMR configura las aplicaciones compatibles para que se autentiquen con la autenticación LDAP en su nombre.
- Amazon EMR configura y mantiene la seguridad de las aplicaciones compatibles con el protocolo Kerberos. No es necesario introducir ningún comando o script.
- Se obtiene un control de acceso detallado (FGAC) mediante la autorización de Apache Ranger para las bases de datos y tablas del metaalmacén de Hive. Para obtener más información, consulte Integración de Amazon EMR con Apache Ranger.
- Cuando necesita credenciales del LDAP para acceder a un clúster, obtiene un control de acceso detallado (FGAC) sobre quién puede acceder a sus clústeres de EMR a través de SSH.

En las siguientes páginas se proporciona información general conceptual, los requisitos previos y los pasos para lanzar un clúster de EMR con la integración del LDAP de Amazon EMR.

Temas

- Descripción general del LDAP con Amazon EMR
- Componentes del LDAP para Amazon EMR
- Compatibilidad de aplicaciones y consideraciones que hay que tener en cuenta sobre el LDAP para Amazon EMR
- Configuración y lanzamiento de un clúster de EMR con el LDAP
- Ejemplos de uso del LDAP con Amazon EMR

Descripción general del LDAP con Amazon EMR

El protocolo ligero de acceso a directorios (LDAP) es un protocolo de software que los administradores de red utilizan para administrar y controlar el acceso a los datos mediante la autenticación de los usuarios dentro de la red de una empresa. El protocolo LDAP almacena la información en una estructura jerárquica de directorios en árbol. Para obtener más información, consulte Basic LDAP Concepts en LDAP.com.

Dentro de la red de una empresa, muchas aplicaciones pueden utilizar el protocolo LDAP para autenticar a los usuarios. Con la integración del LDAP de Amazon EMR, los clústeres de EMR pueden utilizar de forma nativa el mismo protocolo LDAP con una configuración de seguridad adicional.

Amazon EMR admite dos implementaciones principales del protocolo LDAP: Active Directory y OpenLDAP. Si bien son posibles otras implementaciones, la mayoría se ajusta a los mismos protocolos de autenticación que Active Directory u OpenLDAP.

Active Directory (AD)

Active Directory (AD) es un servicio de directorios de Microsoft para redes de dominio de Windows. AD está incluido en la mayoría de los sistemas operativos Windows Server y puede comunicarse con los clientes a través de los protocolos LDAP y LDAPS. Amazon EMR intenta vincular un usuario a su instancia de AD con el nombre principal del usuario (UPN) como nombre distintivo y la contraseña para realizar la autenticación. El UPN utiliza el formato estándar `username@domain_name`.

OpenLDAP

OpenLDAP es una implementación gratuita y de código abierto del protocolo LDAP. Para la autenticación, Amazon EMR intenta vincular un usuario a su instancia de OpenLDAP con el nombre de dominio completo (FQDN) como nombre distintivo y la contraseña. El FQDN utiliza el formato estándar `username_attribute=username,LDAP_user_search_base`. Por lo general, el valor `username_attribute` es `uid` y el valor `LDAP_user_search_base` contiene los atributos del árbol que conduce al usuario. Por ejemplo, `ou=People,dc=example,dc=com`.

Otras implementaciones gratuitas y de código abierto del protocolo LDAP suelen seguir un FQDN similar al de OpenLDAP para los nombres distintivos de sus usuarios.

Componentes del LDAP para Amazon EMR

Puede usar su servidor LDAP para autenticarse con Amazon EMR y cualquier aplicación que el usuario utilice directamente en el clúster de EMR a través de los siguientes componentes.

Agente secreto

El agente secreto es un proceso que se ejecuta en el clúster que autentica todas las solicitudes de los usuarios. El agente secreto crea el enlace de usuario a su servidor LDAP en nombre de las aplicaciones compatibles en el clúster de EMR. El agente secreto se ejecuta como el usuario `emrsecretagent` y escribe registros en el directorio `/emr/secretagent/log`. Estos registros

proporcionan detalles sobre el estado de la solicitud de autenticación de cada usuario y cualquier error que pueda surgir durante la autenticación del usuario.

System Security Services Daemon (SSSD)

SSSD es un daemon que se ejecuta en cada nodo de un clúster de EMR habilitado para el LDAP. SSSD crea y administra un usuario de UNIX para sincronizar su identidad corporativa remota con cada nodo. Las aplicaciones basadas en YARN, como Hive y Spark, requieren que haya un usuario local de UNIX en cada nodo que ejecute una consulta para un usuario.

Compatibilidad de aplicaciones y consideraciones que hay que tener en cuenta sobre el LDAP para Amazon EMR

Aplicaciones compatibles con el LDAP para Amazon EMR

> **Important**
>
> Las aplicaciones que aparecen en esta página son las únicas aplicaciones que Amazon EMR admite para el LDAP. Para garantizar la seguridad del clúster, solo puede incluir aplicaciones compatibles con el LDAP al crear un clúster de EMR con el LDAP habilitado. Si intenta instalar otras aplicaciones no compatibles, Amazon EMR rechazará su solicitud de un nuevo clúster.

Las versiones 6.12 y posteriores de Amazon EMR admiten la integración del LDAP con las siguientes aplicaciones:

- Apache Livy
- Desde Apache Hive hasta HiveServer2 (HS2)
- Trino
- Presto
- Hue

También puede instalar las siguientes aplicaciones en un clúster de EMR y configurarlas para que se ajusten a sus necesidades de seguridad:

- Apache Spark

- Apache Hadoop

Características compatibles con el LDAP para Amazon EMR

Puede utilizar las siguientes características de Amazon EMR con la integración del LDAP:

> **Note**
>
> Para mantener seguras las credenciales del LDAP, debe utilizar el cifrado en tránsito para proteger el flujo de datos dentro y fuera del clúster. Para obtener más información sobre el cifrado en tránsito, consulte Cifrado de datos en reposo y en tránsito.

- Cifrado en reposo y en tránsito (obligatorio)
- Grupos de instancias, flotas de instancias e instancias de spot
- Reconfiguración de aplicaciones en un clúster en ejecución
- Cifrado del servidor (SSE) de EMRFS

Características de no admitidas

Tenga en cuenta las siguientes limitaciones cuando utilice la integración del LDAP de Amazon EMR:

- Amazon EMR deshabilita los pasos para los clústeres con el LDAP habilitado.
- Amazon EMR no admite roles en tiempo de ejecución ni integraciones de AWS Lake Formation en los clústeres con el LDAP habilitado.
- Amazon EMR no admite el LDAP con StartTLS.
- Amazon EMR no admite el modo de alta disponibilidad (clústeres con varios nodos principales) en los clústeres con el LDAP habilitado.
- No puede vincular de forma rotativa las credenciales ni los certificados de los clústeres con el LDAP habilitado. Si alguno de esos campos se ha rotado, le recomendamos que inicie un clúster nuevo con las credenciales o certificados de vinculación actualizados.
- Debe utilizar bases de búsqueda exactas con LDAP. La base de búsqueda de usuarios y grupos de LDAP no admite los filtros de búsqueda de LDAP.

Configuración y lanzamiento de un clúster de EMR con el LDAP

En esta sección, se explica cómo configurar Amazon EMR para su uso con la autenticación LDAP.

Temas

- Agregar permisos de AWS Secrets Manager al rol de instancia de Amazon EMR
- Creación de la configuración de seguridad de Amazon EMR para la integración con el LDAP
- Lanzamiento de un clúster de EMR que se autentique con LDAP

Agregar permisos de AWS Secrets Manager al rol de instancia de Amazon EMR

Amazon EMR utiliza un rol de servicio de IAM para realizar acciones en su nombre para aprovisionar y administrar clústeres. El rol de servicio para instancias de EC2 de clúster, también conocido como el perfil de instancia de EC2 para Amazon EMR, es un tipo especial de rol de servicio que Amazon EMR asigna a cada instancia de EC2 de un clúster en el momento del lanzamiento.

Para definir los permisos para que un clúster de EMR interactúe con los datos de Amazon S3 y otros productos de AWS, defina un perfil de instancia de Amazon EC2 personalizado en lugar de `EMR_EC2_DefaultRole` al lanzar el clúster. Para obtener más información, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2) y Personalización de roles de IAM.

Agregue las siguientes instrucciones al perfil de instancia de EC2 predeterminado para permitir que Amazon EMR etiquete las sesiones y acceda a la instancia de AWS Secrets Manager que almacena los certificados LDAP.

```
    {
      "Sid": "AllowAssumeOfRolesAndTagging",
      "Effect": "Allow",
      "Action": ["sts:TagSession", "sts:AssumeRole"],
      "Resource": [
        "arn:aws:iam::111122223333:role/LDAP_DATA_ACCESS_ROLE_NAME",
        "arn:aws:iam::111122223333:role/LDAP_USER_ACCESS_ROLE_NAME"
      ]
    },
    {
        "Sid": "AllowSecretsRetrieval",
        "Effect": "Allow",
        "Action": "secretsmanager:GetSecretValue",
        "Resource": [
```

```
            "arn:aws:secretsmanager:us-east-1:111122223333:secret:LDAP_SECRET_NAME*",
            "arn:aws:secretsmanager:us-
east-1:111122223333:secret:ADMIN_LDAP_SECRET_NAME*"
        ]
    }
```

> **Note**
>
> Las solicitudes de clúster fallarán si olvida agregar el carácter comodín * al final del nombre del secreto al configurar los permisos de Secrets Manager. El comodín representa las versiones secretas.
> También debe limitar el ámbito de aplicación de la política AWS Secrets Manager a únicamente los certificados que su clúster necesita para aprovisionar instancias.

Creación de la configuración de seguridad de Amazon EMR para la integración con el LDAP

Antes de lanzar un clúster de EMR con integración del LDAP, siga los pasos de Creación de una configuración de seguridad para crear una configuración de seguridad de Amazon EMR para el clúster. Complete las siguientes configuraciones en el bloque `LDAPConfiguration` en `AuthenticationConfiguration` o en los campos correspondientes de la sección Configuraciones de seguridad de la consola de Amazon EMR:

`EnableLDAPAuthentication`

Opción de consola: Protocolo de autenticación: LDAP

Para usar la integración del LDAP, establezca esta opción en `true` o selecciónela como protocolo de autenticación al crear un clúster en la consola. De forma predeterminada, `EnableLDAPAuthentication` es `true` cuando crea una configuración de seguridad en la consola de Amazon EMR.

`LDAPServerURL`

Opción de consola: ubicación del servidor LDAP

La ubicación del servidor LDAP, incluido el prefijo: `ldaps://`*`location_of_server`*.

`BindCertificateARN`

Opción de consola: Certificado SSL del LDAP

El ARN de AWS Secrets Manager que contiene el certificado para firmar el certificado SSL que utiliza el servidor LDAP. Si su servidor LDAP está firmado por una autoridad de certificación (CA) pública, puede proporcionar un ARN de AWS Secrets Manager con un archivo en blanco. Para obtener más información sobre cómo almacenar el certificado en Secrets Manager, consulte Almacenamiento de certificados TLS en AWS Secrets Manager.

`BindCredentialsARN`

Opción de consola: credenciales de vinculación del servidor LDAP

Un ARN de AWS Secrets Manager que contiene las credenciales de vinculación de usuario administrador del LDAP. Las credenciales se almacenan como un objeto JSON. Solo hay un par clave-valor en este secreto; la clave del par es el nombre de usuario y el valor es la contraseña. Por ejemplo, `{"uid=admin,cn=People,dc=example,dc=com": "AdminPassword1"}`. Este campo es opcional, a menos que habilite el inicio de sesión SSH para su clúster de EMR. En muchas configuraciones, las instancias de Active Directory requieren credenciales de vinculación para permitir que SSSD sincronice los usuarios.

`LDAPAccessFilter`

Opción de consola: filtro de acceso del LDAP

Especifica el subconjunto de objetos del servidor LDAP que pueden autenticarse. Por ejemplo, si desea conceder acceso a todos los usuarios con la clase de objeto `posixAccount` de su servidor LDAP, defina el filtro de acceso como `(objectClass=posixAccount)`.

`LDAPUserSearchBase`

Opción de consola: base de búsqueda de usuarios de LDAP

La base de búsqueda a la que pertenecen sus usuarios en su servidor LDAP. Por ejemplo, `cn=People,dc=example,dc=com`.

`LDAPGroupSearchBase`

Opción de consola: base de búsqueda de grupos LDAP

La base de búsqueda a la que pertenecen sus grupos dentro de su servidor LDAP. Por ejemplo, `cn=Groups,dc=example,dc=com`.

`EnableSSHLogin`

Opción de consola: inicio de sesión SSH

Especifica si se permite o no la autenticación por contraseña con credenciales del LDAP. No le recomendamos habilitar esta opción. Los pares de claves son una ruta más segura para permitir el acceso a los clústeres de EMR. Este campo es opcional y de forma predeterminada es `false`.

`LDAPServerType`

Opción de consola: tipo de servidor LDAP

Especifica el tipo de servidor LDAP al que se conecta Amazon EMR. Las opciones compatibles son Active Directory y OpenLDAP. Es posible que otros tipos de servidores LDAP funcionen, pero Amazon EMR no admite oficialmente otros tipos de servidores. Para obtener más información, consulte Componentes del LDAP para Amazon EMR.

`ActiveDirectoryConfigurations`

Un subbloque obligatorio para las configuraciones de seguridad que utilizan el tipo de servidor Active Directory.

`ADDomain`

Opción de consola: dominio de Active Directory

El nombre de dominio utilizado para crear el nombre principal de usuario (UPN) para la autenticación del usuario con configuraciones de seguridad que utilizan el tipo de servidor Active Directory.

Aspectos que hay que tener en cuenta para realizar las configuraciones de seguridad con el LDAP y Amazon EMR

- Para crear una configuración de seguridad con la integración del LDAP de Amazon EMR, debe utilizar el cifrado en tránsito. Para obtener información sobre el cifrado en tránsito, consulte Cifrado de datos en reposo y en tránsito.
- No puede definir la configuración de Kerberos en la misma configuración de seguridad. Amazon EMR aprovisiona un KDC dedicado al KDC de forma automática y administra la contraseña de administrador de este KDC. Los usuarios no pueden acceder a esta contraseña de administrador.
- No puede definir los roles en tiempo de ejecución de IAM ni AWS Lake Formation en la misma configuración de seguridad.
- La `LDAPServerURL` debe tener el protocolo `ldaps://` en su valor.
- El `LDAPAccessFilter` no puede estar vacío.

Uso del LDAP con la integración de Apache Ranger para Amazon EMR

Con la integración del LDAP para Amazon EMR, puede ampliar su integración con Apache Ranger. Cuando extraiga sus usuarios del LDAP en Ranger, podrá asociarlos a un servidor de políticas de Apache Ranger para integrarlos con Amazon EMR y otras aplicaciones. Para ello, defina el campo `RangerConfiguration` de la configuración de seguridad que `AuthorizationConfiguration` va a utilizar con el clúster del LDAP. Para obtener más información acerca de cómo definir la configuración de seguridad, consulte Creación de la configuración de seguridad de EMR.

Cuando utilice el LDAP con Amazon EMR, no tendrá que proporcionar una `KerberosConfiguration` con la integración de Amazon EMR para Apache Ranger.

Lanzamiento de un clúster de EMR que se autentique con LDAP

Siga estos pasos para lanzar un clúster de EMR con el LDAP o Active Directory:

1. Configure su entorno:
   - Asegúrese de que los nodos del clúster de EMR puedan comunicarse con Amazon S3 y AWS Secrets Manager. Para obtener más información sobre cómo modificar el rol de su perfil de instancia de EC2 para comunicarse con estos servicios, consulte Agregar permisos de AWS Secrets Manager al rol de instancia de Amazon EMR.
   - Si planea ejecutar su clúster de EMR en una subred privada, debe usar AWS PrivateLink y puntos de conexión de Amazon VPC o usar la traducción de direcciones de red (NAT) para configurar la VPC de manera que se comunique con S3 y Secrets Manager. Para obtener más información, consulte AWS PrivateLink y puntos de conexión de VPC e Instancias NAT en la Guía de introducción a Amazon VPC.
   - Asegúrese de que haya conectividad de red entre el clúster de EMR y el servidor LDAP. Los clústeres de EMR deben acceder a su servidor LDAP a través de la red. Los nodos principal, central y de tareas del clúster se comunican con el servidor LDAP para sincronizar los datos de los usuarios. Si su servidor LDAP se ejecuta en Amazon EC2, actualice el grupo de seguridad de EC2 para que acepte el tráfico del clúster de EMR. Para obtener más información, consulte Agregar permisos de AWS Secrets Manager al rol de instancia de Amazon EMR.
2. Cree una configuración de seguridad de Amazon EMR para la integración del LDAP. Para obtener más información, consulte Creación de la configuración de seguridad de Amazon EMR para la integración con el LDAP.

3. Una vez completada la configuración, siga los pasos de Lanzar un clúster de Amazon EMR para lanzar el clúster con las siguientes configuraciones:

   - Seleccione Amazon EMR versión 6.12 o superior. Le recomendamos utilizar la última versión de Amazon EMR.
   - Especifique o seleccione únicamente las aplicaciones de su clúster que admitan el LDAP. Para obtener una lista de las aplicaciones que admitan el LDAP con Amazon EMR, consulte Compatibilidad de aplicaciones y consideraciones que hay que tener en cuenta sobre el LDAP para Amazon EMR.
   - Aplique la configuración de seguridad que creó en el paso anterior.

Ejemplos de uso del LDAP con Amazon EMR

Una vez que aprovisione un clúster de EMR que utilice la integración del LDAP, podrá proporcionar sus credenciales de LDAP a cualquier aplicación compatible mediante su mecanismo de autenticación de nombre de usuario y contraseña integrado. Esta página muestra algunos ejemplos.

Uso de la autenticación LDAP con Apache Hive

Example : Apache Hive

El siguiente comando de ejemplo inicia una sesión de Apache Hive a través de HiveServer2 y Beeline:

```
beeline -u "jdbc:hive2://$HOSTNAME:10000/default;ssl=true;sslTrustStore=
$TRUSTSTORE_PATH;trustStorePassword=$TRUSTSTORE_PASS"  -n LDAP_USERNAME -
p LDAP_PASSWORD
```

Uso de la autenticación LDAP con Apache Livy

Example : Apache Livy

El siguiente comando de ejemplo inicia una sesión de Livy a través de cURL. Sustituya *ENCODED-KEYPAIR* por una cadena cifrada en Base64 para `username:password`.

```
curl -X POST --data '{"proxyUser":"LDAP_USERNAME","kind": "pyspark"}' -H "Content-Type:
 application/json" -H "Authorization: Basic ENCODED-KEYPAIR" DNS_OF_PRIMARY_NODE:8998/
sessions
```

Uso de la autenticación LDAP con Presto

Example : Presto

El siguiente comando de ejemplo inicia una sesión de Presto a través de la CLI de Presto:

```
presto-cli --user "LDAP_USERNAME" --password --catalog hive
```

Tras ejecutar este comando, ingrese la contraseña de LDAP cuando se solicite.

Uso de la autenticación LDAP con Trino

Example : Trino

El siguiente comando de ejemplo inicia una sesión de Trino a través de la CLI de Trino:

```
trino-cli --user "LDAP_USERNAME" --password --catalog hive
```

Tras ejecutar este comando, ingrese la contraseña de LDAP cuando se solicite.

Uso de la autenticación LDAP con Hue

Puede acceder a la interfaz de usuario de Hue a través de un túnel SSH que cree en el clúster, o bien puede configurar un servidor proxy para que transmita públicamente la conexión a Hue. Como Hue no se ejecuta en modo HTTPS de forma predeterminada, le recomendamos que utilice una capa de cifrado adicional para garantizar que la comunicación entre los clientes y la interfaz de usuario de Hue esté cifrada con HTTPS. Esto reduce la posibilidad de que exponga accidentalmente las credenciales de usuario en texto sin formato.

Para usar la interfaz de usuario de Hue, abra la interfaz de usuario de Hue en su navegador e ingrese la contraseña y su nombre de usuario del LDAP para iniciar sesión. Si las credenciales son correctas, Hue inicia sesión y usa su identidad para que pueda autenticarse en todas las aplicaciones compatibles.

Uso de SSH para la autenticación por contraseña y tickets de Kerberos para otras aplicaciones

> **Important**
>
> No le recomendamos autenticarse por contraseña para conectarse vía SSH a un clúster de EMR.

Puede usar sus credenciales del LDAP para conectarse vía SSH a un clúster de EMR. Para ello, establezca la configuración `EnableSSHLogin` en `true` en la configuración de seguridad de Amazon EMR que utilice para iniciar el clúster. Cuando se lance el clúster, use el siguiente comando para conectarse vía SSH en el clúster:

```
ssh username@EMR_PRIMARY_DNS_NAME
```

Tras ejecutar este comando, ingrese la contraseña de LDAP cuando se solicite.

Amazon EMR incluye un script en el clúster que permite a los usuarios generar un archivo keytab y un ticket de Kerberos para usarlos con aplicaciones compatibles que no aceptan credenciales del LDAP directamente. Algunas de estas aplicaciones incluyen `spark-submit` Spark SQL y PySpark.

Ejecute `ldap-kinit` y siga las instrucciones. Si la autenticación se realiza correctamente, el archivo keytab de Kerberos aparecerá en su directorio principal con un ticket de Kerberos válido. Utilice el ticket de Kerberos para ejecutar aplicaciones como lo haría en cualquier entorno kerberizado.

Integración de Amazon EMR con AWS Lake Formation

AWS Lake Formation es un servicio administrado que le permite detectar, catalogar, limpiar y proteger los datos de un lago de datos de Amazon Simple Storage Service (S3). Lake Formation proporciona acceso detallado a las columnas de las bases de datos y tablas de Catálogo de datos de AWS Glue. Para obtener más información, consulte ¿Qué es AWS Lake Formation?

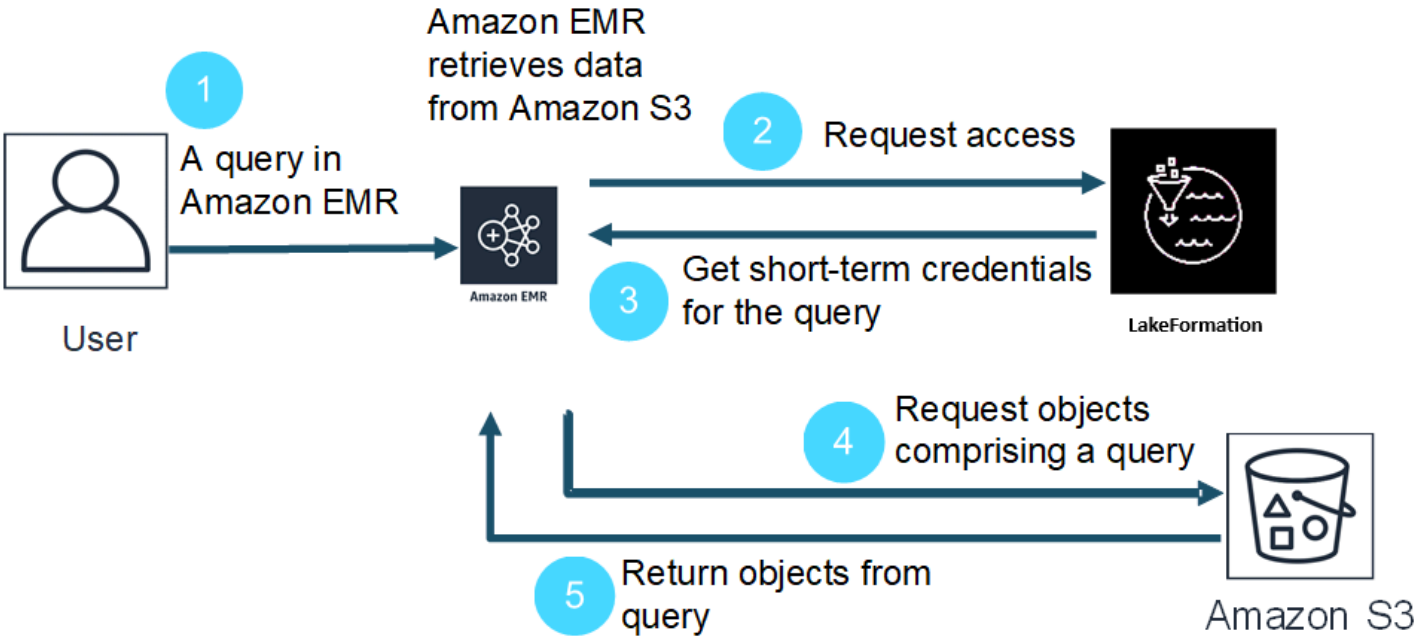
Con las versiones 6.7.0 y posteriores de Amazon EMR, puede aplicar un control de acceso basado en Lake Formation a los trabajos de Spark, Hive y Presto que envíe a los clústeres de Amazon EMR. Para integrar con Lake Formation, debe crear un clúster de EMR con un rol en tiempo de ejecución. Un rol en tiempo de ejecución es un rol de AWS Identity and Access Management (IAM) que se asocia a los trabajos o consultas de Amazon EMR. A continuación, Amazon EMR utiliza este rol para acceder a los recursos de AWS. Para obtener más información, consulte Roles en tiempo de ejecución para los pasos de Amazon EMR.

Cómo funciona Amazon EMR con Lake Formation

Tras integrar Amazon EMR con Lake Formation, puede ejecutar consultas a los clústeres de Amazon EMR con la API de `Step` o con SageMaker Studio. Luego, Lake Formation proporciona acceso a los datos a través de credenciales temporales para Amazon EMR. Este proceso se

denomina “expedición de credenciales”. Para obtener más información, consulte ¿Qué es AWS Lake Formation?

A continuación, se ofrece una descripción general de alto nivel sobre cómo Amazon EMR obtiene acceso a los datos protegidos por las políticas de seguridad de Lake Formation.



1. Un usuario envía una solicitud de datos de Amazon EMR en Lake Formation.
2. Amazon EMR solicita credenciales temporales a Lake Formation para permitir que el usuario acceda a los datos.
3. Lake Formation devuelve credenciales temporales.
4. Amazon EMR envía la solicitud de consulta para obtener datos de Amazon S3.
5. Amazon EMR recibe los datos de Amazon S3, los filtra y devuelve los resultados en función de los permisos de usuario que el usuario definió en Lake Formation.

Para obtener más información sobre cómo agregar usuarios y grupos a las políticas de Lake Formation, consulte Concesión de permisos para el catálogo de datos.

Requisitos previos

Si desea integrar Amazon EMR y Lake Formation, debe cumplir los siguientes requisitos:

- Active la autorización de roles en tiempo de ejecución en el clúster de Amazon EMR.

- Utilice Catálogo de datos de AWS Glue como almacén de metadatos.
- Defina y administre los permisos en Lake Formation para acceder a las bases de datos, tablas y columnas en el Catálogo de datos de AWS Glue. Para obtener más información, consulte ¿Qué es AWS Lake Formation?

Temas

- Habilitación de Lake Formation con Amazon EMR
- Apache Hudi y Lake Formation
- Consideraciones

Habilitación de Lake Formation con Amazon EMR

En esta sección, explicamos cómo crear una configuración de seguridad y cómo configurar Lake Formation para que funcione con Amazon EMR. También explicamos cómo lanzar un clúster con la configuración de seguridad que creó para Lake Formation.

Paso 1: configuración de un rol en tiempo de ejecución para el clúster de EMR

Para usar un rol en tiempo de ejecución para el clúster de EMR, debe crear una configuración de seguridad. Con una configuración de seguridad, puede aplicar opciones consistentes de seguridad, autorización y autenticación en todos sus clústeres.

1. Cree un archivo denominado `lf-runtime-roles-sec-cfg.json` con la siguiente configuración de seguridad.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true,
            "ApplicationScopedIAMRoleConfiguration":{
                "PropagateSourceIdentity":true
            }
        },
        "LakeFormationConfiguration":{
            "AuthorizedSessionTagValue":"Amazon EMR"
        },
  "EncryptionConfiguration": {
            "EnableInTransitEncryption": true,
```

```
            "InTransitEncryptionConfiguration": {
            "TLSCertificateConfiguration": {<Certificate-configuration>}
        }
    }
}
```

2. A continuación, para asegurarse de que la etiqueta de sesión puede autorizar Lake Formation, establezca la propiedad `LakeFormationConfiguration/AuthorizedSessionTagValue` en `Amazon EMR`.
3. Use el siguiente comando para crear la configuración de seguridad de Amazon EMR.

```
aws emr create-security-configuration \
--name 'iamconfig-with-iam-lf' \
--security-configuration file://lf-runtime-roles-sec-cfg.json
```

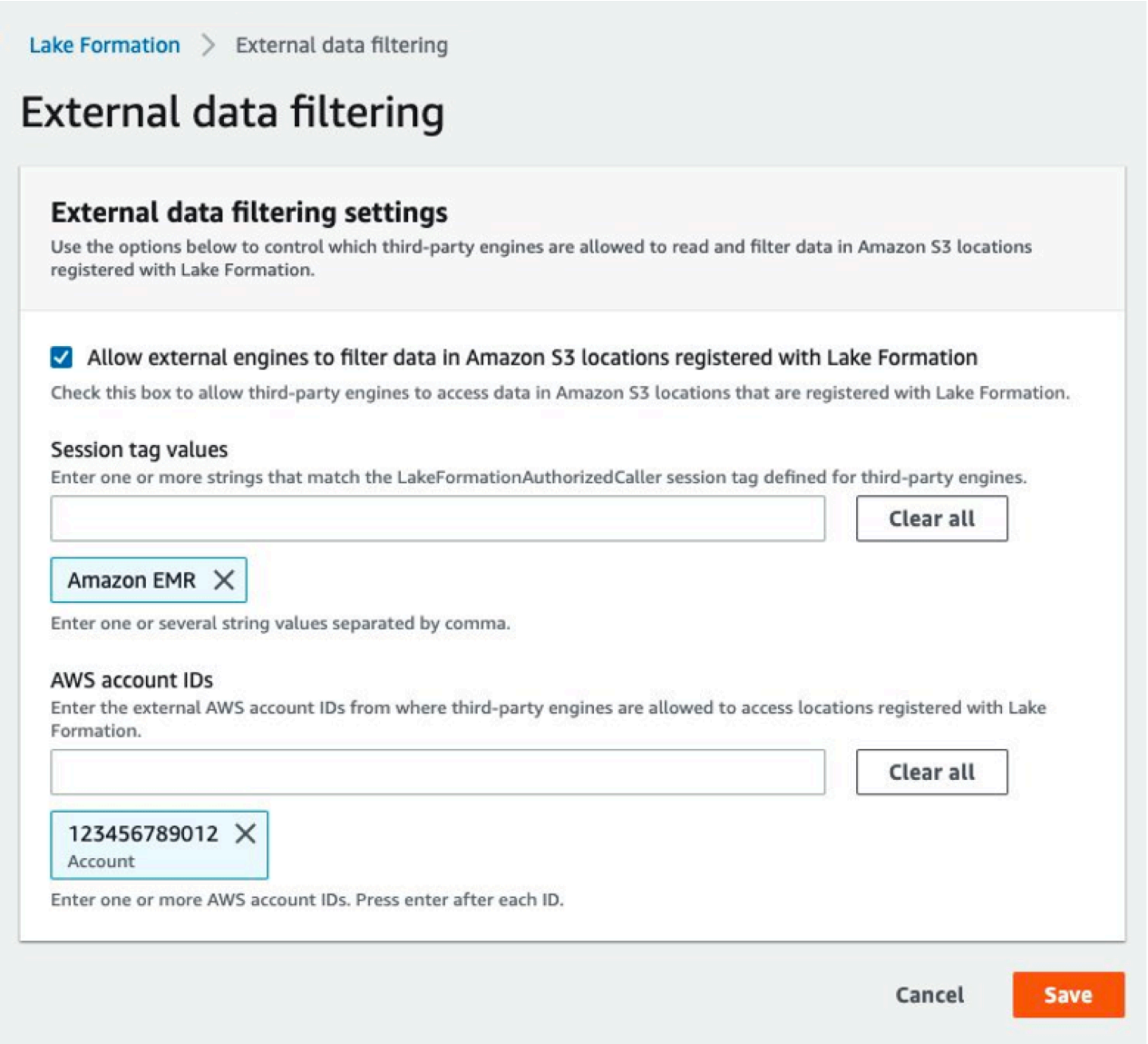
Como alternativa, puede utilizar la consola de Amazon EMR para crear una configuración de seguridad con ajustes personalizados.

Paso 2: lanzar un clúster de Amazon EMR

Ahora tiene todo listo para lanzar un clúster de EMR con la configuración de seguridad que creó en el paso anterior. Para obtener más información sobre las configuraciones de seguridad, consulte Uso de configuraciones de seguridad para configurar la seguridad del clúster y Roles en tiempo de ejecución para los pasos de Amazon EMR.

Paso 4: configuración del control de acceso basado en Lake Formation con los roles en tiempo de ejecución de Amazon EMR

Para aplicar permisos de tabla y columna con Lake Formation, el administrador del lago de datos de Lake Formation debe establecer `Amazon EMR` como valor para la configuración de la etiqueta de sesión, `AuthorizedSessionTagValue`. Lake Formation usa esta etiqueta de sesión para autorizar a los intermediarios y proporcionar acceso al lago de datos. Puede configurar esta etiqueta de sesión en la sección Filtrado de datos externos de la consola de Lake Formation. Sustituya *123456789012* por el ID de su Cuenta de AWS.



Para continuar con la configuración del control de acceso basado en Lake Formation con los roles en tiempo de ejecución de Amazon EMR, debe configurar las concesiones de AWS Glue y Lake Formation para los roles en tiempo de ejecución de Amazon EMR. Para permitir que sus roles en tiempo de ejecución de IAM interactúen con Lake Formation, debe concederles acceso con `lakeformation:GetDataAccess` y `glue:Get*`.

Los permisos de Lake Formation controlan el acceso a los recursos de Catálogo de datos de AWS Glue, a las ubicaciones de Amazon S3 y a los datos subyacentes en esas ubicaciones. Los permisos de IAM controlan el acceso a las API y los recursos de Lake Formation y AWS Glue. Aunque es

posible que tenga el permiso de Lake Formation para acceder a una tabla del catálogo de datos (SELECT), la operación fallará si no tiene el permiso de IAM en la API `glue:Get*`. Para obtener más información sobre el control de acceso de Lake Formation, consulte Información general sobre el control de acceso de Lake Formation.

1. Cree el archivo `emr-runtime-roles-lake-formation-policy.json` con el siguiente contenido.

```
{
    "Version": "2012-10-17",
    "Statement": {
        "Sid": "LakeFormationManagedAccess",
        "Effect": "Allow",
        "Action": [
            "lakeformation:GetDataAccess",
            "glue:Get*",
            "glue:Create*",
            "glue:Update*"
        ],
        "Resource": "*"
    }
}
```

2. Cree la política de IAM relacionada.

```
aws iam create-policy \
--policy-name emr-runtime-roles-lake-formation-policy \
--policy-document file://emr-runtime-roles-lake-formation-policy.json
```

3. Para asignar esta política a sus roles en tiempo de ejecución de IAM, siga los pasos que se indican en Administración de permisos de AWS Lake Formation.

Ahora puede usar roles en tiempo de ejecución y Lake Formation para aplicar permisos de tabla y columna. También puede usar una identidad de origen para controlar las acciones y supervisar las operaciones con AWS CloudTrail. Si desea obtener un ejemplo integral, consulte Introducing runtime roles for Amazon EMR steps.

Apache Hudi y Lake Formation

Las versiones 6.9.0 y posteriores de Amazon EMR incluyen compatibilidad limitada para el control de acceso basado en Lake Formation con Apache Hudi al leer datos con Spark SQL. Amazon EMR

admite consultas SELECT con Spark SQL y se limita al control de acceso de columna. Con esta característica, ahora puede hacer lo siguiente:

- Consultas de instantáneas en tablas de copia y escritura para consultar la última instantánea de la tabla en un instante de confirmación o compactación determinado.
- Consultas optimizadas para la lectura en tablas que se fusionan al leerse para consultar los datos compactados más recientes, que pueden no incluir las actualizaciones más recientes de los archivos de registro que aún no se han compactado.

La siguiente matriz de compatibilidad enumera algunas de las características principales de Apache Hudi con Lake Formation:

	Copiar al escribir	fusionar al leer
Consultas de instantáneas: Spark SQL	Y	N
Consultas automatizadas para la lectura: Spark SQL	No aplicable	Y
Consultas incrementales	N	N
Consultas de viaje en el tiempo	N	N
Consultas de orígenes de datos de Spark	N	N
Escrituras de orígenes de datos de Spark	N	N
DML/DDL	N	N
Tabla de metadatos	N	N

Consulta de tablas de Hudi

En esta sección, se muestra cómo ejecutar las consultas compatibles descritas anteriormente en un clúster habilitado para Lake Formation. La tabla debe ser una tabla de catálogo registrada.

En esta sección se muestra cómo ejecutar las consultas que se admiten en un clúster de Lake Formation, como se indicó anteriormente.

1. Para iniciar el intérprete de comandos de Spark, utilice los siguientes comandos.

```
spark-shell  --jars /usr/lib/hudi/hudi-spark-bundle.jar \
--conf 'spark.serializer=org.apache.spark.serializer.KryoSerializer'
```

```
spark-sql  --jars /usr/lib/hudi/hudi-spark-bundle.jar \
--conf 'spark.serializer=org.apache.spark.serializer.KryoSerializer'
```

2. Para consultar la última instantánea de las tablas que se copian al escribir, utilice los siguientes comandos.

```
select * from <my_hudi_cow_table>
```

```
spark.read.table("<my_hudi_cow_table>")
```

3. Para consultar los datos compactados más recientes de las tablas MOR, puede consultar la tabla optimizada para lectura que tiene el sufijo _ro:

```
SELECT * from <my_hudi_mor_table>_ro
```

```
spark.read.table("<my_hudi_mor_table>_ro")
```

> **Note**
>
> El rendimiento de las lecturas en los clústeres de Lake Formation puede ser más lento debido a las optimizaciones que no se admiten. Estas características incluyen la lista de archivos basada en los metadatos de Hudi y la omisión de datos. Le recomendamos que pruebe el rendimiento de su aplicación para asegurarse de que cumple con su SLA.

Consideraciones

Tenga en cuenta lo siguiente cuando use Amazon EMR con AWS Lake Formation.

- Los usuarios con acceso a una tabla pueden acceder a todas las propiedades de esa tabla. Si tiene un control de acceso basado en Lake Formation en una tabla, revísela para asegurarse de que las propiedades no contengan ningún dato o información confidencial.
- Los clústeres de Amazon EMR con Lake Formation no admiten el uso alternativo de HDFS cuando Spark recopile estadísticas de tablas. Por lo general, esto ayuda a optimizar el rendimiento de las consultas.
- Las operaciones que admiten controles de acceso basados en Lake Formation con tablas no gobernadas de Apache Spark y Apache Hive (de Amazon EMR 6.10.0 y versiones posteriores) incluyen `insert into` y `insert overwrite`.
- Las operaciones que admiten los controles de acceso basados en Lake Formation con Apache Spark y Apache Hive incluyen `select`, `describe`,`show database`, `show table`, `show column` y `show partition`.
- Amazon EMR no admite el acceso de control a las siguientes operaciones basadas en Lake Formation:
  - Escribe en tablas gobernadas
  - El filtro de datos de Lake Formation
  - Instrucciones DDL, como CREATE o ALTER TABLE
- Existen diferencias de rendimiento entre la misma consulta con y sin control de acceso basado en Lake Formation.

Integración de Amazon EMR con Apache Ranger

A partir de Amazon EMR 5.32.0, puede iniciar un clúster que se integre de forma nativa con Apache Ranger. Apache Ranger es un marco de código abierto para habilitar, supervisar y administrar la seguridad integral de los datos en toda la plataforma Hadoop. Para obtener más información, consulte Apache Ranger. Con la integración nativa, puede utilizar su propio Apache Ranger para aplicar un control de acceso a los datos detallado en Amazon EMR.

Esta sección proporciona información general sobre la integración de Amazon EMR con Apache Ranger. También incluye los requisitos previos y los pasos necesarios para lanzar un clúster de Amazon EMR integrado con Apache Ranger.

La integración nativa de Amazon EMR con Apache Ranger ofrece los siguientes beneficios clave:

- Control de acceso detallado a las bases de datos y tablas del metaalmacén de Hive, que le permite definir políticas de filtrado de datos en bases de datos, tablas y columnas para las aplicaciones Apache Spark y Apache Hive. Las aplicaciones de Hive admiten el filtrado de filas y el enmascaramiento de datos.
- La posibilidad de utilizar sus políticas de Hive existentes directamente con las aplicaciones de Amazon EMR para Hive.
- Control de acceso a los datos de prefijos y objetos de Amazon S3, lo que le permite definir políticas de filtrado de datos para acceder a los datos de S3 mediante el sistema de archivos de EMR.
- La capacidad de utilizar Registros de CloudWatch para realizar auditorías centralizadas.
- Amazon EMR instala y administra los complementos de Apache Ranger en su nombre.

Apache Ranger

Apache Ranger es un marco para habilitar, supervisar y administrar la seguridad integral de los datos en toda la plataforma Hadoop.

Apache Ranger tiene las siguientes características:

- Administración de seguridad centralizada para administrar todas las tareas relacionadas con la seguridad en una interfaz de usuario central o mediante API de REST.
- Autorización detallada para realizar una acción u operación específica con un componente o herramienta de Hadoop, administrada a través de una herramienta de administración central.
- Un método de autorización estandarizado para todos los componentes de Hadoop.
- Compatibilidad mejorada para varios métodos de autorización.
- Auditoría centralizada del acceso de los usuarios y de las acciones administrativas (relacionadas con la seguridad) en todos los componentes de Hadoop.

Apache Ranger utiliza dos componentes clave para la autorización:

- Servidor de administración de políticas de Apache Ranger: este servidor le permite definir las políticas de autorización para las aplicaciones de Hadoop. Al llevar a cabo la integración con Amazon EMR, podrá definir y aplicar políticas para que Apache Spark y Hive accedan al metaalmacén de Hive, así como para acceder al Sistema de archivos de EMR (EMRFS) de datos

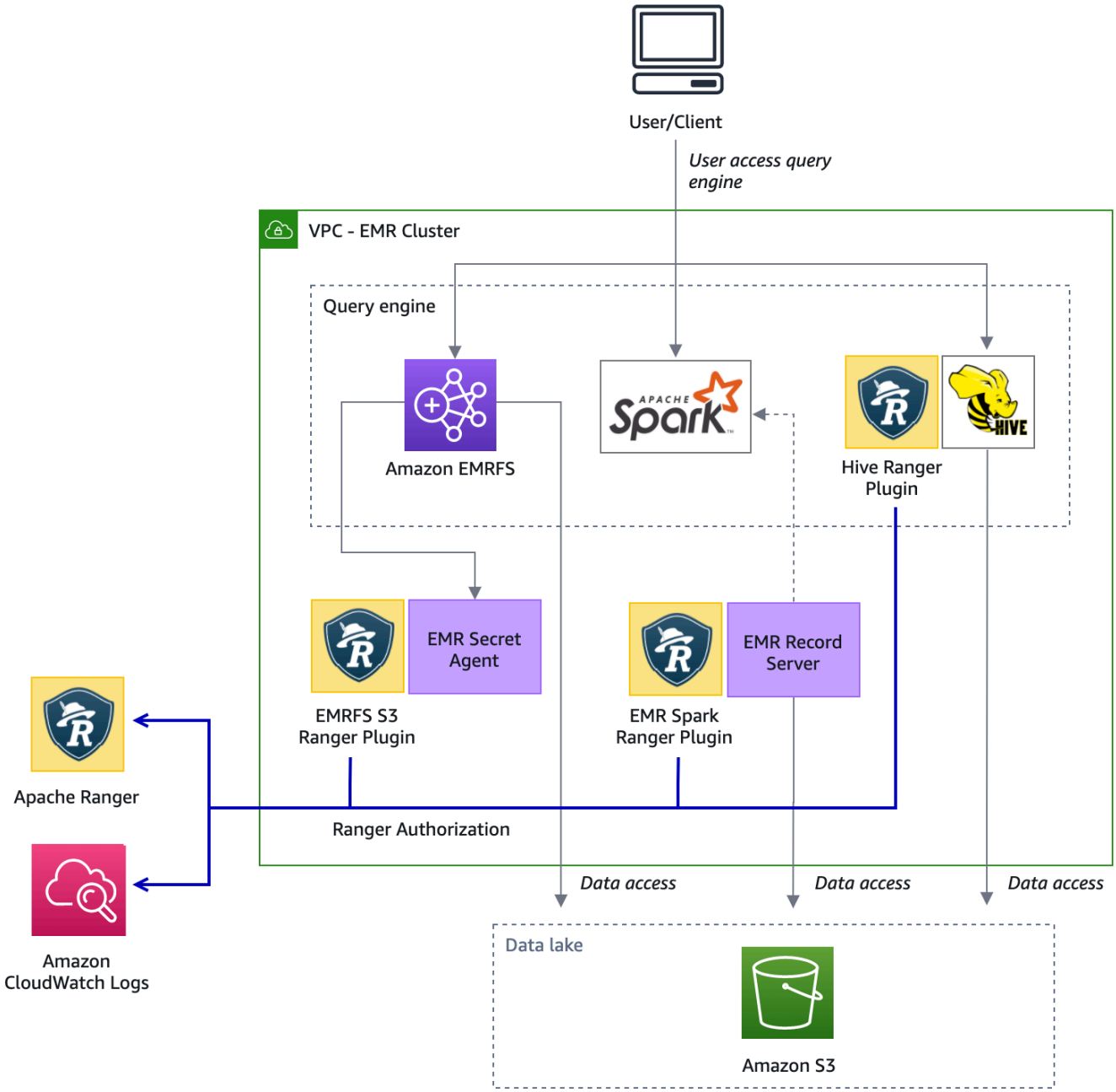
de Amazon S3. Puede configurar un servidor de administración de políticas de Apache Ranger nuevo o utilizar uno existente para integrarlo con Amazon EMR.

- Complemento de Apache Ranger: este complemento valida el acceso de un usuario según las políticas de autorización definidas en el servidor de administración de políticas de Apache Ranger. Amazon EMR instala y configura el complemento Apache Ranger automáticamente para cada aplicación de Hadoop seleccionada en la configuración de Apache Ranger.

Temas

- Arquitectura de la integración de Amazon EMR con Apache Ranger
- Componentes de Amazon EMR

Arquitectura de la integración de Amazon EMR con Apache Ranger



Componentes de Amazon EMR

Amazon EMR permite obtener un control integral del acceso con Apache Ranger mediante los siguientes componentes. Consulte el diagrama de arquitectura para ver una representación visual de estos componentes de Amazon EMR con los complementos de Apache Ranger.

Agente secreto: el agente secreto almacena secretos de forma segura y los distribuye a otros componentes o aplicaciones de Amazon EMR. Los secretos pueden incluir credenciales de usuario temporales, claves de cifrado o tickets de Kerberos. El agente secreto se ejecuta en todos los nodos del clúster e intercepta las llamadas al servicio de metadatos de la instancia. Para las solicitudes de credenciales del rol del perfil de instancia, el agente secreto vende las credenciales en función del usuario solicitante y de los recursos solicitados tras autorizar la solicitud con el complemento S3 Ranger de EMRFS. El agente secreto se ejecuta como el usuario *`emrsecretagent`* y escribe registros en el directorio /emr/secretagent/log. El proceso recurre a un conjunto específico de reglas `iptables` para funcionar. Es importante asegurarse de que `iptables` no esté deshabilitado. Si personaliza la configuración de `iptables`, las reglas de la tabla NAT deben conservarse y no modificarse.

Servidor de registros de EMR: el servidor de registros recibe las solicitudes de Spark para acceder a los datos. A continuación, autoriza las solicitudes reenviando los recursos solicitados al complemento Spark Ranger para Amazon EMR. El servidor de registros lee los datos de Amazon S3 y devuelve los datos filtrados a los que el usuario está autorizado a acceder según la política de Ranger. El servidor de registros se ejecuta en todos los nodos del clúster como usuario emr_record_server y escribe los registros en el directorio /var/log/emr-record-server.

Compatibilidad y limitaciones de la aplicación

Aplicaciones compatibles

La integración entre Amazon EMR y Apache Ranger, en la que EMR instala los complementos de Ranger, actualmente admite las siguientes aplicaciones:

- Apache Spark (disponible con EMR 5.32+ y EMR 6.3+)
- Apache Hive (disponible con EMR 5.32+ y EMR 6.3+)
- Acceso a S3 a través de EMRFS (disponible con EMR 5.32+ y EMR 6.3+)

Las siguientes aplicaciones se pueden instalar en un clúster de EMR y es posible que deban configurarse para satisfacer sus necesidades de seguridad:

- Apache Hadoop (disponible con EMR 5.32+ y EMR 6.3+, incluidos YARN y HDFS)
- Apache Livy (disponible con EMR 5.32+ y EMR 6.3+)
- Apache Zeppelin (disponible con EMR 5.32+ y EMR 6.3+)
- Apache Hue (disponible con EMR 5.32+ y EMR 6.3+)

- Ganglia (disponible con EMR 5.32+ y EMR 6.3+)
- HCatalog (disponible con EMR 5.32+ y EMR 6.3+)
- Mahout (disponible con EMR 5.32+ y EMR 6.3+)
- MXNet (disponible con EMR 5.32+ y EMR 6.3+)
- TensorFlow (disponible con EMR 5.32+ y EMR 6.3+)
- Tez (disponible con EMR 5.32+ y EMR 6.3+)
- Trino (disponible con EMR 6.7+)
- ZooKeeper (disponible con EMR 5.32+ y EMR 6.3+)

> **Important**
>
> Las aplicaciones enumeradas anteriormente son las únicas aplicaciones compatibles actualmente. Para garantizar la seguridad del clúster, se le permite crear un clúster de EMR solo con las aplicaciones de la lista anterior cuando Apache Ranger esté habilitado. Actualmente no se admiten otras aplicaciones. Para garantizar la seguridad de su clúster, este se rechazará si intenta instalar otras aplicaciones.

Características admitidas

Las siguientes características de Amazon EMR se pueden utilizar con Amazon EMR y Apache Ranger:

- Cifrado en reposo y en tránsito
- Autenticación de Kerberos (obligatoria)
- Grupos de instancias, flotas de instancias e instancias de spot
- Reconfiguración de aplicaciones en un clúster en ejecución
- Cifrado del servidor (SSE) de EMRFS

> **Note**
>
> La configuración de cifrado de Amazon EMR rige el SSE. Para obtener más información, consulte Opciones de cifrado.

Limitaciones de la aplicación

Hay varias limitaciones que se deben tener en cuenta al integrar Amazon EMR y Apache Ranger:

- Actualmente, no puede utilizar la consola para crear una configuración de seguridad que especifique la opción de integración con AWS Ranger en la AWS GovCloud (US) Region. La configuración de seguridad se puede llevar a cabo con la CLI.
- Kerberos tiene que estar instalado en el clúster.
- Las IU (interfaces de usuario) de las aplicaciones, como la IU del administrador de recursos de YARN, la IU de NameNode de HDFS y la IU de Livy, no están configuradas con autenticación de forma predeterminada.
- Los permisos predeterminados de HDFS `umask` están configurados de forma que los objetos creados tengan el valor `world wide readable` de forma predeterminada.
- Amazon EMR no admite el modo de alta disponibilidad (varias entidades principales) con Apache Ranger.
- Para ver otras limitaciones, consulte las limitaciones de cada aplicación.

> **Note**
>
> La configuración de cifrado de Amazon EMR rige el SSE. Para obtener más información, consulte Opciones de cifrado.

Limitaciones del complemento

Cada complemento tiene limitaciones específicas. Para ver las limitaciones del complemento Apache Hive, consulte Apache Hive plugin limitations. Para ver las limitaciones del complemento Apache Spark, consulte Apache Spark plugin limitations. Para ver las limitaciones del complemento EMRFS S3, consulte EMRFS S3 plugin limitations.

Configuración de Amazon EMR para Apache Ranger

Antes de instalar Apache Ranger, revise la información de esta sección para asegurarse de que Amazon EMR esté correctamente configurado.

Temas

- Configuración del servidor de Ranger Admin

- Roles de IAM para la integración nativa con Apache Ranger
- Creación de la configuración de seguridad de EMR
- Almacenamiento de certificados TLS en AWS Secrets Manager
- Lanzamiento de un clúster de EMR
- Configuración de Zeppelin para clústeres de Amazon EMR habilitados para Apache Ranger
- Problemas conocidos

Configuración del servidor de Ranger Admin

Para la integración de Amazon EMR, los complementos de la aplicación Apache Ranger deben comunicarse con el servidor de administración mediante TLS/SSL.

Requisito previo: habilitación del protocolo SSL del servidor de Ranger Admin

Apache Ranger en Amazon EMR requiere una comunicación SSL bidireccional entre los complementos y el servidor de Ranger Admin. Para garantizar que los complementos se comuniquen con el servidor de Apache Ranger a través de SSL, habilite el siguiente atributo en ranger-admin-site.xml en el servidor de Ranger Admin.

```
<property>
    <name>ranger.service.https.attrib.ssl.enabled</name>
    <value>true</value>
</property>
```

Además, se requieren las siguientes configuraciones:

```
<property>
    <name>ranger.https.attrib.keystore.file</name>
    <value>_<PATH_TO_KEYSTORE>_</value>
</property>

<property>
    <name>ranger.service.https.attrib.keystore.file</name>
    <value>_<PATH_TO_KEYSTORE>_</value>
</property>

<property>
    <name>ranger.service.https.attrib.keystore.pass</name>
    <value>_<KEYSTORE_PASSWORD>_</value>
```

```
</property>

<property>
    <name>ranger.service.https.attrib.keystore.keyalias</name>
    <value><PRIVATE_CERTIFICATE_KEY_ALIAS></value>
</property>

<property>
    <name>ranger.service.https.attrib.clientAuth</name>
    <value>want</value>
</property>

<property>
    <name>ranger.service.https.port</name>
    <value>6182</value>
</property>
```

Certificados TLS

La integración de Apache Ranger con Amazon EMR requiere que el tráfico desde los nodos de Amazon EMR al servidor de Ranger Admin esté cifrado mediante TLS y que los complementos de Ranger se autentiquen en el servidor de Apache Ranger mediante una autenticación TLS mutua bidireccional. El servicio Amazon EMR necesita el certificado público de su servidor de Ranger Admin (especificado en el ejemplo anterior) y el certificado privado.

Certificados del complemento Apache Ranger

El servidor de Apache Ranger Admin debe poder acceder a los certificados TLS públicos del complemento Apache Ranger para validarlos cuando se conecten los complementos. Hay tres métodos diferentes para hacerlo.

Método 1: configuración de un almacén de confianza en el servidor de Apache Ranger Admin

Complete las siguientes configuraciones en el archivo ranger-admin-site.xml para configurar un almacén de confianza.

```
<property>
    <name>ranger.truststore.file</name>
    <value><LOCATION TO TRUSTSTORE></value>
</property>

<property>
    <name>ranger.truststore.password</name>
```

```
    <value><PASSWORD FOR TRUSTSTORE></value>
</property>
```

Método 2: carga del certificado en el almacén de confianza de certificados CA (cacerts) de Java

Si el servidor de Ranger Admin no especifica un almacén de confianza en sus opciones de JVM, puede colocar los certificados públicos del complemento en el almacén de cacerts predeterminado.

Método 3: creación de un almacén de confianza y configuración como parte de las opciones de JVM

En `{RANGER_HOME_DIRECTORY}/ews/ranger-admin-services.sh`, modifique `JAVA_OPTS` para que incluya "`-Djavax.net.ssl.trustStore=<TRUSTSTORE_LOCATION>`" y "`-Djavax.net.ssl.trustStorePassword=<TRUSTSTORE_PASSWORD>`". Por ejemplo, agregue la siguiente línea después de la variable JAVA_OPTS existente.

```
JAVA_OPTS=" ${JAVA_OPTS} -Djavax.net.ssl.trustStore=${RANGER_HOME}/truststore/
truststore.jck -Djavax.net.ssl.trustStorePassword=changeit"
```

> **Note**
>
> Esta especificación puede exponer la contraseña del almacén de confianza si algún usuario puede iniciar sesión en el servidor de Apache Ranger Admin y ver los procesos en ejecución, por ejemplo, cuando utiliza el comando `ps`.

Uso de certificados autofirmados

Los certificados autofirmados no se recomiendan como certificados. Los certificados autofirmados no se pueden revocar y los certificados autofirmados pueden no cumplir con los requisitos de seguridad internos.

Instalación de definición de servicio

El servidor de Ranger Admin utiliza una definición de servicio para describir los atributos de las políticas de una aplicación. Luego, las políticas se almacenan en un repositorio de políticas para que los clientes las descarguen.

Para poder configurar las definiciones de servicio, las llamadas REST se deben realizar al servidor de Ranger Admin. Consulte Apache Ranger PublicAPISv2 para ver las API necesarias en la siguiente sección.

Instalación de la definición de servicio de Apache Spark

Para instalar la definición de servicio de Apache Spark, consulte Complemento Apache Spark.

Instalación de la definición de servicio de EMRFS

Para instalar la definición de servicio de S3 para Amazon EMR, consulte Complemento EMRFS S3.

Uso de la definición de servicio de Hive

Apache Hive puede usar la definición de servicio de Ranger existente que se incluye con Apache Ranger 2.0 y versiones posteriores. Para obtener más información, consulte Complemento Apache Hive.

Reglas de tráfico de red

Cuando Apache Ranger se integra con su clúster de EMR, el clúster debe comunicarse con servidores adicionales y AWS.

Todos los nodos de Amazon EMR, incluidos los nodos principales y de tareas, deben poder comunicarse con los servidores de Apache Ranger Admin para descargar las políticas. Si su servidor de Apache Ranger Admin se ejecuta en Amazon EC2, debe actualizar el grupo de seguridad para poder captar el tráfico del clúster de EMR.

Además de comunicarse con el servidor de Ranger Admin, todos los nodos deben poder comunicarse con los siguientes productos de AWS.

- Amazon S3
- AWS KMS (si usa EMRFS SSE-KMS)
- Amazon CloudWatch
- AWS STS

Si planea ejecutar su clúster de EMR en una subred privada, configure la VPC para que pueda comunicarse con estos servicios mediante AWS PrivateLink y puntos de conexión de VPC en la Guía del usuario de Amazon VPC o mediante una instancia de traducción de direcciones de red (NAT) de la Guía del usuario de Amazon VPC.

Roles de IAM para la integración nativa con Apache Ranger

La integración entre Amazon EMR y Apache Ranger se basa en tres roles clave que debe crear antes de lanzar el clúster:

- Un perfil de instancia de Amazon EC2 personalizado para Amazon EMR
- Un rol de IAM para los motores de Apache Ranger
- Un rol de IAM para otros productos de AWS

Esta sección proporciona información general sobre estos roles y las políticas que debe incluir en cada rol de IAM. Para obtener más información sobre la creación de estos roles, consulte Configuración del servidor de Ranger Admin.

Perfil de instancia EC2

Amazon EMR utiliza un rol de servicio de IAM para realizar acciones en su nombre para aprovisionar y administrar clústeres. El rol de servicio para instancias de EC2 de clúster, también conocido como el perfil de instancia de EC2 para Amazon EMR, es un tipo especial de rol de servicio asignado a cada instancia de EC2 de un clúster en el momento del lanzamiento.

Para definir los permisos para la interacción del clúster de EMR con los datos de Amazon S3 y con el almacén de metadatos de Hive protegido por Apache Ranger y otros productos de AWS, defina un perfil de instancia de EC2 personalizado para usarlo en lugar del `EMR_EC2_DefaultRole` al lanzar el clúster.

Para obtener más información, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2) y Personalización de roles de IAM.

Debe agregar las siguientes instrucciones al perfil de instancia de EC2 predeterminado de Amazon EMR para poder etiquetar las sesiones y acceder a la instancia de AWS Secrets Manager que almacena los certificados TLS.

```
    {
      "Sid": "AllowAssumeOfRolesAndTagging",
      "Effect": "Allow",
      "Action": ["sts:TagSession", "sts:AssumeRole"],
      "Resource": [
        "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<RANGER_ENGINE-
PLUGIN_DATA_ACCESS_ROLE_NAME>",
        "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<RANGER_USER_ACCESS_ROLE_NAME>"
      ]
    },
    {
        "Sid": "AllowSecretsRetrieval",
```

```
        "Effect": "Allow",
        "Action": "secretsmanager:GetSecretValue",
        "Resource": [

"arn:aws:secretsmanager:<REGION>:<AWS_ACCOUNT_ID>:secret:<PLUGIN_TLS_SECRET_NAME>*",

"arn:aws:secretsmanager:<REGION>:<AWS_ACCOUNT_ID>:secret:<ADMIN_RANGER_SERVER_TLS_SECRET_NAME>
        ]
    }
```

> **Note**
>
> Para los permisos de Secrets Manager, no olvide el comodín ("*") que aparece al final del nombre del secreto o sus solicitudes fallarán. El comodín es para las versiones de los secretos.

> **Note**
>
> Limite el ámbito de aplicación de la política AWS Secrets Manager a solo los certificados necesarios para el aprovisionamiento.

Rol de IAM para Apache Ranger

Este rol proporciona credenciales para que los motores de ejecución de confianza, como Apache Hive y Servidor de registros de Amazon EMR, puedan acceder a los datos de Amazon S3. Utilice únicamente este rol para acceder a los datos de Amazon S3, incluidas las claves de KMS, si utiliza S3 SSE-KMS.

Este rol debe crearse con la política mínima que se indica en el siguiente ejemplo.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "CloudwatchLogsPermissions",
      "Action": [
        "logs:CreateLogGroup",
        "logs:CreateLogStream",
```

```
        "logs:PutLogEvents"
      ],
      "Effect": "Allow",
      "Resource": [

 "arn:aws:logs:<REGION>:<AWS_ACCOUNT_ID>:<CLOUDWATCH_LOG_GROUP_NAME_IN_SECURITY_CONFIGURATION>:
      ]
    },
    {
      "Sid": "BucketPermissionsInS3Buckets",
      "Action": [
        "s3:CreateBucket",
        "s3:DeleteBucket",
        "s3:ListAllMyBuckets",
        "s3:ListBucket"
      ],
      "Effect": "Allow",
      "Resource": [
 *"arn:aws:s3:::bucket1",
        "arn:aws:s3:::bucket2"*
        ]
    },
    {
      "Sid": "ObjectPermissionsInS3Objects",
      "Action": [
        "s3:GetObject",
        "s3:DeleteObject",
        "s3:PutObject"
      ],
      "Effect": "Allow",
      "Resource": [
 *"arn:aws:s3:::bucket1/*",
          "arn:aws:s3:::bucket2/*"
*        ]
    }
  ]
}
```

> **Important**
>
> Se debe incluir el asterisco “*” al final del recurso de registro de CloudWatch para permitir escribir en los flujos de registro.

> **Note**
>
> Si utiliza la visión de consistencia de EMRFS o el cifrado S3-SSE, agregue permisos a las tablas de DynamoDB y a las claves de KMS para que los motores de ejecución puedan interactuar con esos motores.

El rol de IAM de Apache Ranger lo asume el rol de perfil de instancia de EC2. Utilice el siguiente ejemplo para crear una política de confianza que permita que el rol de perfil de instancia de EC2 asuma el rol de IAM de Apache Ranger.

```
   {
     "Sid": "",
     "Effect": "Allow",
     "Principal": {
       "AWS": "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<EC2 INSTANCE PROFILE ROLE NAME eg.
EMR_EC2_DefaultRole>"
     },
     "Action": ["sts:AssumeRole", "sts:TagSession"]
   }
```

El rol de IAM para otros productos de AWS

Este rol proporciona a los usuarios que no son motores de ejecución de confianza credenciales para interactuar con los productos de AWS, si es necesario. No utilice este rol de IAM para permitir el acceso a los datos de Amazon S3, a menos que se trate de datos a los que deberían poder acceder todos los usuarios.

Este rol lo asumirá el rol de perfil de instancia de EC2. Utilice el siguiente ejemplo para crear una política de confianza que permita que el rol de perfil de instancia de EC2 asuma el rol de IAM de Apache Ranger.

```
   {
     "Sid": "",
     "Effect": "Allow",
     "Principal": {
       "AWS": "arn:aws:iam::<AWS_ACCOUNT_ID>:role/<EC2 INSTANCE PROFILE ROLE NAME eg.
EMR_EC2_DefaultRole>"
     },
     "Action": ["sts:AssumeRole", "sts:TagSession"]
```

```
    }
```

Validación de permisos

Consulte Solución de problemas con Apache Ranger para obtener instrucciones sobre cómo validar los permisos.

Creación de la configuración de seguridad de EMR

Creación de una configuración de seguridad de Amazon EMR para Apache Ranger

Antes de lanzar un clúster de Amazon EMR integrado con Apache Ranger, cree una configuración de seguridad.

Console

Para crear una configuración de seguridad que especifique la opción de integración con AWS Ranger

1. En la consola de Amazon EMR, seleccione Configuraciones de seguridad y, a continuación, Crear.
2. Escriba un nombre para la configuración de seguridad en el campo Name (Nombre). Este nombre se utiliza para especificar la configuración de seguridad cuando crea un clúster.
3. En Integración con AWS Ranger, seleccione Habilitar un control de acceso detallado administrado por Apache Ranger.
4. Seleccione un rol de IAM para que Apache Ranger lo aplique. Para obtener más información, consulte Roles de IAM para la integración nativa con Apache Ranger.
5. Seleccione un rol de IAM para que otros servicios de AWS lo apliquen.
6. Configure los complementos para que se conecten al servidor de Ranger Admin ingresando el ARN de Secrets Manager del servidor de Admin y la dirección.
7. Seleccione las aplicaciones para configurar los complementos de Ranger. Complete el ARN de Secrets Manager que contiene el certificado TLS privado del complemento.

   Si no configura Apache Spark ni Apache Hive y los selecciona como aplicaciones para su clúster, la solicitud fallará.
8. Defina otras opciones de configuración de seguridad según corresponda y elija Create (Crear). Debe habilitar la autenticación de Kerberos mediante el KDC dedicado del clúster o un KDC externo.

> **Note**
>
> Actualmente, no puede utilizar la consola para crear una configuración de seguridad que especifique la opción de integración con AWS Ranger en la AWS GovCloud (US) Region. La configuración de seguridad se puede llevar a cabo con la CLI.

CLI

Para crear una configuración de seguridad para la integración de Apache Ranger

1. Sustituya *<ACCOUNT ID>* por su ID de cuenta de AWS.
2. Sustituya *<REGION>* por la región en la que se encuentra el recurso.
3. Especifique un valor para `TicketLifetimeInHours` para determinar el periodo para el que un vale de Kerberos generado por el KDC es válido.
4. Especifique la dirección del servidor de Ranger Admin de `AdminServerURL`.

```
{
    "AuthenticationConfiguration": {
        "KerberosConfiguration": {
            "Provider": "ClusterDedicatedKdc",
            "ClusterDedicatedKdcConfiguration": {
                "TicketLifetimeInHours": 24
            }
        }
    },
    "AuthorizationConfiguration":{
      "RangerConfiguration":{
         "AdminServerURL":"https://_<RANGER ADMIN SERVER IP>_:6182",
         "RoleForRangerPluginsARN":"arn:aws:iam::_<ACCOUNT ID>_:role/_<RANGER PLUGIN
 DATA ACCESS ROLE NAME>_",
         "RoleForOtherAWSServicesARN":"arn:aws:iam::_<ACCOUNT ID>_:role/_<USER
 ACCESS ROLE NAME>_",
         "AdminServerSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES ADMIN SERVERS PUBLIC TLS CERTIFICATE
 WITHOUT VERSION>_",
         "RangerPluginConfigurations":[
            {
               "App":"Spark",
```

```
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES SPARK PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
                "PolicyRepositoryName":"<SPARK SERVICE NAME eg. amazon-emr-spark>"
            },
            {
                "App":"Hive",
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES Hive PLUGIN PRIVATE TLS CERTIFICATE WITHOUT
 VERSION>_",
                "PolicyRepositoryName":"<HIVE SERVICE NAME eg. Hivedev>"
            },
            {
                "App":"EMRFS-S3",
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES EMRFS S3 PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
                "PolicyRepositoryName":"<EMRFS S3 SERVICE NAME eg amazon-emr-emrfs>"
            },
       {
                "App":"Trino",
                "ClientSecretARN":"arn:aws:secretsmanager:_<REGION>_:_<ACCOUNT
 ID>_:secret:_<SECRET NAME THAT PROVIDES TRINO PLUGIN PRIVATE TLS CERTIFICATE
 WITHOUT VERSION>_",
                "PolicyRepositoryName":"<TRINO SERVICE NAME eg amazon-emr-trino>"
            }
         ],
         "AuditConfiguration":{
            "Destinations":{
                "AmazonCloudWatchLogs":{
                   "CloudWatchLogGroup":"arn:aws:logs:<REGION>:_<ACCOUNT ID>_:log-
group:_<LOG GROUP NAME FOR AUDIT EVENTS>_"
                }
            }
         }
      }
   }
}
```

Los PolicyRepositoryNames son los nombres de servicio que se especifican en el servidor de Apache Ranger Admin.

Cree una configuración de seguridad de Amazon EMR con el siguiente contenido. Sustituya la configuración de seguridad por el nombre de su elección. Seleccione esta configuración por el nombre al crear el clúster.

```
aws emr create-security-configuration \
--security-configuration file://./security-configuration.json \
--name security-configuration
```

Configuración de características de seguridad adicionales

Para integrar Amazon EMR de forma segura con Apache Range, configure las siguientes características de seguridad de EMR:

- Habilite la autenticación de Kerberos mediante el KDC dedicado del clúster o un KDC externo. Para obtener instrucciones, consulte Uso de Kerberos para la autenticación con Amazon EMR.
- (Opcional) Habilite el cifrado en tránsito o en reposo. Para obtener más información, consulte Opciones de cifrado.

Para obtener más información, consulte Seguridad en Amazon EMR.

Almacenamiento de certificados TLS en AWS Secrets Manager

Los complementos de Ranger instalados en un clúster de Amazon EMR y el servidor de Ranger Admin deben comunicarse a través del protocolo TLS para garantizar que los datos de la política y otra información enviada no se puedan leer si se interceptan. EMR también exige que los complementos se autentiquen en el servidor de Ranger Admin proporcionando su propio certificado TLS y que realicen una autenticación TLS bidireccional. Esta configuración requería la creación de cuatro certificados: dos pares de certificados TLS públicos y privados. Para obtener instrucciones sobre cómo instalar el certificado en su servidor de Ranger Admin, consulte Configuración del servidor de Ranger Admin. Para completar la configuración, los complementos de Ranger instalados en el clúster de EMR necesitan dos certificados: el certificado TLS público de su servidor de administración y el certificado privado que el complemento utilizará para autenticarse en el servidor de Ranger Admin. Para proporcionar estos certificados TLS, estos deben estar en AWS Secrets Manager y, además, deben proporcionarse en una configuración de seguridad de EMR.

> **Note**
>
> Se recomienda encarecidamente, pero no es obligatorio, crear un par de certificados para cada una de sus aplicaciones a fin de limitar el impacto en caso de que uno de los certificados del complemento se vea comprometido.

> **Note**
>
> Debe realizar un seguimiento de los certificados y rotarlos antes de su fecha de vencimiento.

Formato del certificado

La importación de los certificados a AWS Secrets Manager es igual, independientemente de si se trata del certificado de complemento privado o del certificado de administrador de Ranger público. Antes de importar los certificados TLS, los certificados deben estar en formato PEM 509x.

Un ejemplo de certificado público tiene el siguiente formato:

```
-----BEGIN CERTIFICATE-----
...Certificate Body...
-----END CERTIFICATE-----
```

Un ejemplo de certificado privado tiene el siguiente formato:

```
-----BEGIN PRIVATE KEY-----
...Private Certificate Body...
-----END PRIVATE KEY-----
-----BEGIN CERTIFICATE-----
...Trust Certificate Body...
-----END CERTIFICATE-----
```

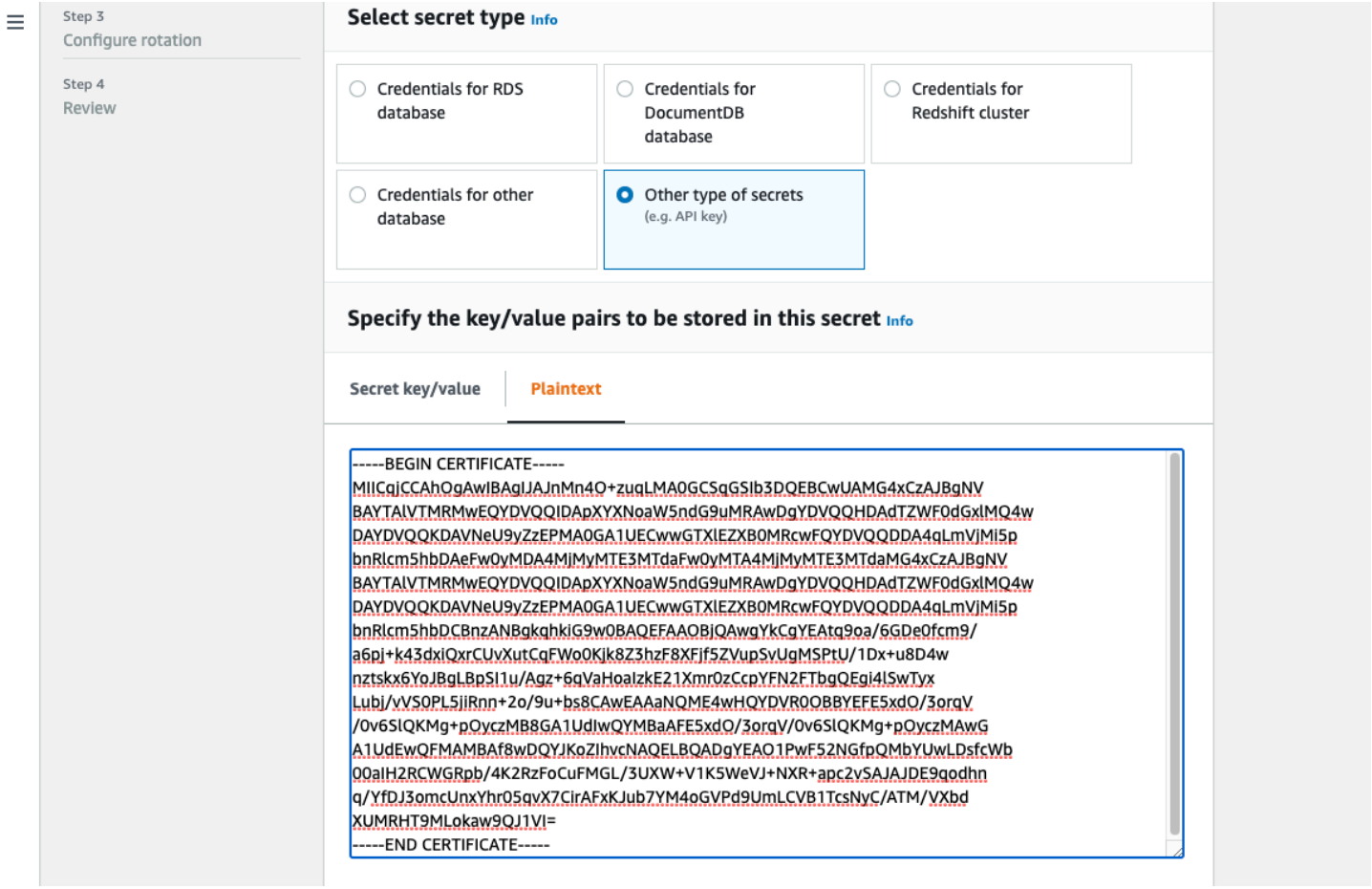
El certificado privado también debe contener un certificado de confianza.

Si desea obtener un certificado, debe ejecutar el siguiente comando:

```
openssl x509 -in <PEM FILE> -text
```

Importación de un certificado a AWS Secrets Manager

Al crear su secreto en Secrets Manager, seleccione Otro tipo de secretos en Tipo de secreto y pegue su certificado cifrado en PEM en el campo Texto no cifrado.



Lanzamiento de un clúster de EMR

Antes de lanzar un clúster de Amazon EMR con Apache Ranger, asegúrese de que cada componente cumpla los siguientes requisitos mínimos de versión:

- Amazon EMR 5.32.0 o posterior, o 6.3.0 o posterior. Le recomendamos que utilice la última versión de Amazon EMR.
- Servidor de Apache Ranger Admin 2.x.

Siga estos pasos:

- Instale Apache Ranger si aún no lo ha hecho. Para obtener más información acerca de la instalación, consulte Apache Ranger 0.5.0 installation.

- Asegúrese de que haya conectividad de red entre el clúster de Amazon EMR y el servidor de Apache Ranger Admin. Consulte Configuración del servidor de Ranger Admin.
- Cree los roles de IAM necesarios. Consulte Roles de IAM para la integración nativa con Apache Ranger.
- Cree una configuración de seguridad de EMR para la instalación de Apache Ranger. Para obtener más información, consulte Creación de la configuración de seguridad de EMR.

Configuración de Zeppelin para clústeres de Amazon EMR habilitados para Apache Ranger

El tema trata sobre cómo configurar Apache Zeppelin para un clúster de Amazon EMR compatible con Apache Ranger, de modo que pueda utilizar Zeppelin como cuaderno para navegar por los datos de forma interactiva. Zeppelin se incluye en las versiones de 5.0.0 de Amazon EMR y posteriores. Las versiones anteriores incluyen Zeppelin como una aplicación de entorno aislado. Para obtener más información, consulte Versiones de lanzamiento de Amazon EMR 4.x en la Guía de publicación de Amazon EMR.

De forma predeterminada, Zeppelin está configurado con una contraseña y un nombre de usuario predeterminados, lo que no es seguro en un entorno multiusuario.

Para configurar Zeppelin, siga los pasos que se describen a continuación:

1. Modifique el mecanismo de autenticación.

   Modifique el archivo `shiro.ini` para implementar el mecanismo de autenticación que prefiera. Zeppelin admite Active Directory, LDAP, PAM y Knox SSO. Consulte Apache Shiro authentication for Apache Zeppelin para obtener más información.
2. Configure Zeppelin para suplantar al usuario final

   Si permite que Zeppelin suplante al usuario final, los trabajos enviados por Zeppelin se pueden ejecutar como ese usuario final. Agregue la siguiente configuración a `core-site.xml`:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.zeppelin.hosts": "*",
      "hadoop.proxyuser.zeppelin.groups": "*"
    },
```

```
    "Configurations": [
    ]
  }
]
```

A continuación, agregue la siguiente configuración a `hadoop-kms-site.xml`, que se encuentra en `/etc/hadoop/conf`:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.proxyuser.zeppelin.hosts": "*",
      "hadoop.kms.proxyuser.zeppelin.groups": "*"
    },
    "Configurations": [
    ]
  }
]
```

También puede agregar estas configuraciones a su clúster de Amazon EMR mediante la consola siguiendo los pasos que se indican en Reconfiguración de un grupo de instancias en la consola.

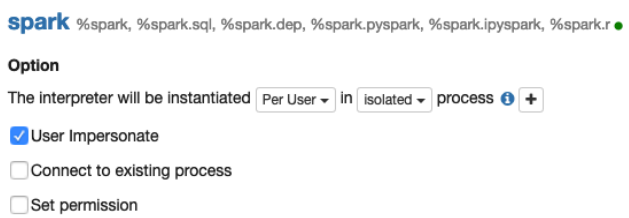
3. Permita que Zeppelin utilice el comando sudo como usuario final

   Cree un archivo `/etc/sudoers.d/90-zeppelin-user` que contenga lo siguiente:

```
zeppelin ALL=(ALL) NOPASSWD:ALL
```

4. Modifique la configuración de los intérpretes para ejecutar las tareas de los usuarios en sus propios procesos.

   Configure todos los intérpretes para que creen instancias de los intérpretes “por usuario” en procesos “aislados”.



5. Modifique **`zeppelin-env.sh`**

Agregue lo siguiente a `zeppelin-env.sh` para que Zeppelin comience a lanzar los intérpretes como usuario final:

```
ZEPPELIN_IMPERSONATE_USER=`echo ${ZEPPELIN_IMPERSONATE_USER} | cut -d @ -f1`
export ZEPPELIN_IMPERSONATE_CMD='sudo -H -u ${ZEPPELIN_IMPERSONATE_USER} bash -c'
```

Agregue lo siguiente a `zeppelin-env.sh` para cambiar los permisos predeterminados del cuaderno a solo lectura únicamente para el creador:

```
export ZEPPELIN_NOTEBOOK_PUBLIC="false"
```

Por último, agregue lo siguiente a `zeppelin-env.sh` para incluir la ruta de la clase RecordServer de EMR después de la primera instrucción `CLASSPATH`:

```
export CLASSPATH="$CLASSPATH:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-connector-common.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-spark-connector.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-
server-client.jar:/usr/share/aws/emr/record-server/lib/aws-emr-record-server-
common.jar:/usr/share/aws/emr/record-server/lib/jars/secret-agent-interface.jar"
```

6. Reinicie Zeppelin.

   Ejecute el siguiente comando para reiniciar Zeppelin:

   ```
   sudo systemctl restart zeppelin
   ```

Problemas conocidos

Problemas conocidos

Existe un problema conocido en la versión 5.32 de Amazon EMR en el que se modificaron los permisos de `hive-site.xml` para que solo los usuarios con privilegios pudieran leerlo, ya que es posible que haya credenciales almacenadas en él. Esto podría impedir que Hue leyera `hive-site.xml` y provocar que las páginas web se recargaran continuamente. Si experimenta este problema, agregue la siguiente configuración para solucionarlo:

```
[
  {
```

```
    "Classification": "hue-ini",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "desktop",
        "Properties": {
          "server_group":"hive_site_reader"
         },
        "Configurations":[
        ]
      }
    ]
  }
]
```

Existe un problema conocido que hace que el complemento EMRFS S3 para Apache Ranger no sea compatible actualmente con la característica Zona de seguridad de Apache Ranger. Las restricciones de control de acceso definidas mediante la característica Zona de seguridad no se aplican a los clústeres de Amazon EMR.

IU de las aplicaciones

De forma predeterminada, las IU de las aplicaciones no realizan la autenticación. Esto incluye la IU de ResourceManager, la IU de NodeManager y la IU de Livy, entre otras. Además, cualquier usuario que tenga la capacidad de acceder a las IU puede ver información sobre los trabajos de todos los demás usuarios.

Si no desea este comportamiento, debe asegurarse de que se utilice un grupo de seguridad para restringir el acceso de los usuarios a las IU de las aplicaciones.

Permisos predeterminados de HDFS

De forma predeterminada, los objetos que los usuarios crean en HDFS reciben permisos legibles en todo el mundo. Esto puede provocar que los usuarios que no deberían tener acceder a los datos puedan leerlos. Para cambiar este comportamiento de forma que los permisos de archivo predeterminados sean de lectura y escritura únicamente para el creador del trabajo, lleve a cabo estos pasos.

Al crear el clúster de EMR, proporcione la siguiente configuración:

```
[
  {
```

```
    "Classification": "hdfs-site",
    "Properties": {
      "dfs.namenode.acls.enabled": "true",
      "fs.permissions.umask-mode": "077",
      "dfs.permissions.superusergroup": "hdfsadmingroup"
    }
  }
]
```

Además, ejecute la siguiente acción de arranque:

```
--bootstrap-actions Name='HDFS UMask Setup',Path=s3://elasticmapreduce/hdfs/umask/
umask-main.sh
```

Complementos de Apache Ranger

Los complementos de Apache Ranger validan el acceso de un usuario según las políticas de autorización definidas en el servidor de administración de políticas de Apache Ranger.

Temas

- Complemento Apache Hive
- Complemento Apache Spark
- Complemento EMRFS S3
- Complemento Trino

Complemento Apache Hive

Apache Hive es un popular motor de ejecución dentro del ecosistema Hadoop. Amazon EMR proporciona un complemento de Apache Ranger para poder proporcionar controles de acceso detallados para Hive. El complemento es compatible con la versión 2.0 y posteriores del servidor de Apache Ranger Admin de código abierto.

Temas

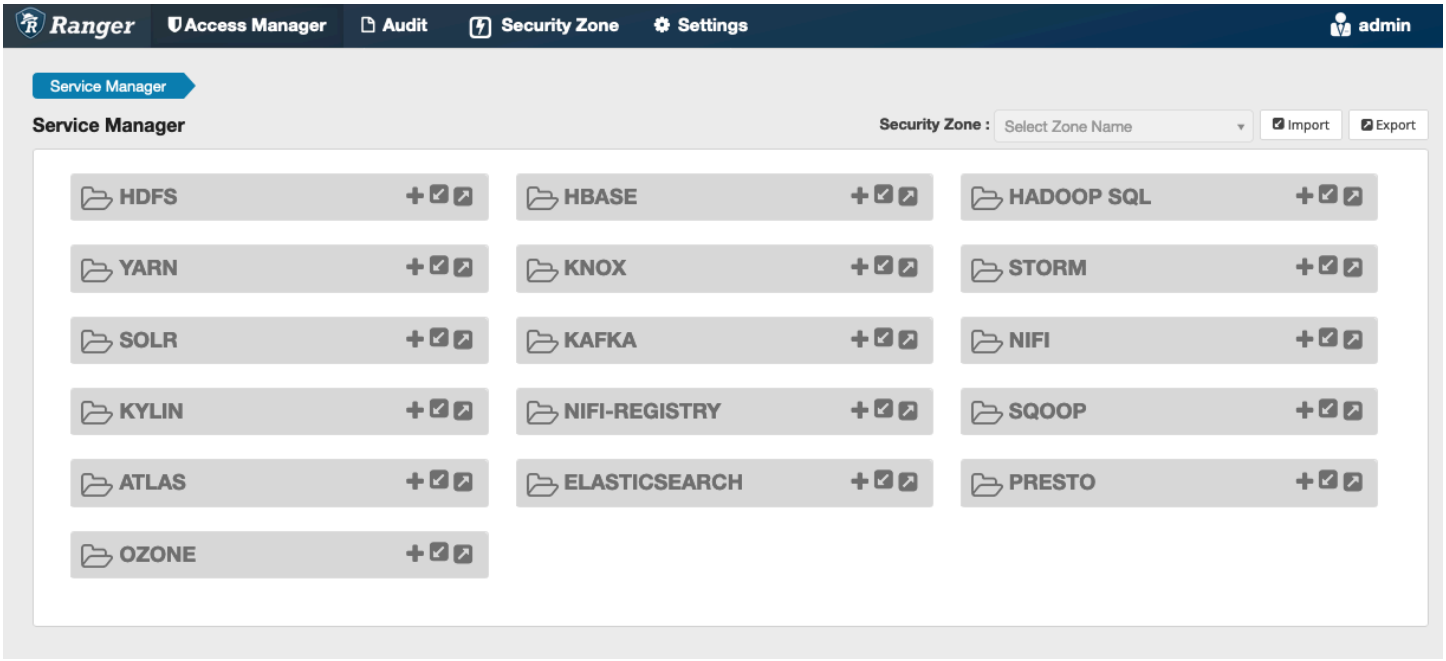
- Características admitidas
- Instalación de la configuración del servicio
- Consideraciones
- Limitaciones

Características admitidas

El complemento Apache Ranger para Hive en EMR admite todas las funciones del complemento de código abierto, que incluye controles de acceso a bases de datos, tablas y columnas, así como el filtrado de filas y el enmascaramiento de datos. Para ver una tabla de comandos de Hive y los permisos de Ranger asociados, consulte Hive commands to Ranger permission mapping.

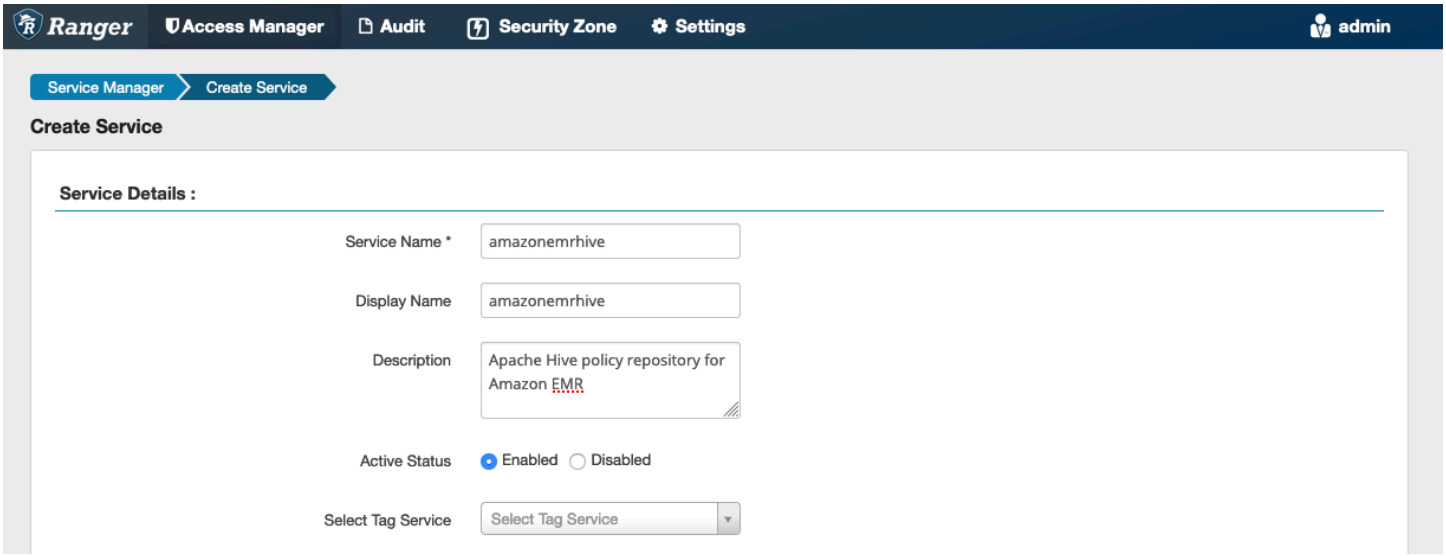
Instalación de la configuración del servicio

El complemento Apache Hive es compatible con la definición de servicio de Hive existente en Apache Hive Hadoop SQL.



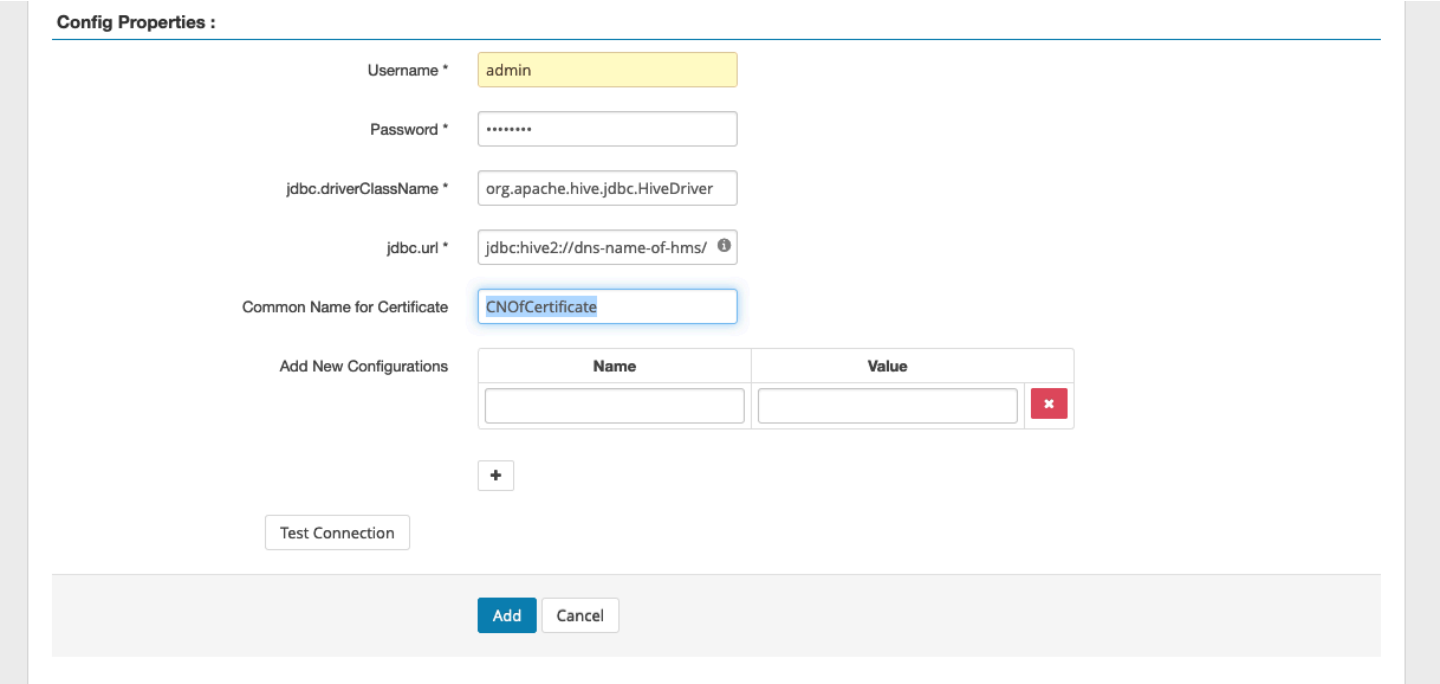
Si no tiene una instancia del servicio en Hadoop SQL, como se muestra arriba, puede crear una. Haga clic en el signo + situado junto a Hadoop SQL.

1. Nombre del servicio (si se muestra): ingrese el nombre del servicio. El valor sugerido es **`amazonemrhive`**. Anote el nombre de este servicio: es necesario al crear una configuración de seguridad de EMR.
2. Nombre público: ingrese el nombre que se mostrará para el servicio. El valor sugerido es **`amazonemrhive`**.

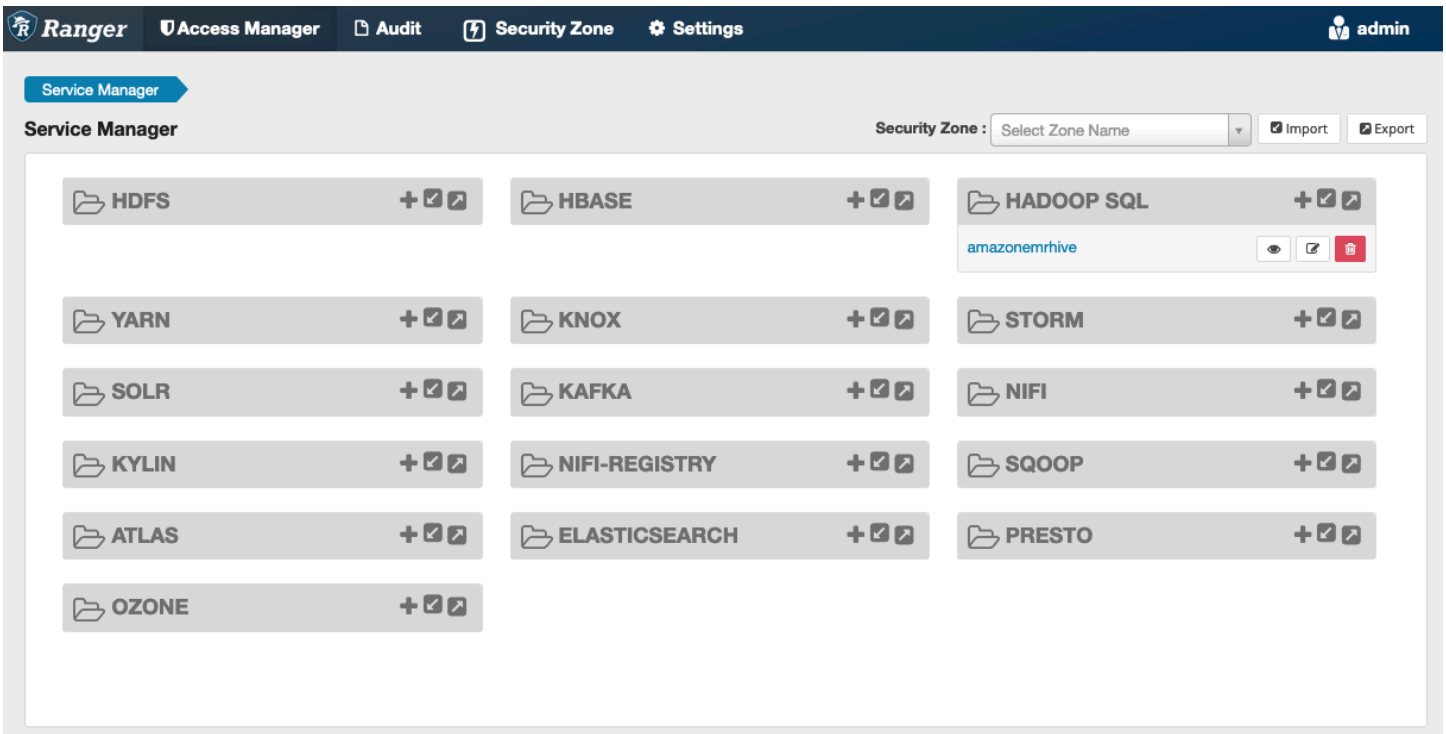


Las propiedades de Apache Hive Config se utilizan para establecer una conexión entre el servidor de Apache Ranger Admin y un HiveServer2 para implementar el autocompletado al crear políticas. No es necesario que las siguientes propiedades sean precisas si no se dispone de un proceso de HiveServer2 persistente. Se pueden rellenar con cualquier información.

- Nombre de usuario: ingrese un nombre de usuario para la conexión JDBC a una instancia de HiveServer2.
- Contraseña: ingrese la contraseña del nombre de usuario anterior.
- jdbc.driver.ClassName: ingrese el nombre de la clase de JDBC para la conectividad con Apache Hive. Puede utilizar el valor predeterminado.
- jdbc.url: ingrese la cadena de conexión JDBC que se utilizará al conectarse a HiveServer2.
- Nombre común del certificado: el campo CN (Nombre común) del certificado que se utiliza para conectarse al servidor de administración desde un complemento cliente. Este valor debe coincidir con el campo CN del certificado TLS que se creó para el complemento.



El botón Probar la conexión comprueba si los valores anteriores se pueden utilizar para conectarse correctamente a la instancia de HiveServer2. Una vez que el servicio se haya creado correctamente, el administrador de servicios debería tener el siguiente aspecto:



Consideraciones

Servidor de metadatos de Hive

Solo motores fiables, específicamente Hive y `emr_record_server`, pueden acceder al servidor de metadatos de Hive para proteger al usuario del acceso no autorizado. Todos los nodos del clúster también acceden al servidor de metadatos de Hive. El puerto 9083 requerido proporciona a todos los nodos acceso al nodo principal.

Autenticación

De forma predeterminada, Apache Hive está configurado para autenticarse mediante Kerberos, tal y como se indicó en la configuración de seguridad de EMR. HiveServer2 también se puede configurar para autenticar a los usuarios mediante LDAP. Consulte Implementing LDAP authentication for Hive on a multi-tenant Amazon EMR cluster para obtener más información.

Limitaciones

Las siguientes son las limitaciones actuales del complemento Apache Hive en Amazon EMR 5.x:

- No se admiten roles de Hive de momento. No se admiten las instrucciones Grant ni Revoke.
- No se admite el uso de la CLI de Hive. JDBC/Beeline es la única forma autorizada de conectar Hive.
- La configuración `hive.server2.builtin.udf.blacklist` debe completarse con las funciones definidas por el usuario (UDF) que considere poco seguras.

Complemento Apache Spark

Amazon EMR ha integrado RecordServer de EMR para proporcionar un control integral del acceso para Spark SQL. RecordServer de EMR es un proceso privilegiado que se ejecuta en todos los nodos de un clúster habilitado para Apache Ranger. Cuando un controlador o ejecutor de Spark ejecuta una instrucción de Spark SQL, todas las solicitudes de metadatos y datos pasan por RecordServer. Para obtener más información sobre RecordServer de EMR, consulte la página Componentes de Amazon EMR.

Temas

- Características admitidas
- Reimplementación de la definición de servicio para usar las instrucciones INSERT, ALTER o DDL

- Instalación de la definición de servicio
- Creación de políticas de Spark SQL
- Consideraciones
- Limitaciones

Características admitidas

Instrucción SQL/Acción de Ranger	STATUS	Versión de EMR compatible
SELECT	Compatible	A partir de la 5.32
SHOW DATABASES	Compatible	A partir de la 5.32
SHOW COLUMNS	Compatible	A partir de la 5.32
SHOW TABLES	Compatible	A partir de la 5.32
MOSTRAR LAS PROPIEDADES DE LA TABLA	Compatible	A partir de la 5.32
DESCRIBE TABLE	Compatible	A partir de la 5.32
INSERT OVERWRITE	Compatible	A partir de las 5.34 y 6.4
INSERT INTO	Compatible	A partir de las 5.34 y 6.4
ALTER TABLE	Compatible	A partir de la 6.4
CREATE TABLE	Compatible	A partir de las 5.35 y 6.7
CREATE DATABASE	Compatible	A partir de las 5.35 y 6.7

Instrucción SQL/Acción de Ranger	STATUS	Versión de EMR compatible
DROP TABLE	Compatible	A partir de las 5.35 y 6.7
DROP DATABASE	Compatible	A partir de las 5.35 y 6.7
DROP VIEW	Compatible	A partir de las 5.35 y 6.7
CREATE VIEW	No es compatible	

Se admiten las siguientes características al usar Spark SQL:

- Control de acceso detallado a las tablas del metaalmacén de Hive, y las políticas se pueden crear para bases de datos, tablas y columnas.
- Las políticas de Apache Ranger pueden incluir políticas de concesión y de denegación a usuarios y grupos.
- Los eventos de auditoría se envían a Registros de CloudWatch.

Reimplementación de la definición de servicio para usar las instrucciones INSERT, ALTER o DDL

> **Note**
>
> A partir de Amazon EMR 6.4, puede usar Spark SQL con las instrucciones: INSERT INTO, INSERT OVERWRITE o ALTER TABLE. A partir de Amazon EMR 6.7, puede usar Spark SQL para crear o eliminar bases de datos y tablas. Si ya tiene una instalación en el servidor de Apache Ranger con las definiciones de servicio de Apache Spark implementadas, utilice el siguiente código para volver a implementar las definiciones de servicio.

```
# Get existing Spark service definition id calling Ranger REST API and JSON
 processor
curl --silent -f -u <admin_user_login>:<password_for_ranger_admin_user> \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
```

```
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef/
name/amazon-emr-spark' | jq .id

# Download the latest Service definition
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/
version-2.0/ranger-servicedef-amazon-emr-spark.json

# Update the service definition using the Ranger REST API
curl -u <admin_user_login>:<password_for_ranger_admin_user> -X PUT -d @ranger-
servicedef-amazon-emr-spark.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/
servicedef/<Spark service definition id from step 1>'
```

Instalación de la definición de servicio

La instalación de la definición de servicio Apache Spark de EMR requiere la configuración del servidor de Ranger Admin. Consulte Configuración del servidor de Ranger Admin.

Siga estos pasos para instalar la definición de servicio de Apache Spark:

Paso 1: inicie sesión mediante SSH en el servidor de Apache Ranger Admin

Por ejemplo:

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Paso 2: descargue la definición de servicio y el complemento del servidor de Apache Ranger Admin

En un directorio temporal, descargue la definición de servicio. Esta definición de servicio es compatible con las versiones 2.x de Ranger.

```
mkdir /tmp/emr-spark-plugin/
cd /tmp/emr-spark-plugin/

wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-spark-plugin-2.x.jar
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-servicedef-amazon-emr-spark.json
```

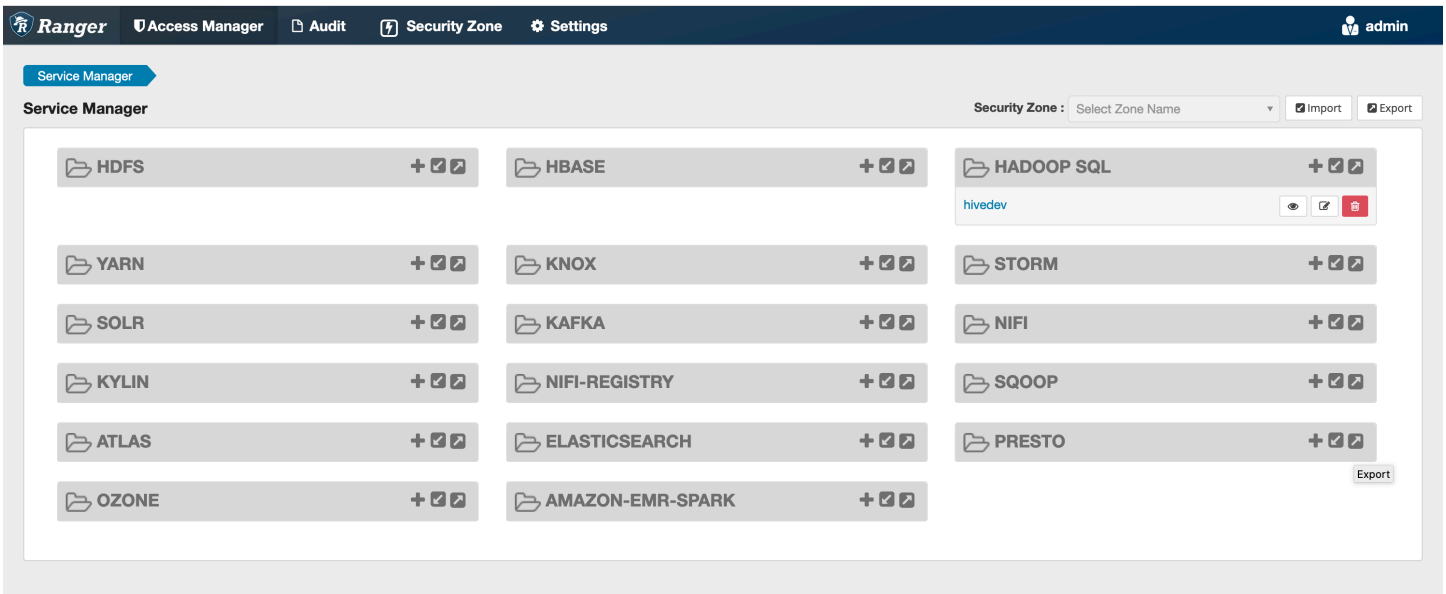
Paso 3: instale el complemento Apache Spark para Amazon EMR

```
export RANGER_HOME=.. # Replace this Ranger Admin's home directory eg /usr/lib/ranger/
ranger-2.0.0-admin
mkdir $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-plugins/amazon-emr-spark
mv ranger-spark-plugin-2.x.jar $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-plugins/
amazon-emr-spark
```

Paso 4: registre la definición de servicio de Apache Spark para Amazon EMR

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_* -X
 POST -d @ranger-servicedef-amazon-emr-spark.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

Si este comando se ejecuta correctamente, verá un nuevo servicio en la IU de Ranger Admin denominado “AMAZON-EMR-SPARK”, como se muestra en la siguiente imagen (se muestra la versión 2.0 de Ranger).

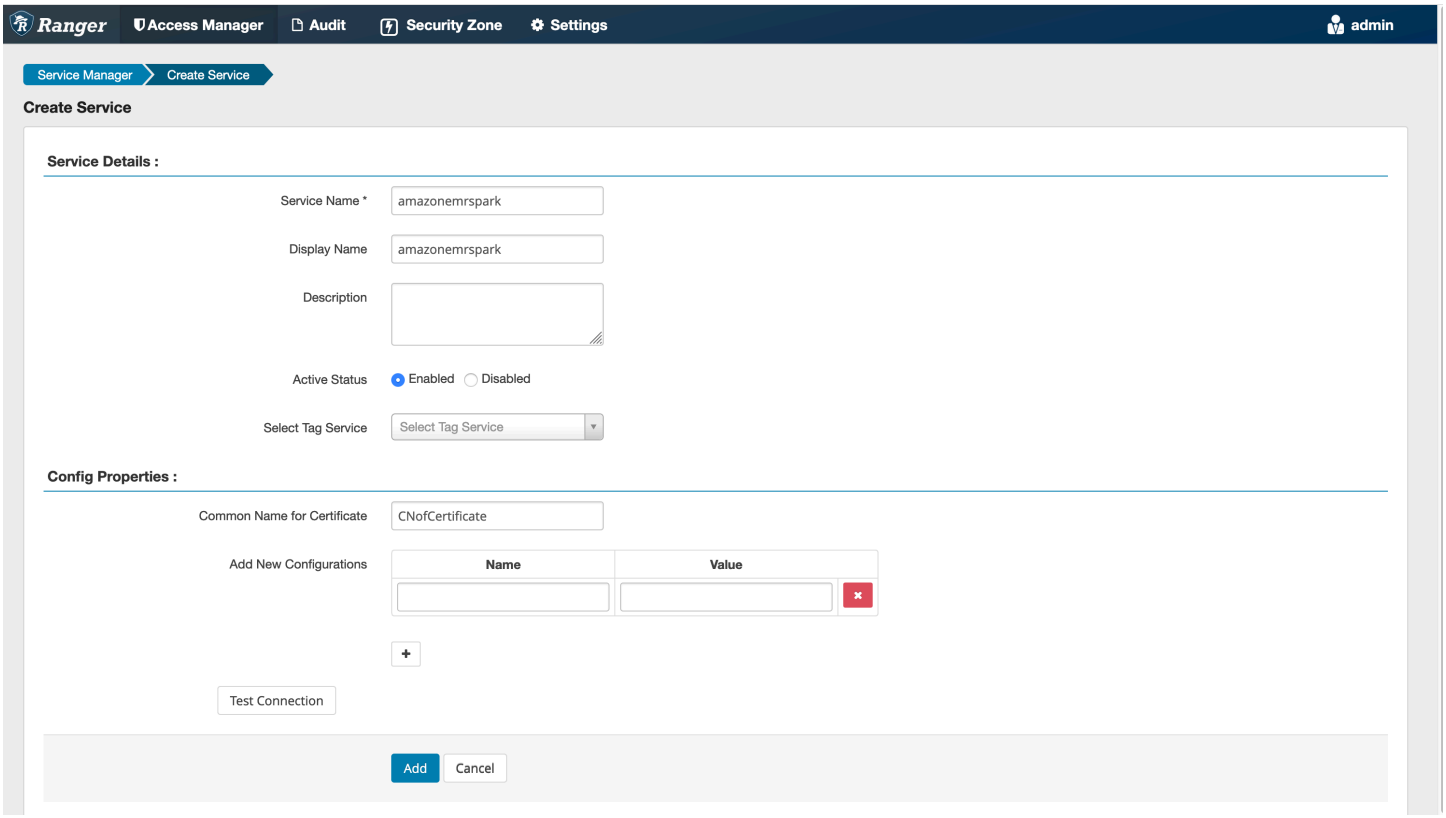


Paso 5: cree una instancia de la aplicación AMAZON-EMR-SPARK

Nombre del servicio (si se muestra): el nombre del servicio que se utilizará. El valor sugerido es **`amazonemrspark`**. Anote el nombre de este servicio, ya que será necesario al crear una configuración de seguridad de EMR.

Nombre público: el nombre que se mostrará para esta instancia. El valor sugerido es **`amazonemrspark`**.

Nombre común del certificado: el campo CN (Nombre común) del certificado que se utiliza para conectarse al servidor de administración desde un complemento cliente. Este valor debe coincidir con el campo CN del certificado TLS que se creó para el complemento.



> **Note**
>
> El certificado TLS de este complemento debería haberse registrado en el almacén de confianza del servidor Ranger Admin. Consulte Certificados TLS para obtener más detalles.

Creación de políticas de Spark SQL

Al crear una nueva política, los campos que hay que rellenar son:

Nombre de la política: el nombre de la política.

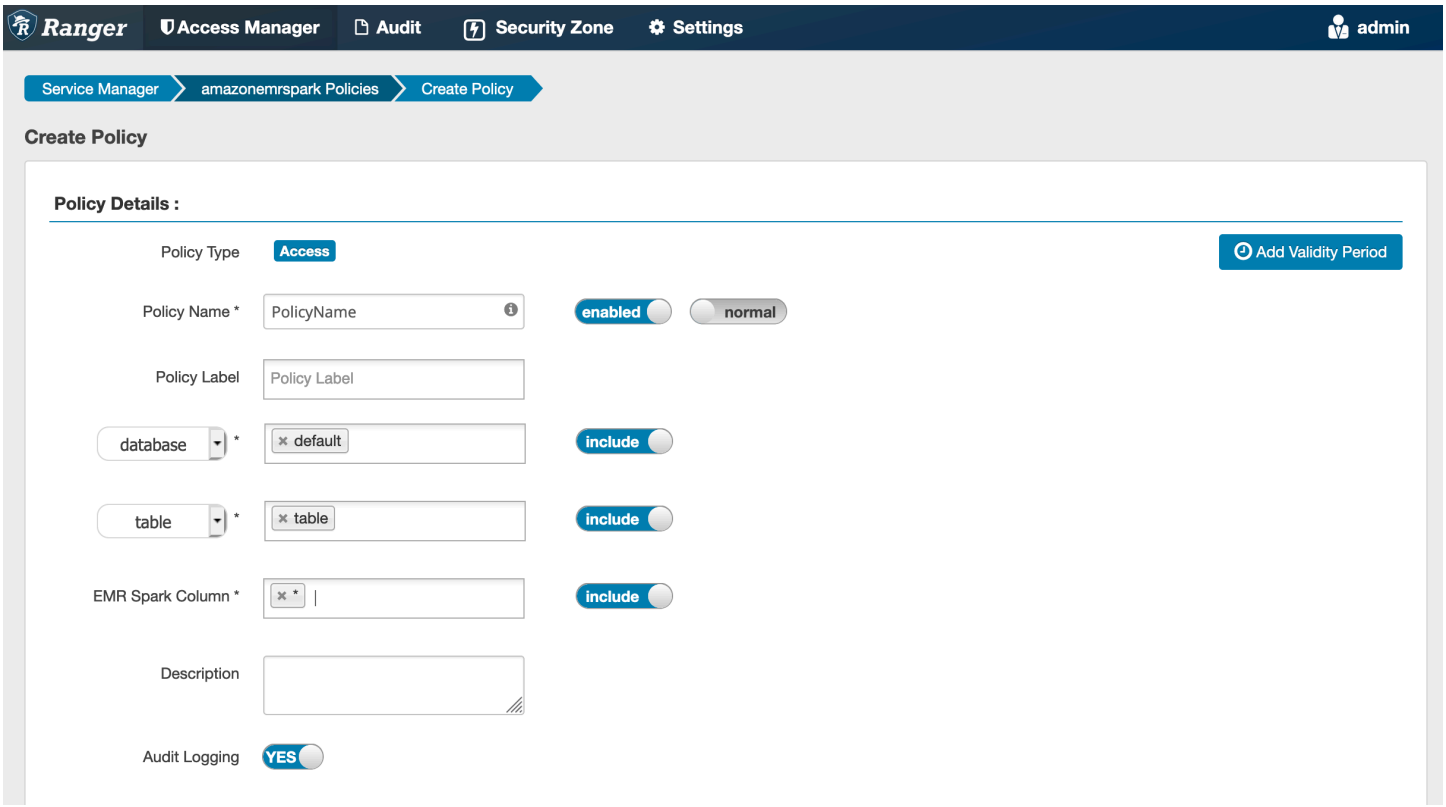
Etiqueta de la política: una etiqueta que puede poner en esta política.

Base de datos: la base de datos a la que se aplica esta política. El comodín “*” representa todas las bases de datos.

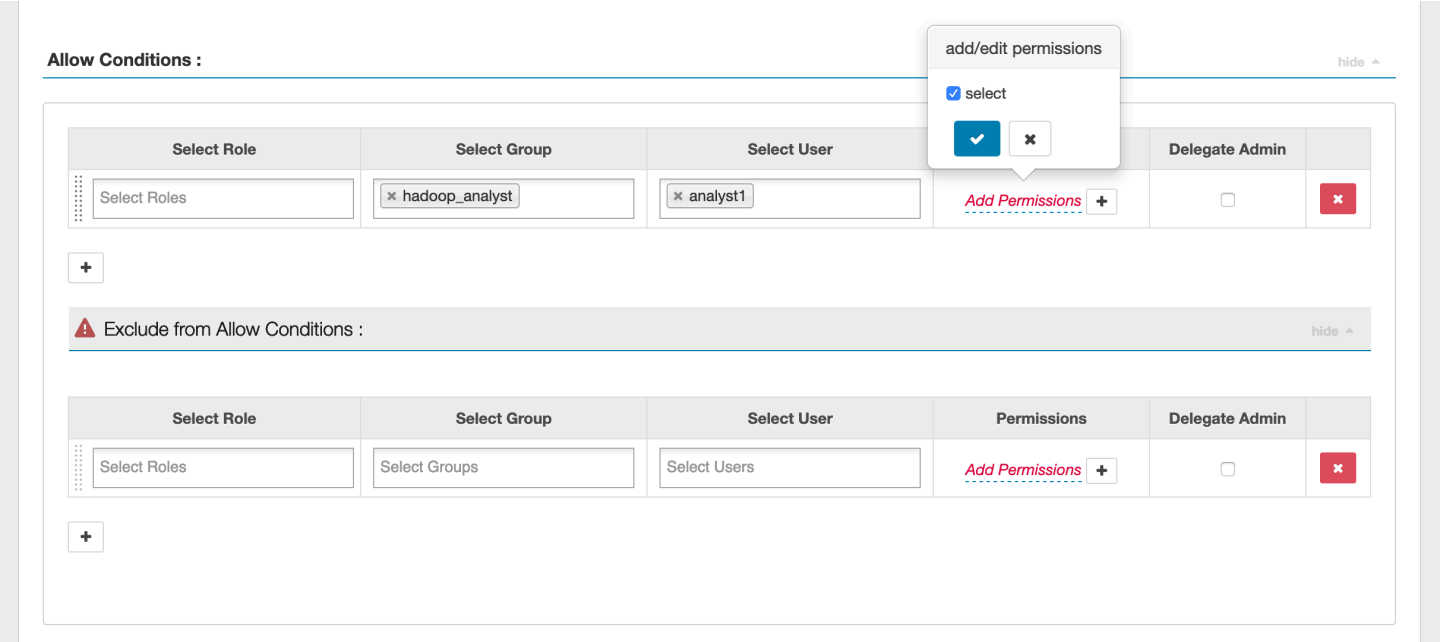
Tabla: las tablas a las que se aplica esta política. El comodín “*” representa todas las tablas.

Columna de EMR Spark: las columnas a las que se aplica esta política. El comodín “*” representa todas las columnas.

Descripción: una descripción de esta política.



Para especificar los usuarios y grupos, ingrese los usuarios y grupos que aparecen a continuación para conceder los permisos. También puede especificar exclusiones para las condiciones de autorización y denegación.



Tras especificar las condiciones de autorización y denegación, haga clic en Guardar.

Consideraciones

Cada nodo del clúster de EMR debe poder conectarse al nodo principal en el puerto 9083.

Limitaciones

Las siguientes son las limitaciones actuales del complemento Apache Spark:

- Servidor de registros siempre se conectará al HMS que se ejecute en un clúster de Amazon EMR. Configure el HMS para que se conecte al modo remoto, si es necesario. No debe incluir valores de configuración en el archivo de configuración Hive-site.xml de Apache Spark.
- Las tablas creadas con orígenes de datos de Spark en CSV o Avro no se pueden leer con RecordServer de EMR. Use Hive para crear y escribir datos, y lea con Record.
- No se admiten las tablas de Delta Lake y Hudi.
- Los usuarios tienen que tener acceso a la base de datos predeterminada. Este es un requisito para el uso de Apache Spark.
- El servidor de Ranger Admin no admite la característica de autocompletar.
- El complemento SparkSQL para Amazon EMR no admite filtros de filas ni enmascaramiento de datos.

- Al utilizar ALTER TABLE con Spark SQL, la ubicación de una partición debe ser el directorio secundario de la ubicación de una tabla. No se admite la inserción de datos en una partición cuya ubicación sea diferente de la ubicación de la tabla.

Complemento EMRFS S3

Para facilitar los controles de acceso a los objetos de S3 en un clúster multiusuario, el complemento EMRFS S3 proporciona controles de acceso a los datos de S3 cuando se accede a ellos a través de EMRFS. Puede permitir el acceso a los recursos de usuarios y grupos de S3.

Para ello, cuando su aplicación intente acceder a los datos de S3, EMRFS envía una solicitud de credenciales al proceso del agente secreto, donde la solicitud se autentica y autoriza mediante un complemento de Apache Ranger. Si la solicitud se autoriza, el agente secreto asume el rol de IAM para los motores Apache Ranger con una política restringida para generar credenciales que solo tienen acceso a la política de Ranger que permitió el acceso. A continuación, las credenciales se devuelven a EMRFS para acceder a S3.

Temas

- Características admitidas
- Instalación de la configuración del servicio
- Creación de políticas de EMRFS S3
- Notas de uso de las políticas de EMRFS S3
- Limitaciones

Características admitidas

El complemento EMRFS S3 proporciona autorización de almacenamiento. Se pueden crear políticas para proporcionar acceso a los usuarios y grupos a los buckets y prefijos de S3. La autorización se realiza únicamente para EMRFS.

Instalación de la configuración del servicio

La instalación de la definición de servicio de Trino requiere la configuración del servidor de Ranger Admin. Para configurar el servidor de Ranger Admin, consulte Configuración del servidor de Ranger Admin.

Siga estos pasos para instalar la definición de servicio de EMRFS.

Paso 1: inicie sesión mediante SSH en el servidor de Apache Ranger Admin.

Por ejemplo:

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Paso 2: descargue la definición de servicio de Amazon EMR y el complemento del servidor de Apache Ranger Admin

En un directorio temporal, descargue la definición de servicio de Amazon EMR. Esta definición de servicio es compatible con las versiones 2.x de Ranger.

```
mkdir /tmp/emr-emrfs-plugin/
cd /tmp/emr-emrfs-plugin/

wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-servicedef-amazon-emr-emrfs.json
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/version-2.0/
ranger-emr-emrfs-plugin-2.x.jar
```

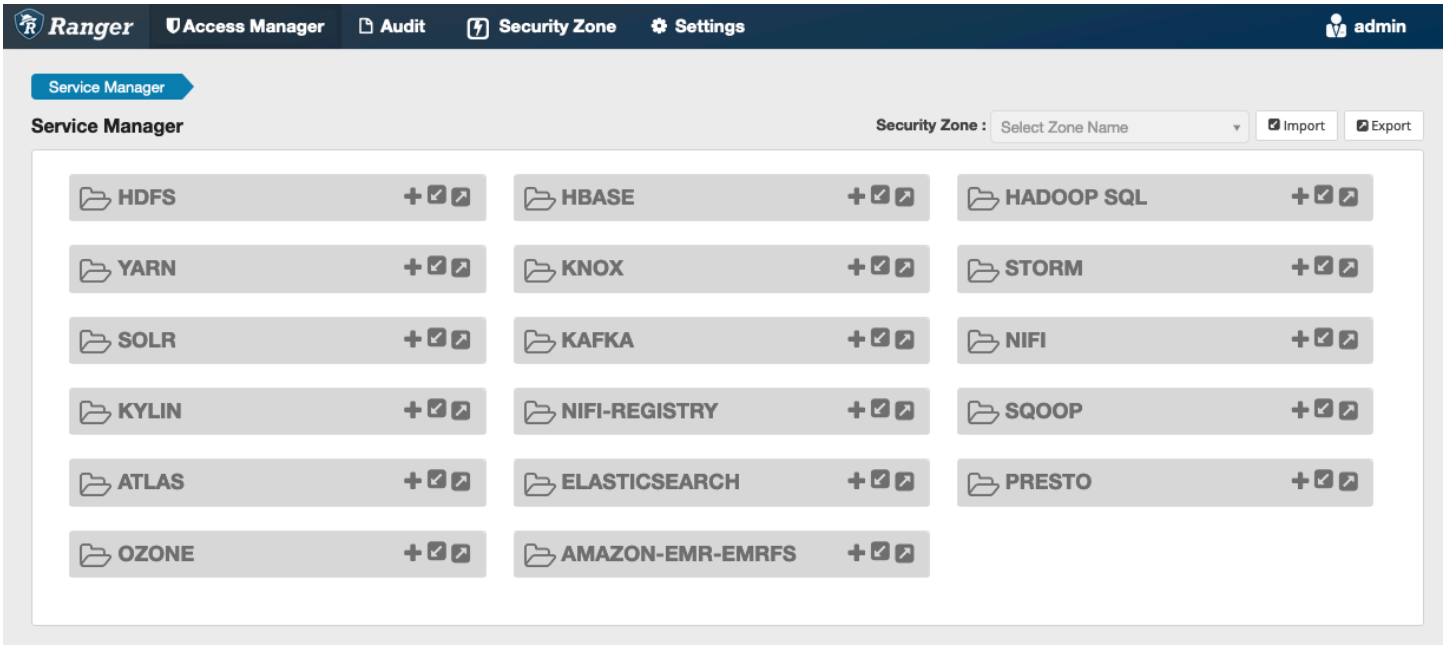
Paso 3: instale el complemento Apache EMRFS S3 para Amazon EMR.

```
export RANGER_HOME=.. # Replace this Ranger Admin's home directory eg /usr/lib/ranger/
ranger-2.0.0-admin
mkdir $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-plugins/amazon-emr-emrfs
mv ranger-emr-emrfs-plugin-2.x.jar $RANGER_HOME/ews/webapp/WEB-INF/classes/ranger-
plugins/amazon-emr-emrfs
```

Paso 4: registre la definición de servicio de EMRFS S3.

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_* -X
 POST -d @ranger-servicedef-amazon-emr-emrfs.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

Si este comando se ejecuta correctamente, verá un nuevo servicio en la IU de Ranger Admin denominado “AMAZON-EMR-S3”, como se muestra en la siguiente imagen (se muestra la versión 2.0 de Ranger).



Paso 5: cree una instancia de la aplicación AMAZON-EMR-EMRFS.

Cree una instancia de la definición de servicio.

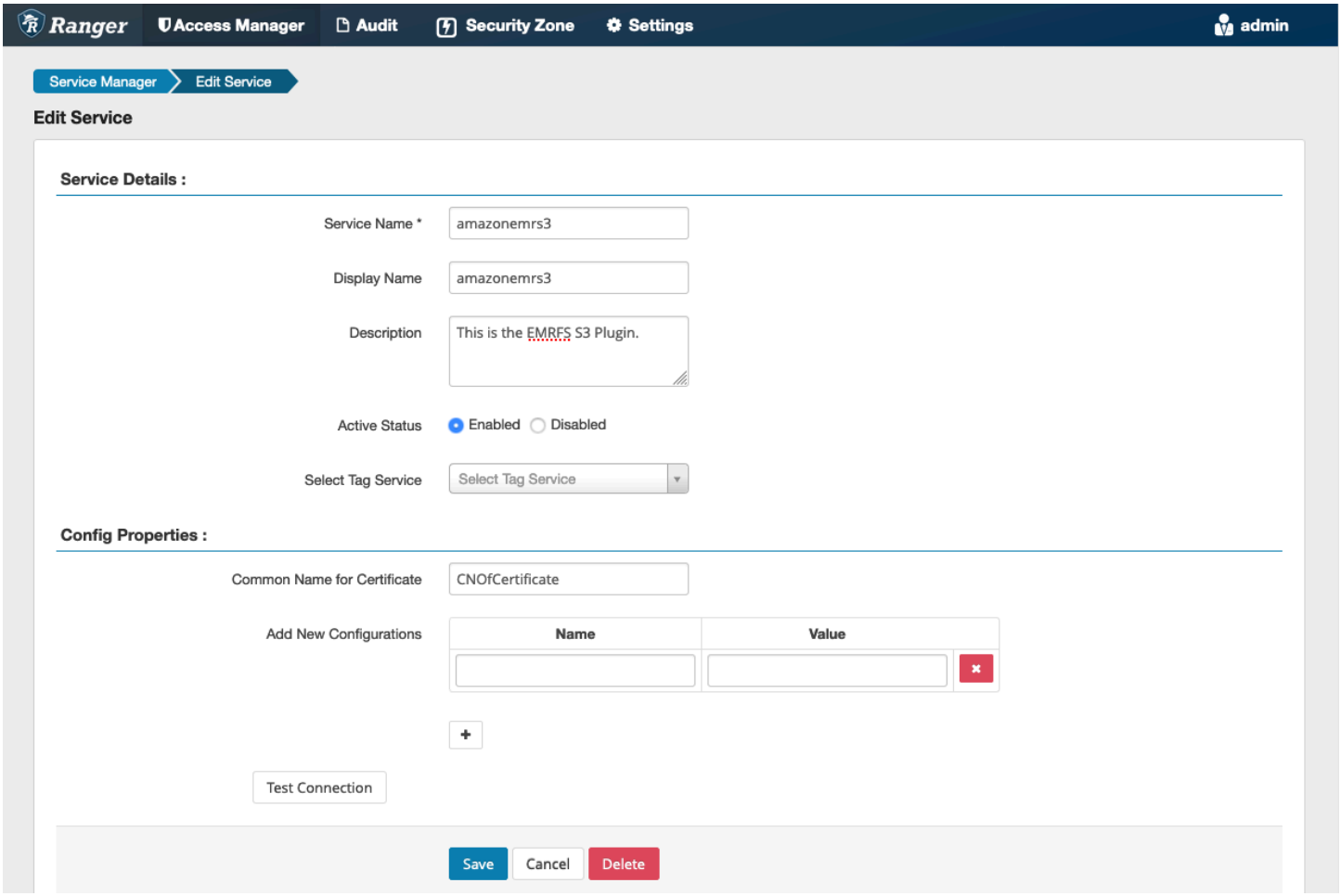
- Haga clic en el signo + situado junto a AMAZON-EMR-EMRFS.

Rellene los siguientes campos:

Nombre del servicio (si se muestra): el valor sugerido es **`amazonemrspark`**. Anote el nombre de este servicio, ya que será necesario al crear una configuración de seguridad de EMR.

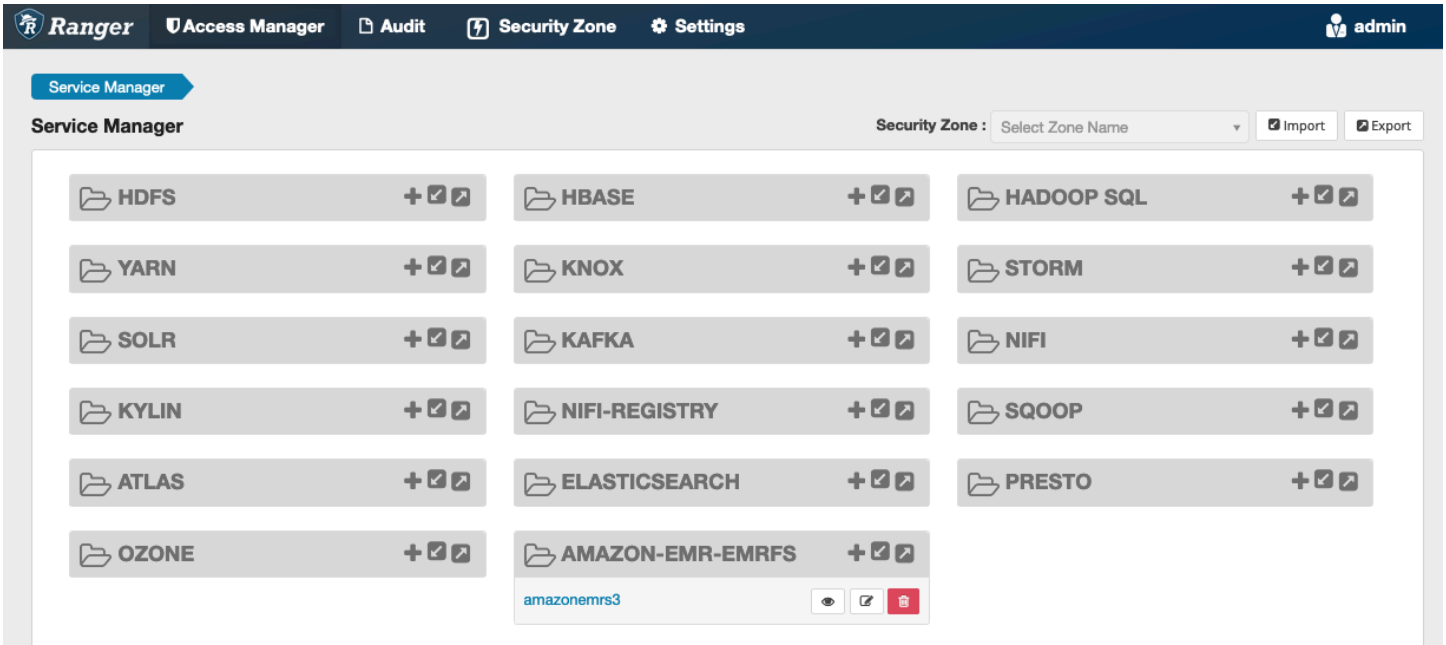
Nombre público: el nombre que se muestra para este servicio. El valor sugerido es **`amazonemrspark`**.

Nombre común del certificado: el campo CN (Nombre común) del certificado que se utiliza para conectarse al servidor de administración desde un complemento cliente. Este valor debe coincidir con el campo CN del certificado TLS que se creó para el complemento.



> **Note**
>
> El certificado TLS de este complemento debería haberse registrado en el almacén de confianza del servidor Ranger Admin. Consulte Certificados TLS para obtener más detalles.

Cuando se crea el servicio, el administrador de servicios incluye “AMAZON-EMR-EMRFS”, como se muestra en la siguiente imagen.



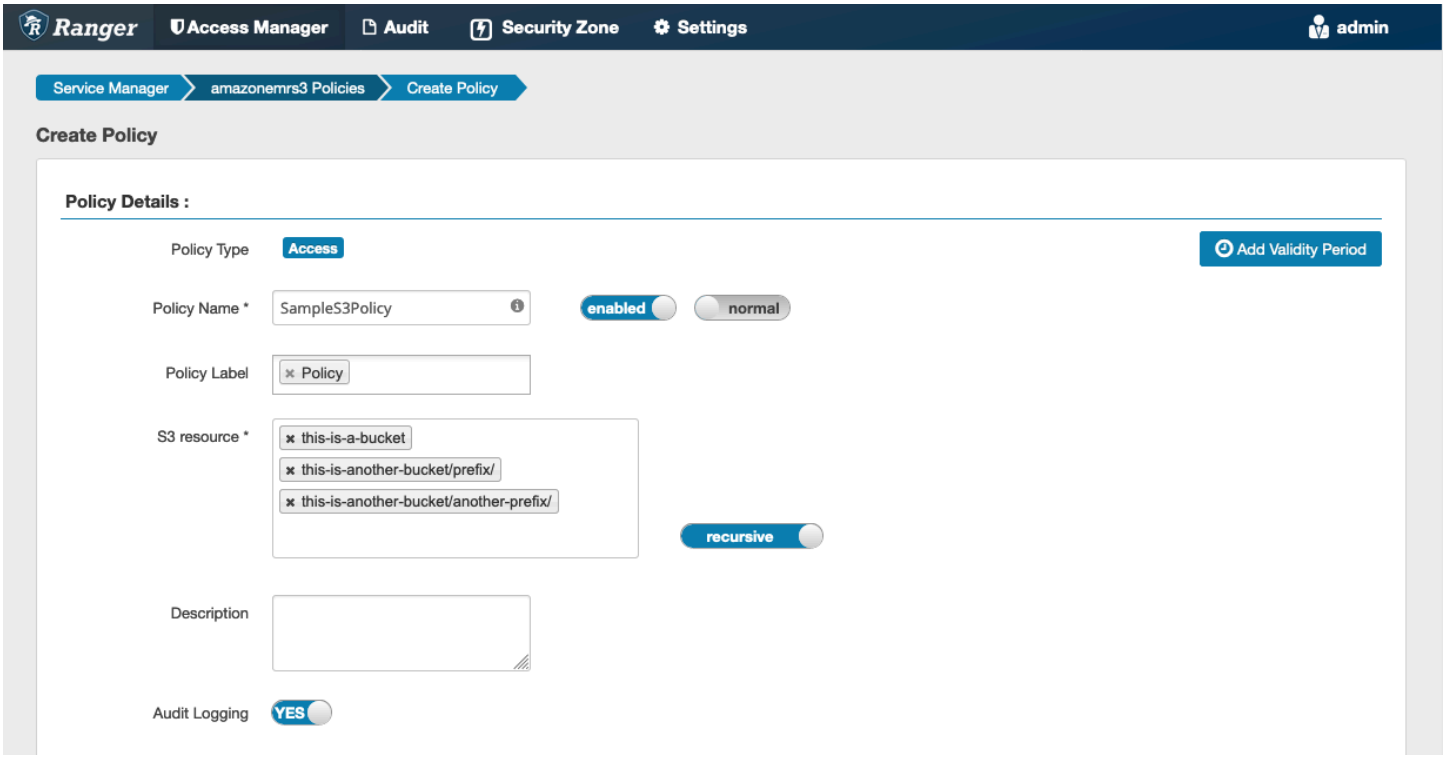
Creación de políticas de EMRFS S3

Para crear una nueva política en la página Crear política del administrador de servicios, rellene los siguientes campos.

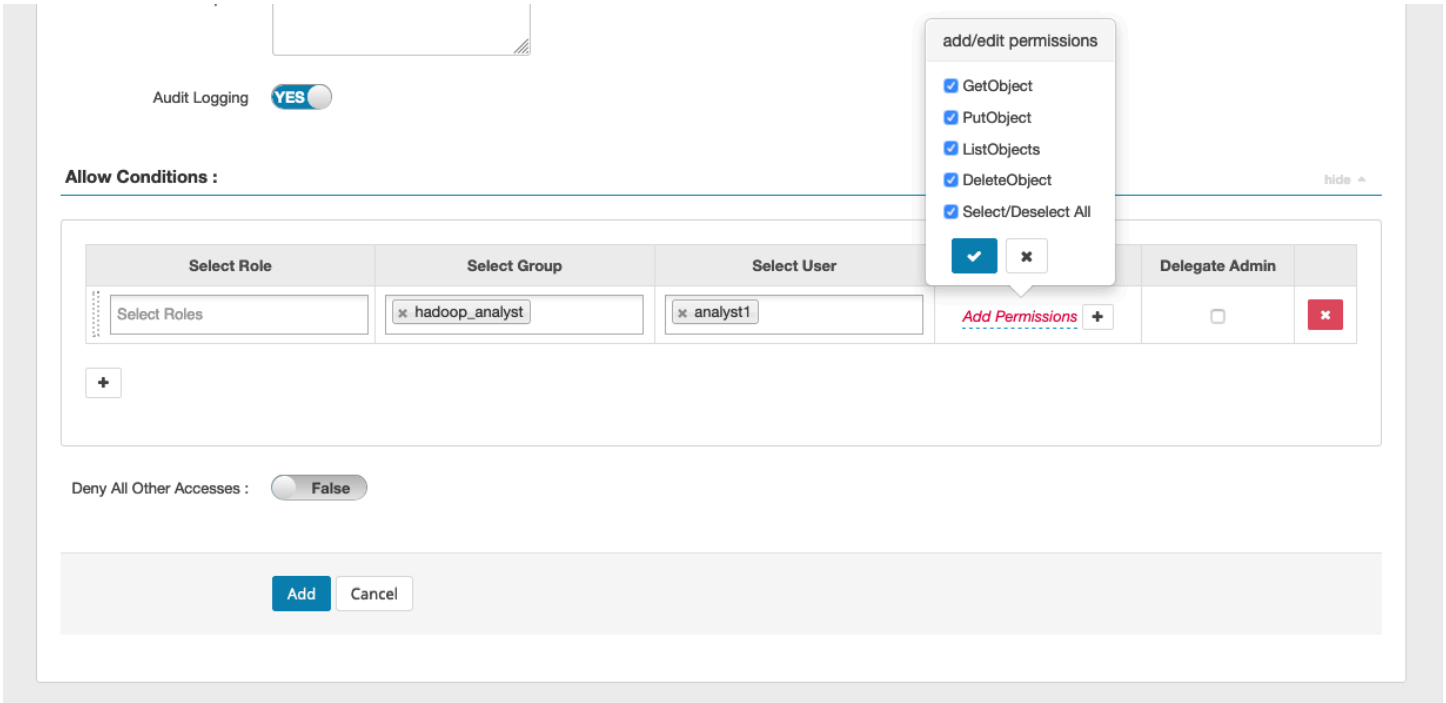
Nombre de la política: el nombre de la política.

Etiqueta de la política: una etiqueta que puede poner en esta política.

Recurso de S3: un recurso que comienza con el bucket y el prefijo opcional. Consulte Notas de uso de las políticas de EMRFS S3 para obtener más información sobre las prácticas recomendadas. Los recursos del servidor Ranger Admin no deben contener **`s3://`**, **`s3a://`** ni **`s3n://`**.



Puede especificar los usuarios y grupos a los que conceder permisos. También puede especificar exclusiones para las condiciones de autorización y denegación.



> **Note**
>
> Se permite un máximo de tres recursos para cada política. Agregar más de tres recursos puede provocar un error cuando se usa esta política en un clúster de EMR. Al agregar más de tres políticas, se muestra un recordatorio sobre el límite de la política.

Notas de uso de las políticas de EMRFS S3

Al crear políticas de S3 en Apache Ranger, hay que tener en cuenta algunas consideraciones de uso.

Permisos para varios objetos de S3

Puede utilizar políticas recursivas y expresiones comodín para conceder permisos a varios objetos de S3 con prefijos comunes. Las políticas recursivas otorgan permisos a todos los objetos con un prefijo común. Las expresiones comodín seleccionan varios prefijos. En conjunto, otorgan permisos a todos los objetos con varios prefijos comunes, como se muestra en los ejemplos siguientes.

Example Uso de una política recursiva

Supongamos que desea obtener permisos para enumerar todos los archivos Parquet de un bucket de S3 organizados de la siguiente manera.

```
s3://sales-reports/americas/
    +- year=2000
    |      +- data-q1.parquet
    |      +- data-q2.parquet
    +- year=2019
    |      +- data-q1.json
    |      +- data-q2.json
    |      +- data-q3.json
    |      +- data-q4.json
    |
    +- year=2020
    |      +- data-q1.parquet
    |      +- data-q2.parquet
    |      +- data-q3.parquet
    |      +- data-q4.parquet
    |      +- annual-summary.parquet
    +- year=2021
```

En primer lugar, considere los archivos Parquet con el prefijo `s3://sales-reports/americas/year=2000`. Puede conceder permisos GetObject a todos ellos de dos maneras:

Uso de políticas no recursivas: una opción es utilizar dos políticas no recursivas independientes, una para el directorio y otra para los archivos.

La primera política concede permiso al prefijo `s3://sales-reports/americas/year=2020` (no hay ningún / final).

```
- S3 resource = "sales-reports/americas/year=2000"
- permission = "GetObject"
- user = "analyst"
```

La segunda política utiliza una expresión comodín para conceder permisos a todos los archivos con prefijo `sales-reports/americas/year=2020/` (tenga en cuenta el / final).

```
- S3 resource = "sales-reports/americas/year=2020/*"
- permission = "GetObject"
- user = "analyst"
```

Uso de una política recursiva: una alternativa más práctica consiste en utilizar una única política recursiva y conceder permisos recursivos al prefijo.

```
- S3 resource = "sales-reports/americas/year=2020"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Hasta ahora, solo se han incluido los archivos Parquet con el prefijo `s3://sales-reports/americas/year=2000`. Ahora también puede incluir los archivos Parquet con un prefijo diferente, `s3://sales-reports/americas/year=2020`, en la misma política recursiva introduciendo una expresión comodín como la siguiente.

```
- S3 resource = "sales-reports/americas/year=20?0"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Políticas para los permisos PutObject y DeleteObject

Escribir políticas para los permisos `PutObject` y `DeleteObject` en archivos en EMRFS requiere un cuidado especial porque, a diferencia de los permisos GetObject, necesitan permisos recursivos adicionales otorgados al prefijo.

Example Políticas para los permisos PutObject y DeleteObject

Por ejemplo, eliminar el archivo `annual-summary.parquet` no solo requiere un permiso DeleteObject para ese archivo concreto.

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "DeleteObject"
- user = "analyst"
```

También requiere una política que conceda permisos `PutObject` y `GetObject` recursivos a su prefijo.

Del mismo modo, modificar el archivo `annual-summary.parquet` no solo requiere un permiso `PutObject` para ese archivo concreto.

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "PutObject"
- user = "analyst"
```

También requiere una política que conceda permisos `GetObject` recursivos a su prefijo.

```
- S3 resource = "sales-reports/americas/year=2020"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Comodines en las políticas

Hay dos áreas en las que se pueden especificar los caracteres comodín. Al especificar un recurso de S3, los caracteres “*” y “?” se puede utilizar. El asterisco “*” sustituye una ruta de S3 y coincide con todo lo que aparece después del prefijo. Tome la siguiente política como ejemplo:

```
S3 resource = "sales-reports/americas/*"
```

Esto coincide con las siguientes rutas de S3.

```
sales-reports/americas/year=2020/
sales-reports/americas/year=2019/
sales-reports/americas/year=2019/month=12/day=1/afile.parquet
sales-reports/americas/year=2018/month=6/day=1/afile.parquet
sales-reports/americas/year=2017/afile.parquet
```

El comodín “?” coincide con cualquier carácter individual. Tome la siguiente política como ejemplo:

```
S3 resource = "sales-reports/americas/year=201?/"
```

Esto coincide con las siguientes rutas de S3.

```
sales-reports/americas/year=2019/
sales-reports/americas/year=2018/
sales-reports/americas/year=2017/
```

Comodines en los usuarios

Hay dos caracteres comodín integrados al asignar usuarios para proporcionar acceso a los usuarios. El primero es el comodín “{USER}”, que proporciona acceso a todos los usuarios. El segundo comodín es “{OWNER}”, que proporciona acceso al propietario de un objeto concreto o directamente. Sin embargo, actualmente no se admite el comodín “{USER}”.

Limitaciones

Las siguientes son las limitaciones actuales del complemento EMRFS S3:

- Las políticas de Apache Ranger pueden tener como máximo tres políticas.
- El acceso a S3 debe realizarse a través de EMRFS y se puede utilizar con aplicaciones relacionadas con Hadoop. No se admite lo siguiente:
  - Bibliotecas Boto3
  - SDK y CLI de AWS
  - Conector de código abierto S3A
- No se admiten las políticas de denegación de Apache Ranger.
- Actualmente, no se admiten las operaciones en S3 con claves con cifrado CSE-KMS.

- No se admite la compatibilidad entre regiones.
- La característica Zona de seguridad de Apache Ranger no es compatible. Las restricciones de control de acceso definidas mediante la característica Zona de seguridad no se aplican a los clústeres de Amazon EMR.
- El usuario de Hadoop no genera ningún evento de auditoría, ya que Hadoop siempre accede al perfil de instancia de EC2.
- Se recomienda deshabilitar la visión de consistencia de Amazon EMR. S3 de tiene un alto grado de consistencia, por lo que ya no es necesario. Consulte Amazon S3 strong consistency para obtener más información.
- El complemento EMRFS S3 realiza numerosas llamadas STS. Se recomienda realizar pruebas de carga en una cuenta de desarrollo y supervisar el volumen de llamadas STS. También se recomienda realizar una solicitud STS para aumentar los límites del servicio de AssumeRole.

Complemento Trino

Trino (anteriormente PrestoSQL) es un motor de consultas SQL que puede utilizar para ejecutar consultas en orígenes de datos como HDFS, almacenamiento de objetos, bases de datos relacionales y bases de datos NoSQL. Elimina la necesidad de migrar los datos a una ubicación central y permite consultarlos desde cualquier lugar en el que se encuentren. Amazon EMR proporciona un complemento de Apache Ranger para proporcionar controles de acceso detallados para Trino. El complemento es compatible con la versión 2.0 y posteriores del servidor de Apache Ranger Admin de código abierto.

Temas

- Características admitidas
- Instalación de la configuración del servicio
- Creación de políticas de Trino
- Consideraciones
- Limitaciones

Características admitidas

El complemento Apache Ranger para Trino en Amazon EMR admite todas las funciones del motor de consultas de Trino, que están protegidas por un control de acceso detallado. Esto incluye controles de acceso a bases de datos, tablas y columnas, así como el filtrado de filas y el enmascaramiento

de datos. Las políticas de Apache Ranger pueden incluir políticas de concesión y de denegación a usuarios y grupos. Los eventos de auditoría también se envían a los registros de CloudWatch.

Instalación de la configuración del servicio

La instalación de la definición de servicio de Trino requiere la configuración del servidor de Ranger Admin. Para configurar el servidor de Ranger Admin, consulte Configuración del servidor de Ranger Admin.

Siga estos pasos para instalar la definición de servicio de Trino.

1. Inicie sesión mediante SSH en el servidor de Apache Ranger Admin.

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

2. Desinstale el complemento del servidor Presto, si existe. Ejecute el siguiente comando. Si se produce un error “Servicio no encontrado”, significa que el complemento del servidor Presto no estaba instalado en su servidor. Continúe con el siguiente paso.

```
curl -f -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin
 user_**_>_* -X DELETE -k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/
v2/api/servicedef/name/presto'
```

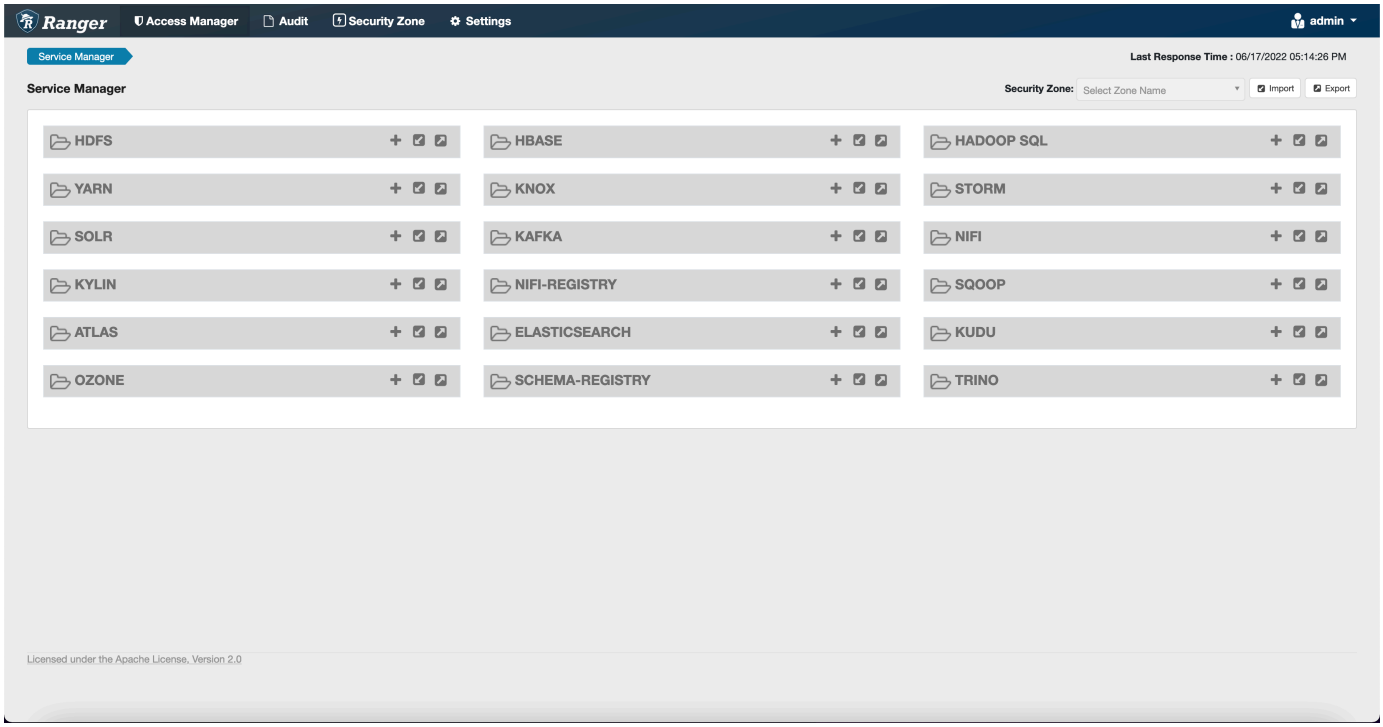
3. Descargue la definición de servicio y el complemento del servidor de Apache Ranger Admin. En un directorio temporal, descargue la definición de servicio. Esta definición de servicio es compatible con las versiones 2.x de Ranger.

```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/
version-2.0/ranger-servicedef-amazon-emr-trino.json
```

4. Registre la definición de servicio de Apache Trino para Amazon EMR.

```
curl -u *<admin users login>*:*_<_**_password_ **_for_** _ranger admin user_**_>_*
 -X POST -d @ranger-servicedef-amazon-emr-trino.json \
-H "Accept: application/json" \
-H "Content-Type: application/json" \
-k 'https://*<RANGER SERVER ADDRESS>*:6182/service/public/v2/api/servicedef'
```

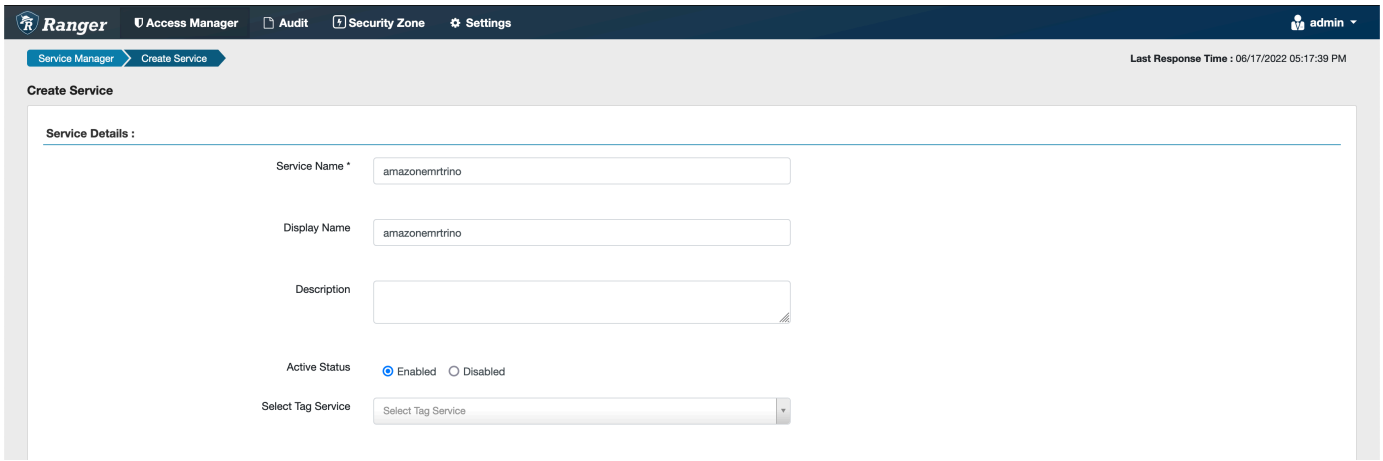
   Si este comando se ejecuta correctamente, verá un nuevo servicio en la IU de Ranger Admin denominado `TRINO`, como se muestra en la siguiente imagen.



5. Cree una instancia de la aplicación `TRINO` e ingrese la siguiente información.

   Nombre del servicio: el nombre del servicio que utilizará. El valor sugerido es `amazonemrtrino`. Anote el nombre de este servicio, ya que será necesario al crear una configuración de seguridad de Amazon EMR.

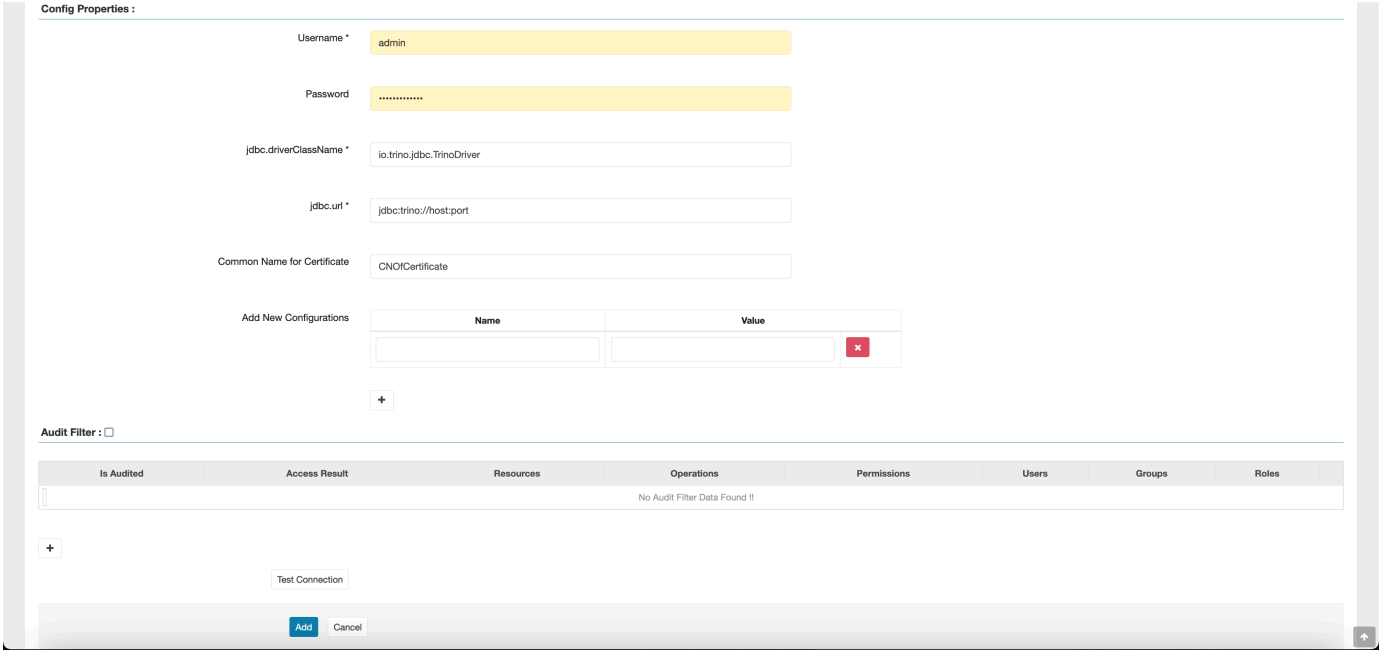
   Nombre público: el nombre que se mostrará para esta instancia. El valor sugerido es `amazonemrtrino`.



   jdbc.driver.ClassName: el nombre de la clase de JDBC para la conectividad con Trino. Puede utilizar el valor predeterminado.

jdbc.url: la cadena de conexión JDBC que se utilizará al conectarse al coordinador de Trino.

Nombre común del certificado: el campo CN (Nombre común) del certificado que se utiliza para conectarse al servidor de administración desde un complemento cliente. Este valor debe coincidir con el campo CN del certificado TLS que se creó para el complemento.



Tenga en cuenta que el certificado TLS de este complemento debería haberse registrado en el almacén de confianza del servidor de Ranger Admin. Para obtener más información, consulte Certificados TLS.

Creación de políticas de Trino

Cuando cree una nueva política, rellena los siguientes campos.

Nombre de la política: el nombre de la política.

Etiqueta de la política: una etiqueta que puede poner en esta política.

Catálogo: el catálogo al que se aplica esta política. El comodín “*” representa todos los catálogos.

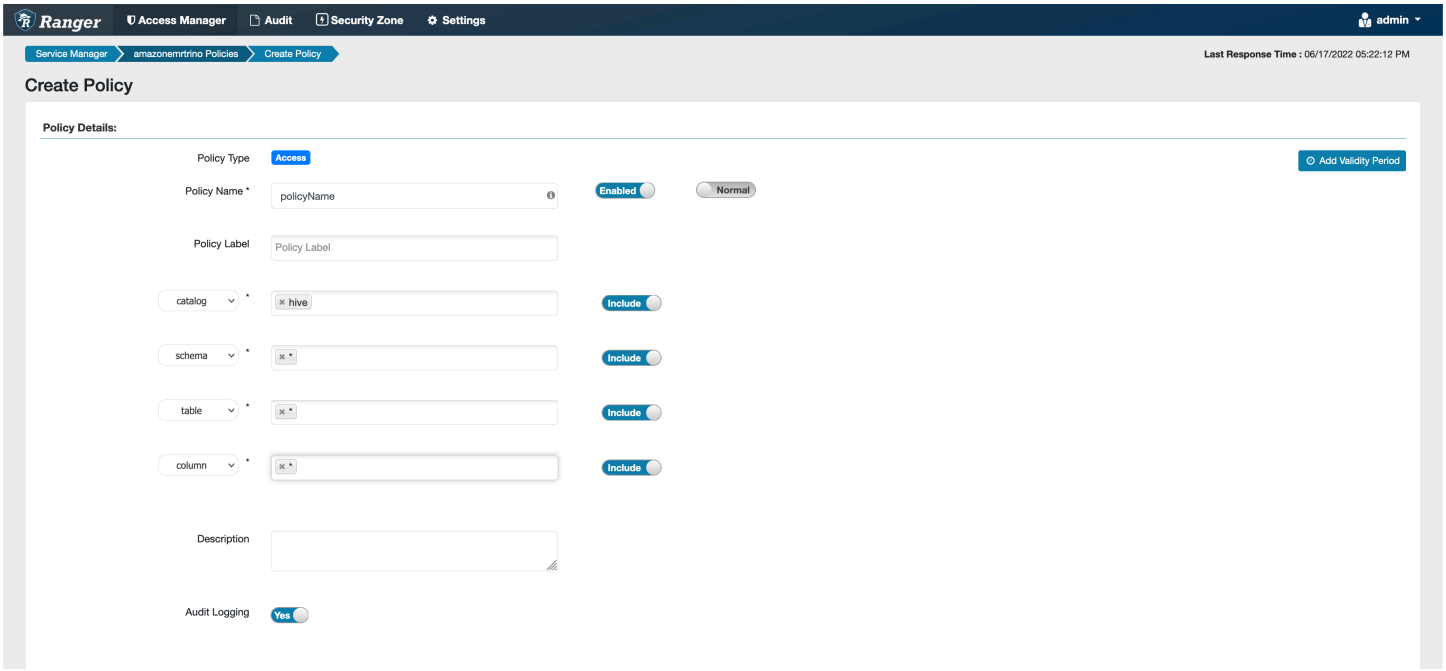
Esquema: los esquemas a los que se aplica esta política. El comodín “*” representa todos los esquemas.

Tabla: las tablas a las que se aplica esta política. El comodín “*” representa todas las tablas.

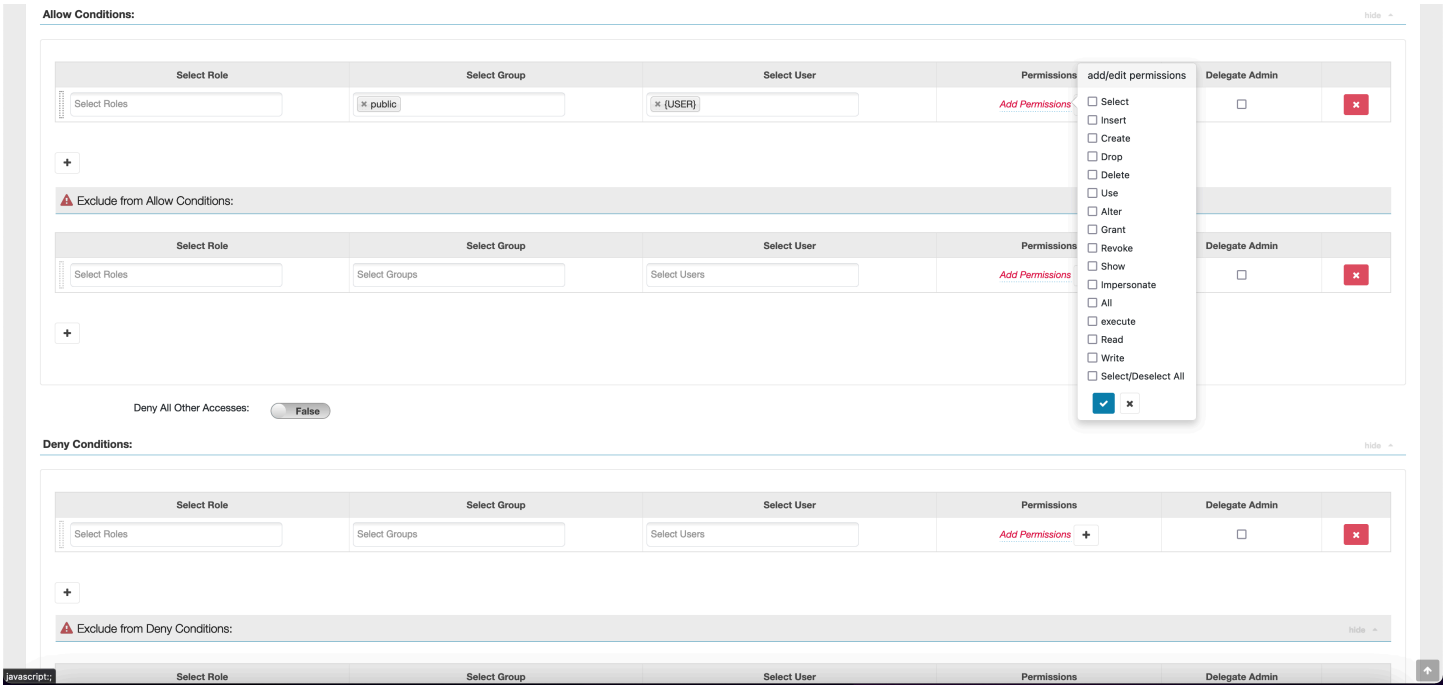
Columna: las columnas a las que se aplica esta política. El comodín “*” representa todas las columnas.

Descripción: una descripción de esta política.

Existen otros tipos de políticas para Usuario de Trino (para el acceso suplantando al usuario), Propiedad del sistema o sesión de Trino (para modificar las propiedades del sistema o la sesión del motor), Funciones/Procedimientos (para permitir las llamadas a funciones o procedimientos) y URL (para conceder acceso de lectura/escritura al motor en las ubicaciones de los datos).



Para conceder permisos a usuarios y grupos específicos, ingrese los usuarios y grupos. También puede especificar exclusiones para las condiciones de autorización y denegación.



Tras especificar las condiciones de autorización y denegación, elija Guardar.

Consideraciones

Al crear políticas de Trino en Apache Ranger, hay que tener en cuenta algunas consideraciones de uso.

Servidor de metadatos de Hive

Solo motores fiables, específicamente el motor Trino, pueden acceder al servidor de metadatos de Hive para proteger al usuario del acceso no autorizado. Todos los nodos del clúster también acceden al servidor de metadatos de Hive. El puerto 9083 requerido proporciona a todos los nodos acceso al nodo principal.

Autenticación

De forma predeterminada, Trino está configurado para autenticarse mediante Kerberos, tal y como se indicó en la configuración de seguridad de Amazon EMR.

Cifrado en tránsito obligatorio

El complemento Trino requiere que tenga activado el cifrado en tránsito en la configuración de seguridad de Amazon EMR. Para habilitar el cifrado, consulte Cifrado en tránsito.

Limitaciones

Las siguientes son las limitaciones actuales del complemento Trino:

- El servidor de Ranger Admin no admite la característica de autocompletar.

Solución de problemas con Apache Ranger

Estos son algunos de los problemas que se diagnostican con frecuencia relacionados con el uso de Apache Ranger.

Recomendaciones

- Haga pruebas con un único clúster de nodo principal: los clústeres maestros de un solo nodo se aprovisionan más rápido que un clúster de varios nodos, lo que puede reducir el tiempo de cada iteración de prueba.
- Configure el modo de desarrollo del clúster. Al iniciar el clúster de EMR, establezca el parámetro `--additional-info"` en:

  ```
  '{"clusterType":"development"}'
  ```

  Este parámetro solo se puede configurar a través de AWS CLI o el SDK de AWS y no está disponible en la consola de Amazon EMR. Cuando se activa este indicador y el clúster maestro no lo aprovisiona, el servicio Amazon EMR mantiene activo el clúster durante algún tiempo antes de retirarlo del servicio. Este tiempo es muy útil para sondear varios archivos de registro antes de que finalice el clúster.

El clúster de EMR no se pudo aprovisionar

Existen varios motivos por los que un clúster de Amazon EMR puede no iniciarse. Las siguientes son algunas formas de diagnosticar el problema.

Compruebe los registros de aprovisionamiento de EMR

Amazon EMR usa Puppet para instalar y configurar aplicaciones en un clúster. Si consulta los registros, obtendrá detalles sobre si se ha producido algún error durante la fase de aprovisionamiento de un clúster. Se puede acceder a los registros en el clúster o en S3 si los registros están configurados para enviarse a S3.

Los registros se almacenan en `/var/log/provision-node/apps-phase/0/{UUID}/puppet.log` en el disco y `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/provision-node/apps-phase/0/{UUID}/puppet.log.gz`.

Mensajes de error comunes

Mensaje de error	Causa
Puppet (err): Systemd start for emr-record-server failed! journalctl log for emr-record-server:	No se pudo iniciar Servidor de registros de EMR. Consulte los registros de Servidor de registros de EMR a continuación.
Puppet (err): Systemd start for emr-record-server failed! journalctl log for emrsecretagent:	Agente secreto de EMR no se pudo iniciar. Consulte Revisión de los registros de Agente secreto a continuación.
/Stage[main]/Ranger_plugins::Ranger_hive_plugin/Ranger_plugins::Prepare_two_way_tls[configure 2-way TLS in Hive plugin]/Exec[create keystore and truststore for Ranger Hive plugin]/returns (notice): 140408606197664:error:0906D06C:PEM routines:PEM_read_bio:no start line:pem_lib.c:707:Expecting: ANY PRIVATE KEY	El certificado TLS privado de Secrets Manager para el certificado del complemento Apache Ranger no tiene el formato correcto o no es un certificado privado. Consulte Certificados TLS para obtener más información sobre los formatos de los certificados.
//Stage[main]/Ranger_plugins::Ranger_s3_plugin/Ranger_plugins::Prepare_two_way_tls[configure 2-way TLS in Ranger s3 plugin]/Exec[create keystore and truststore for Ranger amazon-emr-s3 plugin]/returns (notice): An error occurred (AccessDeniedException) when calling the GetSecretValue operation: User: arn:aws:sts::XXXXXXXXXXX:assumed-role/EMR_EC2_DefaultRole/i-XXXXXXXXXXX is not authorized to perform: secretsmanager:Get	El rol del perfil de instancia de EC2 no tiene los permisos correctos para recuperar los certificados TLS del agente secreto.

Mensaje de error	Causa
SecretValue on resource: arn:aws:secretsmanager:us-east-1:XXXXXXXXXX:secret:AdminServer-XXXXX	

Revisión de los registros de Agente secreto

Los registros de Agente secreto se encuentran en `/emr/secretagent/log/` en un nodo de EMR o en el directorio `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/daemons/secretagent/` de S3.

Mensajes de error comunes

Mensaje de error	Causa
Exception in thread "main" com.amazonaws.services.securitytoken.model.AWSSecurityTokenServiceException: User: arn:aws:sts::XXXXXXXXXXXX:assumed-role/EMR_EC2_DefaultRole/i-XXXXXXXXXXXXXXXX is not authorized to perform: sts:AssumeRole on resource: arn:aws:iam::XXXXXXXXXXXX:role/*RangerPluginDataAccessRole* (Service: AWSSecurityTokenService; Status Code: 403; Error Code: AccessDenied; Request ID: XXXXXXXX-XXXX-XXXX-XXXX-XXXXXXXXXXXX; Proxy: null)	La excepción anterior significa que el rol del perfil de instancia de EC2 de EMR no tiene permisos para asumir el rol RangerPluginDataAccessRole. Consulte Roles de IAM para la integración nativa con Apache Ranger.
ERROR qtp54617902-149: Web App Exception Occurred javax.ws.rs.NotAllowedException: HTTP 405 Method Not Allowed	Estos errores se pueden ignorar.

Compruebe los registros de Servidor de registros (para Spark SQL)

Los registros de Servidor de registros de EMR están disponibles en /var/log/emr-record-server/ en un nodo de EMR. También se pueden encontrar en el directorio s3://<UBICACIÓN DEL REGISTRO>/<ID DEL CLÚSTER>/node/<ID DE LA INSTANCIA DE EC2>/daemons/emr-record-server/ en S3.

Mensajes de error comunes

Mensaje de error	Causa
InstanceMetadataServiceResourceFetcher:105 - [] Fail to retrieve token com.amazonaws.SdkClientException: Failed to connect to service endpoint	Agente secreto de EMR no se inició o está experimentando un problema. Inspeccione los registros de Agente secreto para ver si hay errores y el script de Puppet para determinar si se produjo algún error de aprovisionamiento.

Las consultas experimentan errores inesperados

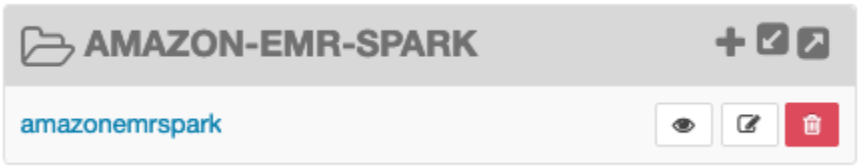
Compruebe los registros del complemento Apache Ranger (registros de Apache Hive, Servidor de registros de EMR, Agente secreto de EMR, etc.).

Esta sección es común a todas las aplicaciones que se integran con el complemento Ranger, como Apache Hive, Servidor de registros de EMR y Agente secreto de EMR.

Mensajes de error comunes

Mensaje de error	Causa
ERROR PolicyRefresher:272 - [] PolicyRefresher(serviceName=policy-repository): failed to find service. Will clean up local cache of policies (-1)	Este mensaje de error significa que el nombre del servicio que proporcionó en la configuración de seguridad de EMR no coincide con un repositorio de políticas de servicio del servidor de Ranger Admin.

Si en el servidor de Ranger Admin su servicio AMAZON-EMR-SPARK tiene el siguiente aspecto, debe ingresar **`amazonemrspark`** como nombre del servicio.



Control del tráfico de red con grupos de seguridad

Los grupos de seguridad funcionan como firewalls virtuales para que las instancias EC2 del clúster controlen el tráfico entrante y saliente. Cada grupo de seguridad tiene un conjunto de reglas que controlan el tráfico entrante y un conjunto de reglas distinto que controlan el tráfico saliente. Para obtener más información, consulte Grupos de seguridad de Amazon EC2 para instancias Linux en la Guía del usuario de Amazon EC2 para instancias de Linux.

Puede utilizar dos clases de grupos de seguridad con Amazon EMR: grupos de seguridad administrados por Amazon EMR y grupos de seguridad adicionales.

Cada clúster tiene asociados grupos de seguridad administrados. Puede utilizar los grupos de seguridad administrados predeterminados que Amazon EMR crea o especificar grupos de seguridad administrados personalizados. En ambos casos, Amazon EMR agrega automáticamente reglas a los grupos de seguridad administrados que un clúster necesita para comunicarse entre las instancias del clúster y los productos de AWS.

Los grupos de seguridad adicionales son opcionales. Puede especificarlos junto con los grupos de seguridad administrados para adaptar el acceso a las instancias de clúster. Los grupos de seguridad adicionales contienen solo las reglas que defina. Amazon EMR no los modifica.

Las reglas que Amazon EMR crea en los grupos de seguridad administrados permiten la comunicación entre los componentes internos del clúster. Para permitir que los usuarios y las aplicaciones obtengan acceso a un clúster desde fuera de este, puede editar las reglas de los grupos de seguridad administrados, crear grupos de seguridad adicionales con reglas adicionales o ambas cosas.

> **Important**
>
> Además, la edición de reglas de los grupos de seguridad administrados puede tener consecuencias no deseadas. Es posible bloquear accidentalmente el tráfico necesario para que los clústeres funcionen correctamente y generar errores debido a que los nodos estén

inaccesibles. Planifique y pruebe cuidadosamente las configuraciones de los grupos de seguridad antes de su implementación.

Solo puede especificar grupos de seguridad al crear un clúster. No se pueden añadir a un clúster ni a las instancias de clúster mientras se está ejecutando un clúster, pero es posible editar, añadir y eliminar reglas de los grupos de seguridad existentes. Las reglas surtirán efecto tan pronto como se guarden.

Los grupos de seguridad son restrictivos de forma predeterminada. A menos que se añada una regla que permita el tráfico, el tráfico se rechaza. Si existe más de una regla que se aplica al mismo tráfico y al mismo origen, se aplica la regla más permisiva. Por ejemplo, si tiene una regla que permite el tráfico SSH desde la dirección IP 192.0.2.12/32 y otra regla que permite el acceso a todo el tráfico TCP desde el rango 192.0.2.0/24, tiene prioridad la regla que permite todo el tráfico TCP desde el rango que incluye 192.0.2.12. En este caso, el cliente en la dirección 192.0.2.12 podría tener más acceso del deseado.

Important

Actúe con precaución al editar las reglas de grupos de seguridad para abrir puertos. Asegúrese de agregar reglas que solo permitan el tráfico desde los clientes de confianza y autenticados para los protocolos y puertos necesarios para ejecutar las cargas de trabajo.

Puede configurar Bloquear acceso público de Amazon EMR en cada región que utilice para evitar la creación de clústeres si una regla permite el acceso público en algún puerto que no haya agregado a una lista de excepciones. En el caso de las cuentas de AWS creadas después de julio de 2019, Bloquear acceso público de Amazon EMR está activado de forma predeterminada. En el caso de las cuentas de AWS que crearon un clúster antes de julio de 2019, Bloquear acceso público de Amazon EMR está desactivado de forma predeterminada. Para obtener más información, consulte Uso de Bloquear el acceso público de Amazon EMR.

Temas

- Uso de grupos de seguridad administrados por Amazon EMR
- Uso de grupos de seguridad adicionales
- Especificación de los grupos de seguridad adicionales administrados por Amazon EMR
- Especificación de grupos de seguridad de EC2 para Cuadernos de Amazon EMR

- Uso de Bloquear el acceso público de Amazon EMR

> **Note**
>
> Amazon EMR tiene como objetivo utilizar alternativas inclusivas para términos industriales potencialmente ofensivos o no inclusivos, como “maestro” y “esclavo”. Hemos adoptado una nueva terminología para fomentar una experiencia más inclusiva y que así pueda entender mejor los componentes del servicio.
> Ahora describimos los “nodos” como instancias y describimos los tipos de instancias de Amazon EMR como instancias principales, básicas y de tareas. Durante la transición, es posible que siga encontrando referencias antiguas a términos obsoletos, como los que se refieren a los grupos de seguridad de Amazon EMR.

Uso de grupos de seguridad administrados por Amazon EMR

> **Note**
>
> Amazon EMR tiene como objetivo utilizar alternativas inclusivas para términos industriales potencialmente ofensivos o no inclusivos, como “maestro” y “esclavo”. Hemos adoptado una nueva terminología para fomentar una experiencia más inclusiva y que así pueda entender mejor los componentes del servicio.
> Ahora describimos los “nodos” como instancias y describimos los tipos de instancias de Amazon EMR como instancias principales, básicas y de tareas. Durante la transición, es posible que siga encontrando referencias antiguas a términos obsoletos, como los que se refieren a los grupos de seguridad de Amazon EMR.

Son varios los grupos de seguridad administrados que están asociados a la instancia principal y con las instancias secundarias y de tareas de un clúster. Se requiere un grupo de seguridad administrado adicional para el acceso al servicio al crear un clúster en una subred privada. Para obtener más información sobre la función de los grupos de seguridad administrados con respecto a la configuración de la red, consulte Opciones de Amazon VPC.

Cuando especifique grupos de seguridad administrados para un clúster, debe utilizar el mismo tipo de grupo de seguridad, predeterminado o personalizado, para todos los grupos. Por ejemplo, no

puede especificar un grupo de seguridad personalizado para la instancia principal y, acto seguido, no especificar un grupo de seguridad personalizado para las instancias secundarias y de tareas.

Si piensa utilizar los grupos de seguridad administrados predeterminados, no es necesario que los especifique al crear un clúster. Amazon EMR utiliza automáticamente los valores predeterminados. Además, si los valores predeterminados aún no existen en la VPC del clúster, Amazon EMR los crea. Amazon EMR también los crea si los especifica de forma explícita y aún no existen.

Puede editar reglas en los grupos de seguridad administrados una vez que se hayan creado los clústeres. Cuando se crea un clúster nuevo, Amazon EMR comprueba las reglas de los grupos de seguridad administrados que se especifican y, a continuación, crea las reglas entrantes que faltan y que el clúster nuevo necesita, junto con las reglas que se hayan podido agregar anteriormente. A menos que se indique lo contrario, cada regla para los grupos de seguridad administrados por Amazon EMR predeterminados también se agrega a los grupos de seguridad administrados por Amazon EMR personalizados que especifique.

Los grupos de seguridad administrados predeterminados son los siguientes:

- ElasticMapReduce-primary

  Para ver las reglas de este grupo de seguridad, consulte Grupo de seguridad administrado por Amazon EMR para la instancia principal (subredes públicas).
- ElasticMapReduce-core

  Para ver las reglas de este grupo de seguridad, consulte Grupo de seguridad administrado por Amazon EMR para las instancias básicas y de tareas (subredes públicas).
- ElasticMapReduce-Primary-Private

  Para ver las reglas de este grupo de seguridad, consulte Grupo de seguridad administrado por Amazon EMR para la instancia principal (subredes privadas).
- ElasticMapReduce-Core-Private

  Para ver las reglas de este grupo de seguridad, consulte Grupo de seguridad administrado por Amazon EMR para las instancias secundarias y de tareas (subredes privadas).
- ElasticMapReduce-ServiceAccess

  Para ver las reglas de este grupo de seguridad, consulte Grupo de seguridad administrado por Amazon EMR para el acceso de los servicios (subredes privadas).

Grupo de seguridad administrado por Amazon EMR para la instancia principal (subredes públicas)

El grupo de seguridad administrado predeterminado de la instancia principal en las subredes públicas tiene el Nombre de grupo ElasticMapReduce-primary. Tiene las siguientes reglas: Si especifica un grupo de seguridad administrado personalizado, Amazon EMR agrega las mismas reglas a su grupo de seguridad personalizado.

Tipo	Protoc	Rango de puerto	Origen	Detalles
Reglas de entrada				
Todo el ICMP IPv4	Todos	N/A	El ID de grupo del grupo de seguridad administrado para la instancia principal. En otras palabras, el mismo grupo de seguridad en el que aparece la regla.	Estas reglas reflexivas permiten el tráfico entrante desde cualquier instancia asociada al grupo de seguridad especificado. El uso del grupo de seguridad predeterminado `ElasticMapReduce-primary` para varios clústeres permite que los nodos secundarios y de tareas de dichos clústeres se comuniquen entre sí a través de ICMP o de cualquier puerto TCP o UDP. Especifique grupos de seguridad administrados personalizados para restringir el acceso entre clústeres.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		
Todo el ICMP IPV4	Todos	N/A	El ID de grupo del grupo de seguridad administrado especificado para los nodos secundarios y de tareas.	Estas reglas permiten todo el tráfico ICMP entrante y el tráfico a través de cualquier puerto TCP o UDP desde cualquier instancia secundaria y de tareas asociada al grupo de seguridad especificado, incluso si las instancias se encuentran en clústeres distintos.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		

Tipo	Protoc	Rango de puerto	Origen	Detalles
Personalizado	TCP	8443	Diversos rangos de direcciones IP de Amazon	Estas reglas permiten que el administrador del clúster se comunique con el nodo principal.

Para conceder a los orígenes de confianza acceso SSH al grupo de seguridad principal con la consola antigua

Para editar los grupos de seguridad, debe tener permiso para administrar los grupos de seguridad de la VPC en la que se encuentra el clúster. Para obtener más información, consulte Cambio de los permisos de un usuario y el Ejemplo de política que permite administrar los grupos de seguridad de EC2 en la Guía del usuario de IAM.

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Clusters (Clústeres). Seleccione el nombre del clúster que desea modificar.
3. Elija el enlace Grupos de seguridad para nodo maestro de Seguridad y acceso.
4. Elija ElasticMapReduce-master en la lista.
5. Elija la pestaña Reglas de entrada y, a continuación, elija Editar reglas de entrada.
6. Compruebe si hay una regla de entrada que permita el acceso público con la siguiente configuración. Si existe, seleccione Eliminar para eliminarla.

   - Tipo

     SSH
   - Puerto

     22
   - Fuente

     0.0.0.0/0 personalizado

> **Warning**
>
> Antes de diciembre de 2020, el grupo de seguridad ElasticMapReduce-master tenía una regla preconfigurada para permitir el tráfico entrante en el puerto 22 desde todos los orígenes. Esta regla se creó para simplificar las conexiones SSH iniciales al nodo principal. Le recomendamos encarecidamente que elimine esta regla de entrada y que restrinja el tráfico a los orígenes de confianza.

7. Desplácese a la parte inferior de la lista de reglas y seleccione Agregar regla.
8. En Type (Tipo), seleccione SSH.

   Al seleccionar SSH, se ingresa automáticamente TCP en Protocolo y 22 en Rango de puertos.
9. Como origen, seleccione Mi IP para agregar automáticamente su dirección IP como dirección de origen. También puede agregar un rango de direcciones IP de clientes de confianza personalizadas o crear reglas adicionales para otros clientes. Muchos entornos de red asignan direcciones IP de forma dinámica, por lo que es posible que en el futuro necesite actualizar las direcciones IP de los clientes de confianza.
10. Seleccione Save.
11. Si lo desea, puede seleccionar ElasticMapReduce-slave en la lista y repetir los pasos anteriores para permitir el acceso de los clientes SSH a los nodos principales y de tarea.

Grupo de seguridad administrado por Amazon EMR para las instancias básicas y de tareas (subredes públicas)

El grupo de seguridad administrado predeterminado de las instancias secundarias y de tareas de las subredes públicas tiene el Nombre de grupo ElasticMapReduce-core. El grupo de seguridad administrado predeterminado tiene las siguientes reglas, y Amazon EMR agrega las mismas reglas si especifica un grupo de seguridad administrado personalizado.

Tipo	Protoc	Rango de puerto	Origen	Detalles
Reglas de entrada				

Tipo	Protoc	Rango de puerto	Origen	Detalles
Todo el ICMP IPV4	Todos	N/A	El ID de grupo del grupo de seguridad administrado para las instancias secundarias y de tareas. En otras palabras, el mismo grupo de seguridad en el que aparece la regla.	Estas reglas reflexivas permiten el tráfico entrante desde cualquier instancia asociada al grupo de seguridad especificado. El uso del grupo de seguridad predeterminado `ElasticMapReduce-core` para varios clústeres permite que las instancias secundarias y de tareas de dichos clústeres se comuniquen entre sí a través de ICMP o de cualquier puerto TCP o UDP. Especifique grupos de seguridad administrados personalizados para restringir el acceso entre clústeres.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		
Todo el ICMP IPV4	Todos	N/A	El ID de grupo del grupo de seguridad administrado para la instancia principal.	Estas reglas permiten todo el tráfico ICMP entrante y el tráfico a través de cualquier puerto TCP o UDP desde cualquier instancia principal asociada al grupo de seguridad especificado, incluso si las instancias se encuentran en clústeres distintos.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		

Grupo de seguridad administrado por Amazon EMR para la instancia principal (subredes privadas)

El grupo de seguridad administrado predeterminado de la instancia principal en las subredes privadas tiene el Nombre de grupo ElasticMapReduce-Primary-Private. El grupo de seguridad administrado predeterminado tiene las siguientes reglas, y Amazon EMR agrega las mismas reglas si especifica un grupo de seguridad administrado personalizado.

Tipo	Protoc	Rango de puerto	Origen	Detalles
Reglas de entrada				
Todo el ICMP IPv4	Todos	N/A	El ID de grupo del grupo de seguridad administrado para la instancia principal. En otras palabras, el mismo grupo de seguridad en el que aparece la regla.	Estas reglas reflexivas permiten el tráfico entrante desde cualquier instancia asociada al grupo de seguridad especificado y que esté accesible desde dentro de la subred privada. El uso del grupo de seguridad predeterminado `ElasticMapReduce-Primary-Private` para varios clústeres permite que los nodos secundarios y de tareas de dichos clústeres se comuniquen entre sí a través de ICMP o de cualquier puerto TCP o UDP. Especifique grupos de seguridad administrados personalizados para restringir el acceso entre clústeres.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		
Todo el ICMP IPV4	Todos	N/A	El ID de grupo del grupo de seguridad administrado para los nodos secundarios y de tareas.	Estas reglas permiten todo el tráfico ICMP entrante y el tráfico a través de cualquier puerto TCP o UDP desde cualquier instancia secundaria y de tareas asociada al grupo de seguridad especificado y que esté accesible desde la subred privada, incluso si las instancias se encuentran en clústeres distintos.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		
HTTPS (8443)	TCP	8443	El ID de grupo del grupo de seguridad administrado para el acceso de los servicios en una subred privada.	Esta regla permite que el administrador del clúster se comunique con el nodo principal.
Reglas de salida				

Tipo	Protoc	Rango de puerto	Origen	Detalles
Todo el tráfico	Todos	Todos	0.0.0.0/0	Permite el acceso de salida a Internet.
TCP personalizada	TCP	9443	El ID de grupo del grupo de seguridad administrado para el acceso de los servicios en una subred privada.	Si se elimina la regla de salida predeterminada “Todo el tráfico” anterior, esta regla es un requisito mínimo para las versiones 5.30.0 y posteriores de Amazon EMR. **Note** Amazon EMR no agrega esta regla cuando utiliza un grupo de seguridad administrado personalizado.
TCP personalizada	TCP	80 (http) o 443 (https)	El ID de grupo del grupo de seguridad administrado para el acceso de los servicios en una subred privada.	Si se elimina la regla de salida predeterminada “Todo el tráfico” anterior, esta regla es un requisito mínimo para que las versiones 5.30.0 y posteriores de Amazon EMR se conecten a Amazon S3 a través de https. **Note** Amazon EMR no agrega esta regla cuando utiliza un grupo de seguridad administrado personalizado.

Grupo de seguridad administrado por Amazon EMR para las instancias secundarias y de tareas (subredes privadas)

El grupo de seguridad administrado predeterminado de las instancias secundarias y de tareas de las subredes privadas tiene el Nombre de grupo ElasticMapReduce-Core-Private. El grupo de seguridad

administrado predeterminado tiene las siguientes reglas, y Amazon EMR agrega las mismas reglas si especifica un grupo de seguridad administrado personalizado.

Tipo	Protoc	Rango de puerto	Origen	Detalles
Reglas de entrada				
Todo el ICMP IPV4	Todos	N/A	El ID de grupo del grupo de seguridad administrado para las instancias secundarias y de tareas. En otras palabras, el mismo grupo de seguridad en el que aparece la regla.	Estas reglas reflexivas permiten el tráfico entrante desde cualquier instancia asociada al grupo de seguridad especificado. El uso del grupo de seguridad predeterminado `ElasticMapReduce-core` para varios clústeres permite que las instancias secundarias y de tareas de dichos clústeres se comuniquen entre sí a través de ICMP o de cualquier puerto TCP o UDP. Especifique grupos de seguridad administrados personalizados para restringir el acceso entre clústeres.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		
Todo el ICMP IPV4	Todos	N/A	El ID de grupo del grupo de seguridad administrado para la instancia principal.	Estas reglas permiten todo el tráfico ICMP entrante y el tráfico a través de cualquier puerto TCP o UDP desde cualquier instancia principal asociada al grupo de seguridad especificado, incluso si las instancias se encuentran en clústeres distintos.
Todos los TCP	TCP	Todos		
Todo el UDP	UDP	Todos		
HTTPS (8443)	TCP	8443	El ID de grupo del grupo de seguridad administrado para el acceso de los servicios en una subred privada.	Esta regla permite que el administrador del clúster se comunique con los nodos secundarios y de tareas.

Tipo	Protoc	Rango de puerto	Origen	Detalles
Reglas de salida				
Todo el tráfico	Todos	Todos	0.0.0.0/0	Consulte Edición de reglas de salida a continuación.
TCP personalizada	TCP	80 (http) o 443 (https)	El ID de grupo del grupo de seguridad administrado para el acceso de los servicios en una subred privada.	Si se elimina la regla de salida predeterminada “Todo el tráfico” anterior, esta regla es un requisito mínimo para que las versiones 5.30.0 y posteriores de Amazon EMR se conecten a Amazon S3 a través de https. **Note** Amazon EMR no agrega esta regla cuando utiliza un grupo de seguridad administrado personalizado.

Edición de reglas de salida

De forma predeterminada, Amazon EMR crea este grupo de seguridad con reglas de salida que permiten todo el tráfico saliente en todos los protocolos y puertos. Se selecciona Permitir todo el tráfico saliente porque varias aplicaciones cliente y de Amazon EMR que se pueden ejecutar en clústeres de Amazon EMR pueden requerir reglas de salida diferentes. Amazon EMR no puede anticipar estos ajustes específicos al crear grupos de seguridad predeterminados. Puede limitar las salidas en sus grupos de seguridad para incluir solo las reglas que se adapten a sus casos de uso y políticas de seguridad. Como mínimo, este grupo de seguridad requiere las siguientes reglas de salida, pero es posible que algunas aplicaciones necesiten una salida adicional.

Tipo	Protoc	Rango de puerto	Destino	Detalles
Todos los TCP	TCP	Todos	pl-*xxxxxxxx*	Lista de prefijos de Amazon S3 administrados `com.amazonaws.` *MyRegion*`.s3`.
All Traffic	Todos	Todos	sg-*xxxxxxxxx xxxxxxxx*	El ID del grupo de seguridad `ElasticMapReduce-Core-Private` .
All Traffic	Todos	Todos	sg-*xxxxxxxxx xxxxxxxx*	El ID del grupo de seguridad `ElasticMapReduce-Primary-Private` .
TCP personalizada	TCP	9443	sg-*xxxxxxxxx xxxxxxxx*	El ID del grupo de seguridad `ElasticMapReduce-ServiceAccess` .

Grupo de seguridad administrado por Amazon EMR para el acceso de los servicios (subredes privadas)

El grupo de seguridad administrado predeterminado para el acceso de los servicios en las subredes privadas tiene el Group Name (Nombre de grupo) ElasticMapReduce-ServiceAccess. Tiene reglas de entrada y reglas de salida que permiten el tráfico a través de HTTPS (puerto 8443 o 9443) hacia los demás grupos de seguridad administrados en las subredes privadas. Estas reglas permiten que el administrador del clúster se comunique con el nodo principal y con nodos principales y de tarea. Se necesitan las mismas reglas si utiliza grupos de seguridad personalizados.

Tipo	Protoc	Rango de puerto	Origen	Detalles
Reglas de entrada necesarias para clústeres de Amazon EMR con versiones 5.30.0 y posteriores.				
TCP personalizada	TCP	9443	El ID de grupo del grupo de seguridad	Esta regla permite la comunicación entre el grupo de seguridad de la instancia principal y el grupo de seguridad de acceso al servicio.

Tipo	Protoc	Rango de puerto	Origen	Detalles
			administrado de la instancia principal.	
Reglas de salida necesarias para todos los clústeres de Amazon EMR				
TCP personalizada	TCP	8443	El ID de grupo del grupo de seguridad administrado de la instancia principal.	Estas reglas permiten que el administrador del clúster se comunique con el nodo principal y con nodos principales y de tarea.
TCP personalizada	TCP	8443	El ID de grupo del grupo de seguridad administrado para las instancias secundarias y de tareas.	Estas reglas permiten que el administrador del clúster se comunique con el nodo principal y con nodos principales y de tarea.

Uso de grupos de seguridad adicionales

Tanto si utiliza los grupos de seguridad administrados predeterminados como si especifica grupos de seguridad administrados personalizados, puede utilizar grupos de seguridad adicionales. Los grupos de seguridad adicionales le proporcionan la flexibilidad necesaria para adaptar el acceso entre diferentes clústeres y desde clientes, aplicaciones y recursos externos.

Considere el siguiente escenario de ejemplo. Dispone de varios clústeres y necesita que se comuniquen entre ellos, pero desea permitir el acceso SSH entrante a la instancia principal solo a un determinado subconjunto de clústeres. Para ello, puede utilizar el mismo conjunto de grupos de seguridad administrados para los clústeres. A continuación, debe crear grupos de seguridad adicionales que permitan el acceso SSH entrante desde los clientes de confianza y especificar los grupos de seguridad adicionales para la instancia principal de cada uno de los clústeres del subconjunto.

Puede aplicar hasta cuatro grupos de seguridad adicionales a la instancia principal, cuatro a las instancias secundarias y de tareas, y cuatro para el acceso al servicio (en subredes privadas).

Si fuera necesario, puede especificar el mismo grupo de seguridad adicional para las instancias principales, las instancias secundarias y de tareas y el acceso al servicio. El número máximo de grupos de seguridad y reglas de la cuenta está sujeto a los límites de la cuenta. Para obtener más información, consulte Límites de los grupos de seguridad en la Guía del usuario de Amazon VPC.

Especificación de los grupos de seguridad adicionales administrados por Amazon EMR

Puede especificar grupos de seguridad mediante la AWS Management Console, la AWS CLI, o la API de Amazon EMR. Si no especifica grupos de seguridad, Amazon EMR crea grupos de seguridad predeterminados. La especificación de grupos de seguridad adicionales es opcional. Puede asignar grupos de seguridad adicionales a las instancias principales, a las instancias secundarias y de tareas y al acceso de los servicios (solo para las subredes privadas).

New console

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

Para especificar grupos de seguridad con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Redes, seleccione la flecha situada junto a Grupos de seguridad de EC2 (firewall) para ampliar esta sección. En Nodo principal y en Nodos principales y de tareas, se seleccionan de forma predeterminada los grupos de seguridad administrados por Amazon EMR predeterminados. Si utiliza una subred privada, también tiene la opción de seleccionar un grupo de seguridad para Acceso de servicio.
4. Para cambiar el grupo de seguridad administrado por Amazon EMR, utilice el menú desplegable Elegir los grupos de seguridad para seleccionar una opción diferente de la lista de opciones del grupo de seguridad administrado por Amazon EMR. Dispone de un grupo

de seguridad administrado por Amazon EMR tanto para Nodo principal como para Nodos principales y de tareas.

5. Para agregar grupos de seguridad personalizados, utilice el mismo menú desplegable Elegir los grupos de seguridad para seleccionar hasta cuatro grupos de seguridad personalizados de la lista de opciones Grupos de seguridad personalizados. Puede tener hasta cuatro grupos de seguridad personalizados tanto para Nodo principal como para Nodos principales y de tareas.
6. Elija cualquier otra opción que se aplique a su clúster.
7. Para lanzar el clúster, elija Crear clúster.

Old console

Para especificar grupos de seguridad con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Crear clúster e Ir a las opciones avanzadas.
3. Elija las opciones para el clúster hasta que llegue al Step 4: Security (Paso 4: Seguridad).
4. Elija EC2 Security Groups (Grupos de seguridad de EC2) para expandir la sección.

   En EMR managed security groups (Grupos de seguridad administrados de EMR), los grupos de seguridad administrados predeterminados están seleccionados de forma predeterminada. Si no existe un valor predeterminado en la VPC para Master (Principal), Core & Task (Secundarias y de tareas) o Service Access (Acceso a los servicios) (solo subredes privadas), Create (Crear) aparece delante del nombre del grupo de seguridad asociado.
5. Si utiliza grupos de seguridad administrados personalizados, selecciónelos en las listas EMR managed security groups (Grupos de seguridad administrados de EMR).

   Si selecciona un grupo de seguridad administrado personalizado, un mensaje le indica que debe seleccionar un grupo de seguridad personalizado para las demás instancias. En un clúster, no puede mezclar grupos de seguridad administrados predeterminados y personalizados, debe elegir unos u otros.
6. Como opción, en Additional security groups (Grupos de seguridad adicionales), elija el icono de lápiz, seleccione un máximo de cuatro grupos de seguridad en la lista y, a continuación, elija Assign security groups (Asignar grupos de seguridad). Repita estos pasos para cada

instancia Master (Principal), Core & Task (Secundaria y de tareas) y Service Access (Acceso a los servicios).

7. Elija Create Cluster (Crear clúster).

Especificación de grupos de seguridad con la AWS CLI

Para especificar los grupos de seguridad mediante la AWS CLI, utilice el comando `create-cluster` con los siguientes parámetros para la opción `--ec2-attributes`:

Parámetro	Descripción
`EmrManagedPrimarySecurityGroup`	Utilice este parámetro para especificar un grupo de seguridad administrado personalizado para la instancia principal. Si se especifica este parámetro, también se deben especificar `EmrManagedCoreSecurityGroup`. Para clústeres en subredes privadas, también se debe especificar `ServiceAccessSecurityGroup`.
`EmrManagedCoreSecurityGroup`	Utilice este parámetro para especificar un grupo de seguridad administrado personalizado para las instancias secundarias y de tareas. Si se especifica este parámetro, también se deben especificar `EmrManagedPrimarySecurityGroup`. Para clústeres en subredes privadas, también se debe especificar `ServiceAccessSecurityGroup`.
`ServiceAccessSecurityGroup`	Utilice este parámetro para especificar un grupo de seguridad administrado personalizado para el acceso de los servicios, que se aplica únicamente a los clústeres de las subredes privadas. El grupo de seguridad que especifiq

Parámetro	Descripción
	ue como `ServiceAccessSecurityGroup` no debe usarse para ningún otro propósito y también debe reservarse para Amazon EMR. Si se especifica este parámetro, también se deben especificar `EmrManagedPrimaryS ecurityGroup` .
`AdditionalPrimarySecurityGroups`	Utilice este parámetro para especificar hasta cuatro grupos de seguridad adicionales para la instancia principal.
`AdditionalCoreSecurityGroups`	Utilice este parámetro para especificar hasta cuatro grupos de seguridad adicionales para las instancias secundarias y de tareas.

Example : especifique grupos de seguridad administrados por Amazon EMR personalizados y grupos de seguridad adicionales

En el siguiente ejemplo, se especifican grupos de seguridad administrados por Amazon EMR personalizados para un clúster en una subred privada, varios grupos de seguridad adicionales para la instancia principal y un único grupo de seguridad adicional para las instancias secundarias y de tareas.

> **Note**
>
> Se incluyen caracteres de continuación de línea de Linux (\) para facilitar la lectura. Se pueden eliminar o utilizar en los comandos de Linux. En Windows, elimínelos o sustitúyalos por un signo de intercalación (^).

```
aws emr create-cluster --name "ClusterCustomManagedAndAdditionalSGs" \
--release-label emr-emr-5.36.1 --applications Name=Hue Name=Hive \
Name=Pig --use-default-roles --ec2-attributes \
SubnetIds=subnet-xxxxxxxxxxxx,KeyName=myKey,\
ServiceAccessSecurityGroup=sg-xxxxxxxxxxxx,\
```

```
EmrManagedPrimarySecurityGroup=sg-xxxxxxxxxxx,\
EmrManagedCoreSecurityGroup=sg-xxxxxxxxxx,\
AdditionalPrimarySecurityGroups=['sg-xxxxxxxxxx',\
'sg-xxxxxxxxxx','sg-xxxxxxxxxx'],\
AdditionalCoreSecurityGroups=sg-xxxxxxxxxx \
--instance-type m5.xlarge
```

Para obtener más información, consulte create-cluster en la Referencia de comandos de la AWS CLI.

Especificación de grupos de seguridad de EC2 para Cuadernos de Amazon EMR

Al crear un cuaderno de EMR, se utilizan dos grupos de seguridad para controlar el tráfico de red entre el cuaderno de EMR y el clúster de Amazon EMR cuando se usa el editor de cuadernos. Los grupos de seguridad predeterminados tienen reglas mínimas que únicamente permiten el tráfico de red entre el servicio Cuadernos de Amazon EMR y los clústeres a los que están asociados los cuadernos.

Un cuaderno EMR utiliza Apache Livy para comunicarse con el clúster mediante un proxy a través del puerto TCP 18888. Cuando crea grupos de seguridad personalizados con reglas adaptadas al entorno, puede limitar el tráfico de la red de forma que solo un subconjunto de cuadernos pueda ejecutar código en el editor de cuadernos en determinados clústeres. El clúster utiliza su seguridad personalizada además de los grupos de seguridad predeterminados del clúster. Para obtener más información, consulte Control del tráfico de red con grupos de seguridad en la Guía de administración de Amazon EMR y Especificación de grupos de seguridad de EC2 para Cuadernos de Amazon EMR.

Grupo de seguridad de EC2 predeterminado para la instancia principal

El grupo de seguridad de EC2 predeterminado para la instancia principal está asociado a esta, junto con los grupos de seguridad del clúster para dicha instancia.

Nombre del grupo: ElasticMapReduceEditors-Livy

Reglas

- Entrada

  Permite el tráfico en el puerto TCP 18888 desde cualquier recurso en el grupo de seguridad de EC2 predeterminado para Cuadernos de Amazon EMR
- Salida

Ninguno

Grupo de seguridad de EC2 predeterminado para Cuadernos de Amazon EMR

El grupo de seguridad de EC2 predeterminado para el cuaderno de EMR está asociado al editor de cualquier cuaderno de EMR al que esté asignado.

Nombre del grupo: ElasticMapReduceEditors-Editor

Reglas

- Entrada

  Ninguno
- Salida

  Permite el tráfico en el puerto TCP 18888 a cualquier recurso en el grupo de seguridad de EC2 predeterminado para Cuadernos de Amazon EMR

Grupo de seguridad de EC2 personalizado para Cuadernos de Amazon EMR al asociar cuadernos con repositorios de Git

Para vincular un repositorio de Git a su cuaderno, el grupo de seguridad del cuaderno de EMR debe incluir una regla de salida para permitir que el cuaderno envíe tráfico a Internet. Se recomienda crear un nuevo grupo de seguridad para este fin. La actualización del grupo de seguridad predeterminado ElasticMapReduceEditors-Editor puede proporcionar las mismas reglas de salida a otros blocs de notas que estén asociados a este grupo de seguridad.

Reglas

- Entrada

  Ninguno
- Salida

  Permita que el cuaderno envíe tráfico a Internet a través del clúster, como se muestra en el siguiente ejemplo: El valor 0.0.0.0/0 se usa solo como ejemplo. Puede modificar esta regla para especificar las direcciones IP de sus repositorios basados en Git.

Tipo	Protocolo	Rango de puerto	Destino
Regla TCP personalizada	TCP	18888	SG-
HTTPS	TCP	443	0.0.0.0/0

Uso de Bloquear el acceso público de Amazon EMR

Bloquear el acceso público (BPA) de Amazon EMR le impide lanzar un clúster en una subred pública si el clúster tiene una configuración de seguridad que permite el tráfico entrante desde direcciones IP públicas en un puerto.

> **Important**
>
> Bloquear el acceso público está habilitado de forma predeterminada. Si desea aumentar la protección de la cuenta, le recomendamos que lo mantenga habilitado.

Explicación de Bloquear el acceso público

Puede utilizar la configuración de cuenta de Bloquear el acceso público para administrar de forma centralizada el acceso a la red pública a los clústeres de Amazon EMR.

Cuando un usuario de su Cuenta de AWS lanza un clúster, Amazon EMR comprueba las reglas de puerto del grupo de seguridad del clúster y las compara con las reglas de tráfico entrante. Si el grupo de seguridad tiene una regla de entrada que abre los puertos a las direcciones IP públicas IPv4 0.0.0.0/0 o IPv6 ::/0, y esos puertos no están especificados como excepciones para su cuenta, Amazon EMR no permite que el usuario cree el clúster.

Si un usuario modifica las reglas del grupo de seguridad de un clúster en ejecución en una subred pública para tener una regla de acceso público que infrinja la configuración de BPA de su cuenta, Amazon EMR revoca la nueva regla si tiene permiso para hacerlo. Si Amazon EMR no tiene permiso para revocar la regla, crea un evento en el panel de AWS Health que describe la infracción. Para conceder el permiso de revocación de la regla a Amazon EMR, consulte Configuración de Amazon EMR para revocar las reglas de los grupos de seguridad.

Bloquear acceso público está habilitado de forma predeterminada para todos los clústeres de cada Región de AWS de su Cuenta de AWS. BPA se aplica a todo el ciclo de vida de un clúster, pero no a los clústeres que se crean en subredes privadas. Puede configurar excepciones a la regla de BPA; el puerto 22 es una excepción de forma predeterminada. Para obtener más información sobre la configuración de excepciones, consulte Configuración de Bloquear el acceso público.

Configuración de Bloquear el acceso público

Puede actualizar los grupos de seguridad y la configuración de Bloquear el acceso público en sus cuentas en cualquier momento.

Puede activar y desactivar la configuración de Bloquear el acceso público (BPA) con la AWS Management Console, AWS Command Line Interface (AWS CLI) y la API de Amazon EMR. La configuración se aplica a toda su cuenta región por región. Para mantener la seguridad del clúster, le recomendamos usar BPA.

New console

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

Para configurar Bloquear el acceso público con la nueva consola

1. Inicie sesión en AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En la barra de navegación superior, seleccione la Región que quiera configurar si aún no está seleccionada.
3. En EMR en EC2, en el panel de navegación izquierdo, elija Bloquear el acceso público.
4. En Block public access settings (Configuración de Block Public Access), complete los pasos siguientes.

Para…	Haga lo siguiente...
Activar o desactivar Block Public Access	

Para…	Haga lo siguiente...
	Seleccione Editar, Activar o Desactiva r según corresponda y, a continuación, seleccione Guardar.
Editar puertos en la lista de excepciones	1. Seleccione Editar y busque la sección Excepciones de rango de puertos. 2. Para añadir puertos a la lista de excepciones, elija Add a port range (Añadir un rango de puertos) y escriba un nuevo puerto o rango de puertos. Repita para cada puerto o rango de puertos que desee añadir. 3. Para eliminar un puerto o rango de puertos, elija Eliminar, junto a la entrada en la lista de rangos de puertos. 4. Seleccione Save.

Old console

Para configurar Bloquear el acceso público con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/emr.
2. En la barra de navegación superior, asegúrese de que la Región que quiera configurar esté seleccionada.
3. Elija Block public access (Bloquear acceso público).
4. En Block public access settings (Configuración de Block Public Access), complete los pasos siguientes.

Para…	Haga lo siguiente...
Activar o desactivar Block Public Access	Elija Change (Cambiar), elija On (Activar) o Off (Desactivar) según corresponda y, a continuación, elija la marca de verificación que desea confirmar.
Editar puertos en la lista de excepciones	1. En Exceptions (Excepciones), elija Edit (Editar). 2. Para añadir puertos a la lista de excepciones, elija Add a port range (Añadir un rango de puertos) y escriba un nuevo puerto o rango de puertos. Repita para cada puerto o rango de puertos que desee añadir. 3. Para eliminar un puerto o rango de puertos, elija la x situado junto a la entrada en la lista Port ranges (Rangos de puertos). 4. Elija Guardar cambios.

AWS CLI

Para configurar Block Public Access mediante la AWS CLI

- Utilice el comando `aws emr put-block-public-access-configuration` para configurar Block Public Access, tal y como se muestra en los siguientes ejemplos.

Para…	Haga lo siguiente...
Activar Block Public Access	Defina `BlockPublicSecurityGroupRules` en `true` como se muestra en el ejemplo siguiente. Para que el clúster se lance, ningún grupo de seguridad asociado a un clúster puede tener una regla de entrada que permita el acceso público. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=true`
Desactivar Block Public Access	Defina `BlockPublicSecurityGroupRules` en `false` como se muestra en el ejemplo siguiente. Los grupos de seguridad asociados a un clúster pueden tener reglas de entrada que permitan el acceso público en cualquier puerto. No recomendamos esta configuración. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=false`

Para…	Haga lo siguiente...
Activar Block Public Access y especificar los puertos como excepciones	En el siguiente ejemplo se activa el bloqueo del acceso público y se especifica el puerto 22 y los puertos 100-101 como excepcion es. Esto permite crear clústeres si un grupo de seguridad asociado tiene una regla de entrada que permite el acceso público en los puertos 22, 100 o 101. `aws emr put-block-public-access-configuration --block-public-access-configuration '{ "BlockPublicSecurityGroupRules": true, "PermittedPublicSecurityGroupRuleRanges": [ { "MinRange": 22, "MaxRange": 22 }, { "MinRange": 100, "MaxRange": 101 } ] }'`

Configuración de Amazon EMR para revocar las reglas de los grupos de seguridad

Amazon EMR necesita permiso para revocar las reglas de los grupos de seguridad y cumplir con la configuración de Bloquear el acceso público. Puede seguir uno de estos métodos para conceder a Amazon EMR el permiso que necesita:

- Recomendado Asocie la política administrada `AmazonEMRServicePolicy_v2` al rol de servicio. Para obtener más información, consulte Rol de servicio para Amazon EMR (rol de EMR).
- Cree una nueva política insertada que permita realizar la acción `ec2:RevokeSecurityGroupIngress` en los grupos de seguridad. Para obtener más información sobre cómo modificar una política de permisos de roles, consulte Modificación de una política de permisos de rol con la consola de IAM, la API de AWS y AWS CLI en la Guía del usuario de IAM.

Resolución de infracciones de Bloquear el acceso público

Si se produce una infracción de Bloquear el acceso público, puede mitigarla con una de las siguientes acciones:

- Si desea acceder a una interfaz web de su clúster, utilice una de las opciones descritas en Ver las interfaces web alojadas en clústeres de Amazon EMR para acceder a la interfaz a través de SSH (puerto 22).
- Para permitir el tráfico al clúster desde direcciones IP específicas en lugar de desde la dirección IP pública, agregue una regla de grupo de seguridad. Para obtener más información, consulte Agregar reglas a un grupo de seguridad en la Guía de introducción de Amazon EC2.
- (No recomendado) Puede configurar las excepciones de BPA de Amazon EMR para que incluyan el puerto o el rango de puertos que desee. Al especificar una excepción de BPA, se crea un riesgo con un puerto desprotegido. Si planea especificar una excepción, debe eliminarla tan pronto como deje de ser necesaria. Para obtener más información, consulte Configuración de Bloquear el acceso público.

Identificación de los clústeres asociados a las reglas de los grupos de seguridad

Es posible que tenga que identificar todos los clústeres asociados a una regla de grupo de seguridad determinada o buscar la regla de grupo de seguridad de un clúster determinado.

- Si conoce el grupo de seguridad, puede identificar sus clústeres asociados si encuentra las interfaces de red del grupo de seguridad. Para obtener más información, consulte ¿Cómo puedo encontrar los recursos asociados a un grupo de seguridad de Amazon EC2? en AWS re:Post. Las instancias de Amazon EC2 que estén conectadas a estas interfaces de red se etiquetarán con el ID del clúster al que pertenecen.
- Si desea buscar los grupos de seguridad de un clúster conocido, siga los pasos que se indican en Ver el estado y los detalles del clúster. Puede encontrar los grupos de seguridad del clúster en el panel Red y seguridad de la consola o en el campo `Ec2InstanceAttributes` de la AWS CLI.

Validación de la conformidad de Amazon EMR

Auditores externos evalúan la seguridad y la conformidad de Amazon EMR en numerosos programas de conformidad de AWS. Estos incluyen SOC, PCI, FedRAMP, HIPAA y otros.

Para obtener una lista de AWS servicios en el ámbito de programas de cumplimiento específicos, consulte los AWS servicios en ámbito por programa de cumplimiento. Para obtener información general, consulte Programas de conformidad de AWS.

Puede descargar los informes de auditoría de terceros utilizando AWS Artifact. Para obtener más información, consulte Descarga de informes en AWS Artifact.

La responsabilidad de conformidad al utilizar Amazon EMR se determina en función de la confidencialidad de los datos, los objetivos de conformidad de la empresa y la legislación y normativa aplicables. Si su uso de Amazon EMR está sujeto a conformidad con normas tales como HIPAA, PCI o FedRAMP, AWS proporciona recursos que pueden servir de ayuda:

- Guías de inicio rápido de seguridad y conformidad: estas guías de implementación tratan consideraciones sobre arquitectura y ofrecen pasos para implementar los entornos de referencia centrados en la seguridad y la conformidad en AWS.
- Documento técnico sobre arquitectura para seguridad y conformidad de HIPAA: en este documento técnico, se describe cómo las empresas pueden utilizar AWS para crear aplicaciones conformes con la HIPAA.
- Recursos de conformidad de AWS: este conjunto de manuales y guías podría aplicarse a su sector y ubicación.
- AWS Config: este servicio de AWS evalúa en qué medida las configuraciones de los recursos cumplen las prácticas internas, las directrices del sector y la normativa.
- AWS Security Hub: este servicio de AWS proporciona una vista integral de su estado de seguridad en AWS que lo ayuda a verificar la conformidad con los estándares y las prácticas recomendadas del sector de seguridad.

Resiliencia en Amazon EMR

La infraestructura global de AWS se compone de regiones y zonas de disponibilidad de AWS. AWS Las regiones proporcionan varias zonas de disponibilidad físicamente independientes y aisladas que se encuentran conectadas mediante redes con un alto nivel de rendimiento y redundancia, además de baja latencia. Con las zonas de disponibilidad, puede diseñar y utilizar aplicaciones y bases de datos que realizan una conmutación por error automática entre zonas de disponibilidad sin interrupciones. Las zonas de disponibilidad tienen una mayor disponibilidad, tolerancia a errores y escalabilidad que las infraestructuras tradicionales de centros de datos únicos o múltiples.

Para obtener más información sobre las zonas de disponibilidad y las regiones de AWS, consulte Infraestructura global de AWS.

Además de la infraestructura global de AWS, Amazon EMR ofrece varias características que le brindan ayuda con sus necesidades de resiliencia y copia de seguridad de los datos.

- Integración con Amazon S3 a través de EMRFS
- Soporte para varios nodos principales

Seguridad de la infraestructura de Amazon EMR

Como servicio administrado, Amazon EMR está protegido por los procedimientos de seguridad de la red global de AWS que se detallan en el documento técnico Amazon Web Services: Overview of security processes.

Puede utilizar llamadas a la API publicadas en AWS para acceder a Amazon EMR a través de la red. Los clientes deben admitir el protocolo de seguridad de la capa de transporte (TLS) 1.2. Los clientes también deben ser compatibles con conjuntos de cifrado con confidencialidad directa total (PFS) tales como Ephemeral Diffie-Hellman (DHE) o Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). La mayoría de los sistemas modernos como Java 7 y posteriores son compatibles con estos modos.

Además, las solicitudes deben estar firmadas mediante un ID de clave de acceso y una clave de acceso secreta que esté asociada a una entidad de seguridad de IAM. También puede utilizar AWS Security Token Service (AWS STS) para generar credenciales de seguridad temporales para firmar solicitudes.

Temas

- Conexión a Amazon EMR mediante un punto de conexión de VPC de tipo interfaz

Conexión a Amazon EMR mediante un punto de conexión de VPC de tipo interfaz

Puede conectarse directamente a Amazon EMR mediante un punto de conexión de VPC de tipo interfaz (AWS PrivateLink) en su nube privada virtual (VPC) en lugar de conectarse a través de Internet. Cuando se utiliza un punto de conexión de VPC de tipo interfaz, la comunicación entre su VPC y Amazon EMR se realiza en su totalidad dentro de la red de AWS. Cada punto de conexión de

VPC está representado por una o varias Interfaces de red elásticas (ENI) con direcciones IP privadas en las subredes de la VPC.

El punto de conexión de VPC de tipo interfaz conecta directamente la VPC con Amazon EMR sin necesidad de contar con una puerta de enlace de Internet, dispositivos NAT, conexiones de VPN ni conexiones de AWS Direct Connect. Las instancias de la VPC no necesitan direcciones IP públicas para comunicarse con la API de Amazon EMR.

Para utilizar Amazon EMR a través de la VPC, debe conectarse desde una instancia que esté dentro de la VPC o conectar su red privada a la VPC a través de una red privada virtual (VPN) de Amazon o AWS Direct Connect. Para obtener más información sobre Amazon VPN, consulte Conexiones VPN en la Guía del usuario de Amazon Virtual Private Cloud. Para obtener información sobre AWS Direct Connect, consulte Creación de una conexión en la Guía del usuario de AWS Direct Connect.

Puede crear un punto de conexión de VPC de tipo interfaz para conectarse a Amazon EMR mediante la consola de AWS o los comandos de AWS Command Line Interface (AWS CLI). Para obtener más información, consulte Creación de un punto de conexión de interfaz.

Después de crear un punto de conexión de VPC de tipo interfaz, si habilita nombres de host DNS privados para el punto de conexión, el punto de conexión predeterminado de Amazon EMR se resuelve en el punto de conexión de VPC. El punto de conexión del nombre de servicio predeterminado de Amazon EMR tiene el siguiente formato:

```
elasticmapreduce.Region.amazonaws.com
```

Si no habilita nombres de host de DNS privados, Amazon VPC proporciona un nombre de punto de conexión de DNS que puede utilizar en el siguiente formato.

```
VPC_Endpoint_ID.elasticmapreduce.Region.vpce.amazonaws.com
```

Para obtener más información, consulte Puntos de enlace de la VPC de tipo interfaz (AWS PrivateLink) en la Guía del usuario de Amazon VPC.

Amazon EMR permite llamar a todas las acciones de la API en su VPC.

Puede adjuntar políticas de punto de conexión de VPC a un punto de conexión de VPC para controlar el acceso de las entidades principales de IAM. También puede asociar grupos de seguridad con un punto de enlace de la VPC para controlar el acceso de entrada y salida en función del origen

y el destino del tráfico de red, como un rango de direcciones IP. Para obtener más información, consulte Control del acceso a los servicios con puntos de conexión de VPC.

Creación de una política de punto de conexión de VPC para Amazon EMR

Puede crear una política para los puntos de conexión de VPC de Amazon EMR y especificar lo siguiente:

- La entidad principal que puede o no puede realizar acciones
- Las acciones que se pueden realizar
- Los recursos en los que se pueden llevar a cabo las acciones

Para obtener más información, consulte Control del acceso a los servicios con puntos de conexión de VPC en la guía del usuario de Amazon VPC.

Example : política de punto de conexión de VPC para denegar el acceso desde una cuenta de AWS especificada.

La siguiente política de punto de conexión de VPC deniega a la cuenta de AWS *123456789012* todo el acceso a los recursos mediante el punto de conexión.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "*",
            "Effect": "Deny",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Example : política de punto de conexión de VPC para permitir el acceso de VPC solo a una entidad principal de IAM especificada (usuario).

La siguiente política de punto de conexión de VPC permite el acceso pleno solo al usuario *lijuan* en la cuenta de AWS *123456789012*. A las demás entidades principales de IAM se les deniega el acceso a través del punto de enlace.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "arn:aws:iam::123456789012:user/lijuan"
                ]
            }
        }]
}
```

Example : política de punto de conexión de VPC para permitir operaciones de EMR de solo lectura.

La siguiente política de punto de conexión de VPC solo permite a la cuenta de AWS *123456789012* realizar las acciones de Amazon EMR especificadas.

Las acciones especificadas proporcionan el equivalente al acceso de solo lectura para Amazon EMR. Las demás acciones en la VPC se deniegan para la cuenta especificada. A las demás cuentas se les deniega el acceso. Para obtener una lista de acciones de Amazon EMR, consulte Acciones, recursos y claves de condición de Amazon EMR.

```
{
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:DescribeSecurityConfiguration",
                "elasticmapreduce:GetBlockPublicAccessConfiguration",
                "elasticmapreduce:ListBootstrapActions",
                "elasticmapreduce:ViewEventsFromAllClustersInConsole",
                "elasticmapreduce:ListSteps",
                "elasticmapreduce:ListInstanceFleets",
                "elasticmapreduce:DescribeCluster",
```

```
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:DescribeStep",
                "elasticmapreduce:ListInstances",
                "elasticmapreduce:ListSecurityConfigurations",
                "elasticmapreduce:DescribeEditor",
                "elasticmapreduce:ListClusters",
                "elasticmapreduce:ListEditors"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Example : política de punto de conexión de VPC que deniega el acceso a un clúster específico.

La siguiente política de punto de conexión de VPC permite el acceso total a todas las cuentas y entidades principales, pero deniega el acceso de la cuenta de AWS *123456789012* a las acciones hechas en el clúster de Amazon EMR con el ID de clúster *j-A1B2CD34EF5G*. Se siguen permitiendo otras acciones de Amazon EMR que no admiten permisos de recursos para los clústeres. Para obtener una lista de acciones de Amazon EMR y su tipo de recurso correspondiente, consulte Acciones, recursos y claves de condición para Amazon EMR.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "*",
            "Effect": "Deny",
            "Resource": "arn:aws:elasticmapreduce:us-west-2:123456789012:cluster/j-
A1B2CD34EF5G",
            "Principal": {
                "AWS": [
```

```
                    "123456789012"
                ]
            }
        }
    ]
}
```

Administrar clústeres

Después de haber lanzado el clúster, puede monitorearlo y administrarlo. Amazon EMR proporciona varias herramientas que puede utilizar para conectar y controlar el clúster.

Temas

- Conectarse a un clúster
- Enviar trabajo a un clúster
- Ver y monitorizar un clúster
- Usar el escalado de clústeres
- Terminar un clúster
- Clonación de un clúster con la consola
- Automatizar clústeres periódicos con AWS Data Pipeline

Conectarse a un clúster

Cuando ejecute un clúster de Amazon EMR, a menudo lo único que tendrá que hacer es ejecutar una aplicación para analizar los datos y, a continuación, recopilar la salida desde un bucket de Amazon S3. En otras ocasiones, es posible que desee interactuar con el nodo principal mientras que el clúster se está ejecutando. Por ejemplo, es posible que desee conectar con el nodo principal para ejecutar consultas interactivas, comprobar archivos de registro, depurar un problema con el clúster, monitorizar el rendimiento mediante una aplicación como Ganglia que se ejecuta en el nodo principal, etc. Las secciones siguientes describen técnicas que puede utilizar para conectarse al nodo principal.

En un clúster de EMR, el nodo principal es una instancia Amazon EC2 que coordina las instancias de EC2 que se ejecutan como nodos principales y de tareas. El nodo principal expone un nombre de DNS pública que puede utilizar para conectarse al mismo. De forma predeterminada, Amazon EMR crea reglas de grupo de seguridad para el nodo principal, los nodos secundarios y los nodos de tarea que determinan cómo se obtiene acceso a los nodos.

> **Note**
>
> Puede conectarse al nodo principal solo mientras se está ejecutando el clúster. Cuando se termina el clúster, la instancia de EC2 que actúa como nodo principal se termina y ya no está

disponible. Para conectarse al nodo principal, también debe autenticarse en el clúster. Puede utilizar Kerberos para la autenticación o especificar una clave privada del par de claves de Amazon EC2 al lanzar el clúster. Para obtener más información sobre la configuración de Kerberos y, a continuación, conectar, consulte Uso de Kerberos para la autenticación con Amazon EMR. Cuando lanza un clúster desde la consola, la clave privada del par de claves de Amazon EC2 se especifica en la sección Seguridad y acceso de la página Crear clúster.

De forma predeterminada, el grupo de seguridad ElasticMapReduce-master no permite el acceso de SSH entrante. Es posible que tenga que añadir una regla entrante que permita el acceso de SSH (TCP puerto 22) desde los orígenes a los que desea tener acceso. Para obtener más información acerca de la modificación de reglas de los grupos de seguridad, consulte Agregar reglas a un grupo de seguridad en la Guía del usuario de Amazon EC2 para instancias de Linux.

Important

No modifique las reglas restantes en el grupo de seguridad ElasticMapReduce-master. La modificación de estas reglas podría interferir con la operación del clúster.

Temas

- Antes de conectarse: autorice el tráfico entrante
- Conectarse al nodo principal mediante SSH

Antes de conectarse: autorice el tráfico entrante

Antes de conectarse a un clúster de Amazon EMR, debe autorizar el tráfico SSH entrante (puerto 22) procedente de clientes de confianza, como la dirección IP de su ordenador. Para ello, edite las reglas del grupo de seguridad administrado de los nodos a los que desee conectarse. Por ejemplo, en las siguientes instrucciones se muestra cómo agregar una regla de entrada para el acceso SSH al grupo de seguridad ElasticMapReduce-master predeterminado.

Para obtener más información acerca del uso de los grupos de seguridad con Amazon EMR, consulte Control del tráfico de red con grupos de seguridad.

New console

Para conceder a los orígenes de confianza acceso SSH al grupo de seguridad principal con la nueva consola

Para editar los grupos de seguridad, debe tener permiso para administrar los grupos de seguridad de la VPC en la que se encuentra el clúster. Para obtener más información, consulte Cambio de los permisos de un usuario y el Ejemplo de política que permite administrar los grupos de seguridad de EC2 en la Guía del usuario de IAM.

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y, a continuación, seleccione el clúster que desee actualizar. Se abrirá la página de detalles del clúster. Se preseleccionará la pestaña Propiedades de esta página.
3. En Redes, en la pestaña Propiedades, seleccione la flecha situada junto a Grupos de seguridad de EC2 (firewall) para ampliar esta sección. En Nodo principal, seleccione el enlace del grupo de seguridad. Se abrirá la consola de EC2.
4. Elija la pestaña Reglas de entrada y, a continuación, Editar reglas de entrada.
5. Compruebe si hay una regla de entrada que permita el acceso público con la siguiente configuración. Si existe, seleccione Eliminar para eliminarla.

   - Tipo

     SSH
   - Puerto

     22
   - Fuente

     0.0.0.0/0 personalizado

> **Warning**
>
> Antes de diciembre de 2020, el grupo de seguridad ElasticMapReduce-master tenía una regla preconfigurada para permitir el tráfico entrante en el puerto 22 desde todos los orígenes. Esta regla se creó para simplificar las conexiones SSH iniciales al nodo

principal. Le recomendamos encarecidamente que elimine esta regla de entrada y que restrinja el tráfico a los orígenes de confianza.

6. Desplácese a la parte inferior de la lista de reglas y seleccione Agregar regla.
7. En Type (Tipo), seleccione SSH. Esta selección introduce automáticamente TCP en Protocolo y 22 en Rango de puertos.
8. Como origen, seleccione Mi IP para agregar automáticamente su dirección IP como dirección de origen. También puede agregar un rango de direcciones IP de clientes de confianza personalizadas o crear reglas adicionales para otros clientes. Muchos entornos de red asignan direcciones IP de forma dinámica, por lo que es posible que en el futuro necesite actualizar las direcciones IP de los clientes de confianza.
9. Seleccione Save.
10. Si lo desea, vuelva al paso 3, seleccione Nodos principales y de tarea y repita los pasos 4 a 8. Esto otorga acceso al cliente SSH a los nodos principales y de tarea.

Old console

Para conceder a los orígenes de confianza acceso SSH al grupo de seguridad principal con la consola antigua

Para editar los grupos de seguridad, debe tener permiso para administrar los grupos de seguridad de la VPC en la que se encuentra el clúster. Para obtener más información, consulte Cambio de los permisos de un usuario y el Ejemplo de política que permite administrar los grupos de seguridad de EC2 en la Guía del usuario de IAM.

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione Clusters (Clústeres). Seleccione el nombre del clúster que desea modificar.
3. Elija el enlace Grupos de seguridad para nodo maestro de Seguridad y acceso.
4. Elija ElasticMapReduce-master en la lista.
5. Elija la pestaña Reglas de entrada y, a continuación, elija Editar reglas de entrada.
6. Compruebe si hay una regla de entrada que permita el acceso público con la siguiente configuración. Si existe, seleccione Eliminar para eliminarla.

   - Tipo

SSH

- Puerto

  22

- Fuente

  0.0.0.0/0 personalizado

> **Warning**
>
> Antes de diciembre de 2020, el grupo de seguridad ElasticMapReduce-master tenía una regla preconfigurada para permitir el tráfico entrante en el puerto 22 desde todos los orígenes. Esta regla se creó para simplificar las conexiones SSH iniciales al nodo principal. Le recomendamos encarecidamente que elimine esta regla de entrada y que restrinja el tráfico a los orígenes de confianza.

7. Desplácese a la parte inferior de la lista de reglas y seleccione Agregar regla.
8. En Type (Tipo), seleccione SSH.

   Al seleccionar SSH, se ingresa automáticamente TCP en Protocolo y 22 en Rango de puertos.
9. Como origen, seleccione Mi IP para agregar automáticamente su dirección IP como dirección de origen. También puede agregar un rango de direcciones IP de clientes de confianza personalizadas o crear reglas adicionales para otros clientes. Muchos entornos de red asignan direcciones IP de forma dinámica, por lo que es posible que en el futuro necesite actualizar las direcciones IP de los clientes de confianza.
10. Seleccione Save.
11. Si lo desea, puede seleccionar ElasticMapReduce-slave en la lista y repetir los pasos anteriores para permitir el acceso de los clientes SSH a los nodos principales y de tarea.

Conectarse al nodo principal mediante SSH

Secure Shell (SSH) es un protocolo de red que puede utilizar para crear una conexión segura a un equipo remoto. Después de realizar una conexión, el terminal en su equipo local se comporta como si se estuviera ejecutando en el equipo remoto. Los comandos que emita a nivel local se ejecutan

en el equipo remoto y la salida de comandos desde el equipo remoto aparece en la ventana del terminal.

Al utilizar SSH con AWS, puede conectarse a una instancia de EC2, que es un servidor virtual que se ejecuta en la nube. Al trabajar con Amazon EMR, el uso más común de SSH consiste en conectarse a la instancia de EC2 que actúa como nodo principal del clúster.

El uso de SSH para conectarse al nodo principal le ofrece la posibilidad de monitorizar e interactuar con el clúster. Puede emitir comandos de Linux en el nodo principal, ejecutar aplicaciones como, por ejemplo, Hive y Pig de forma interactiva, examinar directorios, leer archivos de registro, etc. También puede crear un túnel en la conexión SSH para ver las interfaces web alojadas en el nodo principal. Para obtener más información, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Para conectar con el nodo principal mediante SSH, necesita el nombre de DNS público del nodo principal. Además, el grupo de seguridad asociado al nodo principal debe tener una regla de entrada que permita el tráfico SSH (puerto TCP 22) desde un origen que incluya el cliente donde se origina la conexión SSH. Es posible que tenga que añadir una regla para permitir una conexión SSH de su cliente. Para obtener más información acerca de la modificación de reglas de los grupos de seguridad, consulte Control del tráfico de red con grupos de seguridad y Agregar reglas a un grupo de seguridad en la Guía del usuario de Amazon EC2 para instancias de Linux.

Recuperar el nombre de DNS público del nodo principal

Puede recuperar el nombre de DNS público principal utilizando la consola de Amazon EMR y la AWS CLI.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para recuperar el nombre de DNS público del nodo principal con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.

2. En EMR en EC2, en el panel de navegación izquierdo, seleccione Clústeres y seleccione el clúster en el que desee recuperar el nombre de DNS público.
3. Anote el valor del DNS público del nodo principal en la sección Resumen de la página de detalles del clúster.

Old console

Para recuperar el nombre de DNS público del nodo principal con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. En la página Cluster List (Lista de clústeres), seleccione el enlace para el clúster.
3. Tenga en cuenta el valor del DNS público principal que aparece en la sección Resumen de la página de detalles del clúster.

> **Note**
>
> También puede elegir el enlace SSH para obtener instrucciones sobre cómo crear una conexión SSH con el nodo principal.

CLI

Para recuperar el nombre de DNS público del nodo principal con la AWS CLI

1. Para recuperar el identificador del clúster, escriba el siguiente comando.

```
aws emr list-clusters
```

La salida enumera los clústeres, incluidos los ID de los clústeres. Tenga en cuenta el ID del clúster al que se está conectando.

```
"Status": {
    "Timeline": {
        "ReadyDateTime": 1408040782.374,
        "CreationDateTime": 1408040501.213
    },
```

```
    "State": "WAITING",
    "StateChangeReason": {
        "Message": "Waiting after step completed"
    }
},
"NormalizedInstanceHours": 4,
"Id": "j-2AL4XXXXXX5T9",
"Name": "My cluster"
```

2. Para obtener una lista de las instancias de clúster incluido el nombre de DNS público para el clúster, escriba uno de los siguientes comandos. Sustituya *j-2AL4XXXXXX5T9* por el ID del clúster devuelto por el comando anterior.

```
aws emr list-instances --cluster-id j-2AL4XXXXXX5T9
```

O bien:

```
aws emr describe-cluster --cluster-id j-2AL4XXXXXX5T9
```

La salida enumera las instancias de clúster, incluidos nombres de DNS y las direcciones IP. Anote el valor para `PublicDnsName`.

```
"Status": {
    "Timeline": {
        "ReadyDateTime": 1408040779.263,
        "CreationDateTime": 1408040515.535
    },
    "State": "RUNNING",
    "StateChangeReason": {}
},
"Ec2InstanceId": "i-e89b45e7",
"PublicDnsName": "ec2-###-##-##-###.us-west-2.compute.amazonaws.com"

"PrivateDnsName": "ip-###-##-##-###.us-west-2.compute.internal",
"PublicIpAddress": "##.###.###.##",
"Id": "ci-12XXXXXXXXFMH",
"PrivateIpAddress": "###.##.#.###"
```

Para más información, consulte Comandos de Amazon EMR en la AWS CLI.

Conectarse al nodo principal mediante SSH y una clave privada de Amazon EC2 en Linux, Unix y Mac OS X

Para crear una conexión SSH autenticada con un archivo de clave privada, tendrá que especificar la clave privada del par de claves de Amazon EC2 al lanzar un clúster. Para obtener más información sobre el acceso al par de claves, consulte Pares de claves de Amazon EC2 en la Guía del usuario de Amazon EC2 para instancias de Linux.

Su equipo Linux muy probablemente incluye un cliente SSH de forma predeterminada. Por ejemplo, OpenSSH está instalado en la mayoría de los sistemas operativos Linux, Unix y macOS. Puede comprobar si tiene un cliente SSH escribiendo ssh en la línea de comandos. Si su equipo no reconoce el comando, instale un cliente SSH para conectarse al nodo principal. El proyecto OpenSSH ofrece una implementación gratuita de toda la suite de herramientas de SSH. Para obtener más información, consulte el sitio web OpenSSH.

Las siguientes instrucciones muestran la apertura de una conexión SSH al nodo principal de Amazon EMR en Linux, Unix y Mac OS X.

Para configurar los permisos de archivo de clave privada del par de claves

Antes de poder utilizar la clave privada del par de claves de Amazon EC2 para crear una conexión SSH, debe establecer permisos en el archivo `.pem` de modo que solo el propietario de la clave tenga permiso para acceder a él. Esto es necesario para crear una conexión SSH utilizando el terminal o la AWS CLI.

1. Asegúrese de permitir el tráfico SSH entrante. Para obtener instrucciones, consulte Antes de conectarse: autorice el tráfico entrante.
2. Localice su archivo `.pem`. Estas instrucciones suponen que el archivo se denomina `mykeypair.pem` y que se almacena en el directorio de inicio del usuario actual.
3. Escriba el comando siguiente para definir los permisos. Sustituya *~/mykeypair.pem* por la ruta de acceso completa y el nombre del archivo de clave privada del par de claves. Por ejemplo `C:/Users/<username>/.ssh/mykeypair.pem`.

   ```
   chmod 400 ~/mykeypair.pem
   ```

   Si no establece permisos en el archivo `.pem`, recibirá un error indicando que el archivo de clave no está protegido y la clave se rechazará. Para conectarse, solo tiene que establecer permisos en el archivo de clave privada del par de claves la primera vez que lo use.

Para conectarse al nodo principal utilizando el terminal

1. Abra una ventana de terminal. En Mac OS X, elija Aplicaciones > Utilidades > Terminal. En otras distribuciones de Linux, el terminal normalmente se encuentra en Aplicaciones > Accesorios > Terminal.
2. Para establecer una conexión al nodo principal, escriba el siguiente comando. Sustituya *ec2-###-##-##-###.compute-1.amazonaws.com* por el nombre de DNS público principal del clúster y sustituya *~/mykeypair.pem* por la ruta de acceso completa y el nombre del archivo `.pem`. Por ejemplo `C:/Users/<username>/.ssh/mykeypair.pem`.

   ```
   ssh hadoop@ec2-###-##-##-###.compute-1.amazonaws.com -i ~/mykeypair.pem
   ```

   > **Important**
   >
   > Debe utilizar el nombre de inicio de sesión `hadoop` cuando se conecte al nodo principal de Amazon EMR; de lo contrario, es posible que vea un error similar a `Server refused our key`.

3. Una advertencia indica que la autenticidad del host al que se está conectando no se puede verificar. Escriba `yes` para continuar.
4. Cuando haya terminado de trabajar en el nodo principal, escriba el siguiente comando para cerrar la conexión SSH.

   ```
   exit
   ```

Si tiene dificultades para usar SSH para conectarse a su nodo principal, consulte Solucionar problemas de conexión a su instancia.

Conectarse al nodo principal mediante SSH en Windows

Los usuarios de Windows pueden utilizar un cliente SSH como PuTTY para conectarse al nodo principal. Antes de conectarse al nodo principal de Amazon EMR, debe descargar e instalar PuTTY y PuTTYgen. Puede descargar estas herramientas desde la página de descarga de PuTTY.

PuTTY no admite de forma nativa el formato de archivo de clave privada del par de claves (`.pem`) generado por Amazon EC2. Utilice PuTTYgen para convertir su archivo de clave al formato

de PuTTY requerido (`.ppk`). Debe convertir su clave en este formato (`.ppk`) antes de intentar conectarse al nodo principal utilizando PuTTY.

Para obtener más información sobre la conversión de la clave, consulte Conversión de la clave privada mediante PuTTYgen en la Guía del usuario de Amazon EC2 para instancias de Linux.

Para conectarse al nodo principal utilizando PuTTY

1. Asegúrese de permitir el tráfico SSH entrante. Para obtener instrucciones, consulte Antes de conectarse: autorice el tráfico entrante.
2. Abrir `putty.exe`. También puede lanzar PuTTY desde la lista de programas de Windows.
3. Si es necesario, en la lista Category (Categoría), elija Session (Sesión).
4. En Host Name (or IP address) (Nombre de host (o dirección IP)), escriba `hadoop@`*`MasterPublicDNS`*. Por ejemplo, `hadoop@`*`ec2-###-##-##-###.compute-1.amazonaws.com`*.
5. En la lista Category (Categoría), elija Connection > SSH (Conexión > SSH), Auth (Autenticación).
6. En Private key file for authentication (Archivo de clave privada para la autenticación), elija Browse (Examinar) y seleccione el archivo `.ppk` que ha generado.
7. Selecciona Abrir y, a continuación, Sí para descartar la alerta de seguridad de PuTTY.

> **Important**
>
> Al iniciar sesión en el nodo principal, escriba `hadoop` si se le solicita un nombre de usuario.

8. Cuando haya terminado de trabajar en el nodo principal, puede cerrar la conexión SSH cerrando PuTTY;.

> **Note**
>
> Para evitar que se supere el tiempo de espera de la conexión SSH, puede elegir Connection (Conexión) en la lista Category (Categoría) y seleccionar la opción Enable TCP_keepalives (Habilitar conexiones keepalives de TCP). Si dispone de una sesión de SSH activa en PuTTY, puede cambiar la configuración abriendo el contexto (clic con el botón derecho) en la barra del título de PuTTY y seleccionar Cambiar configuración.

Si tiene dificultades para usar SSH para conectarse a su nodo principal, consulte Solucionar problemas de conexión a su instancia.

Conectarse al nodo principal utilizando la AWS CLI

Puede crear una conexión SSH con el nodo principal utilizando la AWS CLI en Windows y en Linux, Unix y Mac OS X. Independientemente de la plataforma, necesita el nombre de DNS público del nodo principal y la clave privada del par de claves de Amazon EC2. Si utiliza la AWS CLI en Linux, Unix o Mac OS X, también debe establecer permisos en el archivo de clave privada (`.pem` o `.ppk`), tal como se muestra en Para configurar los permisos de archivo de clave privada del par de claves.

Para conectarse al nodo principal utilizando la AWS CLI

1. Asegúrese de permitir el tráfico SSH entrante. Para obtener instrucciones, consulte Antes de conectarse: autorice el tráfico entrante.
2. Para recuperar el identificador del clúster, escriba:

   ```
   aws emr list-clusters
   ```

   La salida enumera los clústeres, incluidos los ID de los clústeres. Tenga en cuenta el ID del clúster al que se está conectando.

   ```
   "Status": {
       "Timeline": {
           "ReadyDateTime": 1408040782.374,
           "CreationDateTime": 1408040501.213
       },
       "State": "WAITING",
       "StateChangeReason": {
           "Message": "Waiting after step completed"
       }
   },
   "NormalizedInstanceHours": 4,
   "Id": "j-2AL4XXXXXX5T9",
   "Name": "AWS CLI cluster"
   ```

3. Escriba el comando siguiente para abrir una conexión SSH al nodo principal. En el ejemplo siguiente, sustituya *j-2AL4XXXXXX5T9* por el ID del clúster y sustituya *~/mykeypair.key* por la ruta de acceso completa y el nombre de archivo del archivo `.pem` (para Linux,

Unix y Mac OS X) o `.ppk` (para Windows). Por ejemplo `C:\Users\<username>\.ssh\mykeypair.pem`.

```
aws emr ssh --cluster-id j-2AL4XXXXXX5T9 --key-pair-file ~/mykeypair.key
```

4. Cuando haya terminado de trabajar en el nodo principal, cierre la ventana de la AWS CLI.

Para más información, consulte Comandos de Amazon EMR en la AWS CLI. Si tiene dificultades para usar SSH para conectarse a su nodo principal, consulte Solucionar problemas de conexión a su instancia.

Puertos de servicio de Amazon EMR

> **Note**
>
> A continuación se muestran las interfaces y los puertos de servicio para los componentes de Amazon EMR. Esta no es una lista completa de los puertos de servicio. Los servicios no predeterminados, como los puertos SSL y los distintos tipos de protocolos, no aparecen en la lista.

> **Important**
>
> Actúe con precaución al editar las reglas de grupos de seguridad para abrir puertos. Asegúrese de agregar reglas que solo permitan el tráfico desde los clientes de confianza y autenticados para los protocolos y puertos necesarios para ejecutar las cargas de trabajo.

Componente	Descripción del servicio	El servicio se ejecuta de forma predeterminada	Port	Clave de configuración
Hadoop	API de REST de KMS HTTP	Sí	9600	hadoop.kms.http.port
HDFS	Interfaz de usuario web de namenode	Sí	9870	dfs.namenode.http-address

Componente	Descripción del servicio	El servicio se ejecuta de forma predeterminada	Port	Clave de configuración
	RPC de namenode	Sí	8020	dfs.namenode.rpc-address
	Interfaz de usuario web de DataNode	Sí	9864	dfs.datanode.http.address
	Datanode HTTP para la transferencia de datos	Sí	9866	dfs.datanode.address
	RPC de datanode para la transferencia de datos	Sí	9867	dfs.datanode.ipc.address
Hive	HiveServer2 Thrift	Sí	10000	hive.server2.thrift.port
	HiveServer2 HTTP	No	10001	hive.server2.thrift.http.port
	Interfaz de usuario web de HiveServer2	Sí	10002	hive.server2.webui.port
	Metaalmacén de Hive	Sí	9083	hive.metastore.port / metastore.thrift.port
	WebHCat	No	50111	templeton.port
	Servicio de administración de daemons (RPC) de LLAP	No	15004	hive.llap.management.rpc.port

Componente	Descripción del servicio	El servicio se ejecuta de forma predeterminada	Port	Clave de configuración
	Puerto de mezclas aleatorias de YARN para las mezclas aleatorias alojadas en daemons de LLAP	No	15551	hive.llap.daemon.yarn.shuffle.port
	RPC del daemon de LLAP	No	Dinámico	hive.llap.daemon.rpc.port
	Interfaz web del daemon de LLAP	No	15002	hive.llap.daemon.web.port
	Servicio de salida del daemon de LLAP	No	15003	hive.llap.daemon.output.service.port
Oozie		Sí	11000	
Tez	Tez UI	Sí	8080	
YARN	Mezclas aleatorias	Sí	13562	mapreduce.shuffle.port
	RPC del localizador	Sí	8040	yarn.nodemanager.localizer.address
		Sí	8041	
	Dirección de la aplicación web de NM	Sí	8042	yarn.nodemanager.webapp.address

Componente	Descripción del servicio	El servicio se ejecuta de forma predeterminada	Port	Clave de configuración
	Aplicación web de RM	Sí	8088	yarn.resourcemanager.webapp.address
		Sí	8025	
	Programador	Sí	8030	yarn.resourcemanager.scheduler.address
	interfaz del administrador de aplicaciones	Sí	8032	yarn.resourcemanager.address
	Interfaz de administración de RM	Sí	8033	yarn.resourcemanager.admin.address
	Interfaz de usuario web del servidor de JobHistory	Sí	19888	mapreduce.jobhistory.webapp.address
	Interfaz de usuario web del administrador de servidor de JobHistory	Sí	10033	mapreduce.jobhistory.admin.address
	Servidor de JobHistory (RPC)	Sí	10020	mapreduce.jobhistory.address

Componente	Descripción del servicio	El servicio se ejecuta de forma predeterminada	Port	Clave de configuración
	Servidor de cronología de aplicaciones (RPC)	Sí	10200	yarn.timeline-service.address
	Interfaz de usuario web HTTP del servidor de cronología de aplicaciones	Sí	8188	yarn.timeline-service.webapp.address
	Interfaz de usuario web HTTPS del servidor de cronología de aplicaciones	No	8190	yarn.timeline-service.webapp.https.address
		Sí	20888	
ZooKeeper	Puerto del cliente	Sí	2181	
		Sí	37301	
		Sí	8341	

Ver las interfaces web alojadas en clústeres de Amazon EMR

Important

Es posible configurar un grupo de seguridad personalizado para permitir el acceso entrante a estas interfaces web. Tenga en cuenta que cualquier puerto en el que permita el tráfico entrante representa una posible vulnerabilidad de seguridad. Revise con atención los grupos

de seguridad personalizados para asegurarse de minimizar las vulnerabilidades. Para obtener más información, consulte Control del tráfico de red con grupos de seguridad.

Hadoop y otras aplicaciones que instale en el clúster de EMR, publican interfaces de usuario como sitios web alojados en el nodo principal. Por motivos de seguridad, cuando se utilizan grupos de seguridad administrados de Amazon EMR, estos sitios web solo están disponibles en el servidor web local del nodo principal. Por ese motivo, debe conectarse al nodo principal para ver las interfaces web. Para obtener más información, consulte Conectarse al nodo principal mediante SSH. Hadoop también publica las interfaces de usuario como sitios web alojados en los nodos secundarios y de tareas. Estos sitios web también se encuentran disponibles solo en servidores web locales en los nodos.

En la tabla siguiente, se muestran las interfaces web que puede ver en las instancias de clúster. Estas interfaces de Hadoop están disponibles en todos los clústeres. En las interfaces de la instancia principal, sustituya *master-public-dns-name* por el DNS público principal que aparece en la pestaña Resumen del clúster de la consola de Amazon EMR. En las interfaces de las instancias secundarias y de tareas, sustituya *coretask-public-dns-name* por el Public DNS name (Nombre DNS público) que aparece para la instancia. Para buscar el nombre DNS público de una instancia, vaya a la consola de Amazon EMR, seleccione el clúster en la lista, seleccione la pestaña Hardware, seleccione el ID del grupo de instancias que contiene la instancia a la que desea conectarse y, a continuación, anote el nombre DNS público que aparece para la instancia.

Nombre de interfaz	URI
Servidor del historial de Flink (versión 5.33 y posteriores de EMR)	http://*master-public-dns-name* :8082/
Ganglia	http://*master-public-dns-name* /ganglia/
NameNode de HDFS de Hadoop (versión anterior a la 6.x de EMR)	https://*master-public-dns-name* :50470/
Hadoop HDFS NameNode	http://*master-public-dns-name* :50070/
Hadoop HDFS DataNode	http://*coretask-public-dns-name* :50075/

Nombre de interfaz	URI
NameNode de HDFS de Hadoop (version 6.x de EMR)	https://*master-public-dns-name* :9870/
DataNode de HDFS de Hadoop (versión anterior a la 6.x de EMR)	https://*coretask-public-dns-name* :50475/
DataNode de HDFS de Hadoop (version 6.x de EMR)	https://*coretask-public-dns-name* :9865/
HBase	http://*master-public-dns-name* :16010/
Hue	http://*master-public-dns-name* :8888/
JupyterHub	https://*master-public-dns-name* :9443/
Livy	http://*master-public-dns-name* :8998/
Spark HistoryServer	http://*master-public-dns-name* :18080/
Tez	http://*master-public-dns-name* :8080/tez-ui
YARN NodeManager	http://*coretask-public-dns-name* :8042/
YARN ResourceManager	http://*master-public-dns-name* :8088/
Zeppelin	http://*master-public-dns-name* :8890/

Dado que hay varias interfaces específicas de aplicación disponibles en el nodo principal que no están disponibles en los nodos secundarios y de tarea, las instrucciones en este documento son específicas del nodo principal de Amazon EMR. El acceso a las interfaces web en los nodos secundarios y de tarea puede realizarse de la misma manera en que se accedería a las interfaces web en el nodo principal.

Existen varias formas en las que puede acceder a las interfaces web en el nodo principal. El método más rápido y más sencillo consiste en utilizar SSH para conectarse al nodo principal y utilizar el navegador basado en texto, Lynx, para ver los sitios web en su cliente SSH. Sin embargo, Lynx es un navegador basado en texto con una interfaz de usuario limitada que no puede mostrar gráficos. El

siguiente ejemplo muestra cómo abrir la interfaz de Hadoop ResourceManager utilizando Lynx (las URL de Lynx también se facilitan al iniciar sesión en el nodo principal mediante SSH).

```
lynx http://ip-###-##-##-###.us-west-2.compute.internal:8088/
```

Existen dos opciones restantes para acceder a las interfaces web en el nodo principal que proporcionan una funcionalidad del navegador completa. Elija una de las siguientes opciones:

- Opción 1 (recomendada para usuarios más técnicos): utilice un cliente SSH para conectarse al nodo principal, configure el túnel SSH con enrutamiento de puertos local y utilice un navegador de Internet para abrir interfaces web alojadas en el nodo principal. Este método le permite configurar el acceso a la interfaz web sin usar un proxy SOCKS.
- Opción 2 (recomendada para nuevos usuarios): utilice un cliente SSH para conectarse al nodo principal, configure el túnel SSH con enrutamiento de puertos local y configure su navegador de Internet para utilizar un complemento como FoxyProxy para Firefox o SwitchyOmega para Chrome para administrar su configuración de proxy SOCKS. Este método le permite filtrar automáticamente las URL en función de los patrones de texto y limitar la configuración del proxy a dominios que coinciden con la forma del nombre de DNS del nodo principal. Para obtener más información acerca de cómo configurar FoxyProxy para Firefox y Google Chrome, consulte Opción 2, parte 2: configurar ajustes de proxy para ver sitios web alojados en el nodo principal.

> **Note**
>
> Si modifica el puerto en el que se ejecuta una aplicación mediante la configuración de clúster, el hipervínculo al puerto no se actualizará en la consola de Amazon EMR. Esto se debe a que la consola no tiene la funcionalidad de leer la configuración de `server.port`.

Con Amazon EMR versión 5.25.0 o posterior, puede acceder a la interfaz de usuario del servidor del historial de Spark desde la consola sin configurar un proxy web a través de una conexión SSH. Para obtener más información, consulte Acceso de un clic al servidor del historial de Spark persistente.

Temas

- Opción 1: configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos local
- Opción 2, parte 1: configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos dinámico

- Opción 2, parte 2: configurar ajustes de proxy para ver sitios web alojados en el nodo principal

Opción 1: configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos local

Para conectarse al servidor web local en el nodo principal, debe crear un túnel SSH entre su equipo y el nodo principal. Esto se conoce como enrutamiento de puertos. Si no desea utilizar un proxy SOCKS, puede configurar un túnel SSH al nodo principal a través del enrutamiento de puertos local. Con el enrutamiento de puertos local, especifique los puertos locales no utilizados que se utilizan para reenviar el tráfico a puertos remotos específicos en el servidor web local del nodo principal.

La configuración de un túnel SSH utilizando el enrutamiento de puerto local requiere el nombre de DNS pública del nodo principal y el archivo de clave privado de par de claves. Para obtener información sobre cómo localizar el nombre de DNS público principal, consulte Para recuperar el nombre de DNS público del nodo principal con la consola antigua. Para obtener más información sobre el acceso al par de claves, consulte Pares de claves de Amazon EC2 en la Guía del usuario de Amazon EC2 para instancias de Linux. Para obtener más información sobre los sitios que podría desear ver en el nodo principal, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos local con OpenSSH

Para configurar un túnel SSH utilizando el enrutamiento de puertos local en terminal

1. Asegúrese de permitir el tráfico SSH entrante. Para obtener instrucciones, consulte Antes de conectarse: autorice el tráfico entrante.
2. Abra una ventana de terminal. En Mac OS X, elija Aplicaciones > Utilidades > Terminal. En otras distribuciones de Linux, el terminal normalmente se encuentra en Aplicaciones > Accesorios > Terminal.
3. Escriba el siguiente comando para abrir un túnel SSH en su máquina local. Este comando de ejemplo accede a la interfaz web de ResourceManager reenviando el tráfico en el puerto local 8157 (un puerto local sin utilizar elegido de forma aleatoria) al puerto 8088 en el servidor web local del nodo principal.

   En el comando, sustituya *~/mykeypair.pem* por la ubicación y el nombre del archivo `.pem` y sustituya *ec2-###-##-##-###.compute-1.amazonaws.com* por el nombre de DNS público principal de su clúster. Para acceder a una interfaz web diferente, sustituya `8088` por el número de puerto correspondiente. Por ejemplo, sustituya `8088` por `8890` para la interfaz de Zeppelin.

```
ssh -i ~/mykeypair.pem -N -L 8157:ec2-###-##-##-
###.compute-1.amazonaws.com:8088 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
```

`-L` hace referencia al uso de enrutamiento de puertos local que le permite especificar un puerto local utilizado para reenviar datos al puerto remoto identificado en el servidor web local del nodo principal.

Después de emitir este comando, el terminal permanece abierto y no devuelve una respuesta.

4. Para abrir la interfaz web de ResourceManager en su navegador, escriba `http://localhost:8157/` en la barra de direcciones.
5. Cuando haya terminado de trabajar con las interfaces web en el nodo principal, cierre las ventanas del terminal.

Opción 2, parte 1: configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos dinámico

Para conectarse al servidor web local en el nodo principal, debe crear un túnel SSH entre su equipo y el nodo principal. Esto se conoce como enrutamiento de puertos. Si crea el túnel SSH con enrutamiento de puertos dinámico, todo el tráfico dirigido a un puerto local especificado sin utilizar se enruta al servidor web local en el nodo principal. Esto crea un proxy SOCKS. A continuación, puede configurar su navegador de Internet para utilizar un complemento como FoxyProxy o SwitchyOmega para administrar los ajustes del proxy SOCKS.

El uso de un complemento de administración de proxy le permite filtrar automáticamente las URL en función de los patrones de texto y limitar la configuración del proxy a dominios que coinciden con la forma del nombre de DNS pública del nodo principal. El complemento del navegador gestiona automáticamente la activación y desactivación del proxy al cambiar entre la visualización de sitios web alojados en el nodo principal y en Internet.

Antes de comenzar, necesita el nombre de DNS pública del nodo principal y el archivo de clave privada del par de claves. Para obtener información sobre cómo localizar el nombre de DNS público principal, consulte Para recuperar el nombre de DNS público del nodo principal con la consola antigua. Para obtener más información sobre el acceso al par de claves, consulte Pares de claves de Amazon EC2 en la Guía del usuario de Amazon EC2 para instancias de Linux. Para obtener más información sobre los sitios que podría desear ver en el nodo principal, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos dinámico con OpenSSH

Para configurar un túnel SSH utilizando el enrutamiento de puertos dinámico con OpenSSH

1. Asegúrese de permitir el tráfico SSH entrante. Para obtener instrucciones, consulte Antes de conectarse: autorice el tráfico entrante.
2. Abra una ventana de terminal. En Mac OS X, elija Aplicaciones > Utilidades > Terminal. En otras distribuciones de Linux, el terminal normalmente se encuentra en Aplicaciones > Accesorios > Terminal.
3. Escriba el siguiente comando para abrir un túnel SSH en su máquina local. Sustituya *~/mykeypair.pem* por la ubicación y el nombre del archivo `.pem`, sustituya *8157* por un número de puerto local sin utilizar y, por último, sustituya *c2-###-##-##-###.compute-1.amazonaws.com* por el nombre de DNS público principal del clúster.

   ```
   ssh -i ~/mykeypair.pem -N -D 8157 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
   ```

   Después de emitir este comando, el terminal permanece abierto y no devuelve una respuesta.

   > **Note**
   >
   > `-D` hace referencia al uso de enrutamiento de puertos dinámico que le permite especificar un puerto local utilizado para reenviar datos a todos los puertos remotos en el servidor web local del nodo principal. El enrutamiento de puertos dinámicos crea una escucha de proxy SOCKS local en el puerto especificado en el comando.

4. Una vez que el túnel está activo, configure un proxy SOCKS para su navegador. Para obtener más información, consulte Opción 2, parte 2: configurar ajustes de proxy para ver sitios web alojados en el nodo principal.
5. Cuando haya terminado de trabajar con las interfaces web en el nodo principal, cierre la ventana del terminal.

Configurar un túnel SSH con el enrutamiento de puertos dinámico con la AWS CLI

Puede crear una conexión SSH con el nodo principal utilizando la AWS CLI en Windows y Linux, Unix y Mac OS X. Si está utilizando la AWS CLI en Linux, Unix o Mac OS X, deberá establecer permisos en el archivo `.pem`, tal y como se muestra en Para configurar los permisos de archivo

de clave privada del par de claves. Si utiliza la AWS CLI en Windows, PuTTY debe aparecer en la variable de entorno de la ruta o podría recibir un error como, por ejemplo, OpenSSH o PuTTY no está disponible.

Para configurar un túnel SSH utilizando el enrutamiento de puertos dinámico con la AWS CLI

1. Asegúrese de permitir el tráfico SSH entrante. Para obtener instrucciones, consulte Antes de conectarse: autorice el tráfico entrante.
2. Crear una conexión SSH con el nodo principal, tal y como se muestra en Conectarse al nodo principal utilizando la AWS CLI.
3. Para recuperar el identificador del clúster, escriba:

   ```
   aws emr list-clusters
   ```

   La salida enumera los clústeres, incluidos los ID de los clústeres. Tenga en cuenta el ID del clúster al que se está conectando.

   ```
   "Status": {
       "Timeline": {
           "ReadyDateTime": 1408040782.374,
           "CreationDateTime": 1408040501.213
       },
       "State": "WAITING",
       "StateChangeReason": {
           "Message": "Waiting after step completed"
       }
   },
   "NormalizedInstanceHours": 4,
   "Id": "j-2AL4XXXXXX5T9",
   "Name": "AWS CLI cluster"
   ```

4. Escriba el siguiente comando para abrir un túnel SSH en el nodo principal mediante el enrutamiento de puertos dinámico. En el ejemplo siguiente, sustituya *j-2AL4XXXXXX5T9* por el ID del clúster y sustituya *~/mykeypair.key* por la ubicación y el nombre de archivo del archivo `.pem` (para Linux, Unix y Mac OS X) o `.ppk` (para Windows).

   ```
   aws emr socks --cluster-id j-2AL4XXXXXX5T9 --key-pair-file ~/mykeypair.key
   ```

> **Note**
>
> El comando de socks configura automáticamente el enrutamiento de puertos dinámico en el puerto local 8157. Actualmente este ajuste no se puede modificar.

5. Una vez que el túnel está activo, configure un proxy SOCKS para su navegador. Para obtener más información, consulte Opción 2, parte 2: configurar ajustes de proxy para ver sitios web alojados en el nodo principal.
6. Cuando haya terminado de trabajar con las interfaces web en el nodo principal, cierre las ventana de la AWS CLI.

   Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Configurar un túnel SSH en el nodo principal mediante PuTTY

Los usuarios de Windows pueden utilizar un cliente SSH como PuTTY para crear un túnel SSH al nodo principal. Antes de conectarse al nodo principal de Amazon EMR, debe descargar e instalar PuTTY y PuTTYgen. Puede descargar estas herramientas desde la página de descarga de PuTTY.

PuTTY no admite de forma nativa el formato de archivo de clave privada del par de claves (`.pem`) generado por Amazon EC2. Utilice PuTTYgen para convertir su archivo de clave al formato de PuTTY requerido (`.ppk`). Debe convertir su clave en este formato (`.ppk`) antes de intentar conectarse al nodo principal utilizando PuTTY.

Para obtener más información sobre la conversión de la clave, consulte Conversión de la clave privada mediante PuTTYgen en la Guía del usuario de Amazon EC2 para instancias de Linux.

Para configurar un túnel SSH utilizando el enrutamiento de puertos dinámico en PuTTY

1. Asegúrese de permitir el tráfico SSH entrante. Para obtener instrucciones, consulte Antes de conectarse: autorice el tráfico entrante.
2. Haga doble clic en `putty.exe` para iniciar PuTTY. También puede lanzar PuTTY desde la lista de programas de Windows.

> **Note**
>
> Si ya tiene una sesión de SSH activa con el nodo principal, puede agregar un túnel pulsando con el botón derecho del ratón en la barra del título de PuTTY y seleccionar Cambiar configuración.

3. Si es necesario, en la lista Category (Categoría), elija Session (Sesión).
4. En el campo Host Name (Nombre de host), escriba **hadoop@***MasterPublicDNS*. Por ejemplo, **hadoop@***ec2-###-##-##-###.compute-1.amazonaws.com*.
5. En la lista Category (Categoría), expanda Connection > SSH (Conexión > SSH) y, a continuación, elija Auth (Autenticación).
6. En Private key file for authentication (Archivo de clave privada para la autenticación), elija Browse (Examinar) y seleccione el archivo `.ppk` que ha generado.

> **Note**
>
> PuTTY no admite de forma nativa el formato de archivo de clave privada del par de claves (`.pem`) generado por Amazon EC2. Utilice PuTTYgen para convertir su archivo de clave al formato de PuTTY requerido (`.ppk`). Debe convertir su clave en este formato (`.ppk`) antes de intentar conectarse al nodo principal utilizando PuTTY.

7. En la lista Category (Categoría), expanda Connection > SSH (Conexión > SSH) y, a continuación, elija Tunnels (Túneles).
8. En el campo Puerto de origen, escriba `8157` (un puerto local no utilizado) y, a continuación, seleccione Agregar.
9. Deje el campo Destination (Destino) en blanco.
10. Seleccione las opciones Dynamic (Dinámico) y Auto.
11. Elija Abrir.
12. Elija Yes (Sí) para descartar la alerta de seguridad de PuTTY.

> **Important**
>
> Cuando inicia sesión en el nodo principal, escriba `hadoop` si se le solicita un nombre de usuario.

13. Una vez que el túnel está activo, configure un proxy SOCKS para su navegador. Para obtener más información, consulte Opción 2, parte 2: configurar ajustes de proxy para ver sitios web alojados en el nodo principal.
14. Cuando haya terminado de trabajar con las interfaces web en el nodo principal, cierre la ventana de PuTTY.

Opción 2, parte 2: configurar ajustes de proxy para ver sitios web alojados en el nodo principal

Si utiliza un túnel de SSH con enrutamiento de puertos dinámico, debe utilizar un complemento de administración de proxy de SOCKS para controlar los ajustes del proxy en el navegador. El uso de una herramienta de administración de proxy de SOCKS le permite filtrar automáticamente las URL en función de los patrones de texto y limitar la configuración del proxy a dominios que coinciden con la forma del nombre de DNS pública del nodo principal. El complemento del navegador gestiona automáticamente la activación y desactivación del proxy al cambiar entre la visualización de sitios web alojados en el nodo principal y en Internet. Para administrar sus ajustes de proxy, configure el navegador para utilizar un complemento como FoxyProxy o SwitchyOmega.

Para obtener más información acerca de la creación de un túnel de SSH, consulte Opción 2, parte 1: configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos dinámico. Para obtener más información acerca de las interfaces web disponibles, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Incluya los siguientes ajustes cuando configure su complemento de proxy:

- Usa localhost como dirección de host.
- Use el mismo número de puerto local que seleccionó para establecer el túnel SSH con el nodo principal en Opción 2, parte 1: configurar un túnel SSH al nodo principal utilizando el enrutamiento de puertos dinámico. Por ejemplo, el puerto *8157*. Este puerto también debe coincidir con el número de puerto que utiliza en PuTTY u en otro emulador de terminal que utilice para conectarse.
- Especifique el protocolo SOCKS v5. SOCKS v5 le permite configurar de forma opcional la autorización de usuario.
- URL Patterns (Patrones de URL)

  Los siguientes patrones de URL deben autorizarse y especificarse con un tipo de patrón comodín:

  - Los patrones *ec2*.*compute*.amazonaws.com* y *10*.amazonaws.com* coinciden con el nombre de DNS público de los clústeres de las regiones de EE. UU.

- Los patrones *ec2*.compute* y *10*.compute* coinciden con el nombre de DNS público de los clústeres de todas las demás regiones.
- El patrón 10.* proporciona acceso a los archivos de registro de JobTracker en Hadoop. Modifique este filtro si entra en conflicto con su plan de acceso de red.
- Los patrones *.ec2.internal* y *.compute.internal* hacen coincidir los nombres de DNS privados (internos) de los clústeres de la región `us-east-1` y de todas las demás regiones, respectivamente.

Ejemplo: configurar FoxyProxy para Firefox

En el siguiente ejemplo se muestra una configuración de FoxyProxy Standard (versión 7.5.1) para Mozilla Firefox.

FoxyProxy proporciona un conjunto de herramientas de administración de proxy. Le permite utilizar un servidor proxy para las URL que coincidan con los patrones correspondientes a los dominios utilizados por las instancias de Amazon EC2 de su clúster de Amazon EMR.

Para instalar y configurar FoxyProxy con Mozilla Firefox

1. En Firefox, vaya a https://addons.mozilla.org/, busque FoxyProxy Standard y siga las instrucciones para agregar FoxyProxy a Firefox.
2. Mediante un editor de texto, cree un archivo JSON denominado `foxyproxy-settings.json` con la siguiente configuración de ejemplo.

```
{
  "k20d21508277536715": {
    "active": true,
    "address": "localhost",
    "port": 8157,
    "username": "",
    "password": "",
    "type": 3,
    "proxyDNS": true,
    "title": "emr-socks-proxy",
    "color": "#0055E5",
    "index": 9007199254740991,
    "whitePatterns": [
      {
        "title": "*ec2*.*compute*.amazonaws.com*",
        "active": true,
```

```
      "pattern": "*ec2*.*compute*.amazonaws.com*",
      "importedPattern": "*ec2*.*compute*.amazonaws.com*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "*ec2*.compute*",
      "active": true,
      "pattern": "*ec2*.compute*",
      "importedPattern": "*ec2*.compute*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "10.*",
      "active": true,
      "pattern": "10.*",
      "importedPattern": "http://10.*",
      "type": 1,
      "protocols": 2
    },
    {
      "title": "*10*.amazonaws.com*",
      "active": true,
      "pattern": "*10*.amazonaws.com*",
      "importedPattern": "*10*.amazonaws.com*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "*10*.compute*",
      "active": true,
      "pattern": "*10*.compute*",
      "importedPattern": "*10*.compute*",
      "type": 1,
      "protocols": 1
    },
    {
      "title": "*.compute.internal*",
      "active": true,
      "pattern": "*.compute.internal*",
      "importedPattern": "*.compute.internal*",
      "type": 1,
      "protocols": 1
```

```
      },
      {
        "title": "*.ec2.internal* ",
        "active": true,
        "pattern": "*.ec2.internal*",
        "importedPattern": "*.ec2.internal*",
        "type": 1,
        "protocols": 1
      }
    ],
    "blackPatterns": []
  },
  "logging": {
    "size": 100,
    "active": false
  },
  "mode": "patterns",
  "browserVersion": "68.12.0",
  "foxyProxyVersion": "7.5.1",
  "foxyProxyEdition": "standard"
}
```

3. Abra la página Administrar sus extensiones de Firefox (vaya a about:addons y seleccione Extensiones).
4. Seleccione FoxyProxy Standard y, a continuación, seleccione el botón de más opciones (el botón que tiene forma de puntos suspensivos).
5. Seleccione Opciones en el menú desplegable.
6. Seleccione Importar ajustes en el menú de la izquierda.
7. En la página Importar ajustes, seleccione Importar ajustes en Importar ajustes desde FoxyProxy 6.0+, busque la ubicación del archivo `foxyproxy-settings.json` que ha creado, selecciónelo y seleccione Abrir.
8. Pulse Aceptar cuando se le pida para sobrescribir los ajustes existentes y guarde la nueva configuración.

Ejemplo: configurar SwitchyOmega para Chrome

En el siguiente ejemplo, se muestra cómo configurar la extensión SwitchyOmega para Google Chrome. SwitchyOmega le permite configurar, administrar y cambiar entre varios proxies.

Para instalar y configurar SwitchyOmega utilizando Google Chrome

1. Vaya a https://chrome.google.com/webstore/category/extensions, busque Proxy SwitchyOmega y agréguelo a Chrome.
2. Seleccione Nuevo perfil e introduzca `emr-socks-proxy` como nombre del perfil.
3. Seleccione Perfil de PAC y, a continuación, Crear. Los archivos de configuración automática de proxy (PAC) le ayudan a definir una lista de permisos para las solicitudes del navegador que deben reenviarse a un servidor proxy web.
4. En el campo Script de PAC, sustituya el contenido por el siguiente script, que define qué URL deben reenviarse a través del servidor proxy web. Si especificó un número de puerto diferente al configurar el túnel SSH, sustituya *8157* por su número de puerto.

```
function FindProxyForURL(url, host) {
    if (shExpMatch(url, "*ec2*.*compute*.amazonaws.com*")) return 'SOCKS5
 localhost:8157';
    if (shExpMatch(url, "*ec2*.compute*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "http://10.*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*10*.compute*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*10*.amazonaws.com*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*.compute.internal*")) return 'SOCKS5 localhost:8157';
    if (shExpMatch(url, "*ec2.internal*")) return 'SOCKS5 localhost:8157';
    return 'DIRECT';
}
```

5. En Acciones, seleccione Aplicar cambios para guardar la configuración del proxy.
6. En la barra de herramientas de Chrome, seleccione SwitchyOmega y selecciona el perfil `emr-socks-proxy`.

Acceder a una interfaz web en el navegador

Para abrir una interfaz web, introduzca el nombre de DNS público de su nodo principal o de núcleo seguido del número de puerto de la interfaz seleccionada en la barra de direcciones del navegador. En el siguiente ejemplo se muestra la URL que debe introducir para conectarse al servidor HistoryServer de Spark.

```
http://master-public-dns-name:18080/
```

Para obtener instrucciones sobre cómo recuperar el nombre de DNS público de un nodo, consulte Recuperar el nombre de DNS público del nodo principal. Para obtener una lista completa de las

direcciones URL de la interfaz web, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Enviar trabajo a un clúster

En esta sección se describen los métodos que puede utilizar para enviar el trabajo a un clúster de Amazon EMR. Para enviar el trabajo, puede agregar pasos o enviar los trabajos de Hadoop de forma interactiva al nodo principal.

Tenga en cuenta las siguientes reglas de comportamiento de los pasos al enviar los pasos a un clúster:

- El ID de paso puede contener un máximo de 256 caracteres.
- Puede tener hasta 256 pasos pendientes y en ejecución en un clúster.
- Incluso si tiene 256 pasos activos en ejecución en el clúster, puede enviar trabajos de forma interactiva al nodo principal. Puede enviar un número ilimitado de pasos a lo largo de la vida útil de un clúster en ejecución prolongada, pero solo puede haber 256 pasos en ejecución o pendientes en un momento dado.
- Con las versiones 4.8.0 y posteriores de Amazon EMR, excepto la versión 5.0.0, puede cancelar los pasos pendientes. Para obtener más información, consulte Cancelación de pasos.
- Con las versiones 5.28.0 y posteriores de Amazon EMR, puede cancelar los pasos pendientes y en ejecución. También puede optar por ejecutar varios pasos en paralelo para mejorar la utilización del clúster y ahorrar costos. Para obtener más información, consulte Consideraciones para ejecutar varios pasos en paralelo.

> **Note**
>
> Para obtener el mejor rendimiento, se recomienda que almacene las acciones de arranque personalizadas, los scripts y otros archivos que desee utilizar con Amazon EMR en un bucket de Amazon S3 que se encuentre en la misma Región de AWS que su clúster.

Temas

- Agregar pasos a un clúster con la consola de administración de Amazon EMR
- Adición de pasos a un clúster con la AWS CLI

- Consideraciones para ejecutar varios pasos en paralelo
- Visualización de pasos
- Cancelación de pasos

Agregar pasos a un clúster con la consola de administración de Amazon EMR

Utilice los siguientes procedimientos para agregar pasos a un clúster con la AWS Management Console. Para obtener información detallada sobre cómo enviar los pasos de aplicaciones de macrodatos específicas, consulte las siguientes secciones de la guía de versiones de Amazon EMR:

- Enviar un paso JAR personalizado
- Enviar un paso de transmisión de Hadoop
- Enviar un paso de Spark
- Enviar un paso de Pig
- Ejecutar un comando o un script como paso
- Transferir los valores a los pasos para ejecutar los scripts de Hive

Agregar pasos durante la creación del clúster

Desde la AWS Management Console, puede agregar pasos al crear un clúster.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para agregar pasos al crear un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.

2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En Pasos, seleccione Agregar paso. Escriba los valores adecuados en los campos del cuadro de diálogo Agregar paso. Para obtener información sobre cómo dar formato a los argumentos de los pasos, consulte Agregar argumentos de paso. Las opciones varían según el tipo de paso. Para agregar el paso y salir del cuadro de diálogo, seleccione Agregar paso.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para agregar pasos al crear un clúster con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/home. Seleccione Crear clúster: opciones avanzadas.
2. En la página Step 1: Software and Steps (Paso 1: Software y pasos), en Steps (optional) (Pasos (opcional)), seleccione Run multiple steps in parallel to improve cluster utilization and save cost (Ejecutar varios pasos en paralelo para mejorar la utilización del clúster y ahorrar costos). El valor predeterminado para el nivel de simultaneidad es 10. Puede elegir entre 2 y 256 pasos que se pueden ejecutar en paralelo.

   > **Note**
   >
   > La ejecución de varios pasos en paralelo solo es compatible con las versiones 5.28.0 y posteriores de Amazon EMR.

3. En After last step completes (Después de completar el último paso), elija Cluster enters waiting state (El cluster entra en estado de espera) o Auto-terminate the cluster (Terminar automáticamente el clúster).
4. Elija Step type (Tipo de paso) y, a continuación, Add step (Añadir paso).
5. Escriba los valores adecuados en los campos del cuadro de diálogo Add Step (Añadir paso). Para obtener información sobre cómo dar formato a los argumentos de los pasos, consulte Agregar argumentos de paso. Las opciones varían según el tipo de paso. Si ha habilitado Ejecutar varios pasos en paralelo para mejorar la utilización del clúster y ahorrar costos, la única opción disponible para Acción en caso de error es Continuar. A continuación, elija Add (Añadir).

Agregar pasos a un clúster en ejecución

Con la AWS Management Console, puede agregar pasos a un clúster con la opción de terminación automática desactivada.

New console

Para agregar pasos a un clúster en ejecución con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar.
3. En la página Pasos de la página de detalles del clúster, seleccione Agregar paso. Para clonar un paso existente, seleccione el menú desplegable Acciones y seleccione Clonar paso.
4. Escriba los valores adecuados en los campos del cuadro de diálogo Agregar paso. Las opciones varían según el tipo de paso. Para agregar el paso y salir del cuadro de diálogo, elija Agregar paso.

Old console

Para agregar pasos a un clúster en ejecución con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/home. En la página Cluster List (Lista de clústeres), seleccione el enlace para el clúster.
2. En la página Cluster Details (Detalles del clúster), seleccione la pestaña Steps (Pasos).
3. En la pestaña Steps (Pasos) elija Add step (Añadir paso).
4. Escriba los valores adecuados en los campos del cuadro de diálogo Add Step (Añadir paso) y, a continuación, elija Add (Añadir). Las opciones varían según el tipo de paso.

Modificar el nivel de simultaneidad de pasos en un clúster en ejecución

Con la AWS Management Console, puede modificar el nivel de simultaneidad de pasos en un clúster en ejecución.

> **Note**
>
> Solo puede ejecutar varios pasos en paralelo con las versiones 5.28.0 y posteriores de Amazon EMR.

New console

Para modificar la simultaneidad de pasos en un clúster en ejecución con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar. El clúster debe estar en ejecución para cambiar su atributo de simultaneidad.
3. En la pestaña Pasos de la página de detalles del clúster, busque la sección Atributos. Seleccione Editar para cambiar la simultaneidad. Escriba un valor comprendido entre 1 y 256.

Old console

Para modificar la simultaneidad de pasos en un clúster en ejecución con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/home. En la página Cluster List (Lista de clústeres), seleccione el enlace para el clúster.
2. En la página Cluster Details (Detalles del clúster), seleccione la pestaña Steps (Pasos).
3. En Concurrency (Simultaneidad), elija Change (Cambiar). Seleccione un nuevo valor para el nivel de simultaneidad de pasos y, a continuación, guarde el cambio.

Agregar argumentos de paso

Si utiliza la AWS Management Console para agregar un paso a su clúster, puede especificar los argumentos para ese paso en el campo Argumentos. Debe separar los argumentos con espacios en blanco y rodear los argumentos de cadena que constan de caracteres y espacios en blanco con comillas.

Example : argumentos correctos

Los argumentos de ejemplo siguientes tienen el formato correcto para la AWS Management Console, con el último argumento de cadena entre comillas.

```
bash -c "aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

También puede colocar cada argumento en una línea independiente para facilitar la lectura, como se muestra en el siguiente ejemplo.

```
bash
-c
"aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

Example : argumentos incorrectos

Los siguientes argumentos de ejemplo tienen un formato incorrecto para la AWS Management Console. Observe que el último argumento de cadena, `aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .`, contiene espacios en blanco y no está entre comillas.

```
bash -c aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .
```

Adición de pasos a un clúster con la AWS CLI

En los siguientes procedimientos se muestra cómo agregar pasos a un clúster recién creado y a un clúster en ejecución con la AWS CLI. En ambos ejemplos, el subcomando `--steps` se utiliza para agregar pasos al clúster.

Para añadir pasos durante la creación del clúster

- Escriba el siguiente comando para crear un clúster y añadir un paso de Apache Pig. Asegúrese de sustituir *myKey* por el nombre de su par de claves de Amazon EC2.

```
aws emr create-cluster --name "Test cluster" \
--applications Name=Spark \
--use-default-roles \
--ec2-attributes KeyName=myKey \
--instance-groups InstanceGroupType=PRIMARY,InstanceCount=1,InstanceType=m5.xlarge
 InstanceGroupType=CORE,InstanceCount=2,InstanceType=m5.xlarge \
--steps '[{"Args":["spark-submit","--deploy-mode","cluster","--
class","org.apache.spark.examples.SparkPi","/usr/lib/spark/examples/jars/spark-
```

```
examples.jar","5"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"command-
runner.jar","Properties":"","Name":"Spark application"}]'
```

> **Note**
>
> La lista de argumentos cambia en función del tipo de paso.

De forma predeterminada, el nivel de simultaneidad de pasos es 1. Puede establecer el nivel de simultaneidad de pasos con el parámetro `StepConcurrencyLevel` al crear un clúster.

La salida de un identificador de clúster es similar a la siguiente.

```
{
    "ClusterId": "j-2AXXXXXXGAPLF"
}
```

Para añadir un paso a un clúster en ejecución

- Escriba el siguiente comando para añadir un paso a un clúster en ejecución. Sustituya *j-2AXXXXXXGAPLF* por el ID de su propio clúster.

```
aws emr add-steps --cluster-id j-2AXXXXXXGAPLF \
--steps '[{"Args":["spark-submit","--deploy-mode","cluster","--
class","org.apache.spark.examples.SparkPi","/usr/lib/spark/examples/jars/spark-
examples.jar","5"],"Type":"CUSTOM_JAR","ActionOnFailure":"CONTINUE","Jar":"command-
runner.jar","Properties":"","Name":"Spark application"}]'
```

La salida es un identificador de paso similar al siguiente.

```
{
    "StepIds": [
 "s-Y9XXXXXXAPMD"
    ]
}
```

Para modificar StepConcurrencyLevel en un clúster en ejecución

1. En un clúster en ejecución, puede modificar `StepConcurrencyLevel` con la API de `ModifyCluster`. Por ejemplo, escriba el siguiente comando para aumentar `StepConcurrencyLevel` a `10`. Sustituya *`j-2AXXXXXXGAPLF`* por el ID de su clúster.

   ```
   aws emr modify-cluster --cluster-id j-2AXXXXXXGAPLF --step-concurrency-level 10
   ```

2. El resultado es similar al siguiente.

   ```
   {
   "StepConcurrencyLevel": 10
   }
   ```

Para obtener más información acerca del uso de los comandos de Amazon EMR en la AWS CLI, consulte la Referencia de los comandos de la AWS CLI.

Consideraciones para ejecutar varios pasos en paralelo

- Los pasos que se ejecutan en paralelo pueden completarse en cualquier orden, pero los pasos pendientes de la cola pasan al estado de ejecución en el orden en que se enviaron.
- Al seleccionar un nivel de simultaneidad de pasos para el clúster, debe considerar si el tipo de instancia del nodo principal cumple o no los requisitos de memoria de las cargas de trabajo de usuario. El proceso de ejecución del paso principal se ejecuta en el nodo principal para cada paso. Para ejecutar varios pasos en paralelo, se requiere más memoria y utilización de CPU en el nodo principal que si los pasos se ejecutan de uno en uno.
- Para gestionar los recursos y realizar las programaciones complejas de los pasos simultáneos, puede utilizar las funciones de programación de YARN como `FairScheduler` o `CapacityScheduler`. Por ejemplo, puede utilizar `FairScheduler` con un conjunto `queueMaxAppsDefault` para evitar que se ejecuten más de un determinado número de trabajos a la vez.
- El nivel de simultaneidad de pasos está sujeto a las configuraciones de los administradores de recursos. Por ejemplo, si YARN está configurado con un paralelismo de solo `5`, entonces únicamente podrá tener cinco aplicaciones YARN ejecutándose en paralelo, aunque `StepConcurrencyLevel` se haya establecido en `10`. Para obtener más información sobre la configuración de los administradores de recursos, consulte Configurar aplicaciones en la Guía de publicación de Amazon EMR.

- No puede agregar un paso con un valor de `ActionOnFailure` que no sea CONTINUE mientras el nivel de simultaneidad de pasos del clúster sea superior a 1.
- Si el nivel de simultaneidad de pasos de un clúster es superior a uno, la característica `ActionOnFailure` del paso no se activará.
- Si un clúster tiene un nivel de simultaneidad de pasos de 1, pero tiene varios pasos en ejecución, es posible que `TERMINATE_CLUSTER ActionOnFailure` se active, pero `CANCEL_AND_WAIT ActionOnFailure` no se activará. Este caso de periferia se presenta cuando el nivel de simultaneidad de los pasos del clúster era superior a uno, pero se reducía cuando se ejecutaban varios pasos.
- Puede usar el escalado automático de EMR para escalar en sentido ascendente o descendente en función de los recursos de YARN y así evitar la contención de recursos. Para obtener información, consulte Uso del escalado automático con una política personalizada para grupos de instancias en la Guía de administración de Amazon EMR.
- Cuando se disminuye el nivel de simultaneidad de pasos, EMR permite que se completen todos los pasos en ejecución antes de reducir el número de pasos. Si los recursos se agotan porque el clúster está ejecutando demasiados pasos simultáneos, recomendamos cancelar manualmente los pasos en ejecución para liberar recursos.

Visualización de pasos

El número total de registros de pasos que puede ver (independientemente de su estado) es de 1 000. Este total incluye tanto los pasos del sistema como los pasos enviados por el usuario. Cuando el estado de los pasos enviados por el usuario cambia a COMPLETO o ERROR, los pasos adicionales enviados por el usuario se pueden agregar al clúster hasta que se alcance el límite de 1 000 pasos. Después de que los 1 000 pasos se hayan añadido a un clúster, el envío de pasos adicionales provoca la eliminación de los registros de pasos más antiguos enviados por el usuario. Estos registros no se eliminan de los archivos de registro. Se eliminan en la pantalla de la consola y no aparecen cuando se utiliza la AWS CLI o la API para recuperar la información del clúster. Los registros de paso del sistema no se eliminan nunca.

La información de paso que puede ver depende del mecanismo utilizado para recuperar la información del clúster. La tabla siguiente indica la información de paso devuelta por cada una de las opciones disponibles.

Opción	DescribeJobFlow o --describe --jobflow	ListSteps o list-steps
SDK	256 pasos	1 000 pasos
CLI de Amazon EMR	256 pasos	N/A
AWS CLI	N/A	1 000 pasos
API	256 pasos	1 000 pasos

Cancelación de pasos

Puede cancelar los pasos pendientes y en ejecución desde la AWS Management Console, la AWS CLI o la API de Amazon EMR.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para cancelar los pasos con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, seleccione Clústeres y, a continuación, seleccione el clúster que desee actualizar.
3. En la pestaña Pasos de la página de detalles del clúster, active la casilla de verificación situada junto al paso que desee cancelar. Seleccione el menú desplegable Acciones y, a continuación, seleccione Cancelar pasos.

4. En el cuadro de diálogo Cancelar el paso, seleccione si desea cancelar el paso y esperar a que se cierre o cancelar el paso y forzar el cierre. A continuación, seleccione Confirm (Confirmar).
5. El estado de los pasos de la tabla Pasos cambia a `CANCELLED`.

Old console

Para cancelar los pasos con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. En la página Cluster Details (Detalles del clúster), expanda la sección Steps (Pasos).
3. Para cada paso que desee cancelar, seleccione selecciónelo en la lista Steps (Pasos). A continuación, seleccione Cancel step (Cancelar paso).
4. En el cuadro de diálogo Cancel step (Cancelar paso) deje la opción predeterminada Cancel the step and wait for it to exit (Cancelar el paso y esperar a que salga). Si desea finalizar el paso inmediatamente sin esperar a que se complete ningún proceso, elija Cancel the step and force it to exit (Cancelar el paso y forzarlo a salir).
5. Seleccione Cancel step (Cancelar paso).

CLI

Para cancelar con la AWS CLI

- Utilice el comando `aws emr cancel-steps`, especificando el clúster y los pasos que desea cancelar. El siguiente ejemplo muestra un comando de AWS CLI para cancelar dos pasos.

```
aws emr cancel-steps --cluster-id j-2QUAXXXXXXXXX \
--step-ids s-3M8DXXXXXXXXX s-3M8DXXXXXXXXX \
--step-cancellation-option SEND_INTERRUPT
```

Con la versión 5.28.0 de Amazon EMR, puede seleccionar una de las dos siguientes opciones de cancelación para el parámetro `StepCancellationOption` al cancelar pasos.

- `SEND_INTERRUPT`: esta es la opción predeterminada. Cuando se recibe una solicitud de cancelación de pasos, EMR envía una señal `SIGTERM` al paso. Agregue un controlador de señales `SIGTERM` a su lógica de pasos para atrapar esta señal y terminar los procesos de pasos descendientes o espere a que se completen.
- `TERMINATE_PROCESS`: cuando se selecciona esta opción, EMR envía una señal `SIGKILL` al paso y a todos sus procesos descendientes, que los termina inmediatamente.

Consideraciones para cancelar los pasos

- Al cancelar un paso pendiente o en ejecución, ese paso se elimina del recuento de pasos activos.
- Si se cancela un paso en ejecución, no se permite que un paso pendiente comience a ejecutarse, suponiendo que no haya cambios en `stepConcurrencyLevel`.
- La cancelación de un paso en ejecución no activa el paso `ActionOnFailure`.
- Para las versiones 5.32.0 y posteriores de EMR, `SEND_INTERRUPT StepCancellationOption` envía una señal `SIGTERM` al proceso secundario del paso. Debe estar atento a esta señal y realizar una limpieza para apagarlo correctamente. `TERMINATE_PROCESS StepCancellationOption` envía una señal `SIGKILL` al proceso secundario del paso y a todos sus procesos descendientes; sin embargo, los procesos asíncronos no se ven afectados.

Ver y monitorizar un clúster

Amazon EMR proporciona varias herramientas que puede utilizar para recopilar información sobre su clúster. Puede acceder a información sobre el clúster desde la consola, la CLI o mediante programación. Las interfaces web estándar de Hadoop y los archivos de registro están disponibles en el nodo principal. También puede utilizar servicios de monitorización como CloudWatch y Ganglia para realizar un seguimiento del rendimiento del clúster.

El historial de aplicaciones también está disponible desde la consola mediante las interfaces de usuario de aplicación «persistentes» para Spark History Server a partir de Amazon EMR 5.25.0. Con Amazon EMR 6.x, el servidor de YARN Timeline persistente y las interfaces de usuario de Tez también están disponibles. Estos servicios están alojados fuera del clúster, por lo que puede acceder al historial de aplicaciones durante 30 días después de que termine el clúster, sin necesidad de una conexión SSH o proxy web. Consulte Ver el historial de aplicaciones.

Temas

- Ver el estado y los detalles del clúster

- Depuración de pasos mejorada
- Ver el historial de aplicaciones
- Ver archivos de registro de
- Ver instancias del clúster en Amazon EC2
- Eventos y métricas de CloudWatch
- Ver métricas de aplicaciones de clúster con Ganglia
- Registro de llamadas a la API de Amazon EMR en AWS CloudTrail

Ver el estado y los detalles del clúster

Después de crear un clúster, puede monitorizar su estado y obtener información detallada acerca de su ejecución y los errores que puedan haberse producido, incluso después de que se haya terminado. Amazon EMR guarda los metadatos de los clústeres terminados para su referencia durante dos meses, después de los cuales se eliminan. No puede eliminar clústeres del historial de clústeres, pero con la AWS Management Console, puede utilizar la función Filter (Filtro), y con la AWS CLI, puede utilizar opciones con el comando `list-clusters` para centrarse en los clústeres que le interesen.

Puede acceder al historial de aplicaciones almacenado en el clúster durante una semana desde el momento en que se registra, independientemente de si el clúster se está ejecutando o ha terminado. Además, las interfaces de usuario de aplicaciones persistentes almacenan el historial de aplicaciones fuera del clúster durante 30 días después de que termine un clúster. Consulte Ver el historial de aplicaciones.

Para obtener más información sobre los estados de los clústeres, como los estados En espera y En ejecución, consulte Descripción del ciclo de vida del clúster.

Ver los detalles de un clúster mediante la AWS Management Console

En la lista Clústeres de https://console.aws.amazon.com/emr, se muestran todos los clústeres de su cuenta y su región de AWS, incluidos los clústeres terminados. En la lista, se muestra lo siguiente para cada clúster: el nombre y el ID, el estado, los detalles del estado, la hora de creación, el tiempo transcurrido desde que se ejecutó el clúster y las horas de instancia normalizadas que han acumulado todas las instancias de EC2 del clúster. Esta lista es el punto de partida para monitorear el estado de los clústeres. Se ha diseñado para que pueda profundizar hasta los detalles de cada clúster para su análisis y resolución de problemas.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para ver la información del clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, seleccione Clústeres y seleccione el clúster que desee ver.
3. Utilice el panel Resumen para ver los aspectos básicos de la configuración del clúster, como el estado del clúster, las aplicaciones de código abierto que Amazon EMR instaló en el clúster y la versión de Amazon EMR que utilizó para crear el clúster. Utilice les pestañas que aparecen debajo del resumen para ver la información que se describe en la siguiente tabla.

Old console

Para ver la información del clúster con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Para ver un breve resumen de la información del clúster, seleccione la flecha hacia abajo situada junto al enlace del clúster en Nombre. La fila del clúster se expande para proporcionar más información sobre el clúster, el hardware, los pasos y las acciones de arranque. Utilice los enlaces de esta sección para examinar los detalles. Por ejemplo, haga clic en un enlace en Steps (Pasos) para obtener acceso a los archivos de registro del paso, ver el archivo JAR asociado al paso, consultar los trabajos y las tareas del paso y obtener acceso a los archivos de registro.
3. Para ver la información del clúster en detalle, seleccione el enlace del clúster debajo de Nombre para abrir la página de detalles del clúster. La siguiente información está disponible en la página de detalles del clúster de la consola anterior:

Pestaña (consola antigua)	Descripción (consola antigua)
Properties	Utilice esta pestaña para ver el sistema operativo del clúster, las configuraciones de seguridad y terminación del clúster, la información de la VPC y la subred y dónde almacena los registros en Amazon S3.
Acciones de arranque	Utilice esta pestaña para ver el estado de las acciones de arranque que ejecuta el clúster cuando se lanza. Las acciones de arranque se utilizan para las instalaciones y la configuración avanzada del software personalizado. Para obtener más informaci ón, consulte Crear acciones de arranque para instalar software adicional.
Supervisión	Utilice esta pestaña para ver las métricas clave del funcionamiento del clúster. Puede ver datos de nivel de clúster, datos de nivel de nodo e información sobre las operaciones E/S y almacenamiento de datos.
Instancias	Utilice esta pestaña para ver información acerca de los nodos del clúster, incluidos los identificadores de instancias de EC2, los nombres de DNS, los volúmenes de EBS, etc.
Pasos	Utilice esta pestaña para ver el estado y obtener acceso a los archivos logs de los pasos emitidos. Para obtener más informaci ón acerca de los pasos, consulte Enviar trabajo a un clúster.
Aplicaciones	Utilice esta pestaña para ver el servidor de YARN Timeline persistente fuera del clúster y los detalles de la aplicación de Tez UI. También puede ver información sobre las

Pestaña (consola antigua)	Descripción (consola antigua)
	aplicaciones instaladas, las configuraciones del clúster y los grupos de instancias. Las interfaces de usuario de aplicaciones en el clúster están disponibles mientras se ejecuta el clúster.
Eventos	Utilice esta pestaña para ver los registros de eventos del clúster. Para obtener más información, consulte Monitorización de eventos de Amazon EMR con CloudWatch.
Etiquetas	Use esta pestaña para ver las etiquetas que haya aplicado al clúster.

Ver los detalles de un clúster mediante la AWS CLI

Los siguientes ejemplos muestran cómo recuperar detalles del clúster utilizando la AWS CLI. Para obtener más información sobre los comandos disponibles, consulte la Referencia de comandos de la AWS CLI de Amazon EMR. Puede usar el comando describe-cluster para ver detalles de nivel de clúster como el estado, la configuración de hardware y de software, los ajustes de VPC, las acciones de arranque, los grupos de instancias, etc. Para obtener más información acerca de los estados del clúster, consulte Descripción del ciclo de vida del clúster. El siguiente ejemplo ilustra cómo usar el comando `describe-cluster`, seguido de ejemplos del comando list-clusters.

Example Ver el estado del clúster

Para utilizar el comando `describe-cluster`, necesita el ID del clúster. Este ejemplo muestra cómo obtener una lista de los clústeres creados en un intervalo de fechas determinado y cómo usar uno de los ID de clúster devueltos para obtener más información sobre el estado de un determinado clúster.

El siguiente comando describe el clúster *`j-1K48XXXXXXHCB`*, que debe sustituirse por el ID del clúster que se desee.

```
aws emr describe-cluster --cluster-id j-1K48XXXXXXHCB
```

La salida de este comando es similar a la siguiente:

```
{
    "Cluster": {
        "Status": {
            "Timeline": {
                "ReadyDateTime": 1438281058.061,
                "CreationDateTime": 1438280702.498
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Waiting for steps to run"
            }
        },
        "Ec2InstanceAttributes": {
            "EmrManagedMasterSecurityGroup": "sg-cXXXXX0",
            "IamInstanceProfile": "EMR_EC2_DefaultRole",
            "Ec2KeyName": "myKey",
            "Ec2AvailabilityZone": "us-east-1c",
            "EmrManagedSlaveSecurityGroup": "sg-example"
        },
        "Name": "Development Cluster",
        "ServiceRole": "EMR_DefaultRole",
        "Tags": [],
        "TerminationProtected": false,
        "ReleaseLabel": "emr-4.0.0",
        "NormalizedInstanceHours": 16,
        "InstanceGroups": [
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1438281058.101,
                        "CreationDateTime": 1438280702.499
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "CORE",
                "InstanceGroupType": "CORE",
                "Id": "ig-2EEXAMPLEXXP",
                "Configurations": [],
                "InstanceType": "m5.xlarge",
```

```
                "Market": "ON_DEMAND",
                "RunningInstanceCount": 1
            },
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1438281023.879,
                        "CreationDateTime": 1438280702.499
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "MASTER",
                "InstanceGroupType": "MASTER",
                "Id": "ig-2A1234567XP",
                "Configurations": [],
                "InstanceType": "m5.xlarge",
                "Market": "ON_DEMAND",
                "RunningInstanceCount": 1
            }
        ],
        "Applications": [
            {
                "Version": "1.0.0",
                "Name": "Hive"
            },
            {
                "Version": "2.6.0",
                "Name": "Hadoop"
            },
            {
                "Version": "0.14.0",
                "Name": "Pig"
            },
            {
                "Version": "1.4.1",
                "Name": "Spark"
            }
        ],
        "BootstrapActions": [],
        "MasterPublicDnsName": "ec2-X-X-X-X.compute-1.amazonaws.com",
```

```
        "AutoTerminate": false,
        "Id": "j-jobFlowID",
        "Configurations": [
            {
                "Properties": {
                    "hadoop.security.groups.cache.secs": "250"
                },
                "Classification": "core-site"
            },
            {
                "Properties": {
                    "mapreduce.tasktracker.reduce.tasks.maximum": "5",
                    "mapred.tasktracker.map.tasks.maximum": "2",
                    "mapreduce.map.sort.spill.percent": "90"
                },
                "Classification": "mapred-site"
            },
            {
                "Properties": {
                    "hive.join.emit.interval": "1000",
                    "hive.merge.mapfiles": "true"
                },
                "Classification": "hive-site"
            }
        ]
    }
}
```

Example Mostrar clústeres por fecha de creación

Para recuperar clústeres creados en un intervalo de fechas específico, use el comando `list-clusters` con los parámetros `--created-after` y `--created-before`.

El siguiente comando muestra todos los clústeres creados entre el 9 de octubre de 2019 y el 12 de octubre de 2019.

```
aws emr list-clusters --created-after 2019-10-09T00:12:00 --created-
before 2019-10-12T00:12:00
```

Example Mostrar clústeres por estado

Para mostrar clústeres por estado, use el comando `list-clusters` con el parámetro `--cluster-states`. Los estados de clúster válidos incluyen: STARTING, BOOTSTRAPPING, RUNNING, WAITING, TERMINATING, TERMINATED y TERMINATED_WITH_ERRORS.

```
aws emr list-clusters --cluster-states TERMINATED
```

También puede utilizar los siguientes parámetros de acceso directo para mostrar todos los clústeres con los estados especificados:

- El parámetro `--active` filtra los clústeres por los estados STARTING, BOOTSTRAPPING, RUNNING, WAITING o TERMINATING.
- El parámetro `--terminated` filtra los clústeres por el estado TERMINATED.
- El parámetro `--failed` filtra los clústeres por el estado TERMINATED_WITH_ERRORS.

Los siguientes comandos devuelven el mismo resultado.

```
aws emr list-clusters --cluster-states TERMINATED
```

```
aws emr list-clusters --terminated
```

Para obtener más información acerca de los estados del clúster, consulte Descripción del ciclo de vida del clúster.

Depuración de pasos mejorada

Si un paso de Amazon EMR devuelve un error y envió el trabajo utilizando la operación de la API de pasos con la versión 5.x o posterior de la AMI, Amazon EMR puede identificar y devolver la causa raíz del error del paso en algunos casos, junto con el nombre del archivo de registro relevante y una parte del registro de seguimiento de la pila de la aplicación a través de la API. Por ejemplo, puede identificar los siguientes errores:

- Un error de Hadoop común como, por ejemplo, el directorio de salida ya existe, el directorio de entrada no existe o una aplicación se queda sin memoria.
- Errores de Java como, por ejemplo, una aplicación que se ha compilado con una versión incompatible de Java o se ha ejecutado con una clase principal que no se encuentra.

- Un problema al acceder a objetos almacenados en Amazon S3.

Esta información está disponible utilizando las operaciones DescribeStep y ListSteps de la API. El campo FailureDetails de StepSummary devuelto por dichas operaciones. Para acceder a la información de FailureDetails, utilice la AWS CLI, la consola o el SDK de AWS.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Tenga en cuenta que la nueva consola de Amazon EMR no ofrece la depuración de pasos. Sin embargo, puede ver los detalles de la terminación del clúster siguiendo estos pasos.

Para ver los detalles del error con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, seleccione Clústeres y, a continuación, seleccione el clúster que desee ver.
3. Anote el valor del estado de la sección Resumen de la página de detalles del clúster. Si el estado es terminado con errores, pase el ratón sobre el texto para ver los detalles de los errores del clúster.

Old console

Para ver los detalles del error con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Cluster List (Lista de clústeres) y seleccione un clúster.

3. Seleccione el icono de flecha junto a cada paso para ver más detalles. Si el paso ha devuelto un error y Amazon EMR puede identificar la causa raíz, verá los detalles del error.

CLI

Para ver los detalles del error con la AWS CLI

- Para obtener los detalles del error de un paso con la AWS CLI, utilice el comando `describe-step`.

```
aws emr describe-step --cluster-id j-1K48XXXXXHCB --step-id s-3QM0XXXXXM1W
```

El resultado tendrá un aspecto similar al siguiente:

```
{
  "Step": {
    "Status": {
      "FailureDetails": {
        "LogFile": "s3://myBucket/logs/j-1K48XXXXXHCB/steps/s-3QM0XXXXXM1W/
stderr.gz",
        "Message": "org.apache.hadoop.mapred.FileAlreadyExistsException: Output
 directory s3://myBucket/logs/beta already exists",
        "Reason": "Output directory already exists."
      },
      "Timeline": {
        "EndDateTime": 1469034209.143,
        "CreationDateTime": 1469033847.105,
        "StartDateTime": 1469034202.881
      },
      "State": "FAILED",
      "StateChangeReason": {}
    },
    "Config": {
      "Args": [
        "wordcount",
        "s3://myBucket/input/input.txt",
        "s3://myBucket/logs/beta"
      ],
      "Jar": "s3://myBucket/jars/hadoop-mapreduce-examples-2.7.2-amzn-1.jar",
      "Properties": {}
    },
```

```
    "Id": "s-3QM0XXXXXM1W",
    "ActionOnFailure": "CONTINUE",
    "Name": "ExampleJob"
  }
}
```

Ver el historial de aplicaciones

Puede ver los detalles de aplicación del servidor del historial de Spark y de YARN Timeline Service en la página de detalles del clúster en la consola. El historial de aplicaciones de Amazon EMR le permite solucionar problemas y analizar más fácilmente los trabajos activos y el historial de trabajos.

New console

La sección Interfaces de usuario de aplicaciones de la pestaña Aplicaciones ofrece varias opciones de visualización, en función del estado del clúster y de las aplicaciones que haya instalado en él.

- Acceso fuera del clúster a las interfaces de usuario de aplicaciones persistentes: a partir de la versión 5.25.0 de Amazon EMR, los enlaces a las interfaces de usuario de aplicaciones persistentes están disponibles para la interfaz de usuario de Spark y el servicio de historial de Spark. Con las versiones 5.30.1 y posteriores de Amazon EMR, Tez UI y el servidor de YARN Timeline también tienen interfaces de usuario de aplicaciones persistentes. El servidor de YARN Timeline y Tez UI son aplicaciones de código abierto que proporcionan métricas y herramientas visuales para clústeres activos y terminados. La interfaz de usuario de Spark ofrece detalles acerca de las etapas y tareas del programador, tamaños de RDD y uso de memoria, información medioambiental e información acerca de los ejecutores en ejecución. Las interfaces de usuario de aplicaciones persistentes se ejecutan fuera del clúster, por lo que la información del clúster y los registros están disponibles durante 30 días después de que termine una aplicación. A diferencia de las interfaces de usuario de aplicaciones en clúster, las interfaces de usuario de aplicaciones persistentes no requieren que configure un proxy web a través de una conexión SSH.
- Interfaces de usuario de aplicaciones en el clúster: hay una variedad de interfaces de usuario del historial de aplicaciones que se pueden ejecutar en un clúster. Las interfaces de usuario en el clúster están alojadas en el nodo maestro y requieren que configure una conexión SSH al servidor web. Las interfaces de usuario de aplicaciones en el clúster mantienen el historial de aplicaciones durante una semana después de que termine una aplicación. Para obtener más

información e instrucciones sobre cómo configurar un túnel SSH, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Con la excepción del servidor del historial de Spark, el servidor de YARN Timeline y las aplicaciones de Hive, el historial de aplicaciones en el clúster solo se puede consultar mientras el clúster está en ejecución.

Old console

La pestaña Interfaces de usuario de aplicaciones proporciona varias opciones de visualización:

- Acceso fuera del clúster a las interfaces de usuario de aplicaciones persistentes: a partir de la versión 5.25.0 de Amazon EMR, los enlaces a las interfaces de usuario de aplicaciones persistentes están disponibles para Spark. Con las versiones 5.30.1 y posteriores de Amazon EMR, Tez UI y el servidor de YARN Timeline también tienen interfaces de usuario de aplicaciones persistentes. El servidor de YARN Timeline y Tez UI son aplicaciones de código abierto que proporcionan métricas y herramientas visuales para clústeres activos y terminados. La interfaz de usuario de Spark ofrece detalles acerca de las etapas y tareas del programador, tamaños de RDD y uso de memoria, información medioambiental e información acerca de los ejecutores en ejecución. Las interfaces de usuario de aplicaciones persistentes se ejecutan fuera del clúster, por lo que la información del clúster y los registros están disponibles durante 30 días después de que termine una aplicación. A diferencia de las interfaces de usuario de aplicaciones en clúster, las interfaces de usuario de aplicaciones persistentes no requieren que configure un proxy web a través de una conexión SSH.
- Interfaces de usuario de aplicaciones en el clúster: hay una variedad de interfaces de usuario del historial de aplicaciones que se pueden ejecutar en un clúster. Las interfaces de usuario en el clúster están alojadas en el nodo maestro y requieren que configure una conexión SSH al servidor web. Las interfaces de usuario de aplicaciones en el clúster mantienen el historial de aplicaciones durante una semana después de que termine una aplicación. Para obtener más información e instrucciones sobre cómo configurar un túnel SSH, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

  Con la excepción del servidor del historial de Spark, el servidor de YARN Timeline y las aplicaciones de Hive, el historial de aplicaciones en el clúster solo se puede consultar mientras el clúster está en ejecución.
- Historial de aplicaciones de alto nivel: con las versiones de 5.8.0 a 5.36.0 y de 6.x a 6.8.0 de Amazon EMR, puede ver un resumen del historial de aplicaciones en la consola antigua de

Amazon EMR, incluidas las métricas clave de las tareas de fase y los ejecutores. A partir del 23 de enero de 2023, Amazon EMR retirará el historial de aplicaciones de alto nivel para todas las versiones. Si utiliza la versión 5.25.0 de Amazon EMR o una superior, le recomendamos que utilice en su lugar la interfaz de usuario de aplicaciones persistentes.

Ver interfaces de usuario de aplicaciones persistentes

A partir de la versión 5.25.0 de Amazon EMR, puede conectarse a los detalles persistentes de la aplicación del servidor del historial de Spark alojados fuera del clúster mediante la página Resumen del clúster o la pestaña Interfaces de usuario de aplicaciones de la consola. Las interfaces de aplicación persistentes de Tez UI y YARN Timeline están disponibles a partir de la versión 5.30.1 de Amazon EMR. El acceso de enlace con un solo clic al historial de aplicaciones persistentes ofrece las siguientes ventajas:

- Puede analizar y solucionar rápidamente los trabajos activos y el historial de trabajos sin configurar un proxy web a través de una conexión SSH.
- Puede acceder al historial de aplicaciones y a los archivos de registro pertinentes para los clústeres activos y terminados. Los registros están disponibles durante 30 días desde que finalice la aplicación.

En la pestaña Interfaces de usuario de aplicaciones o en la página Resumen del clúster de la consola antigua de las versiones 5.30.1 o 6.x de Amazon EMR, seleccione el enlace Servidor de YARN Timeline, Tez UI o Servidor del historial de Spark.

La interfaz de usuario de la aplicación se abre en una nueva pestaña del navegador. Para obtener más información, consulte Monitoring and Instrumentation.

Puede ver los registros de contenedores de YARN a través de los enlace en Spark History Server, el servidor de YARN Timeline y Tez UI.

> **Note**
>
> Para obtener acceso a los registros de contenedor de YARN desde el servidor del historial de Spark, el servidor de YARN Timeline y Tez UI, debe habilitar el registro en Amazon S3 para su clúster. Si el registro no está habilitado, los enlaces a los registros de contenedor de YARN no funcionarán.

Recopilación de registros

Para habilitar el acceso con un solo clic a las interfaces de usuario de aplicaciones persistentes, Amazon EMR recopila dos tipos de registros:

- Los registros de eventos de aplicación se recopilan en un bucket del sistema EMR. Los registros de eventos se cifran en reposo mediante el cifrado del lado del servidor con claves administradas de Amazon S3 (SSE-S3). Si utiliza una subred privada para el clúster, asegúrese de incluir `"arn:aws:s3:::prod.MyRegion.appinfo.src/*"` en la lista de recursos de la política de Amazon S3 para la subred privada. Para obtener más información, consulte Política de Amazon S3 mínima para subred privada.
- Los registros de contenedores de YARN se recopilan en un bucket de Amazon S3 propiedad de usted. Debe habilitar el registro para que su clúster acceda a los registros de contenedores de YARN. Para más información, consulte Configurar el registro y la depuración de un clúster.

Si necesita deshabilitar esta característica por motivos de privacidad, puede detener el daemon mediante un script de arranque al crear un clúster, como se muestra en el siguiente ejemplo.

```
aws emr create-cluster --name "Stop Application UI Support" --release-label emr-5.36.1
 \
--applications Name=Hadoop Name=Spark --ec2-attributes KeyName=<myEMRKeyPairName> \
--instance-groups InstanceGroupType=MASTER,InstanceCount=1,InstanceType=m3.xlarge
 InstanceGroupType=CORE,InstanceCount=1,InstanceType=m3.xlarge
 InstanceGroupType=TASK,InstanceCount=1,InstanceType=m3.xlarge \
--use-default-roles --bootstrap-actions Path=s3://region.elasticmapreduce/bootstrap-
actions/run-if,Args=["instance.isMaster=true","echo Stop Application UI | sudo tee /
etc/apppusher/run-apppusher; sudo systemctl stop apppusher || exit 0"]
```

Después de ejecutar este script de arranque, Amazon EMR no recopilará ningún registro de eventos del servidor del historial de Spark ni del servidor de YARN Timeline en el bucket del sistema de EMR. No habrá información del historial de aplicaciones disponible en la pestaña Interfaces de usuario de aplicaciones y perderá el acceso a todas las interfaces de usuario de la aplicación desde la consola.

Archivos de registro de eventos de Spark de gran tamaño

En algunos casos, los trabajos de Spark de larga duración, como las transmisiones de Spark, y los trabajos de gran tamaño, como las consultas SQL de Spark, pueden generar registros de eventos de gran tamaño. Con los registros de eventos de gran tamaño, puede ocupar rápidamente el espacio en disco de las instancias de computación y encontrar errores `OutOfMemory` al cargar las interfaces

de usuario persistentes. Para evitar estos problemas, recomendamos que active la característica de acumulación y compactación de registros de eventos de Spark. Esta característica solo está disponible en las versiones emr-6.1.0 y posteriores de Amazon EMR. Para obtener más información sobre la acumulación y la compactación, consulte Applying compaction on rolling event log files en la documentación de Spark.

Para activar la característica de acumulación y compactación de registros de eventos de Spark, active los siguientes ajustes en la configuración de Spark.

- `spark.eventLog.rolling.enabled`: activa la acumulación de registros de eventos en función del tamaño. Este ajuste está desactivado de forma predeterminada.
- `spark.eventLog.rolling.maxFileSize`: cuando se activa la acumulación, especifica el tamaño máximo del archivo de registros de eventos antes de que se acumule. El valor predeterminado es 128 MB.
- `spark.history.fs.eventLog.rolling.maxFilesToRetain`: especifica el número máximo de archivos de registros de eventos no compactados que se deben retener. De forma predeterminada, se retienen todos los archivos de registros de eventos. Configúrelo en un número inferior para compactar los registros de eventos más antiguos. El valor más bajo es 1.

Tenga en cuenta que la compactación intenta excluir los eventos con archivos de registros de eventos desactualizados, como los siguientes. Si descarta los eventos, dejará de verlos en la interfaz de usuario del servidor del historial de Spark.

- Eventos para trabajos terminados y eventos relacionados de etapas o tareas.
- Eventos para ejecutores terminados.
- Eventos para consultas SQL completadas y eventos relacionados de trabajos, etapas y tareas.

Para lanzar un clúster con la acumulación y la compactación activadas

1. Cree un archivo `spark-configuration.json` con la siguiente configuración.

```
[
   {
     "Classification": "spark-defaults",
        "Properties": {
           "spark.eventLog.rolling.enabled": true,
           "spark.history.fs.eventLog.rolling.maxFilesToRetain": 1
        }
```

```
    }
]
```

2. Cree su clúster con la configuración de compactación de acumulación de Spark de la siguiente manera.

```
aws emr create-cluster \
--release-label emr-6.6.0 \
--instance-type m4.large \
--instance-count 2 \
--use-default-roles \
--configurations file://spark-configuration.json
```

Condiciones y limitaciones

El acceso con un solo clic a las interfaces de usuario de aplicaciones persistentes actualmente tiene las siguientes limitaciones.

- Habrá un retraso de al menos dos minutos cuando los detalles de la aplicación aparezcan en la interfaz de usuario del servidor del historial de Spark.
- Esta característica solo funciona cuando el directorio de registros de eventos de la aplicación está en HDFS. De forma predeterminada, Amazon EMR almacena los registros de eventos en un directorio de HDFS. Si cambia el directorio predeterminado a un sistema de archivos diferente, como Amazon S3, esta característica no funcionará.
- Esta función no está disponible actualmente para clústeres de EMR con varios nodos maestros ni para clústeres de EMR integrados con AWS Lake Formation.
- Para habilitar el acceso con un solo clic a las interfaces de usuario de aplicaciones persistentes, debe tener permiso para la acción `DescribeCluster` de Amazon EMR. Si deniega el permiso de una entidad principal de IAM a esta acción, el cambio de permiso tarda aproximadamente cinco minutos en propagarse.
- Si vuelve a configurar las aplicaciones en un clúster en ejecución, el historial de aplicaciones no estará disponible a través de la interfaz de usuario de la aplicación.
- Para cada Cuenta de AWS, el límite predeterminado de las interfaces de usuario de aplicaciones activas es de 200.
- Puede acceder a las interfaces de usuario de aplicaciones desde la consola en las regiones Este de EE. UU. (Norte de Virginia), Oeste de EE. UU. (Norte de California), Canadá (centro), UE (Fráncfort, Irlanda, Londres, París, Estocolmo), Asia-Pacífico (Bombay, Seúl, Singapur, Sídney y

Tokio), América del Sur (São Paulo), China (Pekín) operada por Sinnet y China (Ningxia) operada por NWCD.

Ver el historial de una aplicación de alto nivel

> **Note**
>
> Recomendamos que utilice la interfaz de aplicaciones persistentes para mejorar la experiencia de usuario y retener el historial de las aplicaciones durante un máximo de 30 días. El historial de aplicaciones de alto nivel que se describe en esta página no está disponible en la nueva consola de Amazon EMR (https://console.aws.amazon.com/emr). Para obtener más información, consulte Ver interfaces de usuario de aplicaciones persistentes.

Con las versiones de 5.8.0 a 5.36.0 y de 6.x a 6.8.0 de Amazon EMR, puede ver el historial de aplicaciones de alto nivel desde la pestaña Interfaces de usuario de aplicaciones de la consola antigua de Amazon EMR. La interfaz de usuario de aplicaciones de Amazon EMR conserva el resumen del historial de la aplicación durante 7 días después de que se haya completado la solicitud.

Condiciones y limitaciones

Tenga en cuenta las siguientes limitaciones cuando utilice la pestaña Interfaces de usuario de aplicaciones en la antigua consola de Amazon EMR.

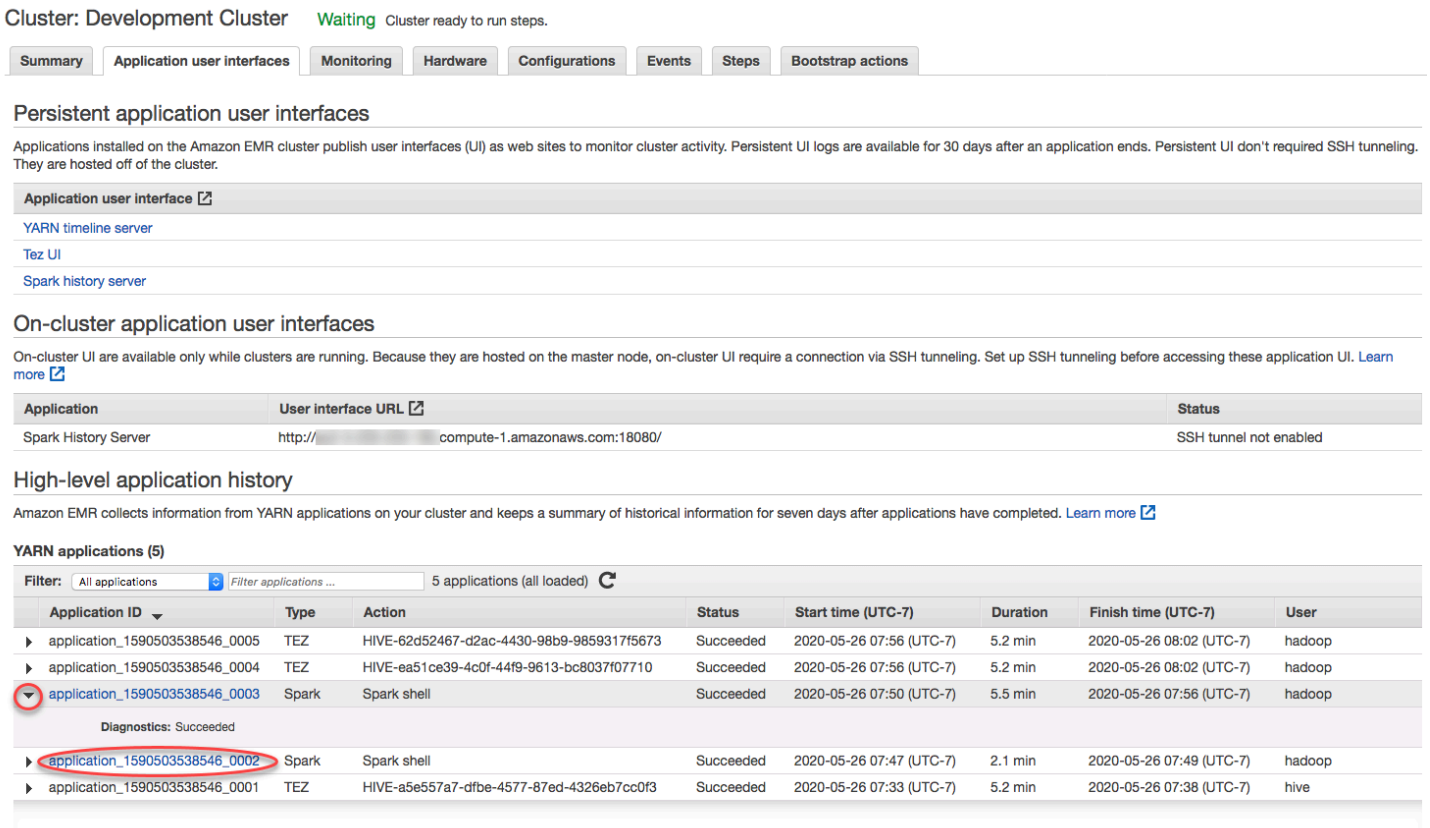
- Solo puede acceder a la característica del historial de aplicaciones de alto nivel cuando utiliza las versiones de 5.8.0 a 5.36.0 y de 6.x a 6.8.0 de Amazon EMR. A partir del 23 de enero de 2023, Amazon EMR retirará el historial de aplicaciones de alto nivel para todas las versiones. Si utiliza la versión 5.25.0 de Amazon EMR o una superior, le recomendamos que utilice en su lugar la interfaz de usuario de aplicaciones persistentes.
- La característica del historial de aplicaciones de alto nivel no es compatible con las aplicaciones de Spark Streaming.
- El acceso con un solo clic a las interfaces de usuario de aplicaciones persistentes no está disponible actualmente para los clústeres de Amazon EMR con varios nodos maestros ni para los clústeres de Amazon EMR integrados con AWS Lake Formation.

Ejemplo: ver el historial de una aplicación de alto nivel

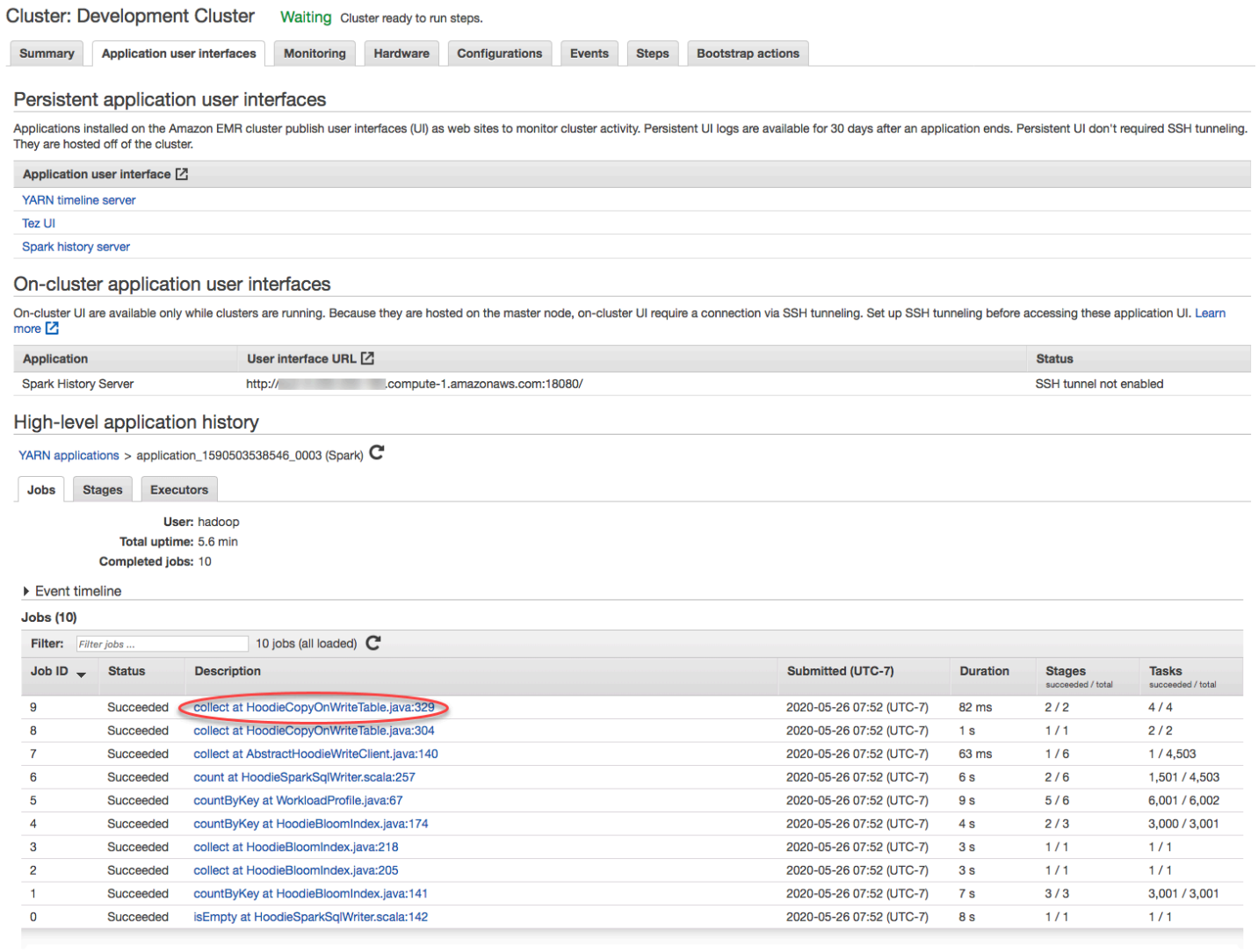
En la siguiente secuencia se muestra cómo desplazarse por los detalles de los trabajos en las aplicaciones de Spark o YARN utilizando la opción Interfaces de usuario de aplicaciones de la página de detalles del clúster de la consola antigua.

Para ver los detalles del clúster, seleccione Nombre en la lista Clústeres. Para ver información sobre los registros de contenedor de YARN, debe habilitar el registro para el clúster. Para más información, consulte Configurar el registro y la depuración de un clúster. Para el historial de aplicaciones de Spark, la información proporcionada en la tabla de resumen es solo un subconjunto de la información disponible a través de la interfaz de usuario del servidor del historial de Spark.

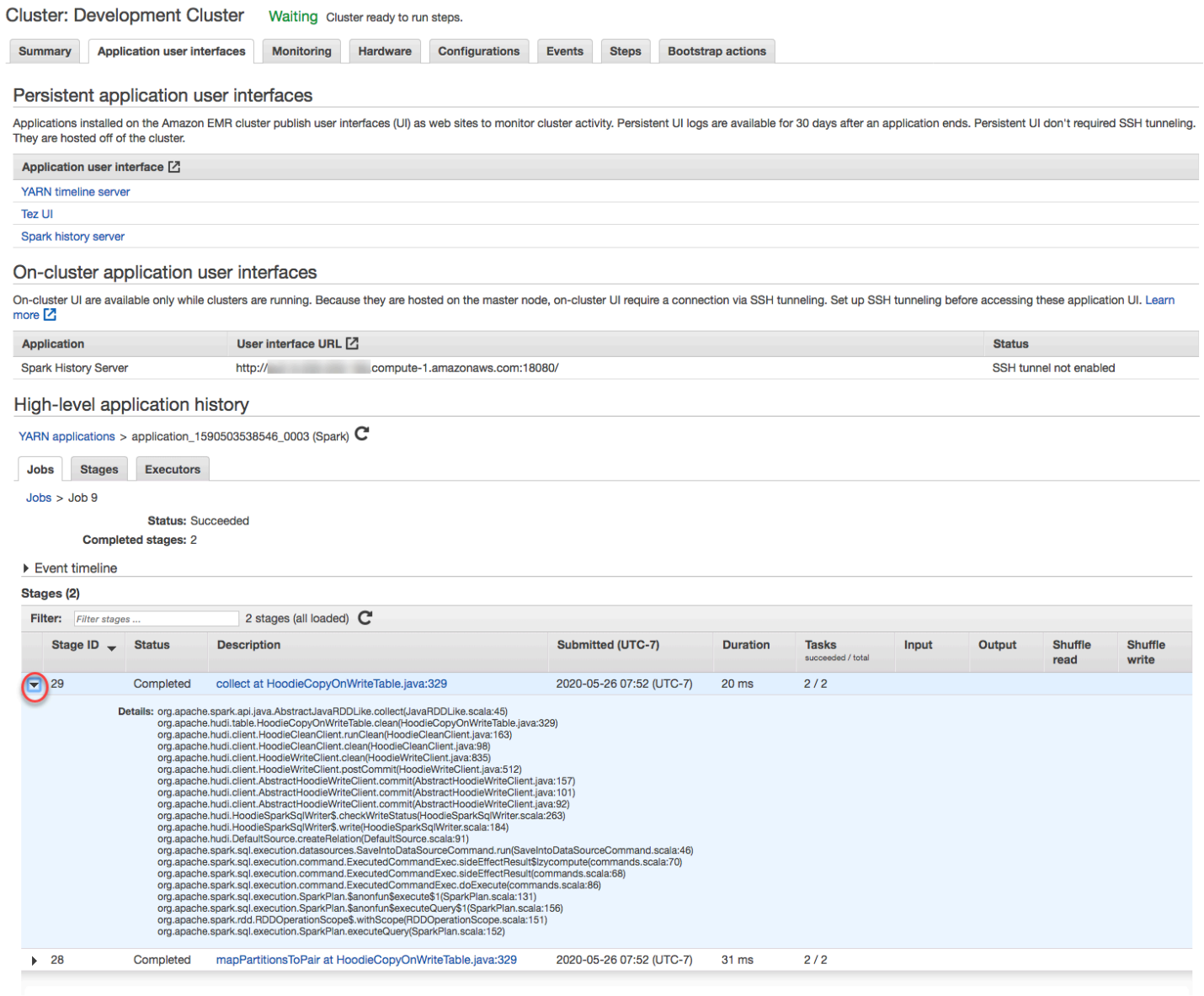
En la pestaña Interfaces de usuario de aplicaciones de Historial de aplicaciones de alto nivel, puede expandir una fila para mostrar el resumen del diagnóstico de una aplicación de Spark o seleccionar un enlace con el identificador de la aplicación para ver los detalles de otra aplicación.



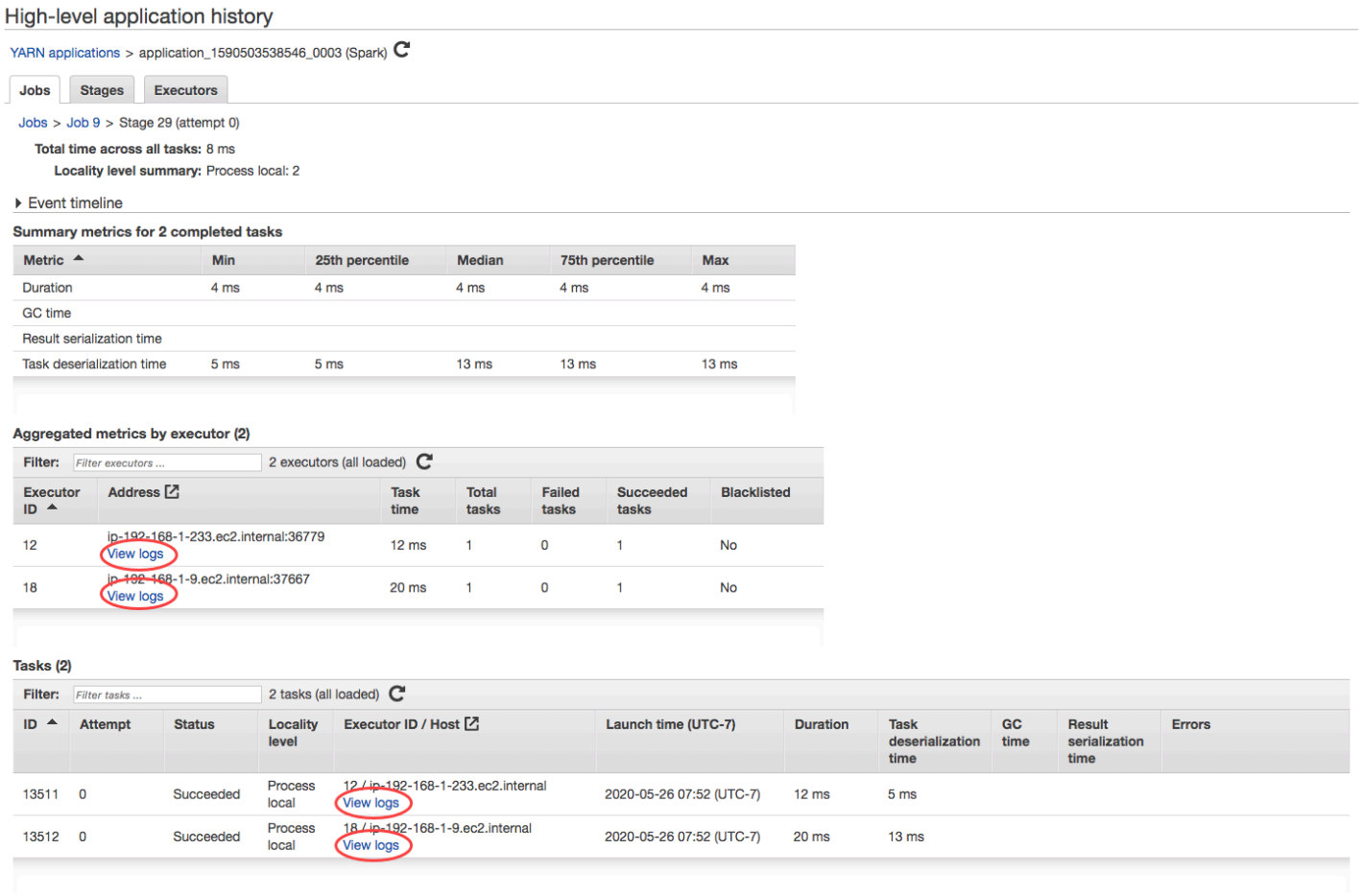
Al seleccionar un enlace con un ID de aplicación, la interfaz de usuario cambia para mostrar los detalles de la aplicación YARN de esa aplicación. En la pestaña Trabajos de los detalles de la aplicación YARN, puede seleccionar el enlace Descripción de un trabajo para mostrar los detalles de ese trabajo.



En la página de detalles del trabajo, puede expandir la información sobre etapas específicas del trabajo y, a continuación, seleccione el enlace Descripción para ver los detalles de la etapa.



En la página de detalles de la etapa, puede ver las métricas clave de las tareas y los ejecutores de la etapa. También puede ver los registros de tareas y ejecutores mediante los enlaces Ver registros.



Ver archivos de registro de

Amazon EMR y Hadoop producen archivos de registro que notifican el estado en el clúster. De forma predeterminada, se escriben en el nodo principal del directorio `/mnt/var/log/`. En función de cómo configure el clúster al lanzarlo, estos registros también pueden archivarse en Amazon S3 y pueden ser visibles a través de la herramienta de depuración gráfica.

Existen muchos tipos de registros escritos en el nodo principal. Amazon EMR escribe registros de estado de instancias, pasos y acciones de arranque. Apache Hadoop escribe registros para notificar el procesamiento de los trabajos, de las tareas y de los intentos de tareas. Hadoop también guarda registros de sus daemons. Para obtener más información sobre los registros escritos por Hadoop, consulte http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/ClusterSetup.html.

Ver archivos de registro en el nodo principal

En la siguiente tabla se enumeran algunos de los archivos de registro que encontrará en el nodo principal.

Location	Descripción
/emr/instance-controller/log/bootstrap-actions	Los registros escritos durante el procesamiento de las acciones de arranque.
/mnt/var/log/hadoop-state-pusher	Los registros escritos por el proceso state pusher de Hadoop.
/emr/instance-controller/log	Registros de controlador de instancia.
/emr/instance-state	Registros de estado de instancia. Contienen información sobre la CPU, el estado de la memoria y los subprocesos del recolector de elementos no utilizados del nodo.
/emr/service-nanny	Los registros escritos por el proceso nanny de servicio.
/mnt/var/log/*aplicación*	Registros específicos de una aplicación como, por ejemplo, Hadoop, Spark o Hive.
/mnt/var/log/hadoop/steps/*N*	Registros de paso que contienen información sobre el procesamiento del paso. El valor de *N* indica el stepId asignado por Amazon EMR. Por ejemplo, un clúster tiene dos pasos: `s-1234ABCDEFGH` y `s-5678IJKLMNOP`. El primer paso está ubicado en `/mnt/var/log/hadoop/steps/s-1234ABCDEFGH/` y el segundo paso en `/mnt/var/log/hadoop/steps/s-5678IJKLMNOP/`. Los registros de pasos escritos por Amazon EMR son los siguientes.

Location	Descripción
	• controlador: información sobre el procesamiento del paso. Si se produce un error en el paso durante la carga, puede encontrar el registro de seguimiento de la pila en este registro. • syslog: describe la ejecución de los trabajos de Hadoop en el paso. • stderr: el canal de error estándar de Hadoop mientras procesa el paso. • stdout: el canal de salida estándar de Hadoop mientras procesa el paso.

Para ver los archivos de registro en el nodo principal con la AWS CLI.

1. Utilice SSH para conectarse al nodo principal como se describe en Conectarse al nodo principal mediante SSH.
2. Vaya al directorio que contiene la información del archivo de registro que desea ver. La tabla anterior ofrece una lista de los tipos de archivos de registro que están disponibles y donde los encontrará. El siguiente ejemplo muestra el comando para acceder al registro de paso con un ID, `s-1234ABCDEFGH`.

   ```
   cd /mnt/var/log/hadoop/steps/s-1234ABCDEFGH/
   ```

3. Utilice el visor de archivos que desee para ver el archivo de log. En el siguiente ejemplo se usa el comando `less` de Linux para ver el archivo de log `controller`.

   ```
   less controller
   ```

Ver los archivos de registro archivados en Amazon S3

De forma predeterminada, los clústeres de Amazon EMR que se lanzan utilizando la consola archivan automáticamente los archivos de registro en Amazon S3. Puede especificar su propia ruta de registro o bien puede permitir que la consola genere automáticamente una ruta de registro por

usted. En los clústeres que se lanzan con la CLI o la API, debe configurar el archivado de registros de Amazon S3 manualmente.

Cuando Amazon EMR está configurado para archivar los archivos de registro en Amazon S3, almacena los archivos en la ubicación de S3 que se haya especificado, en la carpeta /*cluster-id*/, donde *cluster-id* es el identificador del clúster.

En la siguiente tabla se enumeran algunos de los archivos de registro que encontrará en Amazon S3.

Location	Descripción
/*cluster-id* /node/	Los registros de nodo, incluida la acción de arranque, el estado de la instancia y los registros de aplicación para el nodo. Los registros para cada nodo se almacenan en una carpeta etiquetada con el identificador de la instancia EC2 de ese nodo.
/*cluster-id* /node/*instance-id* /*application*	Los registros creados por cada aplicación o daemon asociado con una aplicación. Por ejemplo, el registro del servidor de Hive se encuentra en *cluster-id* /node/*instance-id* /hive/hive-server.log .
/*cluster-id* /steps/*step-id*/	Registros de paso que contienen información sobre el procesamiento del paso. El valor de *step-id* indica el ID de paso asignado por Amazon EMR. Por ejemplo, un clúster tiene dos pasos: s-1234ABCDEFGH y s-5678IJKLMNOP . El primer paso está ubicado en /mnt/var/log/hadoop/steps/s-1234ABCDEFGH/ y el segundo paso en /mnt/var/log/hadoop/steps/s-5678IJKLMNOP/ . Los registros de pasos escritos por Amazon EMR son los siguientes.

Location	Descripción
	• controlador: información sobre el procesamiento del paso. Si se produce un error en el paso durante la carga, puede encontrar el registro de seguimiento de la pila en este registro. • syslog: describe la ejecución de los trabajos de Hadoop en el paso. • stderr: el canal de error estándar de Hadoop mientras procesa el paso. • stdout: el canal de salida estándar de Hadoop mientras procesa el paso.
/*cluster-id* /containers	Registros de contenedor de aplicaciones. Los registros para cada aplicación YARN se almacenan en estas ubicaciones.
/*cluster-id* /hadoop-mapreduce/	Los registros que contienen información acerca de los detalles de configuración y el historial de trabajos de MapReduce.

Para ver los archivos de registro archivados en Amazon S3 con la consola de Amazon S3

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon S3 en https://console.aws.amazon.com/s3/.
2. Abra el bucket de S3 que especificó al configurar el clúster para archivar los archivos de registro en Amazon S3.
3. Acceda al archivo de registro que contiene la información que se va a mostrar. La tabla anterior ofrece una lista de los tipos de archivos de registro que están disponibles y donde los encontrará.
4. Descargue el objeto del archivo de registro para verlo. Para obtener instrucciones, consulte Descarga de un objeto.

Ver archivos de registro en la herramienta de depuración

Amazon EMR no habilita automáticamente la herramienta de depuración. Debe configurarla al lanzar el clúster. Tenga en cuenta que la nueva consola de Amazon EMR no ofrece la herramienta de depuración.

Para ver los registros del clúster con la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. En la página Lista de clústeres, seleccione el icono de detalles situado junto al clúster que desee ver.

   Se abrirá la página Detalles del clúster. En la sección Pasos, los enlaces situados a la derecha de cada paso muestran los distintos tipos de registros disponibles para el paso. Amazon EMR genera estos registros.
3. Para ver una lista de los trabajos de Hadoop asociados a un paso determinado, seleccione el enlace Ver trabajos situado a la derecha del paso.
4. Para ver una lista de las tareas de Hadoop asociadas a un trabajo determinado, seleccione el enlace Ver tareas situado a la derecha del trabajo.
5. Para ver una lista de los intentos que ha realizado una tarea determinada al intentar completarla, seleccione el enlace Ver intentos situado a la derecha de la tarea.
6. Para ver los registros generados por un intento de tarea, seleccione los enlaces stderr, stdout y syslog situados a la derecha del intento de la tarea.

La herramienta de depuración muestra enlaces a los archivos de registro después de que Amazon EMR cargue los archivos de registro en el bucket de Amazon S3. Dado que los archivos de registro se cargan en Amazon S3 cada cinco minutos, puede tardar algunos minutos en que las cargas de archivo de registro se completen una vez completo el paso.

Amazon EMR actualiza periódicamente el estado de los trabajos, las tareas y los intentos de tareas de Hadoop en la herramienta de depuración. Puede hacer clic en Refresh List (Actualizar lista) en los paneles de depuración para obtener el estado más actualizado de estos elementos.

Ver instancias del clúster en Amazon EC2

Para ayudarle a gestionar los recursos, Amazon EC2 le permite asignar metadatos a los recursos en forma de etiquetas. Cada etiqueta de Amazon EC2 consta de una clave y un valor. Las etiquetas le permiten clasificar los recursos de Amazon EC2 de diversas maneras: por ejemplo, según su finalidad, propietario o entorno.

Puede buscar y filtrar los recursos en función de las etiquetas. Las etiquetas que asigne a los recursos a través de su cuenta de AWS solo están disponibles para usted. Otras cuentas que compartan el mismo recurso no pueden ver sus etiquetas.

Amazon EMR etiqueta automáticamente cada instancia de EC2 que lanza con pares clave-valor. Las claves identifican el clúster y el grupo de instancias al que pertenece la instancia. Esto facilita la tarea de filtrar sus instancias de EC2 para mostrar, por ejemplo, únicamente aquellas instancias pertenecientes a un determinado clúster o para mostrar todas las instancias en ejecución actualmente en el grupo de instancias de tarea. Esto resulta especialmente útil si va a ejecutar varios clústeres simultáneamente o a gestionar un gran número de instancias de EC2.

Estos son los pares clave-valor predefinidos que asigna Amazon EMR:

Clave	Valor	Definición de los valores
aws:elasticmapreduce:job-flow-id	*job-flow-identifier*	El ID del clúster para el que se aprovisiona la instancia. Aparece en el formato `j-XXXXXXXXXXXXX` y puede tener una longitud de hasta 256 caracteres.
aws:elasticmapreduce:instance-group-role	*group-role*	El tipo de grupo de instancias, introducido como uno de los siguientes valores: `master`, `core` o `task`.

Puede ver y filtrar por las etiquetas que agrega Amazon EMR. Para obtener más información, consulte Uso de etiquetas en la Guía del usuario de Amazon EC2 para instancias de Linux. Dado que las etiquetas definidas por Amazon EMR son etiquetas del sistema y no se pueden editar ni eliminar, las secciones sobre mostrar y filtrar las etiquetas son las más importantes.

> **Note**
>
> Amazon EMR agrega etiquetas a la instancia de EC2 cuando su estado se actualiza a En ejecución. Si la latencia se produce entre el momento en que se aprovisiona la instancia de EC2 y el momento en que su estado se establece en En ejecución, las etiquetas que establezca Amazon EMR aparecerán una vez que se inicie la instancia. Si no ve las etiquetas, espere algunos minutos y actualice la vista.

Eventos y métricas de CloudWatch

Utilice eventos y métricas para realizar un seguimiento de la actividad y el estado de un clúster de Amazon EMR. Los eventos son útiles para monitorizar la incidencia específica dentro de un clúster; por ejemplo, cuando un clúster cambia de estado desde el inicio a la ejecución. Las métricas son útiles para monitorizar un valor específico; por ejemplo, el porcentaje de espacio en disco disponible que utiliza HDFS dentro de un clúster.

Para obtener más información acerca de Eventos de CloudWatch, consulte la Guía del usuario de Eventos de Amazon CloudWatch. Para obtener más información sobre las métricas de CloudWatch, consulte Uso de las métricas de Amazon CloudWatch y Creación de alarmas de Amazon CloudWatch en la Guía del usuario de Amazon CloudWatch.

Temas

- Monitorización de métricas de Amazon EMR con CloudWatch
- Monitorización de eventos de Amazon EMR con CloudWatch
- Responder a los eventos de CloudWatch

Monitorización de métricas de Amazon EMR con CloudWatch

Las métricas se actualizan cada cinco minutos y se recogen automáticamente y se envían a CloudWatch por cada clúster de Amazon EMR. Este intervalo no se puede configurar. No se aplica ningún cargo por las métricas de Amazon EMR que se comunican en CloudWatch. Estas métricas de puntos de datos de cinco minutos se archivan durante 63 días, tras lo cual se descartan los datos.

¿Cómo utilizo las métricas de Amazon EMR?

En la siguiente tabla, se muestran los usos más comunes de las métricas notificadas por Amazon EMR. Se trata de sugerencias que puede usar como punto de partida y no de una lista

completa. Para ver una lista completa de las métricas notificadas por Amazon EMR, consulte Métricas notificadas por Amazon EMR en CloudWatch.

¿Cómo?	Métricas relevantes
Realizar un seguimiento del progreso de mi clúster	Consulte las métricas `RunningMapTasks` , `RemainingMapTasks` , `RunningReduceTasks` y `RemainingReduceTasks` .
Detectar clústeres que están inactivos	La métrica `IsIdle` realiza un seguimiento de si el clúster está disponible, pero actualmente no está ejecutando ninguna tarea. Puede configurar una alarma para que se active cuando el clúster haya estado inactivo durante un periodo de tiempo determinado, como, por ejemplo, treinta minutos.
Detectar si un nodo se queda sin espacio de almacenamiento	La métrica `MRUnhealthyNodes` registra cuándo uno o más nodos principales o de tarea se quedan sin almacenamiento en disco local y pasan a un estado `UNHEALTHY` de YARN. Por ejemplo, los nodos principales o de tarea se están quedando sin espacio en disco y no podrán ejecutar tareas.
Detectar si un clúster se queda sin espacio de almacenamiento	La métrica `HDFSUtilization` monitorea la capacidad de HDFS combinada del clúster y puede requerir el redimensionamiento del clúster para agregar más nodos principales. Por ejemplo, el uso de HDFS es elevado, lo que podría afectar a los trabajos y al estado del clúster.
Detectar cuándo un clúster se está ejecutando a una capacidad reducida	La métrica `MRLostNodes` registra cuándo uno o más nodos principales o de tarea no pueden comunicarse con el nodo maestro. Por

¿Cómo?	Métricas relevantes
	ejemplo, el nodo maestro no puede acceder al nodo principal o de tarea.

Para obtener más información, consulte El clúster termina con NO_SLAVE_LEFT y los nodos principales con FAILED_BY_MASTER y AWSSupport-AnalyzeEMRLogs.

Acceder a las métricas de CloudWatch para Amazon EMR

Puede ver las métricas que Amazon EMR notifica a CloudWatch mediante la consola de Amazon EMR o la consola de CloudWatch. También puede recuperar métricas mediante el comando `mon-get-stats` de la CLI de CloudWatch o la API `GetMetricStatistics` de CloudWatch. Para obtener más información sobre cómo visualizar y obtener las métricas de Amazon EMR con CloudWatch, consulte la Guía del usuario de Amazon CloudWatch.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para ver métricas con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, seleccione Clústeres y, a continuación, seleccione el clúster del que desee ver las métricas. Se abrirá la página de detalles del clúster.
3. Seleccione la pestaña Monitorización en la página de detalles del clúster. Seleccione cualquiera de las opciones denominadas Estado del clúster, Estado del nodo o Entradas y salidas para cargar los informes acerca del progreso y el estado del clúster.
4. Después de seleccionar una métrica para verla, puede modificar el tamaño de cada gráfico. Para filtrar el periodo de tiempo del gráfico, seleccione una opción preconfigurada o seleccione Personalizado.

Old console

Para ver métricas con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/.
2. Para ver las métricas de un clúster, seleccione un clúster para mostrar el panel Summary (Resumen).
3. Elija Monitoring (Monitorización) para ver información sobre dicho clúster. Seleccione cualquiera de las pestañas denominadas Estado del clúster, Asignar/Reducir, Estado del nodo o E/S para cargar los informes acerca del progreso y el estado del clúster.
4. Después de elegir una métrica que ver, puede seleccionar un tamaño de gráfico. Edite los campos Start (Inicio) y End (Finalización) para filtrar las métricas según un marco temporal específico.

Métricas notificadas por Amazon EMR en CloudWatch

En la siguiente tabla se muestran todas las métricas que Amazon EMR notifica en la consola y envía a CloudWatch.

Métricas de Amazon EMR

Amazon EMR envía los datos de varias métricas a CloudWatch. Todos los clústeres de Amazon EMR envían automáticamente métricas en intervalos de cinco minutos. Las métricas se archivan durante dos semanas; después de ese periodo, los datos se descartan.

El espacio de nombres de `AWS/ElasticMapReduce` incluye las siguientes métricas.

> **Note**
>
> Amazon EMR extrae métricas de un clúster. Si un clúster deja de estar disponible, no se registra ninguna métrica hasta que el clúster vuelve a estar disponible.

Están disponibles las siguientes métricas para los clústeres que ejecutan las versiones 2.x de Hadoop.

Métrica	Descripción
Estado del clúster	
IsIdle	Indica que un clúster ya no está funcionando, pero sigue activo y acumulando cargos. Se establece en 1 si no se ejecuta ninguna tarea ni ningún trabajo; en caso contrario , se establece en 0. Este valor se comprueba a intervalo s de cinco minutos, y un valor de 1 indica que el clúster estaba inactivo cuando se comprobó, no que estuvo inactivo durante los cinco minutos. Para evitar falsos positivos, debe activar una alarma cuando este valor sea 1 durante más de una comprobación consecutiva de cinco minutos. Por ejemplo, puede activar una alarma cuando este valor sea 1 durante treinta minutos o más. Caso de uso: monitorizar el rendimiento del clúster Unidades: booleano
ContainerAllocated	El número de contenedores de recursos asignados por ResourceManager. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
ContainerReserved	El número de contenedores reservados. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
ContainerPending	El número de contenedores en la cola que aún no se han asignado. Caso de uso: monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
ContainerPendingRatio	La proporción entre contenedores pendientes y contenedo res asignados (ContainerPendingRatio = Container Pending / ContainerAllocated). Si ContainerAllocated = 0, entonces ContainerPendingRatio = ContainerPending. El valor de ContainerPendingRatio representa un número, no un porcentaje. Este valor es útil para escalar recursos del clúster en función del comportamiento de asignación de contenedores. Unidades: recuento
AppsCompleted	El número de aplicaciones enviadas a YARN que se han completado. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
AppsFailed	El número de aplicaciones enviadas a YARN que no se han podido completar. Caso de uso: monitorizar el progreso del clúster, monitorizar el estado del clúster Unidades: recuento
AppsKilled	El número de aplicaciones enviadas a YARN que se han cancelado. Caso de uso: monitorizar el progreso del clúster, monitorizar el estado del clúster Unidades: recuento

Métrica	Descripción
AppsPending	El número de aplicaciones enviadas a YARN que están en estado pendiente. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
AppsRunning	El número de aplicaciones enviadas a YARN que se están ejecutando. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
AppsSubmitted	El número de aplicaciones enviadas a YARN. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
Estado del nodo	
CoreNodesRunning	El número de nodos secundarios en funcionamiento. Los puntos de datos de esta métrica solo se registran cuando existe un grupo de instancias correspondiente. Caso de uso: monitorizar el estado del clúster Unidades: recuento

Métrica	Descripción
CoreNodesPending	El número de nodos secundarios en espera de ser asignados. Puede que no todos los nodos secundarios solicitados estén disponibles inmediatamente; esta métrica registra las solicitudes pendientes. Los puntos de datos de esta métrica solo se registran cuando existe un grupo de instancias correspondiente. Caso de uso: monitorizar el estado del clúster Unidades: recuento
LiveDataNodes	El porcentaje de nodos de datos que reciben trabajo de Hadoop. Caso de uso: monitorizar el estado del clúster Unidades: porcentaje
MRTotalNodes	El número de nodos actualmente disponibles para trabajos de MapReduce. Equivalente a la métrica YARN `mapred.resourcemanager.TotalNodes` . Caso de uso: monitorizar el progreso del clúster Unidades: recuento
MRActiveNodes	El número de nodos que se están ejecutando actualmente en tareas o trabajos de MapReduce. Equivalente a la métrica YARN `mapred.resourcemanager.NoOfActiveNodes` . Caso de uso: monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
MRLostNodes	El número de nodos asignados a MapReduce que se han marcado con el estado LOST. Equivalente a la métrica YARN `mapred.resourcemanager.NoOfLostNodes` . Caso de uso: monitorizar el estado del clúster, monitorizar el progreso del clúster Unidades: recuento
MRUnhealthyNodes	El número de nodos disponibles para trabajos de MapReduce marcados con el estado UNHEALTHY. Equivalente a la métrica YARN `mapred.resourceman ager.NoOfUnhealthyNodes` . Caso de uso: monitorizar el progreso del clúster Unidades: recuento
MRDecommissionedNodes	El número de nodos asignados a aplicaciones de MapReduce que se han marcado con el estado DECOMMISSIONED. Equivalente a la métrica YARN `mapred.resourcemanager.NoOfDecommiss ionedNodes` . Caso de uso: monitorizar el estado del clúster, monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
MRRebootedNodes	El número de nodos disponibles parea MapReduce que se han reiniciado y marcado con el estado REBOOTED. Equivalente a la métrica YARN `mapred.resourceman ager.NoOfRebootedNodes` . Caso de uso: monitorizar el estado del clúster, monitorizar el progreso del clúster Unidades: recuento
MultiMasterInstanceGroupNod esRunning	El número de nodos principales en ejecución. Caso de uso: monitorizar errores y sustituciones del nodo principal Unidades: recuento
MultiMasterInstanceGroupNod esRunningPercentage	El porcentaje de nodos principales que se están ejecutand o por encima del recuento de instancias del nodo principal solicitadas. Caso de uso: monitorizar errores y sustituciones del nodo principal Unidades: porcentaje
MultiMasterInstanceGroupNod esRequested	El número de nodos principales solicitados. Caso de uso: monitorizar errores y sustituciones del nodo principal Unidades: recuento
E/S	

Métrica	Descripción
S3BytesWritten	Número de bytes escritos en Amazon S3. Esta métrica solo agrega trabajos de MapReduce y no se aplica a otras cargas de trabajo de Amazon EMR. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento
S3BytesRead	Número de bytes leídos desde Amazon S3. Esta métrica solo agrega trabajos de MapReduce y no se aplica a otras cargas de trabajo de Amazon EMR. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento
HDFSUtilization	El porcentaje de almacenamiento HDFS usado actualmente. Caso de uso: analizar el rendimiento del clúster Unidades: porcentaje
HDFSBytesRead	El número de bytes leídos de HDFS. Esta métrica solo agrega trabajos de MapReduce y no se aplica a otras cargas de trabajo de Amazon EMR. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
HDFSBytesWritten	El número de bytes escritos en HDFS. Esta métrica solo agrega trabajos de MapReduce y no se aplica a otras cargas de trabajo de Amazon EMR. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento
MissingBlocks	El número de bloques en los que HDFS no tiene réplicas. Pueden tratarse de bloques dañados. Caso de uso: monitorizar el estado del clúster Unidades: recuento
CorruptBlocks	El número de bloques que HDFS registra como dañados. Caso de uso: monitorizar el estado del clúster Unidades: recuento
TotalLoad	El número total de transferencias de datos simultáneas, Caso de uso: monitorizar el estado del clúster Unidades: recuento
MemoryTotalMB	La cantidad de memoria total del clúster. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
MemoryReservedMB	La cantidad de memoria reservada. Caso de uso: monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
MemoryAvailableMB	La cantidad de memoria disponible para asignar. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
YARNMemoryAvailablePercentage	El porcentaje de memoria restante disponible para YARN (YARNMemoryAvailablePercentage = MemoryAvailableMB / MemoryTotalMB). Este valor es útil para escalar recursos del clúster en función del uso de memoria de YARN. Unidades: porcentaje
MemoryAllocatedMB	La cantidad de memoria asignada al clúster. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
PendingDeletionBlocks	El número de bloques marcados para eliminación. Caso de uso: monitorizar el progreso del clúster, monitorizar el estado del clúster Unidades: recuento
UnderReplicatedBlocks	El número de bloques que necesitan replicarse una o varias veces. Caso de uso: monitorizar el progreso del clúster, monitorizar el estado del clúster Unidades: recuento

Métrica	Descripción
DfsPendingReplicationBlocks	El estado de replicación del bloque: bloques que se están replicando, antigüedad de las solicitudes de replicación y solicitudes replicadas correctamente. Caso de uso: monitorizar el progreso del clúster, monitorizar el estado del clúster Unidades: recuento
CapacityRemainingGB	La cantidad de capacidad de disco HDFS restante. Caso de uso: monitorizar el progreso del clúster, monitorizar el estado del clúster Unidades: recuento

A continuación se indican las métricas de Hadoop 1:

Métrica	Descripción
Estado del clúster	
IsIdle	Indica que un clúster ya no está funcionando, pero sigue activo y acumulando cargos. Se establece en 1 si no se ejecuta ninguna tarea ni ningún trabajo; en caso contrario , se establece en 0. Este valor se comprueba a intervalo s de cinco minutos, y un valor de 1 indica que el clúster estaba inactivo cuando se comprobó, no que estuvo inactivo durante los cinco minutos. Para evitar falsos positivos, debe activar una alarma cuando este valor sea 1 durante más de una comprobación consecutiva de cinco minutos. Por ejemplo, puede activar una alarma cuando este valor sea 1 durante treinta minutos o más. Caso de uso: monitorizar el rendimiento del clúster

Métrica	Descripción
	Unidades: booleano
JobsRunning	El número de trabajos del clúster que se encuentran actualmente en ejecución. Caso de uso: monitorizar el estado del clúster Unidades: recuento
JobsFailed	El número de trabajos del clúster que han producido un error. Caso de uso: monitorizar el estado del clúster Unidades: recuento
Map/Reduce	
MapTasksRunning	Número de tareas de asignación en ejecución de cada trabajo. Si tiene un programador instalado y varios trabajos en ejecución, se generan varios gráficos. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
MapTasksRemaining	Número de tareas de asignación pendientes de cada trabajo. Si tiene un programador instalado y varios trabajos en ejecución, se generan varios gráficos. Una tarea de asignación pendiente es aquella que no tiene ninguno de los siguientes estados: Running, Killed o Completed. Caso de uso: monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
MapSlotsOpen	La capacidad de la tarea de asignación no utilizada. Se calcula como el número máximo de tareas de asignación de un clúster determinado menos el número total de tareas de asignación que se están ejecutando actualmente en el clúster. Caso de uso: analizar el rendimiento del clúster Unidades: recuento
RemainingMapTasksPerSlot	La proporción entre el total de tareas de asignación pendientes y el total de slots de asignación disponibles en el clúster. Caso de uso: analizar el rendimiento del clúster Unidades: proporción
ReduceTasksRunning	Número de tareas de reducción en ejecución de cada trabajo. Si tiene un programador instalado y varios trabajos en ejecución, se generan varios gráficos. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
ReduceTasksRemaining	Número de tareas de reducción pendientes de cada trabajo. Si tiene un programador instalado y varios trabajos en ejecución, se generan varios gráficos. Caso de uso: monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
ReduceSlotsOpen	La capacidad de la tarea de reducción no utilizada. Se calcula como la capacidad máxima de la tarea de reducción de un clúster determinado menos el número de tareas de reducción que se están ejecutando actualmente en el clúster. Caso de uso: analizar el rendimiento del clúster Unidades: recuento
Estado del nodo	
CoreNodesRunning	El número de nodos secundarios en funcionamiento. Los puntos de datos de esta métrica solo se registran cuando existe un grupo de instancias correspondiente. Caso de uso: monitorizar el estado del clúster Unidades: recuento
CoreNodesPending	El número de nodos secundarios en espera de ser asignados. Puede que no todos los nodos secundarios solicitados estén disponibles inmediatamente; esta métrica registra las solicitudes pendientes. Los puntos de datos de esta métrica solo se registran cuando existe un grupo de instancias correspondiente. Caso de uso: monitorizar el estado del clúster Unidades: recuento
LiveDataNodes	El porcentaje de nodos de datos que reciben trabajo de Hadoop. Caso de uso: monitorizar el estado del clúster Unidades: porcentaje

Métrica	Descripción
TaskNodesRunning	El número de nodos de tareas en funcionamiento. Los puntos de datos de esta métrica solo se registran cuando existe un grupo de instancias correspondiente. Caso de uso: monitorizar el estado del clúster Unidades: recuento
TaskNodesPending	El número de nodos de tareas en espera de ser asignados . Puede que no todos los nodos de tareas solicitados estén disponibles inmediatamente; esta métrica registra las solicitudes pendientes. Los puntos de datos de esta métrica solo se registran cuando existe un grupo de instancias correspondiente. Caso de uso: monitorizar el estado del clúster Unidades: recuento
LiveTaskTrackers	El porcentaje de rastreadores de tareas que están operativo s. Caso de uso: monitorizar el estado del clúster Unidades: porcentaje
E/S	
S3BytesWritten	Número de bytes escritos en Amazon S3. Esta métrica solo agrega trabajos de MapReduce y no se aplica a otras cargas de trabajo de Amazon EMR. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
S3BytesRead	Número de bytes leídos desde Amazon S3. Esta métrica solo agrega trabajos de MapReduce y no se aplica a otras cargas de trabajo de Amazon EMR. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento
HDFSUtilization	El porcentaje de almacenamiento HDFS usado actualmente. Caso de uso: analizar el rendimiento del clúster Unidades: porcentaje
HDFSBytesRead	El número de bytes leídos de HDFS. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento
HDFSBytesWritten	El número de bytes escritos en HDFS. Caso de uso: analizar el rendimiento del clúster, monitorizar el progreso del clúster Unidades: recuento
MissingBlocks	El número de bloques en los que HDFS no tiene réplicas. Pueden tratarse de bloques dañados. Caso de uso: monitorizar el estado del clúster Unidades: recuento

Métrica	Descripción
TotalLoad	Número total actual de lectores y escritores registrados por todos los DataNodes de un clúster. Caso de uso: Diagnosticar en qué medida un nivel elevado de E/S podría estar contribuyendo a reducir el rendimien to en la ejecución de trabajos. Los nodos de procesos de trabajo que ejecutan el daemon DataNode también deben llevar a cabo tareas de mapeo y reducción. Valores sistemáticamente altos de TotalLoad durante un tiempo prolongado pueden indicar que un nivel elevado de E/ S podría ser uno de los factores que está empeorando el rendimiento. Los picos ocasionales en este valor son normales y no necesariamente son indicativos de un problema. Unidades: recuento

Métricas de capacidad del clúster

Las siguientes métricas indican la capacidad actual o prevista de un clúster. Estas métricas solo están disponibles cuando el escalado administrado o la terminación automática están habilitados.

Para los clústeres compuestos por flotas de instancias, las métricas de capacidad del clúster se miden en `Units`. Para los clústeres compuestos por grupos de instancias, las métricas de capacidad del clúster se miden en `Nodes` o en `VCPU` en función del tipo de unidad utilizado en la política de escalado administrado. Para obtener más información, consulte Uso del escalado administrado de EMR en la Guía de administración de Amazon EMR.

Métrica	Descripción
• `TotalUnitsRequested` • `TotalNodesRequested` • `TotalVCPURequested`	El número total previsto de unidades, nodos o vCPU en un clúster según lo determine el escalado administrado. Unidades: recuento

Métrica	Descripción
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	El número total actual de unidades, nodos o vCPU disponibles en un clúster en ejecución. Cuando se solicita un cambio de tamaño del clúster, esta métrica se actualizará después de agregar o quitar las nuevas instancias del clúster. Unidades: recuento
• `CoreUnitsRequested` • `CoreNodesRequested` • `CoreVCPURequested`	El número previsto de unidades, nodos o vCPU CORE en un clúster según lo determine el escalado administrado. Unidades: recuento
• `CoreUnitsRunning` • `CoreNodesRunning` • `CoreVCPURunning`	El número actual de unidades, nodos o vCPU CORE que se ejecutan en un clúster. Unidades: recuento
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	El número previsto de unidades, nodos o vCPU TASK en un clúster según lo determine el escalado administrado. Unidades: recuento

Métrica	Descripción
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	El número actual de unidades, nodos o vCPU TASK que se ejecutan en un clúster. Unidades: recuento

Amazon EMR emite las siguientes métricas con una granularidad de un minuto cuando se habilita la terminación automática mediante una política de terminación automática. Algunas métricas solo están disponibles para las versiones 6.4.0 y posteriores de Amazon EMR. Para obtener más información acerca de la terminación automática, consulte Uso de una política de terminación automática.

Métrica	Descripción
`TotalNotebookKernels`	El número total de kernels de cuadernos en ejecución e inactivos del clúster. Esta métrica solo está disponible para las versiones 6.4.0 y posteriores de Amazon EMR.
`AutoTerminationIsClusterIdle`	Indica si el clúster está en uso. Un valor de 0 indica que el clúster está siendo utilizado activamente por uno de los siguientes componentes: • Una aplicación YARN • HDFS • Un cuaderno •

Métrica	Descripción
	Una interfaz de usuario integrada en el clúster, como el servidor del historial de Spark Un valor de 1 indica que el clúster está inactivo. Amazon EMR comprueba la inactividad continua de los clústeres (`AutoTerminationIsClusterIdle` = 1). Cuando el tiempo de inactividad de un clúster es igual al valor de `IdleTimeout` de su política de terminación automática, Amazon EMR termina el clúster.

Dimensiones para las métricas de Amazon EMR

Los datos de Amazon EMR se pueden filtrar mediante alguna de las dimensiones de la tabla siguiente.

Dimensión	Descripción
JobFlowId	El mismo que el ID del clúster, que es un identificador único de un clúster con el formato `j-XXXXXXXXXXXXX` . Puede encontrar este valor haciendo clic en el clúster en la consola de Amazon EMR.

Monitorización de eventos de Amazon EMR con CloudWatch

Amazon EMR realiza un seguimiento de eventos y mantiene información acerca de ellos durante un máximo de siete días en la consola de Amazon EMR. Amazon EMR registra los eventos cuando se produce un cambio en el estado de los clústeres, los grupos de instancias, las flotas de instancias, las políticas de escalado automático o los pasos. Los eventos capturan la fecha y la hora en que se produjo el evento, los detalles sobre los elementos afectados y otros puntos de datos críticos.

En la siguiente tabla se muestran eventos de Amazon EMR, junto con el estado o cambio de estado que indica el evento, la gravedad del evento, el tipo del evento, el código del evento y los mensajes de eventos. Amazon EMR representa los eventos como objetos JSON y los envía automáticamente a una secuencia de eventos. El objeto JSON es importante cuando se configuran reglas para el procesamiento de eventos mediante Eventos de CloudWatch porque las reglas buscan asignar patrones en el objeto JSON. Para obtener más información, consulte Eventos y patrones de eventos y Eventos de Amazon EMR en la Guía del usuario de Eventos de Amazon CloudWatch.

> **Note**
>
> A fin de asegurarnos de que reciba la información más pertinente, mejoramos continuamente nuestros mensajes de error. Por ese motivo, le recomendamos que no analice el texto de los mensajes para iniciar las siguientes acciones en su flujo de trabajo.

Eventos de inicio de clústeres

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
`CREATING`	`WARN`	Aprovisionamiento de la flota de instancias de Amazon EMR	Aprovisionamiento de EC2: capacidad de instancia insuficiente	No se puede crear su clúster de Amazon EMR `ClusterId (ClusterName)` para la flota de instancias `InstanceFleetID` . Amazon EC2 tiene una capacidad de spot insuficiente para el tipo de instancia `[Instance type1,`

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
				`Instancetype2]` y una capacidad bajo demanda insuficiente para el tipo de instancia `[Instancetype3, Instancetype4]` en la zona de disponibilidad `[AvailabilityZone1, Avaliabilityzone2]`. Consulte aquí la documentación para obtener más información sobre cómo responder a este evento.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
`CREATING`	`WARN`	Aprovisionamiento del grupo de instancias de Amazon EMR	Aprovisionamiento de EC2: capacidad de instancia insuficiente	No se puede crear su clúster de Amazon EMR `ClusterId (ClusterName)` para el grupo de instancias `InstancegroupID`. Amazon EC2 tiene una capacidad de `[Spot or On-Demand]` insuficiente para el tipo de instancia `Instancetype` en la zona de disponibilidad `AvailabilityZone`. Consulte aquí la documentación para obtener más información sobre cómo responder a este evento.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
`STARTING`	`INFO`	Cambio de estado del clúster de EMR	ninguno	El clúster de Amazon EMR `ClusterId (ClusterName)` se solicitó a las `Time` y se está creando.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
`STARTING`	`INFO`	Cambio de estado del clúster de EMR	ninguno	**Note** Sólo se aplica a clústeres con configuración de flotas de instancia y varias zonas de disponibilidad seleccionadas dentro de Amazon EC2. El clúster de Amazon EMR `ClusterId (ClusterName)` se está creando en la zona (`AvailabilityZoneID`), que se ha elegido a partir de las opciones

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
				de zona de disponibilidad especificadas.
`STARTING`	`INFO`	Cambio de estado del clúster de EMR	ninguno	El clúster de Amazon EMR `ClusterId (ClusterName)` comenzó a ejecutar los pasos a las `Time`.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
`WAITING`	`INFO`	Cambio de estado del clúster de EMR	ninguno	El clúster de Amazon EMR `ClusterId (ClusterName)` se creó a las `Time` y está listo para su uso. - o bien - El clúster de Amazon EMR `ClusterId (ClusterName)` terminó de ejecutar todos los pasos pendientes a las `Time`. **Note** Un clúster en estado `WAITING` todavía podría estar procesando trabajos.

> **Note**
>
> Los eventos con código de evento `EC2 provisioning - Insufficient Instance Capacity`, se emiten periódicamente cuando su clúster de EMR detecta un error de capacidad insuficiente de Amazon EC2 para su flota de instancias o grupo de instancias durante la creación del clúster o la operación de cambio de tamaño. Para obtener información sobre cómo responder a estos eventos, consulte Responder a eventos de capacidad de instancias insuficiente en el clúster de Amazon EMR.

Eventos de terminación de clústeres

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
`TERMINATED`	La gravedad depende del motivo por el que cambia el estado, tal y como se muestra a continuación: • **`CRITICAL`** si el clúster terminó con cualquiera de los siguientes motivos de cambio de estado: `INTERNAL_ERROR` , `VALIDATION_ERROR` , `INSTANCE_`	Cambio de estado del clúster de EMR	ninguno	El clúster de Amazon EMR `ClusterId (ClusterName)` terminó a las `Time` con el siguiente motivo: `StateChangeReason: Code` .

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
	`FAILURE` , `BOOTSTRAP_FAILURE` o `STEP_FAILURE` . • **`INFO`** si el clúster terminó con cualquiera de los siguientes motivos de cambio de estado: `USER_REQUEST` o `ALL_STEPS_COMPLETED` .			
`TERMINATED_WITH_ERRORS`	`CRITICAL`	Cambio de estado del clúster de EMR	ninguno	El clúster de Amazon EMR `ClusterId (ClusterName)` terminó con errores a las `Time` con el siguiente motivo: `StateChangeReason: Code` .

Eventos de cambio de estado de la flota de instancias

> **Note**
>
> La configuración de las flotas de instancias está disponible solo en las versiones 4.8.0 y posteriores de Amazon EMR, excluidas las versiones 5.0.0 y 5.0.3.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
Desde `PROVISIONING` hasta `WAITING`	`INFO`		ninguno	El aprovisionamiento para la flota de instancias `InstanceFleetID` del clúster de Amazon EMR `ClusterId (ClusterName)` está completo. El aprovisionamiento se inició a las `Time` y tardó `Num` minutos. La flota de instancias ahora tiene una capacidad bajo demanda de `Num` y una capacidad de spot de `Num`. La capacidad bajo demanda de destino era `Num`,

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
				y la capacidad de spot de destino era Num.
Desde WAITING hasta RESIZING	INFO		ninguno	Se inició un cambio de tamaño de la flota de instancias InstanceFleetID del clúster de Amazon EMR ClusterId (ClusterName) se inició a las Time. La flota de instancias está cambiando de tamaño desde una capacidad bajo demanda de Num a un destino de Num, y de la capacidad de spot de Num a un destino de Num.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
Desde `RESIZING` hasta `WAITING`	`INFO`		ninguno	La operación de cambio de tamaño de la flota de instancias `InstanceFleetID` del clúster de Amazon EMR `ClusterId (ClusterName)` se ha completado. El cambio de tamaño se inició a las `Time` y tardó `Num` minutos. La flota de instancias ahora tiene una capacidad bajo demanda de `Num` y una capacidad de spot de `Num`. La capacidad bajo demanda de destino era `Num` y la capacidad de spot de destino era `Num`.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
Desde `RESIZING` hasta `WAITING`	`INFO`		ninguno	La operación de cambio de tamaño de la flota de instancias `InstanceFleetID` del clúster de Amazon EMR `ClusterId` (`ClusterName`) ha alcanzado el tiempo de espera y se ha parado. El cambio de tamaño se inició a las `Time` y se paró después de `Num` minutos. La flota de instancias ahora tiene una capacidad bajo demanda de `Num` y una capacidad de spot de `Num`. La capacidad bajo demanda de destino era `Num` y la capacidad de spot de destino era `Num`.

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
SUSPENDED	ERROR		ninguno	La flota de instancias InstanceFleetID del clúster de Amazon EMR ClusterId (ClusterName) se detuvo a las Time por el siguiente motivo: ReasonDesc .
RESIZING	WARNING		ninguno	La operación de cambio de tamaño de la flota de instancias InstanceFleetID del clúster de Amazon EMR ClusterId (ClusterName) está bloqueada por el siguiente motivo: ReasonDesc .

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
`WAITING` o `Running`	`INFO`		ninguno	La operación de redimensionamiento de la flota de instancias `InstanceFleetID` en el clúster de Amazon EMR `ClusterId (ClusterName)` no se pudo completar mientras Amazon EMR agregaba capacidad de spot en la zona de disponibilidad `AvailabilityZone`. Hemos cancelado su solicitud de aprovisionamiento de capacidad de spot adicional. Para ver las acciones recomendadas, consulte Prácticas recomenda

Estado o cambio de estado	Gravedad	Tipo de evento	Código de evento	Mensaje
				das para la flexibilidad de las instancias y las zonas de disponibilidad e inténtelo de nuevo.
`WAITING` o `Running`	`INFO`		ninguno	`Entity` inició una operación de cambio de tamaño de la flota de instancias `InstanceFleetID` del clúster de Amazon EMR `ClusterId (ClusterName)` a las `Time`.

Eventos de cambio de tamaño de la flota de instancias

Tipo de evento	Gravedad	Código de evento	Mensaje
Cambio de tamaño de la flota de instancias de Amazon EMR	`ERROR`	Tiempo de espera de aprovisionamiento de spot	La operación de cambio de tamaño de la flota de instancias `InstanceFleetID` del clúster de Amazon EMR `ClusterId`

Tipo de evento	Gravedad	Código de evento	Mensaje
			`(Cluster Name)` no se pudo completar al adquirir capacidad de spot en la zona de disponibilidad `AvailabilityZone` . Hemos cancelado su solicitud y hemos dejado de intentar aprovisionar capacidad de spot adicional. La flota de instancias tiene una capacidad de spot aprovisionada de `num`. La capacidad de spot de destino era de `num`. Para obtener más información y ver las acciones recomendadas, consulte la página de documentación aquí e inténtelo de nuevo.

Tipo de evento	Gravedad	Código de evento	Mensaje
Cambio de tamaño de la flota de instancias de Amazon EMR	ERROR	Tiempo de espera del aprovisionamiento bajo demanda	La operación de cambio de tamaño de la flota de instancia s `InstanceF leetID` del clúster de Amazon EMR `ClusterId (ClusterName)` no se pudo completar al adquirir capacidad bajo demanda en la zona de disponibi lidad `Availabil ityZone` . Hemos cancelado su solicitud y hemos dejado de intentar aprovisio nar capacidad bajo demanda adicional. La flota de instancias tiene una capacidad bajo demanda aprovisionada de `num`. La capacidad bajo demanda de destino era de `num`. Para obtener más información y ver las acciones recomenda das, consulte la página de documenta ción aquí e inténtelo de nuevo.

Tipo de evento	Gravedad	Código de evento	Mensaje
Cambio de tamaño de la flota de instancias de Amazon EMR	`WARNING`	Aprovisionamiento de EC2: capacidad de instancia insuficiente	No se puede completar la operación de cambio de tamaño de la flota de instancia s `InstanceF leetID` del clúster de EMR `ClusterId (Cluster Name)` porque Amazon EC2 tiene una capacidad de spot insuficiente para los tipos de instancias `[Instancetype1, Instancetype2]` y una capacidad bajo demanda insuficiente para los tipos de instancias `[Instancetype3, Instancet ype4]` en la zona de disponibi lidad `[Availabi lityZone1]` . Hasta ahora, la flota de instancias tiene una capacidad bajo demanda aprovisio nada de `num` y la capacidad bajo demanda de destino era de `num`. La

Tipo de evento	Gravedad	Código de evento	Mensaje
			capacidad de spot aprovisionada es de num y la capacidad de spot de destino es de num. Consulte aquí la documentación para obtener más información sobre cómo responder a este evento.

Tipo de evento	Gravedad	Código de evento	Mensaje
Cambio de tamaño de la flota de instancias de Amazon EMR	`WARNING`	Tiempo de espera de aprovisionamiento de spot: continuación del cambio de tamaño	Seguimos aprovisionando capacidad de spot para la operación de cambio de tamaño de la flota de instancias que se inició a las `time` en la flota de instancias con el ID `InstanceFleetID` del clúster de Amazon EMR `ClusterId (ClusterName)` para `[Instancetype1, Instancetype2]` en la zona de disponibilidad `AvailabilityZone`. En la operación de redimensionamiento anterior que se inició a las `time`, se agotó el tiempo de espera, por lo que Amazon EMR dejó de aprovisionar la capacidad de spot tras agregar `num` de las `num` instancias solicitadas a su flota de instancias. Para obtener más información, consulte la página

Tipo de evento	Gravedad	Código de evento	Mensaje
			de documentación aquí.

Tipo de evento	Gravedad	Código de evento	Mensaje
Cambio de tamaño de la flota de instancias de Amazon EMR	`WARNING`	Tiempo de espera de aprovisionamiento bajo demanda: continuación del cambio de tamaño	Seguimos aprovisionando capacidad bajo demanda para la operación de cambio de tamaño de la flota de instancias que se inició a las `time` en la flota de instancias con el ID `InstanceFleetID` del clúster de Amazon EMR `ClusterId (ClusterName)` para `[Instancetype1, Instancetype2]` en la zona de disponibilidad `AvailabilityZone`. En la operación de redimensionamiento anterior que se inició a las `time`, se agotó el tiempo de espera, por lo que Amazon EMR dejó de aprovisionar la capacidad bajo demanda tras agregar `num` de las `num` instancias solicitadas a su flota de instancias. Para obtener más informaci

Tipo de evento	Gravedad	Código de evento	Mensaje
			ón, consulte la página de documentación aquí.

> **Note**
>
> Los eventos de tiempo de espera de aprovisionamiento se emiten cuando Amazon EMR deja de aprovisionar capacidad de spot o bajo demanda para la flota una vez transcurrido el tiempo de espera. Para obtener información sobre cómo responder a estos eventos, consulte Respuesta a eventos de tiempo de espera agotado en el cambio de tamaño de la flota de instancias de clústeres de Amazon EMR.

Eventos de grupo de instancias

Tipo de evento	Gravedad	Código de evento	Mensaje
Desde `RESIZING` hasta `Running`	`INFO`	ninguno	La operación de cambio de tamaño del grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` se ha completado. Ahora tiene un recuento de instancias de `Num`. El cambio de tamaño se inició a las `Time` y tardó `Num` minutos en completarse.
Desde `RUNNING` hasta `RESIZING`	`INFO`	ninguno	Se inició un cambio de tamaño del

Tipo de evento	Gravedad	Código de evento	Mensaje
			grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` a las `Time`. Está cambiando de tamaño de un recuento de instancias de `Num` a `Num`.
`SUSPENDED`	`ERROR`	ninguno	El grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` se detuvo a las `Time` por el siguiente motivo: `ReasonDesc` .
`RESIZING`	`WARNING`	ninguno	La operación de cambio de tamaño del grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` está bloqueada por el siguiente motivo: `ReasonDesc` .

Tipo de evento	Gravedad	Código de evento	Mensaje
Cambio de tamaño de un grupo de instancias de Amazon EMR	`WARNING`	Aprovisionamiento de EC2: capacidad de instancia insuficiente	No se puede completar la operación de cambio de tamaño que se inició a las `time` en el grupo de instancias `InstanceGroupID` del clúster de EMR `ClusterId (ClusterName)`, ya que Amazon EC2 no tiene suficiente capacidad `Spot/On Demand` para el tipo de instancia `[Instance type]` en la zona de disponibilidad `[AvailabilityZone1]`. Hasta ahora, el grupo de instancias tiene un recuento de instancias de `num` y el número de instancias solicitadas era de `num`. Consulte aquí la documentación para obtener más información sobre cómo responder a este evento.

Tipo de evento	Gravedad	Código de evento	Mensaje
Desde `RUNNING` hasta `RESIZING`	`INFO`	ninguno	`Entity` inició un cambio de tamaño del grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` a las `Time`.

> **Note**
>
> Con la versión 5.21.0 y posteriores de Amazon EMR, puede anular las configuraciones de clúster y especificar las clasificaciones de configuración adicionales para cada grupo de instancias en un clúster en ejecución. Para ello, utilice la consola de Amazon EMR, la AWS Command Line Interface (AWS CLI) o el AWS SDK. Para obtener más información, consulte Suministrar una configuración para un grupo de instancias en un clúster en ejecución.

En la siguiente tabla se muestran eventos de Amazon EMR, junto con el estado o cambio de estado que indica el evento, la gravedad del evento y los mensajes de eventos.

Estado o cambio de estado	Gravedad	Mensaje
`RUNNING`	`INFO`	Un usuario inició un cambio de configuración del grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` a las `Time`. La versión de configuración solicitada es `Num`.

Estado o cambio de estado	Gravedad	Mensaje
Desde RECONFIGURING hasta Running	INFO	La operación de cambio de configuración del grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterName) se ha completado. El cambio de configuración se inició a las Time y tardó Num minutos en completarse. La versión de la configuración actual es Num.
Desde RUNNING hasta RECONFIGURING in	INFO	Se inició un cambio de configuración del grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterName) a las Time. Se configura de la versión número Num a la versión número Num.
RESIZING	INFO	La operación de cambio de configuración hacia la versión de configuración Num para el grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterName) se ha bloqueado temporalmente a las Time porque el grupo de instancias se encuentra en State.

Estado o cambio de estado	Gravedad	Mensaje
RECONFIGURING	INFO	La operación de cambio de tamaño hacia el recuento de instancias Num para el grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterName) se ha bloqueado temporalmente a las Time porque el grupo de instancias se encuentra en State.
RECONFIGURING	WARNING	Se produjo un error en la operación de cambio de configuración del grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterName) a las Time y tardó Num minutos en producirse el error. La versión de configuración de error es Num.
RECONFIGURING	INFO	Las configuraciones se revierten al número de versión correcta anterior Num para el grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterName) a las Time. La nueva versión de configuración es Num.

Estado o cambio de estado	Gravedad	Mensaje
Desde RECONFIGURING hasta Running	INFO	Las configuraciones se han revertido correctamente a la versión correcta anterior Num para el grupo de instancia s InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterN ame) a las Time. La nueva versión de configuración es Num.
Desde RECONFIGURING hasta SUSPENDED	CRITICAL	No se pudo revertir a la versión correcta anterior Num para el grupo de instancia s InstanceGroupID del clúster de Amazon EMR ClusterId (ClusterN ame) a las Time.

Eventos de la política de escalado automático

Estado o cambio de estado	Gravedad	Mensaje
PENDING	INFO	Se ha agregado una política de escalado automátic o al grupo de instancias InstanceGroupID en el clúster de Amazon EMR ClusterId (Cluster Name) a las Time. La política todavía no se ha asociado. - o bien -

Estado o cambio de estado	Gravedad	Mensaje
		Se ha actualizado la política de escalado automático del grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (Cluster Name)` a las `Time`. La política todavía no se ha asociado.
`ATTACHED`	`INFO`	Se ha adjuntado la política de escalado automático del grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` a las `Time`.
`DETACHED`	`INFO`	Se ha separado la política de escalado automático del grupo de instancias `InstanceGroupID` del clúster de Amazon EMR `ClusterId (ClusterName)` a las `Time`.

Estado o cambio de estado	Gravedad	Mensaje
FAILED	ERROR	No se pudo adjuntar la política de escalado automático del grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (Cluster Name) y se produjo un error a las Time. - o bien - No se pudo separar la política de escalado automático del grupo de instancias InstanceGroupID del clúster de Amazon EMR ClusterId (Cluster Name) y se produjo un error a las Time.

Eventos de paso

Estado o cambio de estado	Gravedad	Mensaje
PENDING	INFO	El paso StepID (StepName) se agregó al clúster de Amazon EMR ClusterId (ClusterName) a las Time y está pendiente de ejecución.
CANCEL_PENDING	WARN	El paso StepID (StepName) del clúster de Amazon EMR ClusterId (ClusterName) se canceló

Estado o cambio de estado	Gravedad	Mensaje
		a las Time y está pendiente de cancelación.
RUNNING	INFO	El paso StepID (StepName) del clúster de Amazon EMR ClusterId (ClusterName) comenzó a ejecutarse a las Time.
COMPLETED	INFO	El paso StepID (StepName) del clúster de Amazon EMR ClusterId (ClusterName) completó la ejecución a las Time. El paso empezó a ejecutarse a las Time y tardó Num minutos en completarse.
CANCELLED	WARN	La solicitud de cancelación se ha realizado correctam ente para el paso de clúster StepID (StepName) del clúster de Amazon EMR ClusterId (Cluster Name) a las Time, y el paso ya está cancelado.
FAILED	ERROR	Se produjo un error en el paso StepID (StepName) del clúster de Amazon EMR ClusterId (ClusterN ame) a las Time.

Visualización de eventos con la consola de Amazon EMR

Para cada clúster, puede ver una lista sencilla de eventos en el panel de detalles, que enumera los eventos en orden de aparición descendente. También puede ver todos los eventos de todos los clústeres en una región en orden de aparición descendente.

Si no desea que un usuario vea todos los eventos de clústeres de una región, añada una declaración que deniegue permiso (`"Effect": "Deny"`) para la acción `elasticmapreduce:ViewEventsFromAllClustersInConsole` a una política asociada al usuario.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para ver los eventos de todos los clústeres de una región con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 en el panel de navegación izquierdo, seleccione Eventos.

Para ver los eventos de un clúster determinado con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 en el panel de navegación izquierdo, seleccione Clústeres y, a continuación, seleccione un clúster.
3. Para ver todos los eventos, seleccione la pestaña Eventos en la página de detalles del clúster.

Old console

Para ver los eventos de todos los clústeres de una región con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/.
2. Elija Events (Eventos).

Para ver los eventos de un clúster determinado con la consola antigua

1. Abra la consola de Amazon EMR enhttps://console.aws.amazon.com/elasticmapreduce/.
2. Elija Cluster List (Lista de clústeres), seleccione un clúster y, a continuación, elija View details (Ver detalles).
3. Elija Events (Eventos) en el panel de detalles del clúster.

Responder a los eventos de CloudWatch

En esta sección, se describen varias formas en las que puede responder a los eventos procesables que Amazon EMR emite como mensajes de eventos de CloudWatch.

Temas

- Creación de reglas para los eventos de Amazon EMR con CloudWatch
- Configuración de alarmas en las métricas de CloudWatch
- Responder a eventos de capacidad de instancias insuficiente en el clúster de Amazon EMR
- Respuesta a eventos de tiempo de espera agotado en el cambio de tamaño de la flota de instancias de clústeres de Amazon EMR

Creación de reglas para los eventos de Amazon EMR con CloudWatch

Amazon EMR envía eventos automáticamente a una secuencia de eventos de CloudWatch. Puede crear reglas que asignen eventos de acuerdo con un patrón especificado y dirijan los eventos a destinos para tomar medidas, por ejemplo, enviar una notificación por correo electrónico. Los patrones se asocian al objeto JSON del evento. Para obtener más información sobre los detalles de los eventos de Amazon EMR, consulte Eventos de Amazon EMR en la Guía del usuario de Eventos de Amazon CloudWatch.

Para obtener información acerca de la configuración de reglas de eventos de CloudWatch, consulte Creación de una regla de CloudWatch que se activa en función de un evento.

Configuración de alarmas en las métricas de CloudWatch

Amazon EMR envía métricas a Amazon CloudWatch. En respuesta, puede usar CloudWatch para configurar alarmas en sus métricas de Amazon EMR. Por ejemplo, puede configurar una alarma en CloudWatch para enviar un correo electrónico siempre que el uso de HDFS supere el 80 %. Para obtener instrucciones detalladas, consulte Creación o edición de una alarma de CloudWatch en la Guía del usuario de Amazon CloudWatch.

Responder a eventos de capacidad de instancias insuficiente en el clúster de Amazon EMR

Información general

Los clústeres de Amazon EMR devuelven el código del evento `EC2 provisioning - Insufficient Instance Capacity` cuando la zona de disponibilidad seleccionada no tiene suficiente capacidad para cumplir con la solicitud de inicio o cambio de tamaño del clúster. El evento se emite periódicamente tanto con los grupos de instancias como con las flotas de instancias si Amazon EMR encuentra repetidamente excepciones de capacidad insuficiente y no puede cumplir con su solicitud de aprovisionamiento para una operación de inicio o cambio de tamaño del clúster.

En esta página, se describe la mejor manera de responder a este tipo de evento cuando se produce en su clúster de EMR.

Respuesta recomendada a un evento de capacidad insuficiente

Le recomendamos que responda a un evento de capacidad insuficiente de una de las siguientes maneras:

- Espere a que se recupere la capacidad. La capacidad cambia con frecuencia, por lo que una excepción de capacidad insuficiente puede recuperarse por sí sola. Sus clústeres se iniciarán o terminarán el cambio de tamaño en cuanto haya capacidad de Amazon EC2 disponible.
- Como alternativa, puede terminar el clúster, modificar las configuraciones del tipo de instancia y crear un nuevo clúster con la solicitud de configuración del clúster actualizada. Para obtener más información, consulte Prácticas recomendadas para la flexibilidad de las instancias y las zonas de disponibilidad.

También puede configurar reglas o respuestas automatizadas a un evento de capacidad insuficiente, como se describe en la siguiente sección.

Recuperación automática de un evento de capacidad insuficiente

Puede crear una automatización en respuesta a los eventos de Amazon EMR, como los que tienen el código de evento `EC2 provisioning - Insufficient Instance Capacity`. Por ejemplo, la siguiente función de AWS Lambda termina un clúster de EMR con un grupo de instancias que usa instancias bajo demanda y, a continuación, crea un nuevo clúster de EMR con un grupo de instancias que contiene tipos de instancias diferentes a los de la solicitud original.

Las siguientes condiciones activan el proceso automatizado:

- El evento de capacidad insuficiente se ha estado emitiendo en los nodos principales o de núcleo durante más de 20 minutos.
- El clúster no está en estado LISTO o EN ESPERA. Para obtener más información acerca de los estados del clúster de EMR, consulte Descripción del ciclo de vida del clúster.

> **Note**
>
> Al crear un proceso automatizado para una excepción de capacidad insuficiente, debe tener en cuenta que el evento de capacidad insuficiente es recuperable. La capacidad cambia con frecuencia y sus clústeres reanudarán el cambio de tamaño o comenzarán a funcionar tan pronto como la capacidad de Amazon EC2 esté disponible.

Example función para responder a un evento de capacidad insuficiente

```
// Lambda code with Python 3.10 and handler is lambda_function.lambda_handler
// Note: related IAM role requires permission to use Amazon EMR

import json
import boto3
import datetime
from datetime import timezone

INSUFFICIENT_CAPACITY_EXCEPTION_DETAIL_TYPE = "EMR Instance Group Provisioning"
INSUFFICIENT_CAPACITY_EXCEPTION_EVENT_CODE = (
    "EC2 provisioning - Insufficient Instance Capacity"
)
ALLOWED_INSTANCE_TYPES_TO_USE = [
    "m5.xlarge",
    "c5.xlarge",
```

```
    "m5.4xlarge",
    "m5.2xlarge",
    "t3.xlarge",
]
CLUSTER_START_ACCEPTABLE_STATES = ["WAITING", "RUNNING"]
CLUSTER_START_SLA = 20

CLIENT = boto3.client("emr", region_name="us-east-1")

# checks if the incoming event is 'EMR Instance Fleet Provisioning' with eventCode 'EC2
 provisioning - Insufficient Instance Capacity'
def is_insufficient_capacity_event(event):
    if not event["detail"]:
        return False
    else:
        return (
            event["detail-type"] == INSUFFICIENT_CAPACITY_EXCEPTION_DETAIL_TYPE
            and event["detail"]["eventCode"]
            == INSUFFICIENT_CAPACITY_EXCEPTION_EVENT_CODE
        )


# checks if the cluster is eligible for termination
def is_cluster_eligible_for_termination(event, describeClusterResponse):
    # instanceGroupType could be CORE, MASTER OR TASK
    instanceGroupType = event["detail"]["instanceGroupType"]
    clusterCreationTime = describeClusterResponse["Cluster"]["Status"]["Timeline"][
        "CreationDateTime"
    ]
    clusterState = describeClusterResponse["Cluster"]["Status"]["State"]

    now = datetime.datetime.now()
    now = now.replace(tzinfo=timezone.utc)
    isClusterStartSlaBreached = clusterCreationTime < now - datetime.timedelta(
        minutes=CLUSTER_START_SLA
    )

    # Check if instance group receiving Insufficient capacity exception is CORE or
 PRIMARY (MASTER),
    # and it's been more than 20 minutes since cluster was created but the cluster
 state and the cluster state is not updated to RUNNING or WAITING
    if (
        (instanceGroupType == "CORE" or instanceGroupType == "MASTER")
        and isClusterStartSlaBreached
```

```
        and clusterState not in CLUSTER_START_ACCEPTABLE_STATES
    ):
        return True
    else:
        return False


# Choose item from the list except the exempt value
def choice_excluding(exempt):
    for i in ALLOWED_INSTANCE_TYPES_TO_USE:
        if i != exempt:
            return i


# Create a new cluster by choosing different InstanceType.
def create_cluster(event):
    # instanceGroupType cloud be CORE, MASTER OR TASK
    instanceGroupType = event["detail"]["instanceGroupType"]

    # Following two lines assumes that the customer that created the cluster already
 knows which instance types they use in original request
    instanceTypesFromOriginalRequestMaster = "m5.xlarge"
    instanceTypesFromOriginalRequestCore = "m5.xlarge"

    # Select new instance types to include in the new createCluster request
    instanceTypeForMaster = (
        instanceTypesFromOriginalRequestMaster
        if instanceGroupType != "MASTER"
        else choice_excluding(instanceTypesFromOriginalRequestMaster)
    )
    instanceTypeForCore = (
        instanceTypesFromOriginalRequestCore
        if instanceGroupType != "CORE"
        else choice_excluding(instanceTypesFromOriginalRequestCore)
    )

    print("Starting to create cluster...")
    instances = {
        "InstanceGroups": [
            {
                "InstanceRole": "MASTER",
                "InstanceCount": 1,
                "InstanceType": instanceTypeForMaster,
                "Market": "ON_DEMAND",
```

```
                "Name": "Master",
            },
            {
                "InstanceRole": "CORE",
                "InstanceCount": 1,
                "InstanceType": instanceTypeForCore,
                "Market": "ON_DEMAND",
                "Name": "Core",
            },
        ]
    }
    response = CLIENT.run_job_flow(
        Name="Test Cluster",
        Instances=instances,
        VisibleToAllUsers=True,
        JobFlowRole="EMR_EC2_DefaultRole",
        ServiceRole="EMR_DefaultRole",
        ReleaseLabel="emr-6.10.0",
    )

    return response["JobFlowId"]


# Terminated the cluster using clusterId received in an event
def terminate_cluster(event):
    print("Trying to terminate cluster, clusterId: " + event["detail"]["clusterId"])
    response = CLIENT.terminate_job_flows(JobFlowIds=[event["detail"]["clusterId"]])
    print(f"Terminate cluster response: {response}")


def describe_cluster(event):
    response = CLIENT.describe_cluster(ClusterId=event["detail"]["clusterId"])
    return response


def lambda_handler(event, context):
    if is_insufficient_capacity_event(event):
        print(
            "Received insufficient capacity event for instanceGroup, clusterId: "
            + event["detail"]["clusterId"]
        )

        describeClusterResponse = describe_cluster(event)
```

```
        shouldTerminateCluster = is_cluster_eligible_for_termination(
            event, describeClusterResponse
        )
        if shouldTerminateCluster:
            terminate_cluster(event)

            clusterId = create_cluster(event)
            print("Created a new cluster, clusterId: " + clusterId)
        else:
            print(
                "Cluster is not eligible for termination, clusterId: "
                + event["detail"]["clusterId"]
            )

    else:
        print("Received event is not insufficient capacity event, skipping")
```

Respuesta a eventos de tiempo de espera agotado en el cambio de tamaño de la flota de instancias de clústeres de Amazon EMR

Información general

Los clústeres de Amazon EMR emiten eventos al ejecutar la operación de cambio de tamaño en los clústeres de flotas de instancias. Los eventos de tiempo de espera de aprovisionamiento se emiten cuando Amazon EMR deja de aprovisionar capacidad de spot o bajo demanda para la flota una vez transcurrido el tiempo de espera. El usuario puede configurar la duración del tiempo de espera como parte de las especificaciones de cambio de tamaño de las flotas de instancias. En escenarios de cambios de tamaño consecutivos para la misma flota de instancias, Amazon EMR emite los eventos `Spot provisioning timeout - continuing resize` o `On-Demand provisioning timeout - continuing resize` cuando vence el tiempo de espera de la operación de cambio de tamaño actual. A continuación, comienza a aprovisionar capacidad para la siguiente operación de cambio de tamaño de la flota.

Respuesta a eventos de tiempo de espera agotado en el cambio de tamaño de la flota de instancias

Le recomendamos que responda a un evento de tiempo de espera de aprovisionamiento de una de las siguientes maneras:

- Revisite las especificaciones de cambio de tamaño y vuelva a intentar la operación de cambio de tamaño. Como la capacidad cambia con frecuencia, los clústeres cambiarán de tamaño correctamente en cuanto haya capacidad de Amazon EC2 disponible. Recomendamos a los

clientes que configuren valores más bajos para la duración del tiempo de espera en los trabajos que requieren acuerdos de nivel de servicio más estrictos.

- Como alternativa, puede:
  - Lance un nuevo clúster con tipos de instancias diversificados según las prácticas recomendadas para la flexibilidad de las instancias y las zonas de disponibilidad o
  - Lanzar un clúster con capacidad bajo demanda
- En el caso de un evento “Tiempo de espera de aprovisionamiento: continuación del cambio de tamaño”, también puede esperar a que se procesen las operaciones de cambio de tamaño. Amazon EMR seguirá procesando secuencialmente las operaciones de cambio de tamaño activadas para la flota, respetando las especificaciones de cambio de tamaño configuradas.

También puede configurar reglas o respuestas automatizadas a este evento, como se describe en la siguiente sección.

Recuperación automática de un evento de tiempo de espera de aprovisionamiento

Puede crear una automatización en respuesta a los eventos de Amazon EMR con el código de evento `Spot Provisioning timeout`. Por ejemplo, la siguiente función de AWS Lambda apaga un clúster de EMR con una flota de instancias que usa instancias de spot para los nodos de tarea y, a continuación, crea un nuevo clúster de EMR con una flota de instancias que contiene tipos de instancias más diversificados que la solicitud original. En este ejemplo, el evento `Spot Provisioning timeout` emitido para los nodos de tarea activará la ejecución de la función de Lambda.

Example Ejemplo de función para responder a un evento **`Spot Provisioning timeout`**

```
// Lambda code with Python 3.10 and handler is lambda_function.lambda_handler
// Note: related IAM role requires permission to use Amazon EMR

import json
import boto3
import datetime
from datetime import timezone

SPOT_PROVISIONING_TIMEOUT_EXCEPTION_DETAIL_TYPE = "EMR Instance Fleet Resize"
SPOT_PROVISIONING_TIMEOUT_EXCEPTION_EVENT_CODE = (
    "Spot Provisioning timeout"
)
```

```
CLIENT = boto3.client("emr", region_name="us-east-1")

# checks if the incoming event is 'EMR Instance Fleet Resize' with eventCode 'Spot
 provisioning timeout'
def is_spot_provisioning_timeout_event(event):
    if not event["detail"]:
        return False
    else:
        return (
            event["detail-type"] == SPOT_PROVISIONING_TIMEOUT_EXCEPTION_DETAIL_TYPE
            and event["detail"]["eventCode"]
            == SPOT_PROVISIONING_TIMEOUT_EXCEPTION_EVENT_CODE
        )


# checks if the cluster is eligible for termination
def is_cluster_eligible_for_termination(event, describeClusterResponse):
    # instanceFleetType could be CORE, MASTER OR TASK
    instanceFleetType = event["detail"]["instanceFleetType"]

    # Check if instance fleet receiving Spot provisioning timeout event is TASK
    if (instanceFleetType == "TASK"):
        return True
    else:
        return False


# create a new cluster by choosing different InstanceType.
def create_cluster(event):
    # instanceFleetType cloud be CORE, MASTER OR TASK
    instanceFleetType = event["detail"]["instanceFleetType"]

    # the following two lines assumes that the customer that created the cluster
 already knows which instance types they use in original request
    instanceTypesFromOriginalRequestMaster = "m5.xlarge"
    instanceTypesFromOriginalRequestCore = "m5.xlarge"

    # select new instance types to include in the new createCluster request
    instanceTypesForTask = [
        "m5.xlarge",
        "m5.2xlarge",
        "m5.4xlarge",
        "m5.8xlarge",
        "m5.12xlarge"
```

```
    ]

    print("Starting to create cluster...")
    instances = {
        "InstanceFleets": [
            {
                "InstanceFleetType":"MASTER",
                "TargetOnDemandCapacity":1,
                "TargetSpotCapacity":0,
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesFromOriginalRequestMaster,
                        "WeightedCapacity":1,
                    }
                ]
            },
            {
                "InstanceFleetType":"CORE",
                "TargetOnDemandCapacity":1,
                "TargetSpotCapacity":0,
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesFromOriginalRequestCore,
                        "WeightedCapacity":1,
                    }
                ]
            },
            {
                "InstanceFleetType":"TASK",
                "TargetOnDemandCapacity":0,
                "TargetSpotCapacity":100,
                "LaunchSpecifications":{},
                "InstanceTypeConfigs":[
                    {
                        'InstanceType': instanceTypesForTask[0],
                        "WeightedCapacity":1,
                    },
                    {
                        'InstanceType': instanceTypesForTask[1],
                        "WeightedCapacity":2,
                    },
                    {
                        'InstanceType': instanceTypesForTask[2],
                        "WeightedCapacity":4,
```

```
                    },
                    {
                        'InstanceType': instanceTypesForTask[3],
                        "WeightedCapacity":8,
                    },
                    {
                        'InstanceType': instanceTypesForTask[4],
                        "WeightedCapacity":12,
                    }
                ],
                "ResizeSpecifications": {
                    "SpotResizeSpecification": {
                        "TimeoutDurationMinutes": 30
                    }
                }
            }
        ]
    }
    response = CLIENT.run_job_flow(
        Name="Test Cluster",
        Instances=instances,
        VisibleToAllUsers=True,
        JobFlowRole="EMR_EC2_DefaultRole",
        ServiceRole="EMR_DefaultRole",
        ReleaseLabel="emr-6.10.0",
    )

    return response["JobFlowId"]


# terminated the cluster using clusterId received in an event
def terminate_cluster(event):
    print("Trying to terminate cluster, clusterId: " + event["detail"]["clusterId"])
    response = CLIENT.terminate_job_flows(JobFlowIds=[event["detail"]["clusterId"]])
    print(f"Terminate cluster response: {response}")


def describe_cluster(event):
    response = CLIENT.describe_cluster(ClusterId=event["detail"]["clusterId"])
    return response


def lambda_handler(event, context):
    if is_spot_provisioning_timeout_event(event):
```

```
        print(
            "Received spot provisioning timeout event for instanceFleet, clusterId: "
            + event["detail"]["clusterId"]
        )

        describeClusterResponse = describe_cluster(event)

        shouldTerminateCluster = is_cluster_eligible_for_termination(
            event, describeClusterResponse
        )
        if shouldTerminateCluster:
            terminate_cluster(event)

            clusterId = create_cluster(event)
            print("Created a new cluster, clusterId: " + clusterId)
        else:
            print(
                "Cluster is not eligible for termination, clusterId: "
                + event["detail"]["clusterId"]
            )

    else:
        print("Received event is not spot provisioning timeout event, skipping")
```

Ver métricas de aplicaciones de clúster con Ganglia

Ganglia está disponible con las versiones 4.2 y superiores de Amazon EMR. Ganglia es un proyecto de código que es un sistema distribuido y escalable diseñado para monitorizar clústeres y redes al mismo tiempo que minimiza el impacto en su rendimiento. Al habilitar Ganglia en su clúster, puede generar informes y ver el rendimiento del clúster en su conjunto, además de inspeccionar el rendimiento de cada una de las instancias de nodo individuales. Ganglia también está configurado para adquirir y visualizar las métricas de Hadoop y Spark. Para obtener más información, consulte Ganglia en la Guía de publicación de Amazon EMR.

Registro de llamadas a la API de Amazon EMR en AWS CloudTrail

Amazon EMR se integra con AWS CloudTrail, un servicio que proporciona un registro de las acciones hechas por un usuario, un rol o un servicio de AWS en Amazon EMR. CloudTrail captura todas las llamadas a la API para Amazon EMR como eventos. Las llamadas capturadas incluyen las llamadas desde la consola de Amazon EMR y las llamadas desde el código a las operaciones de la API de Amazon EMR. Si crea un registro de seguimiento, puede habilitar la entrega continua

de eventos de CloudTrail a un bucket de Amazon S3, incluidos los eventos de Amazon EMR. Si no configura un registro de seguimiento, puede ver los eventos más recientes de la consola de CloudTrail en el Historial de eventos. Mediante la información recopilada por CloudTrail, puede determinar la solicitud que se realizó a Amazon EMR, la dirección IP de origen desde la que se realizó la solicitud, quién realizó la solicitud, cuándo se realizó y otros detalles adicionales.

Para obtener más información acerca de CloudTrail, consulte la Guía del usuario de AWS CloudTrail.

Información de Amazon EMR en CloudTrail

CloudTrail se habilita en su cuenta de AWS cuando la crea. Cuando se produce una actividad en Amazon EMR, dicha actividad se registra en un evento de CloudTrail junto con los eventos de los demás servicios de AWS en Historial de eventos. Puede ver, buscar y descargar los últimos eventos de la cuenta de AWS. Para obtener más información, consulte Visualización de eventos con el historial de eventos de CloudTrail.

Para mantener un registro continuo de los eventos de la cuenta de AWS, incluidos los eventos de Amazon EMR, cree un registro de seguimiento. Un registro de seguimiento permite a CloudTrail enviar archivos de registro a un bucket de Amazon S3. De manera predeterminada, cuando se crea un registro de seguimiento en la consola, el registro de seguimiento se aplica a todas las regiones de AWS. El registro de seguimiento registra los eventos de todas las regiones de la partición de AWS y envía los archivos de registro al bucket de Amazon S3 especificado. También es posible configurar otros servicios de AWS para analizar en profundidad y actuar en función de los datos de eventos recopilados en los registros de CloudTrail. Para obtener más información, consulte los siguientes temas:

- Introducción a la creación de registros de seguimiento
- Servicios e integraciones compatibles con CloudTrail
- Configuración de notificaciones de Amazon SNS para CloudTrail
- Recibir archivos de registro de CloudTrail de varias regiones y Recibir archivos de registro de CloudTrail de varias cuentas

Todas las acciones de Amazon EMR se registran en CloudTrail y están documentadas en la referencia de la API de Amazon EMR. Por ejemplo, las llamadas a las acciones `RunJobFlow`, `ListCluster` y `DescribeCluster` generan entradas en los archivos de registros de CloudTrail.

Cada entrada de registro o evento contiene información sobre quién generó la solicitud. La información de identidad del usuario le ayuda a determinar lo siguiente:

- Si la solicitud se realizó con credenciales de usuario AWS Identity and Access Management (IAM) o credenciales de usuario raíz.
- Si la solicitud se realizó con credenciales de seguridad temporales de un rol o fue un usuario federado.
- Si la solicitud la realizó otro servicio de AWS.

En el caso de que un proceso, en lugar de un usuario, cree un clúster, puede usar el identificador `principalId` para determinar el usuario asociado a la creación del clúster. Para obtener más información, consulte el elemento userIdentity de CloudTrail.

Ejemplo: entradas del archivo de registros de Amazon EMR

Un registro de seguimiento es una configuración que permite la entrega de eventos como archivos de registros en un bucket de Amazon S3 que especifique. Los archivos log de CloudTrail pueden contener una o varias entradas de log. Un evento representa una solicitud específica realizada desde un origen y contiene información sobre la acción solicitada, la fecha y la hora de la acción, los parámetros de la solicitud, etc. Los archivos de registro de CloudTrail no rastrean el orden en la pila de las llamadas públicas a la API, por lo que estas no aparecen en ningún orden específico.

En el siguiente ejemplo, se muestra una entrada de registro de CloudTrail que ilustra la acción RunJobFlow.


```
{
 "Records": [
 {
        "eventVersion":"1.01",
        "userIdentity":{
           "type":"IAMUser",
           "principalId":"EX_PRINCIPAL_ID",
           "arn":"arn:aws:iam::123456789012:user/temporary-user-xx-7M",
           "accountId":"123456789012",
           "userName":"temporary-user-xx-7M"
        },
        "eventTime":"2018-03-31T17:59:21Z",
        "eventSource":"elasticmapreduce.amazonaws.com",
        "eventName":"RunJobFlow",
        "awsRegion":"us-west-2",
        "sourceIPAddress":"192.0.2.1",
```

```
         "userAgent":"aws-sdk-java/unknown-version Linux/xx Java_HotSpot(TM)_64-
Bit_Server_VM/xx",
         "requestParameters":{
            "tags":[
               {
                  "value":"prod",
                  "key":"domain"
               },
               {
                  "value":"us-west-2",
                  "key":"realm"
               },
               {
                  "value":"VERIFICATION",
                  "key":"executionType"
               }
            ],
            "instances":{
               "slaveInstanceType":"m5.xlarge",
               "ec2KeyName":"emr-integtest",
               "instanceCount":1,
               "masterInstanceType":"m5.xlarge",
               "keepJobFlowAliveWhenNoSteps":true,
               "terminationProtected":false
            },
            "visibleToAllUsers":false,
            "name":"MyCluster",
            "ReleaseLabel":"emr-5.16.0"
         },
         "responseElements":{
            "jobFlowId":"j-2WDJCGEG4E6AJ"
         },
         "requestID":"2f482daf-b8fe-11e3-89e7-75a3d0e071c5",
         "eventID":"b348a38d-f744-4097-8b2a-e68c9b424698"
      },
  ...additional entries
   ]
}
```

Usar el escalado de clústeres

Puede ajustar el número de instancias Amazon EC2 disponibles para un clúster de Amazon EMR automáticamente o manualmente en respuesta a cargas de trabajo que tienen diferentes demandas. Para utilizar el escalado automático, tiene dos opciones. Puede habilitar Escalado administrado de Amazon EMR o crear una política de escalado automático personalizada. En la siguiente tabla se describen las diferencias entre las dos opciones.

	Escalado administrado de Amazon EMR	Escalado automático personalizado
Políticas y reglas de escalado	No se requiere ninguna política. Amazon EMR administra la actividad de escalado automático evaluando continuamente las métricas del clúster y tomando decisiones de escalado optimizado.	Debe definir y administrar las políticas y reglas de escalado automático, como las condiciones específicas que desencadenan actividades de escalado, periodos de evaluación, periodos de recuperación, etc.
Versiones compatibles de Amazon EMR	Versiones 5.30.0 y posteriores de Amazon EMR (excepto la versión 6.0.0 de Amazon EMR)	Versiones 4.0.0 y posteriores de Amazon EMR
Composición del clúster admitida	Grupos de instancias o flotas de instancias	Solo grupos de instancias
Configuración de límites de escalado	Los límites de escalado se configuran para todo el clúster.	Los límites de escalado solo se pueden configurar para cada grupo de instancias.
Frecuencia de evaluación de las métricas	Cada 5-10 segundos Una evaluación más frecuente de las métricas permite a Amazon EMR tomar decisiones de escalado más precisas.	Puede definir los periodos de evaluación solo en incrementos de cinco minutos.

	Escalado administrado de Amazon EMR	Escalado automático personalizado
Aplicaciones compatibles	Solo se admiten aplicaciones YARN, como Spark, Hadoop, Hive y Flink. Escalado administrado de Amazon EMR no admite aplicaciones que no estén basadas en YARN, como Presto o HBase.	Puede elegir qué aplicaciones se admiten cuando defina las reglas de escalado automático.

Consideraciones

- Un clúster de Amazon EMR siempre consta de uno o tres nodos principales. Una vez que haya configurado inicialmente el clúster, solo podrá escalar los nodos principales y de tareas. No puede escalar el número de nodos principales del clúster.
- En el caso de los grupos de instancias, las operaciones de reconfiguración y cambio de tamaño se realizan de forma consecutiva y no simultánea. Si inicia una reconfiguración mientras se está cambiando el tamaño de un grupo de instancias, la reconfiguración comienza una vez que el grupo de instancias complete el cambio de tamaño en curso. Por el contrario, si inicia una operación de cambio de tamaño mientras se está reconfigurando un grupo de instancias, se cambiará el tamaño en el momento.

Uso del escalado administrado en Amazon EMR

Important

Le recomendamos encarecidamente que utilice la versión más reciente de Amazon EMR (6.14.0) para el escalado administrado. En las versiones anteriores, era posible que se produjeran errores intermitentes en las aplicaciones o retrasos en el escalado. Amazon EMR resolvió este problema en las versiones 5.x: (5.30.2, 5.31.1, 5.32.1, 5.33.1 y posteriores) y las versiones 6.x (6.1.1, 6.2.1, 6.3.1 y posteriores). Para obtener más información sobre la disponibilidad de las versiones y las regiones, consulte Disponibilidad de Escalado administrado.

Información general

Con las versiones 5.30.0 y posteriores de Amazon EMR (excepto la versión 6.0.0 de Amazon EMR), puede habilitar el Escalado administrado de Amazon EMR. El escalado administrado le permite aumentar o disminuir automáticamente el número de instancias o unidades del clúster en función de la carga de trabajo. Amazon EMR evalúa continuamente las métricas del clúster para tomar decisiones de escalado que optimicen los clústeres en cuanto al costo y la velocidad. El escalado administrado está disponible para clústeres compuestos por grupos de instancias o flotas de instancias.

Disponibilidad de Escalado administrado

- En Asia-Pacífico (Yakarta), el Escalado administrado de Amazon EMR está disponible con Amazon EMR 6.14.0 y versiones posteriores.
- El Escalado administrado de Amazon EMR está disponible con Amazon EMR 5.30.0, 6.1.0 y versiones posteriores en las siguientes Regiones de AWS:

  Este de EE. UU. (Norte de Virginia y Ohio), Oeste de EE. UU. (Oregón y Norte de California), América del Sur (São Paulo), Europa (Fráncfort, Irlanda, Londres, Milán, París y Estocolmo), Canadá (centro), Asia-Pacífico (Hong Kong, Bombay, Seúl, Singapur, Sídney y Tokio), Medio Oriente (Baréin), África (Ciudad del Cabo), AWS GovCloud (Este de EE. UU.), AWS GovCloud (Oeste de EE. UU.), China (Pekín) operada por Sinnet y China (Ningxia) operada por NWCD.
- Escalado administrado de Amazon EMR solo funciona con aplicaciones YARN, como Spark, Hadoop, Hive y Flink. No es compatible con aplicaciones que se basen en YARN, como Presto y HBase.

Parámetros de escalado administrado

Debe configurar los siguientes parámetros para el escalado administrado. El límite solo se aplica a los nodos principales y de tareas. No puedes escalar el nodo principal después de la configuración inicial.

- Mínimo (`MinimumCapacityUnits`): el límite inferior de la capacidad de EC2 permitida en un clúster. Se mide mediante núcleos de unidades de procesamiento central virtual (vCPU) o instancias para grupos de instancias. Se mide mediante unidades para flotas de instancias.

- Máximo (`MaximumCapacityUnits`): el límite superior de la capacidad de EC2 permitida en un clúster. Se mide mediante núcleos de unidades de procesamiento central virtual (vCPU) o instancias para grupos de instancias. Se mide mediante unidades para flotas de instancias.
- Límite bajo demanda (`MaximumOnDemandCapacityUnits`) (opcional): límite superior de la capacidad de EC2 permitida para el tipo de mercado bajo demanda de un clúster. Si no se especifica este parámetro, se establece en el valor predeterminado de `MaximumCapacityUnits`.
  - Este parámetro se utiliza para dividir la asignación de capacidad entre instancias bajo demanda e instancias de spot. Por ejemplo, si establece el parámetro mínimo en 2 instancias, el parámetro máximo en 100 instancias y el límite bajo demanda en 10 instancias, Escalado administrado de Amazon EMR escala hasta 10 instancias bajo demanda y asigna la capacidad restante a instancias de spot. Para obtener más información, consulte Escenarios de asignación de nodos.
- Máximo de nodos principales (`MaximumCoreCapacityUnits`) (opcional): límite superior de la capacidad de EC2 permitida para el tipo de nodo principal de un clúster. Si no se especifica este parámetro, se establece en el valor predeterminado de `MaximumCapacityUnits`.
  - Este parámetro se utiliza para dividir la asignación de capacidad entre los nodos principales y de tarea. Por ejemplo, si establece el parámetro mínimo en 2 instancias, el máximo en 100 instancias y el máximo de nodos principales en 17 instancias, Escalado administrado de Amazon EMR escala hasta 17 nodos principales y asigna las 83 instancias restantes a los nodos de tarea. Para obtener más información, consulte Escenarios de asignación de nodos.

Para obtener más información sobre los parámetros de escalado administrado, consulte `ComputeLimits`.

Consideraciones sobre el Escalado administrado de Amazon EMR

- El Escalado administrado se admite en algunas Regiones de AWS y versiones de Amazon EMR. Para obtener más información, consulte Disponibilidad de Escalado administrado.
- Debe configurar los parámetros obligatorios para Escalado administrado de Amazon EMR. Para obtener más información, consulte Parámetros de escalado administrado.
- Para utilizar el escalado administrado, el proceso de recopilación de métricas debe poder conectarse al punto de conexión de la API pública para el escalado administrado en API Gateway. Si utiliza un nombre de DNS privado con Amazon Virtual Private Cloud, el escalado administrado no funcionará correctamente. Para garantizar que el escalado administrado funcione, se recomienda que realice una de las siguientes acciones:

- Elimine el punto de conexión de VPC de la interfaz de API Gateway de su Amazon VPC.
- Siga las instrucciones de ¿Por qué aparece el error HTTP 403 Prohibido al conectarme a mis API de API Gateway desde una VPC? para deshabilitar la configuración del nombre de DNS privado.
- En su lugar, lance su clúster en una subred privada. Para más información, consulte el tema sobre Subredes privadas.

- Si sus trabajos de YARN se ralentizan de forma intermitente durante la reducción vertical, y los registros de YARN Resource Manager indican que la mayoría de sus nodos estaban en la lista de denegación durante ese tiempo, puede ajustar el límite de tiempo de espera para la retirada.

  Reduzca el valor de `spark.blacklist.decommissioning.timeout` de una hora a un minuto para que el nodo esté disponible para que otros contenedores pendientes puedan continuar con el procesamiento de las tareas.

  También debe establecer `YARN.resourcemanager.nodemanager-graceful-decommission-timeout-secs` en un valor mayor para garantizar que Amazon EMR no fuerce la terminación del nodo mientras la “Tarea de Spark” más larga siga ejecutándose en el nodo. El valor predeterminado actual son 60 minutos, lo que significa que YARN fuerza la terminación del contenedor transcurridos 60 minutos una vez que el nodo entra en el estado de retirada.

  En el siguiente ejemplo de la línea de registro de YARN Resource Manager, se muestran los nodos agregado al estado de retirada:

```
2021-10-20 15:55:26,994 INFO
 org.apache.hadoop.YARN.server.resourcemanager.DefaultAMSProcessor
 (IPC Server handler 37 on default port 8030): blacklist are updated in
 Scheduler.blacklistAdditions: [ip-10-10-27-207.us-west-2.compute.internal,
 ip-10-10-29-216.us-west-2.compute.internal, ip-10-10-31-13.us-
west-2.compute.internal, ... , ip-10-10-30-77.us-west-2.compute.internal],
 blacklistRemovals: []
```

  Consulte más información sobre cómo Amazon EMR se integra con la lista de denegación de YARN durante la retirada de nodos, los casos en los que los nodos de Amazon EMR se pueden agregar a la lista de denegación y cómo configurar el comportamiento de retirada de nodos de Spark.

- La sobreutilización de los volúmenes de EBS puede provocar problemas de Escalado administrado. Le recomendamos que mantenga el volumen de EBS por debajo del 90 % de utilización. Para obtener más información, consulte Almacenamiento de la instancia.
- Las métricas de Amazon CloudWatch son fundamentales para que Escalado administrado de Amazon EMR funcione. Le recomendamos que monitorice de cerca las métricas de Amazon CloudWatch para asegurarse de que no falten datos. Para obtener más información sobre cómo configurar las alarmas de CloudWatch para detectar las métricas que faltan, consulte Uso de las alarmas de Amazon CloudWatch.
- Las operaciones de escalado administrado en los clústeres 5.30.0 y 5.30.1 sin Presto instalado pueden provocar errores en las aplicaciones o provocar que un grupo de instancias o una flota de instancias uniformes permanezcan en estado `ARRESTED`, especialmente cuando una operación de reducción vertical va seguida inmediatamente de una operación de escalado vertical.

  Como solución alternativa, elija Presto como aplicación para instalar cuando cree un clúster con las versiones 5.30.0 y 5.30.1 de Amazon EMR, incluso si su trabajo no requiere Presto.
- Al establecer el nodo principal máximo y el límite bajo demanda de Escalado administrado de Amazon EMR, tenga en cuenta las diferencias entre los grupos de instancias y las flotas de instancias. Cada grupo de instancias se compone del mismo tipo de instancia y las mismas opciones de compra para las instancias: bajo demanda o de spot. Para cada flota de instancias, puede especificar hasta cinco tipos de instancia que se pueden aprovisionar como instancias bajo demanda e instancias de spot. Para obtener más información, consulte Crear un clúster con las flotas de instancias o grupos de instancias uniformes, Opciones de las flotas de instancias y Escenarios de asignación de nodos.
- Con Amazon EMR 5.30.0 y versiones posteriores, si quita la regla de salida predeterminada Permitir todo a 0.0.0.0/ para el grupo de seguridad principal, debe agregar una regla que permita la conectividad TCP de salida a su grupo de seguridad para el acceso al servicio en el puerto 9443. El grupo de seguridad para el acceso al servicio también debe permitir el tráfico TCP entrante en el puerto 9443 desde el grupo de seguridad principal. Para más información sobre la configuración de grupos de seguridad, consulte Grupo de seguridad administrado por Amazon EMR para la instancia principal (subredes privadas).
- El escalado administrado no admite la característica de etiquetas de nodos de YARN. Evite usar etiquetas de nodos en clústeres con el escalado administrado. Por ejemplo, no permita que los ejecutores se ejecuten solo en los nodos de tarea. Cuando utiliza etiquetas de nodos en sus clústeres de Amazon EMR, es posible que observe que su clúster no se escala verticalmente, lo que puede provocar una ralentización de su aplicación.

- Puede utilizar AWS CloudFormation para configurar Escalado administrado de Amazon EMR. Para obtener más información, consulte AWS::EMR::Cluster en la Guía del usuario de AWS CloudFormation.

Historial de características

En esta tabla se enumeran las actualizaciones de la capacidad de Escalado administrado de Amazon EMR.

Fecha de lanzamiento de la nueva versión	Capability	Versiones de Amazon EMR
10 de octubre de 2023	El Escalado administrado está disponible en la región de Asia-Pacífico (Yakarta) `ap-southeast-3` .	6.14.0 y versiones posteriores
28 de julio de 2023	Mejora del escalado administrado para cambiar a un grupo de instancias de tareas diferente al escalar verticalmente cuando Amazon EMR experimenta un retraso en el escalado vertical con el grupo de instancias actual.	5.34.0 y versiones posteriores, 6.4.0 y versiones posteriores
16 de junio de 2023	Mejora del escalado administrado para detectar los nodos que ejecutan la aplicación maestra, de modo que esos nodos no se reduzcan verticalmente. Para obtener más información, consulte Comprender la estrategia y los escenarios de asignación de nodos.	5.34.0 y versiones posteriores, 6.4.0 y versiones posteriores

Fecha de lanzamiento de la nueva versión	Capability	Versiones de Amazon EMR
21 de marzo de 2022	Se agregó el reconocimiento de datos de mezclas aleatorias de Spark que se usa al reducir verticalmente los clústeres. En el caso de los clústeres de Amazon EMR con Apache Spark y la característica de escalado administrado habilitada, Amazon EMR supervisa de forma continua los ejecutores de Spark y las ubicaciones de datos de mezclas aleatorias intermedias. Con esta información, Amazon EMR solo reduce verticalmente las instancias infrautilizadas que no contienen datos de mezclas aleatorias utilizados activamente. Esto evita el recálculo de los datos de mezclas aleatorias perdidos, lo que ayuda a reducir los costos y a mejorar el rendimiento laboral. Para más información, consulte Spark Programming Guide.	5.34.0 y versiones posteriores, 6.4.0 y versiones posteriores

Configuración del escalado administrado para Amazon EMR

En las siguientes secciones se explica cómo lanzar un clúster de EMR que utilice el escalado administrado con la AWS Management Console, el AWS SDK for Java o la AWS Command Line Interface.

Temas

- Usar la AWS Management Console para configurar el escalado administrado
- Usar la AWS CLI para configurar el escalado administrado
- Usar el AWS SDK for Java para configurar el escalado administrado

Usar la AWS Management Console para configurar el escalado administrado

Puede usar la consola de Amazon EMR para configurar el escalado administrado al crear un clúster o para cambiar una política de escalado administrado para un clúster en ejecución.

New console

Para configurar el escalado administrado al crear un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. Seleccione la versión emr-5.30.0 o posterior de Amazon EMR, excepto la versión emr-6.0.0.
4. En Opción de escalado y aprovisionamiento del clúster, seleccione Usar escalado administrado de EMR. Especifique la cantidad mínima y máxima de instancias, la cantidad máxima de instancias de nodos principales y la cantidad máxima de instancias bajo demanda.
5. Elija cualquier otra opción que se aplique a su clúster.
6. Para lanzar el clúster, elija Crear clúster.

Para configurar el escalado administrado en un clúster existente con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar.
3. En la pestaña Instancias de la página de detalles del clúster, busque la sección Configuración del grupo de instancias. Seleccione Editar el escalado del clúster para especificar nuevos valores para la cantidad mínima y máxima de instancias y el límite bajo demanda.

Old console

Al crear un clúster en la consola antigua, puede configurar el escalado administrado mediante las opciones rápidas o las opciones avanzadas de configuración del clúster. También puede crear o cambiar una política de escalado administrado para un clúster en ejecución modificando la configuración de Escalado administrado en la página Resumen o Hardware .

Para utilizar las opciones rápidas para configurar el escalado administrado al crear un clúster con la consola antigua

1. Abra la consola de Amazon EMR, seleccione Crear clúster y abra Crear clúster: opciones rápidas.
2. En la sección Configuración de hardware situada junto a Opción de escalado y aprovisionamiento del clúster, seleccione la casilla de verificación para habilitar el escalado de los nodos del clúster en función de la carga de trabajo.
3. En Unidades principales y de tarea, especifique la cantidad mínima y máxima de las instancias principales y de tarea.

Para utilizar la opción avanzada para configurar el escalado administrado al crear un clúster con la consola antigua

1. En la consola de Amazon EMR, seleccione Crear clúster, seleccione Ir a las opciones avanzadas, seleccione las opciones en Paso 1: Software y pasos y después vaya a Paso 2: Configuración de hardware.
2. En la sección Composición del clúster seleccione Flotas de instancias o Grupos de instancias uniformes.
3. En Opción de escalado y aprovisionamiento del clúster, seleccione Habilitar el escalado del clúster. A continuación, seleccione Usar escalado administrado de EMR. En Unidades principales y de tarea, especifique el número mínimo y máximo de instancias o unidades de flota de instancias, el límite bajo demanda y el número máximo de nodos principales.

   Para los clústeres compuestos por grupos de instancias, también puede elegir Crear una política de escalado automático personalizada si desea definir políticas de escalado automática personalizadas para cada grupo de instancias. Para obtener más información, consulte Uso del escalado automático con una política personalizada para grupos de instancias .

Para modificar el escalado administrado en un clúster existente con la consola antigua

1. Abra la consola de Amazon EMR, seleccione el clúster de la lista de clústeres y, a continuación, seleccione la pestaña Hardware.
2. En la sección Opción de escalado y aprovisionamiento del clúster, seleccione Editar en Escalado administrado de Amazon EMR.
3. En la sección Opción de escalado y aprovisionamiento del clúster, especifique nuevos valores para la cantidad mínima y máxima de las instancias y el límite bajo demanda.

Usar la AWS CLI para configurar el escalado administrado

Puede utilizar comandos de la AWS CLI para que Amazon EMR configure el escalado administrado al crear un clúster. Puede utilizar una sintaxis abreviada, especificando la configuración JSON insertada dentro de los comandos pertinentes o puede hacer referencia a un archivo que contenga la configuración JSON. También puede aplicar una política de escalado administrado a un clúster existente y eliminar una política de escalado administrado que se haya aplicado anteriormente. Además, puede recuperar detalles de una configuración de política de escalado desde un clúster en ejecución.

Habilitar el escalado administrado durante el lanzamiento del clúster

Puede habilitar el escalado administrado durante el lanzamiento del clúster, como se muestra en el siguiente ejemplo.

```
aws emr create-cluster \
 --service-role EMR_DefaultRole \
 --release-label emr-5.36.1 \
 --name EMR_Managed_Scaling_Enabled_Cluster \
 --applications Name=Spark Name=Hbase \
 --ec2-attributes KeyName=keyName,InstanceProfile=EMR_EC2_DefaultRole \
 --instance-groups InstanceType=m4.xlarge,InstanceGroupType=MASTER,InstanceCount=1
 InstanceType=m4.xlarge,InstanceGroupType=CORE,InstanceCount=2 \
 --region us-east-1 \
 --managed-scaling-policy
 ComputeLimits='{MinimumCapacityUnits=2,MaximumCapacityUnits=4,UnitType=Instances}'
```

También puede especificar una configuración de política administrada mediante la opción --managed-scaling-policy cuando utilice `create-cluster`.

Aplicación de una política de escalado administrado a un clúster existente

Puede aplicar una política de escalado administrado a un clúster existente, como se muestra en el siguiente ejemplo.

```
aws emr put-managed-scaling-policy
--cluster-id j-123456
--managed-scaling-policy ComputeLimits='{MinimumCapacityUnits=1,
MaximumCapacityUnits=10,  MaximumOnDemandCapacityUnits=10, UnitType=Instances}'
```

También puede aplicar una política de escalado administrado a un clúster existente mediante el comando `aws emr put-managed-scaling-policy`. En el siguiente ejemplo se utiliza una referencia a un archivo JSON, `managedscaleconfig.json`, que especifica la configuración de política de escalado administrado.

```
aws emr put-managed-scaling-policy --cluster-id j-123456 --managed-scaling-policy
 file://./managedscaleconfig.json
```

En el ejemplo siguiente se muestra el contenido del archivo `managedscaleconfig.json`, que define la política de escalado administrado.

```
{
    "ComputeLimits": {
        "UnitType": "Instances",
        "MinimumCapacityUnits": 1,
        "MaximumCapacityUnits": 10,
        "MaximumOnDemandCapacityUnits": 10
    }
}
```

Recuperación de una configuración de política de escalado administrado

El comando `GetManagedScalingPolicy` recupera la configuración de la política. Por ejemplo, el comando siguiente recupera la configuración para el clúster con un ID de clúster de `j-123456`.

```
aws emr get-managed-scaling-policy --cluster-id j-123456
```

El comando produce el siguiente resultado de ejemplo.

```
{
   "ManagedScalingPolicy": {
      "ComputeLimits": {
         "MinimumCapacityUnits": 1,
```

```
        "MaximumOnDemandCapacityUnits": 10,
        "MaximumCapacityUnits": 10,
        "UnitType": "Instances"
      }
   }
}
```

Para más información sobre el uso de comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Eliminación de una política de escalado administrado

El comando `RemoveManagedScalingPolicy` elimina la configuración de política. Por ejemplo, el comando siguiente recupera la configuración del clúster con un ID de clúster de `j-123456`.

```
aws emr remove-managed-scaling-policy --cluster-id j-123456
```

Usar el AWS SDK for Java para configurar el escalado administrado

En el siguiente fragmento de programa se muestra cómo configurar el escalado administrado mediante AWS SDK for Java:

```
package com.amazonaws.emr.sample;

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.regions.Regions;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduce;
import com.amazonaws.services.elasticmapreduce.AmazonElasticMapReduceClientBuilder;
import com.amazonaws.services.elasticmapreduce.model.Application;
import com.amazonaws.services.elasticmapreduce.model.ComputeLimits;
import com.amazonaws.services.elasticmapreduce.model.ComputeLimitsUnitType;
import com.amazonaws.services.elasticmapreduce.model.InstanceGroupConfig;
import com.amazonaws.services.elasticmapreduce.model.JobFlowInstancesConfig;
import com.amazonaws.services.elasticmapreduce.model.ManagedScalingPolicy;
import com.amazonaws.services.elasticmapreduce.model.RunJobFlowRequest;
import com.amazonaws.services.elasticmapreduce.model.RunJobFlowResult;
```

```
public class CreateClusterWithManagedScalingWithIG {

 public static void main(String[] args) {
  AWSCredentials credentialsFromProfile = getCreadentials("AWS-Profile-Name-Here");

  /**
   * Create an Amazon EMR client with the credentials and region specified in order to
 create the cluster
   */
  AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
   .withCredentials(new AWSStaticCredentialsProvider(credentialsFromProfile))
   .withRegion(Regions.US_EAST_1)
   .build();

  /**
   * Create Instance Groups - Primary, Core, Task
   */
  InstanceGroupConfig instanceGroupConfigMaster = new InstanceGroupConfig()
    .withInstanceCount(1)
    .withInstanceRole("MASTER")
    .withInstanceType("m4.large")
    .withMarket("ON_DEMAND");

  InstanceGroupConfig instanceGroupConfigCore = new InstanceGroupConfig()
   .withInstanceCount(4)
   .withInstanceRole("CORE")
   .withInstanceType("m4.large")
   .withMarket("ON_DEMAND");

  InstanceGroupConfig instanceGroupConfigTask = new InstanceGroupConfig()
   .withInstanceCount(5)
   .withInstanceRole("TASK")
   .withInstanceType("m4.large")
   .withMarket("ON_DEMAND");

  List<InstanceGroupConfig> igConfigs = new ArrayList<>();
  igConfigs.add(instanceGroupConfigMaster);
  igConfigs.add(instanceGroupConfigCore);
  igConfigs.add(instanceGroupConfigTask);

        /**
         *  specify applications to be installed and configured when Amazon EMR creates
 the cluster
         */
```

```
 Application hive = new Application().withName("Hive");
 Application spark = new Application().withName("Spark");
 Application ganglia = new Application().withName("Ganglia");
 Application zeppelin = new Application().withName("Zeppelin");

 /**
  * Managed Scaling Configuration -
        * Using UnitType=Instances for clusters composed of instance groups
  *
        * Other options are:
        * UnitType = VCPU ( for clusters composed of instance groups)
        * UnitType = InstanceFleetUnits ( for clusters composed of instance fleets)
        **/
 ComputeLimits computeLimits = new ComputeLimits()
   .withMinimumCapacityUnits(1)
   .withMaximumCapacityUnits(20)
   .withUnitType(ComputeLimitsUnitType.Instances);

 ManagedScalingPolicy managedScalingPolicy = new ManagedScalingPolicy();
 managedScalingPolicy.setComputeLimits(computeLimits);

 // create the cluster with a managed scaling policy
 RunJobFlowRequest request = new RunJobFlowRequest()
         .withName("EMR_Managed_Scaling_TestCluster")
         .withReleaseLabel("emr-5.36.1")          // Specifies the version label for
the Amazon EMR release; we recommend the latest release
         .withApplications(hive,spark,ganglia,zeppelin)
         .withLogUri("s3://path/to/my/emr/logs")  // A URI in S3 for log files is
required when debugging is enabled.
         .withServiceRole("EMR_DefaultRole")      // If you use a custom IAM service
role, replace the default role with the custom role.
         .withJobFlowRole("EMR_EC2_DefaultRole")  // If you use a custom Amazon EMR
role for EC2 instance profile, replace the default role with the custom Amazon EMR
role.
         .withInstances(new JobFlowInstancesConfig().withInstanceGroups(igConfigs)
             .withEc2SubnetId("subnet-123456789012345")
             .withEc2KeyName("my-ec2-key-name")
             .withKeepJobFlowAliveWhenNoSteps(true))
         .withManagedScalingPolicy(managedScalingPolicy);
    RunJobFlowResult result = emr.runJobFlow(request);

    System.out.println("The cluster ID is " + result.toString());
 }
```

```
 public static AWSCredentials getCredentials(String profileName) {
  // specifies any named profile in .aws/credentials as the credentials provider
  try {
   return new ProfileCredentialsProvider("AWS-Profile-Name-Here")
     .getCredentials();
        } catch (Exception e) {
            throw new AmazonClientException(
                    "Cannot load credentials from .aws/credentials file. " +
                    "Make sure that the credentials file exists and that the profile
 name is defined within it.",
                    e);
        }
 }

 public CreateClusterWithManagedScalingWithIG() { }
}
```

Comprender la estrategia y los escenarios de asignación de nodos

En esta sección se ofrece información general sobre la estrategia de asignación de nodos y los escenarios de escalado comunes que puede utilizar con Escalado administrado de Amazon EMR.

Estrategia de asignación de nodos

Escalado administrado de Amazon EMR asigna los nodos principales y de tarea en función de las siguientes estrategias de escalado y reducción verticales:

Estrategia de escalado vertical

- Escalado administrado de Amazon EMR primero agrega capacidad a los nodos principales y, después, a los nodos de tarea hasta que se alcanza la capacidad máxima permitida o hasta que se alcanza la capacidad de destino de escalado vertical deseada.
- Cuando Amazon EMR experimenta un retraso en el escalado vertical con el grupo de instancias actual, los clústeres que usan el escalado administrado cambian automáticamente a un grupo de instancias de tareas diferente.
- Si el parámetro `MaximumCoreCapacityUnits` está establecido, Amazon EMR escala los nodos principales hasta que las unidades principales alcanzan el límite máximo permitido. Toda la capacidad restante se agrega a los nodos de tarea.
- Si el parámetro `MaximumOnDemandCapacityUnits` está establecido, Amazon EMR escala el clúster mediante las instancias bajo demanda hasta que las unidades bajo demanda alcanzan el límite máximo permitido. Toda la capacidad restante se agrega mediante instancias de spot.

- Si se establecen los parámetros `MaximumCoreCapacityUnits` y `MaximumOnDemandCapacityUnits`, Amazon EMR tiene en cuenta ambos límites durante el escalado.

  Por ejemplo, si `MaximumCoreCapacityUnits` es inferior a `MaximumOnDemandCapacityUnits`, Amazon EMR primero escala los nodos principales hasta alcanzar el límite de capacidad principal. Para la capacidad restante, Amazon EMR utiliza primero las instancias bajo demanda para escalar los nodos de tarea hasta alcanzar el límite bajo demanda y, a continuación, utiliza las instancias de spot para los nodos de tarea.

Estrategia de reducción vertical

- Las versiones 5.34.0 y posteriores, y 6.4.0 y posteriores de Amazon EMR admiten el escalado administrado compatible con los datos de mezclas aleatorias de Spark (datos que Spark redistribuye entre las particiones para realizar operaciones específicas). Para más información sobre las operaciones de mezclas aleatorias, consulte Spark Programming Guide. El escalado administrado solo reduce verticalmente las instancias que están infrautilizadas y que no contienen datos de mezclas aleatorias utilizados activamente. Este escalado inteligente evita la pérdida involuntaria de datos de mezclas aleatorias, lo que evita la necesidad de volver a intentar el trabajo y volver a calcular los datos intermedios.
- Escalado administrado de Amazon EMR elimina primero los nodos de tarea y, a continuación, elimina los nodos principales hasta que se alcanza la capacidad de destino de reducción vertical deseada. El clúster nunca se escala por debajo de las restricciones mínimas de la política de escalado administrado.
- Dentro de cada tipo de nodo (ya sean nodos principales o de tarea), Escalado administrado de Amazon EMR elimina primero las instancias de spot y, a continuación, las instancias bajo demanda.
- En el caso de los clústeres que se lanzan con las versiones 5.x, 5.34.0 y posteriores, y las versiones 6.x, 6.4.0 y posteriores de Amazon EMR, el Escalado administrado de Amazon EMR no reduce verticalmente los nodos que tienen `ApplicationMaster` para Apache Spark en ejecución. Esto minimiza los errores y los reintentos en los trabajos, lo que ayuda a mejorar el rendimiento de los trabajos y a reducir los costos. Para confirmar qué nodos del clúster se ejecutan `ApplicationMaster`, visite el servidor del historial de Spark y filtre el controlador en la pestaña Ejecutores del ID de aplicación de Spark.

Si el clúster no tiene ninguna carga, Amazon EMR cancela la adición de nuevas instancias de una evaluación anterior y realiza operaciones de reducción vertical. Si el clúster tiene una carga elevada, Amazon EMR cancela la eliminación de instancias y realiza operaciones de escalado vertical.

Consideraciones sobre la asignación de nodos

Le recomendamos que utilice la opción de compra bajo demanda para los nodos principales a fin de evitar la pérdida de datos de HDFS en caso de recuperación de spot. Puede utilizar la opción de compra puntual para los nodos de tarea a fin de reducir los costos y agilizar la ejecución de los trabajos cuando se agreguen más instancias de spot a los nodos de tarea.

Escenarios de asignación de nodos

Puede crear varios escenarios de escalado en función de sus necesidades configurando los parámetros máximo, mínimo, límite bajo demanda y máximo de nodos principales en diferentes combinaciones.

Escenario 1: escalar únicamente los nodos principales

Para escalar únicamente los nodos principales, los parámetros de escalado administrado deben cumplir los siguientes requisitos:

- El límite bajo demanda es igual al límite máximo.
- El máximo de nodos principales es igual al límite máximo.

Cuando no se especifican los parámetros del límite bajo demanda y el máximo de nodos principales, ambos parámetros se establecen de forma predeterminada en el límite máximo.

En los siguientes ejemplos, se muestra el escenario de escalado de nodos principales únicamente.

Estado inicial del clúster	Parámetros de escalado	Comportamiento del escalado
Grupos de instancias Principal: 1 bajo demanda	`UnitType`: instancias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20	Escale entre 1 y 20 instancias o unidades de flota de instancia

Estado inicial del clúster	Parámetros de escalado	Comportamiento del escalado
Tarea: 1 bajo demanda y 1 de spot	`MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	s en los nodos principales mediante el tipo bajo demanda. No se puede escalar en los nodos de tarea.
Flotas de instancias Principal: 1 bajo demanda Tarea: 1 bajo demanda y 1 de spot	UnitType: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	

Escenario 2: escalar únicamente los nodos de tarea

Para escalar únicamente los nodos de tarea, los parámetros de escalado administrado deben cumplir el siguiente requisito:

- El máximo de nodos principales debe ser igual al límite mínimo.

En los siguientes ejemplos, se muestra el escenario de escalado de nodos de tarea únicamente.

Estado inicial del clúster	Parámetros de escalado	Comportamiento del escalado
Grupos de instancias Principal: 2 bajo demanda Tarea: 1 de spot	`UnitType`: instancias `MinimumCapacityUnits` : 2 `MaximumCapacityUnits` : 20	Mantenga los nodos principal es estables en 2 y escale únicamente los nodos de

Estado inicial del clúster	Parámetros de escalado	Comportamiento del escalado
	`MaximumCoreCapacityUnits` : 2	tarea entre 0 y 18 instancias o unidades de flota de instancias. La capacidad entre los límites mínimo y máximo se agrega únicamente a los nodos de tarea.
Flotas de instancias Principal: 2 bajo demanda Tarea: 1 de spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 2 `MaximumCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 2	

Escenario 3: solo las instancias bajo demanda del clúster

Para tener únicamente instancias bajo demanda, el clúster y los parámetros de escalado administrado deben cumplir el siguiente requisito:

- El límite bajo demanda es igual al límite máximo.

  Si no se especifica el límite bajo demanda, el valor del parámetro se establece de forma predeterminada en el límite máximo. El valor predeterminado indica que Amazon EMR escala únicamente las instancias bajo demanda.

Si el máximo de nodos principales es inferior al límite máximo, el parámetro del máximo de nodos principales se puede utilizar para dividir la asignación de capacidad entre los nodos principales y de tarea.

Para habilitar este escenario en un clúster compuesto por grupos de instancias, todos los grupos de nodos del clúster deben usar el tipo de compra bajo demanda durante la configuración inicial.

En los siguientes ejemplos, se muestra el escenario en el que hay instancias bajo demanda en todo el clúster.

Estado inicial del clúster	Parámetros de escalado	Comportamiento del escalado
Grupos de instancias Principal: 1 bajo demanda Tarea: 1 baja demanda	`UnitType`: instancias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	Escale entre 1 y 12 instancias o unidades de flota de instancias en los nodos principales mediante el tipo bajo demanda. Escale la capacidad restante mediante el tipo bajo demanda en los nodos de tarea. No se puede escalar con instancias de spot.
Flotas de instancias Principal: 1 bajo demanda Tarea: 1 baja demanda	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	

Escenario 4: solo las instancias de spot del clúster

Para tener únicamente instancias de spot, los parámetros de escalado administrado deben cumplir el siguiente requisito:

- El límite bajo demanda está establecido en 0.

Si el máximo de nodos principales es inferior al límite máximo, el parámetro del máximo de nodos principales se puede utilizar para dividir la asignación de capacidad entre los nodos principales y de tarea.

Para habilitar este escenario en un clúster compuesto por grupos de instancias, el grupo de instancias principales debe usar la opción de compra de spot durante la configuración inicial. Si no

hay ninguna instancia de spot en el grupo de instancias de las tareas, Escalado administrado de Amazon EMR crea un grupo de tareas que utiliza instancias de spot cuando es necesario.

En los siguientes ejemplos, se muestra el escenario en el que hay instancias de spot en todo el clúster.

Estado inicial del clúster	Parámetros de escalado	Comportamiento del escalado
Grupos de instancias Principal: 1 de spot Tarea: 1 de spot	`UnitType`: instancias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	Escale entre 1 y 20 instancias o unidades de flota de instancias en los nodos principales mediante el tipo de spot. No se puede escalar con el tipo bajo demanda.
Flotas de instancias Principal: 1 de spot Tarea: 1 de spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	

Escenario 5: escalar las instancias bajo demanda en los nodos principales y las instancias de spot en los nodos de tarea

Para escalar las instancias bajo demanda en los nodos principales y las instancias de spot en los nodos de tarea, los parámetros de escalado administrado deben cumplir los siguientes requisitos:

- El límite bajo demanda debe ser igual al máximo de nodos principales.
- Tanto el límite bajo demanda como el máximo de nodos principales deben ser inferiores al límite máximo.

Para habilitar este escenario en un clúster compuesto por grupos de instancias, el grupo de nodos principales debe usar la opción de compra bajo demanda.

En los siguientes ejemplos, se muestra el escenario en el que se escalan las instancias bajo demanda en los nodos principales y las instancias de spot en los nodos de tarea.

Estado inicial del clúster	Parámetros de escalado	Comportamiento del escalado
Grupos de instancias Principal: 1 bajo demanda Tarea: 1 bajo demanda y 1 de spot	`UnitType`: instancias `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	Escale hasta 6 unidades bajo demanda en el nodo principal, ya que ya hay una unidad bajo demanda en el nodo de tarea y el límite máximo de unidades bajo demanda es de 7. A continuación, escale hasta 13 unidades de spot en los nodos de tarea.
Flotas de instancias Principal: 1 bajo demanda Tarea: 1 bajo demanda y 1 de spot	`UnitType`: InstanceFleetUnits `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	

Descripción de las métricas de escalado administrado

Amazon EMR publica métricas de alta resolución con datos en una granularidad de un minuto cuando se habilita el escalado administrado para un clúster. Puede ver los eventos de cada inicio y terminación de cambio de tamaño controlado por el escalado administrado con la consola de Amazon EMR o la consola de Amazon CloudWatch. Las métricas de CloudWatch son fundamentales para que Escalado administrado de Amazon EMR funcione. Le recomendamos que monitorice

de cerca las métricas de CloudWatch para asegurarse de que no falten datos. Para obtener más información sobre cómo configurar las alarmas de CloudWatch para detectar las métricas que faltan, consulte Uso de las alarmas de Amazon CloudWatch. Para obtener más información acerca de los eventos de CloudWatch con Amazon EMR, consulte Supervisar eventos de CloudWatch.

Las siguientes métricas indican la capacidad actual o prevista de un clúster. Estas métricas solo están disponibles cuando el escalado administrado está habilitado. Para los clústeres compuestos por flotas de instancias, las métricas de capacidad del clúster se miden en `Units`. Para los clústeres compuestos por grupos de instancias, las métricas de capacidad del clúster se miden en `Nodes` o en `vCPU` en función del tipo de unidad utilizado en la política de escalado administrado.

Métrica	Descripción
• `TotalUnitsRequested` • `TotalNodesRequested` • `TotalVCPURequested`	El número total previsto de unidades, nodos o vCPU en un clúster según lo determine el escalado administrado. Unidades: recuento
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	El número total actual de unidades, nodos o vCPU disponibles en un clúster en ejecución. Cuando se solicita un cambio de tamaño del clúster, esta métrica se actualizará después de agregar o quitar las nuevas instancias del clúster. Unidades: recuento
• `CoreUnitsRequested` • `CoreNodesRequested` • `CoreVCPURequested`	El número previsto de unidades, nodos o vCPU CORE en un clúster según lo determine el escalado administrado. Unidades: recuento
• `CoreUnitsRunning` •	El número actual de unidades, nodos o vCPU CORE que se ejecutan en un clúster.

Métrica	Descripción
• `CoreNodesRunning` • `CoreVCPURunning`	Unidades: recuento
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	El número previsto de unidades, nodos o vCPU TASK en un clúster según lo determine el escalado administrado. Unidades: recuento
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	El número actual de unidades, nodos o vCPU TASK que se ejecutan en un clúster. Unidades: recuento

Las siguientes métricas indican el estado de uso del clúster y las aplicaciones. Estas métricas están disponibles para todas las características de Amazon EMR, pero se publican con una resolución más alta con datos y una granularidad de un minuto cuando se habilita el escalado administrado para un clúster. Puede comparar las siguientes métricas con las métricas de capacidad del clúster de la tabla anterior para conocer las decisiones de escalado administrado.

Métrica	Descripción
`AppsCompleted`	El número de aplicaciones enviadas a YARN que se han completado. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
`AppsPending`	

Métrica	Descripción
	El número de aplicaciones enviadas a YARN que están en estado pendiente. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
`AppsRunning`	El número de aplicaciones enviadas a YARN que se están ejecutando. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
`ContainerAllocated`	El número de contenedores de recursos asignados por ResourceManager. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
`ContainerPending`	El número de contenedores en la cola que aún no se han asignado. Caso de uso: monitorizar el progreso del clúster Unidades: recuento

Métrica	Descripción
ContainerPendingRatio	La proporción entre contenedores pendientes y contenedores asignados (ContainerPendingRatio = ContainerPending / ContainerAllocated). Si ContainerAllocated = 0, entonces ContainerPendingRatio = ContainerPending. El valor de ContainerPendingRatio representa un número, no un porcentaje. Este valor es útil para escalar recursos del clúster en función del comportamiento de asignación de contenedores. Unidades: recuento
HDFSUtilization	El porcentaje de almacenamiento HDFS usado actualmente. Caso de uso: analizar el rendimiento del clúster Unidades: porcentaje
IsIdle	Indica que un clúster ya no está funcionando, pero sigue activo y acumulando cargos. Se establece en 1 si no se ejecuta ninguna tarea ni ningún trabajo; en caso contrario, se establece en 0. Este valor se comprueba a intervalos de cinco minutos, y un valor de 1 indica que el clúster estaba inactivo cuando se comprobó, no que estuvo inactivo durante los cinco minutos. Para evitar falsos positivos, debe activar una alarma cuando este valor sea 1 durante más de una comprobación consecutiva de cinco minutos. Por ejemplo, puede activar una alarma cuando este valor sea 1 durante treinta minutos o más. Caso de uso: monitorizar el rendimiento del clúster Unidades: booleano

Métrica	Descripción
`MemoryAvailableMB`	La cantidad de memoria disponible para asignar. Caso de uso: monitorizar el progreso del clúster Unidades: recuento
`MRActiveNodes`	El número de nodos que se están ejecutando actualmente en tareas o trabajos de MapReduce. Equivalente a la métrica YARN `mapred.resourcemanager.NoOfActiveNodes` . Caso de uso: monitorizar el progreso del clúster Unidades: recuento
`YARNMemoryAvailablePercentage`	El porcentaje de memoria restante disponible para YARN (YARNMemoryAvailablePercentage = MemoryAvailableMB / MemoryTotalMB). Este valor es útil para escalar recursos del clúster en función del uso de memoria de YARN. Unidades: porcentaje

Diagramación de métricas de escalado administrado

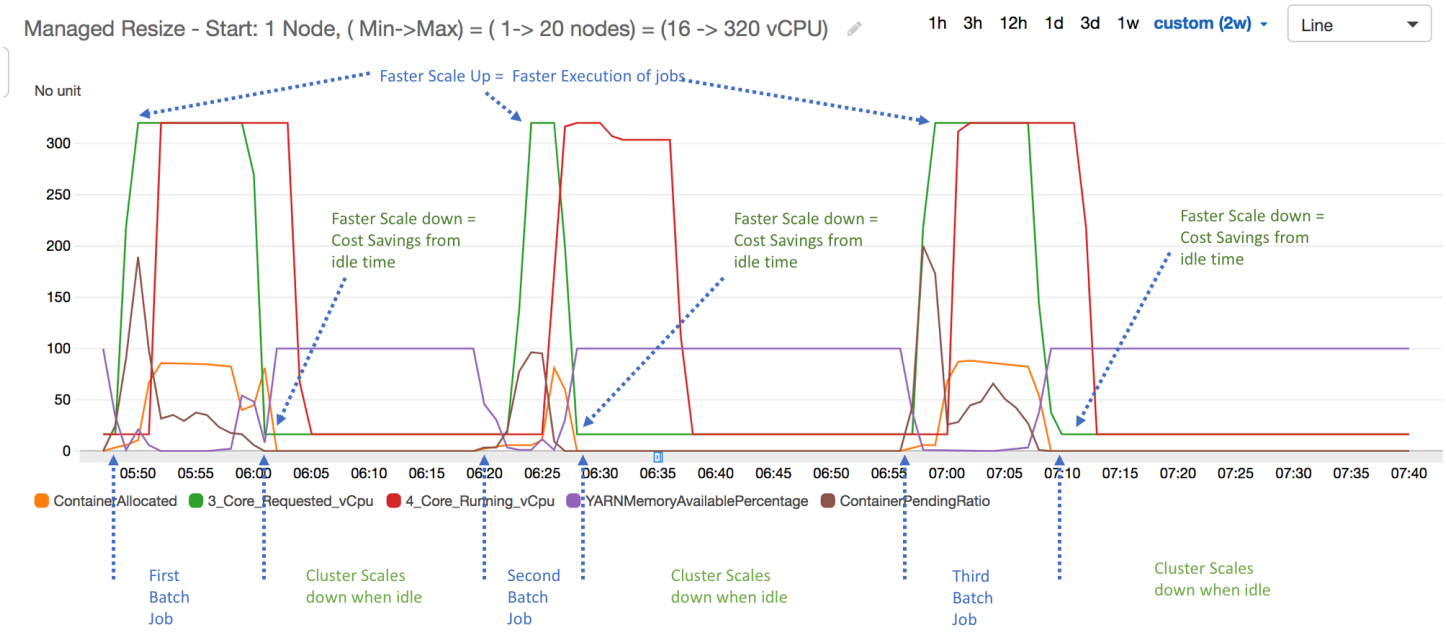
Puede diagramar las métricas para ver los patrones de carga de trabajo del clúster y las decisiones de escalado correspondientes tomadas por Escalado administrado de Amazon EMR, como se muestra en los pasos siguientes.

Para diagramar métricas de escalado administrado en la consola de CloudWatch

1. Abra la consola de CloudWatch.
2. En el panel de navegación, seleccione Amazon EMR. Puede buscar el identificador del clúster que desea monitorizar.
3. Desplácese hacia abajo hasta la métrica que desea representar gráficamente. Abra una métrica para mostrar el gráfico.

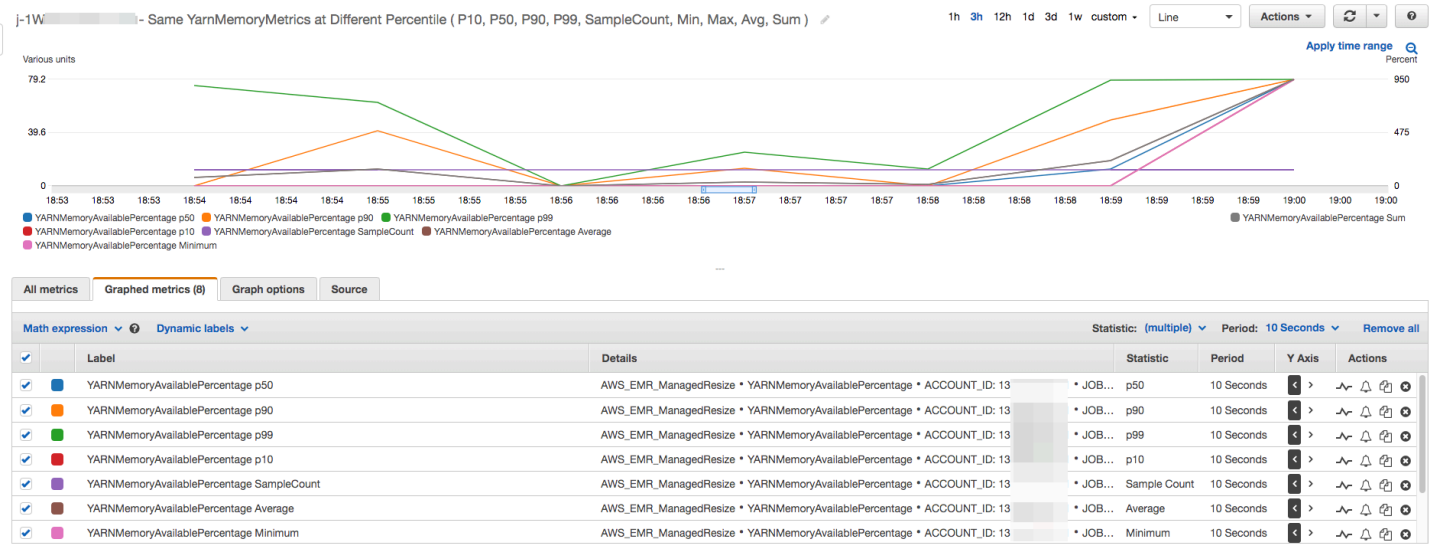
4. Para representar gráficamente una o varias métricas, seleccione la casilla de verificación junto a cada métrica.

En el siguiente ejemplo, se ilustra la actividad de Escalado administrado de Amazon EMR de un clúster. El gráfico muestra tres periodos de reducción de capacidad automática, que ahorran costos cuando hay una carga de trabajo menos activa.



Todas las métricas de capacidad y uso del clúster se publican a intervalos de un minuto. La información estadística adicional también se asocia a cada dato de un minuto, lo que le permite diagramar varias funciones como `Percentiles`, `Min`, `Max`, `Sum`, `Average`, `SampleCount`.

Por ejemplo, el siguiente gráfico muestra la misma métrica `YARNMemoryAvailablePercentage` en percentiles diferentes, P10, P50, P90, P99, junto con `Sum`, `Average`, `Min`, `SampleCount`.



Uso del escalado automático con una política personalizada para grupos de instancias

El escalado automático con una política personalizada en las versiones 4.0 y posteriores de Amazon EMR le permite escalar o reducir horizontalmente los nodos principales y los nodos de tarea en función de una métrica de CloudWatch y otros parámetros especificados en una política de escalado. El escalado automático con una política personalizada está disponible con la configuración de grupos de instancias, pero no con las flotas de instancias. Para obtener más información sobre los grupos de instancias y las flotas de instancias, consulte Crear un clúster con flotas de instancias o grupos de instancias uniformes.

> **Note**
>
> Para utilizar la característica de escalado automático con una política personalizada en Amazon EMR, debe establecer en `true` el parámetro `VisibleToAllUsers` al crear un clúster. Para obtener más información, consulte SetVisibleToAllUsers.

La política de escalado forma parte de la configuración de un grupo de instancias. Puede especificar una política durante la configuración de un grupo de instancias o modificando un grupo de instancias en un clúster existente, incluso cuando dicho grupo de instancias está activo. Cada grupo de instancia de un clúster, excepto el grupo de instancias principales, puede tener su propia política de escalado, que se compone de reglas de escalado horizontal y reducción horizontal. Las reglas de

escalado ascendente y descendente se pueden configurar de manera independiente, con distintos parámetros para cada regla.

Puede configurar las políticas de escalado con la AWS Management Console, la AWS CLI o la API de Amazon EMR. Al utilizar la AWS CLI o la API de Amazon EMR, debe especificar la política de escalado en formato JSON. Además, si utiliza la AWS CLI o la de API de Amazon EMR, puede especificar métricas de CloudWatch personalizadas. Las métricas personalizadas no se pueden seleccionar con la AWS Management Console. Al crear inicialmente una política de escalado con la consola, se preconfigura una política predeterminada adecuada para muchas aplicaciones para ayudarle a comenzar. Puede eliminar o modificar las reglas predeterminadas.

Aunque el escalado automático le permite ajustar la capacidad del clúster de EMR sobre la marcha, debe tener en cuenta los requisitos de carga de trabajo de referencia y planificar las configuraciones del nodo y de los grupos de instancias. Para obtener más información, consulte Directrices de configuración del clúster.

> **Note**
>
> Para la mayoría de las cargas de trabajo, es deseable la configuración de las reglas de escalado ascendente y descendente para optimizar el uso de los recursos. Definir una regla sin la otra significa que tendrá que cambiar manualmente el tamaño del recuento de instancias después de una actividad de escalado. En otras palabras, esto configura una política de escalado ascendente o descendente automática "unidireccional" con un restablecimiento manual.

Creación del rol de IAM; para el escalado automático

El escalado automático en Amazon EMR requiere un rol de IAM con permisos para agregar y terminar instancias cuando se activan las actividades de escalado. Para este fin se dispone de un rol predeterminado, `EMR_AutoScaling_DefaultRole`, configurado con la política de confianza y la política de roles adecuadas. Al crear un clúster con una política de escalado por primera vez utilizando la AWS Management Console, Amazon EMR crea el rol predeterminado y asocia la política administrada predeterminada para los permisos, `AmazonElasticMapReduceforAutoScalingRole`.

Al crear un clúster con un política de escalado automático con la AWS CLI, primero debe asegurarse de que existe el rol de IAM predeterminado o de que dispone de un rol de IAM personalizado con

una política asociada que proporciona los permisos necesarios. Para crear el rol predeterminado, puede ejecutar el comando `create-default-roles` antes de crear un clúster. A continuación, puede especificar la opción `--auto-scaling-role EMR_AutoScaling_DefaultRole` al crear el clúster. También puede crear un rol de escalado automático personalizado y luego especificarlo al crear un clúster, por ejemplo, `--auto-scaling-role` *`MyEMRAutoScalingRole`*. Si crea un rol de escalado automático personalizado para Amazon EMR, le recomendamos que base las políticas de permisos de dicho rol en la política administrada. Para obtener más información, consulte Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS.

Descripción de las reglas de escalado automático

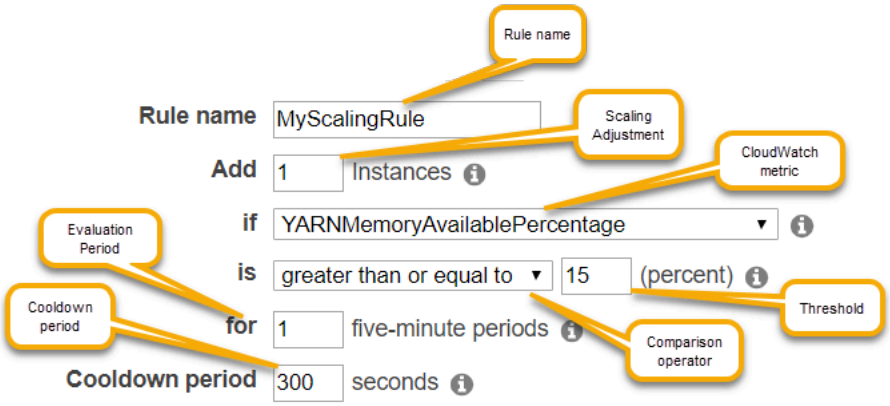
Cuando una regla de escalado horizontal activa una actividad de escalado para un grupo de instancias, las instancias de Amazon EC2 se agregan al grupo de instancias de acuerdo con sus reglas. Aplicaciones como Apache Spark, Apache Hive y Presto pueden utilizar los nodos nuevos en cuanto la instancia Amazon EC2 entre en el estado `InService`. También puede configurar una regla de escalado descendente que termina instancias y elimina nodos. Para obtener más información sobre el ciclo de vida de las instancias de Amazon EC2 que se escalan automáticamente, consulte Ciclo de vida de Auto Scaling en la Guía del usuario de Amazon EC2 Auto Scaling.

Puede configurar cómo termina un clúster las instancias de Amazon EC2. Puede optar por terminar en el límite de instancia de Amazon EC2 por hora para la facturación o una vez completadas las tareas. Esta configuración se aplica tanto al escalado automático como a las operaciones manuales de cambio de tamaño. Para obtener más información acerca de esta configuración, consulte Reducción vertical del clúster.

Los siguientes parámetros para cada regla de una política determinan el comportamiento de escalado automático.

> **Note**
>
> Los parámetros que se enumeran aquí se basan en la AWS Management Console para Amazon EMR. Al utilizar la AWS CLI o la API de Amazon EMR, hay disponibles más opciones de configuración avanzada. Para obtener más información sobre las opciones avanzadas, consulte SimpleScalingPolicyConfiguration en la Referencia de la API de Amazon EMR.

- Número máximo de instancias y número mínimo de instancias. La limitación Máximo de instancias especifica el número máximo de instancias de Amazon EC2 que puede haber en el grupo de instancias y se aplica a todas las reglas de escalado horizontal. Del mismo modo, la limitación Mínimo de instancias especifica el número mínimo de instancias de Amazon EC2 y se aplica a todas las reglas de reducción horizontal.
- El Rule name (Nombre de la regla), que debe ser único en la política.
- El ajuste de escalado, que determina el número de instancias EC2 que desea añadir (para las reglas de escalado ascendente) o terminar (para las reglas de escalado descendente) durante la actividad de escalado activada por la regla.
- La CloudWatch metric (Métrica de CloudWatch), que se vigila para detectar condiciones de alarma.
- Un operador de comparaciones, que se utiliza para comparar la métrica de CloudWatch con el valor de Threshold (Umbral) y determinar una condición de disparador.
- Un periodo de evaluación, en incrementos de cinco minutos, durante el cual la métrica de CloudWatch debe estar en condición de disparador antes de activar la actividad de escalado.
- Un Cooldown period (período de recuperación) en segundos, que determina la cantidad de tiempo que debe transcurrir entre una actividad de escalado iniciada por una regla y el inicio de la próxima actividad de escalado, con independencia de la regla que la activa. Cuando un grupo de instancias ha finalizado una actividad de escalado y alcanza su estado posterior al escalado, el periodo de recuperación ofrece una oportunidad para las métricas de CloudWatch que podrían disparar actividades de escalado siguientes para estabilizar. Para obtener más información, consulte Periodos de recuperación de Auto Scaling en la Guía del usuario de Amazon EC2 Auto Scaling.



Condiciones y limitaciones

- Las métricas de Amazon CloudWatch son fundamentales para que Escalado automático de Amazon EMR funcione. Le recomendamos que monitorice de cerca las métricas de Amazon

CloudWatch para asegurarse de que no falten datos. Para obtener más información sobre cómo configurar las alarmas de Amazon CloudWatch para detectar las métricas que faltan, consulte Uso de las alarmas de Amazon CloudWatch.

- La sobreutilización de los volúmenes de EBS puede provocar problemas de Escalado administrado. Le recomendamos que supervise de cerca el uso de los volúmenes de EBS para asegurarse de que el volumen de EBS esté por debajo del 90 % de utilización. Consulte Almacenamiento de instancias para obtener información sobre cómo especificar volúmenes de EBS adicionales.
- El escalado automático con una política personalizada en las versiones 5.18 a 5.28 de Amazon EMR puede experimentar un error de escalado debido a la ausencia intermitente de datos en las métricas de Amazon CloudWatch. Para mejorar el escalado automático, le recomendamos que utilice las versiones más recientes de Amazon EMR. También puede ponerse en contacto con AWS Support para solicitar un parche si tiene que utilizar una versión de Amazon EMR que esté entre la 5.18 y la 5.28.

Uso de la AWS Management Console para configurar el escalado automático

Al crear un clúster, se configura una política de escalado para grupos de instancias con las opciones avanzadas de configuración de clúster. También puede crear o modificar una política de escalado para un grupo de instancias en servicio modificando los grupos de instancias en la configuración de Hardware de un clúster existente.

> **Note**
>
> La nueva consola de Amazon EMR (https://console.aws.amazon.com/emr) utiliza el escalado administrado en lugar del escalado automático. Para utilizar el escalado automático, asegúrese de haber iniciado sesión en la consola antigua en https://console.aws.amazon.com/elasticmapreduce.

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Si está creando un clúster, en la consola de Amazon EMR, seleccione Crear clúster, seleccione Ir a las opciones avanzadas, seleccione las opciones en Paso 1: Software y pasos y, a continuación, acceda a Paso 2: Configuración de hardware.

- o bien -

Si está modificando un grupo de instancias en un clúster en ejecución, seleccione el clúster en la lista de clústeres y, a continuación, expanda la sección Hardware.

3. En la sección Opción de escalado y aprovisionamiento del clúster, seleccione Habilitar el escalado del clúster. A continuación, seleccione Crear una política de escalado automático personalizada.

   En la tabla Políticas de escalado automático personalizadas, haga clic en el icono de lápiz que aparece en la fila del grupo de instancias que desea configurar. Se abre la pantalla de reglas de Auto Scaling.

4. Escriba el valor de Maximum instances (Número máximo de instancias) que desea que contenga el grupo de instancias después del escalado ascendente y el valor de Minimum instances (Número mínimo de instancias) que desea que contenga el grupo de instancias después del escalado descendente.

5. Haga clic en el icono de lápiz para editar los parámetros de regla, haga clic en la X para eliminar una regla de la política y, a continuación, haga clic en Add rule (Añadir regla) para añadir reglas adicionales.

6. Elija los parámetros de la configuración, tal como se describió anteriormente en este tema. Para obtener descripciones de las métricas de CloudWatch disponibles para Amazon EMR, consulte Métricas y dimensiones de Amazon EMR en la Guía del usuario de Amazon CloudWatch.

Uso de la AWS CLI para configurar el escalado automático

Puede utilizar comandos de la AWS CLI para Amazon EMR si desea configurar el escalado automático al crear un clúster y al crear un grupo de instancias. Puede utilizar una sintaxis abreviada, especificando la configuración JSON insertada dentro de los comandos pertinentes o puede hacer referencia a un archivo que contenga la configuración JSON. También puede aplicar una política de escalado automático a un grupo de instancias existente y eliminar la política de escalado automático que se aplicó anteriormente. Además, puede recuperar detalles de una configuración de política de escalado desde un clúster en ejecución.

> **Important**
>
> Al crear un clúster con una política de escalado automático, es necesario utilizar el comando `--auto-scaling-role` *`MyAutoScalingRole`* para especificar el rol de IAM para el

escalado automático. El rol predeterminado es *EMR_AutoScaling_DefaultRole* y se puede crear mediante el comando `create-default-roles`. El rol solo se puede añadir cuando se crea el clúster y no se puede añadir a un clúster existente.

Para obtener una descripción detallada de los parámetros disponibles a la hora de configurar una política de escalado automático, consulte PutAutoScalingPolicy en la Referencia de la API de Amazon EMR.

Creación de un clúster con una política de escalado automático aplicada a un grupo de instancias

Puede especificar una configuración de escalado automático dentro de la opción `--instance-groups` del comando `aws emr create-cluster`. El siguiente ejemplo ilustra un comando create-cluster donde se proporciona una política insertada de escalado automático para el grupo de instancias secundarias. El comando crea una configuración de escalado equivalente a la política predeterminada de escalado horizontal que aparece al crear una política de escalado automático con la AWS Management Console para Amazon EMR. Por razones de brevedad, no se muestra una política de escalado descendente. No recomendamos la creación de una regla de escalado ascendente sin una regla de escalado descendente.

```
aws emr create-cluster --release-label emr-5.2.0 --service-role
 EMR_DefaultRole --ec2-attributes InstanceProfile=EMR_EC2_DefaultRole
 --auto-scaling-role EMR_AutoScaling_DefaultRole  --instance-groups
 Name=MyMasterIG,InstanceGroupType=MASTER,InstanceType=m5.xlarge,InstanceCount=1
 'Name=MyCoreIG,InstanceGroupType=CORE,InstanceType=m5.xlarge,InstanceCount=2,AutoScalingPolicy
scale-out,Description=Replicates the default scale-out rule in the
 console.,Action={SimpleScalingPolicyConfiguration={AdjustmentType=CHANGE_IN_CAPACITY,ScalingAd
ElasticMapReduce,Period=300,Statistic=AVERAGE,Threshold=15,Unit=PERCENT,Dimensions=[{Key=JobFlo
```

El siguiente comando ilustra el uso de la línea de comandos para proporcionar la definición de política de escalado automático como parte de un archivo de configuración de grupo de instancias denominado *instancegroupconfig.json*.

```
aws emr create-cluster --release-label emr-5.2.0 --service-role EMR_DefaultRole --ec2-
attributes InstanceProfile=EMR_EC2_DefaultRole --instance-groups file://your/path/to/
instancegroupconfig.json --auto-scaling-role EMR_AutoScaling_DefaultRole
```

Con el contenido del archivo de configuración siguiente:

```
[
{
  "InstanceCount": 1,
  "Name": "MyMasterIG",
  "InstanceGroupType": "MASTER",
  "InstanceType": "m5.xlarge"
},
{
  "InstanceCount": 2,
  "Name": "MyCoreIG",
  "InstanceGroupType": "CORE",
  "InstanceType": "m5.xlarge",
  "AutoScalingPolicy":
    {
     "Constraints":
      {
       "MinCapacity": 2,
       "MaxCapacity": 10
      },
     "Rules":
     [
      {
       "Name": "Default-scale-out",
       "Description": "Replicates the default scale-out rule in the console for YARN
 memory.",
       "Action":{
        "SimpleScalingPolicyConfiguration":{
          "AdjustmentType": "CHANGE_IN_CAPACITY",
          "ScalingAdjustment": 1,
          "CoolDown": 300
        }
       },
       "Trigger":{
        "CloudWatchAlarmDefinition":{
          "ComparisonOperator": "LESS_THAN",
          "EvaluationPeriods": 1,
          "MetricName": "YARNMemoryAvailablePercentage",
          "Namespace": "AWS/ElasticMapReduce",
          "Period": 300,
          "Threshold": 15,
          "Statistic": "AVERAGE",
          "Unit": "PERCENT",
```

```
          "Dimensions":[
             {
               "Key" : "JobFlowId",
               "Value" : "${emr.clusterId}"
             }
          ]
        }
       }
      }
     ]
   }
}
]
```

Agregar un grupo de instancias con una política de escalado automático a un clúster

Puede especificar una configuración de política de escalado con la opción `--instance-groups` con el comando `add-instance-groups` de la misma forma que al utilizar `create-cluster`. En el siguiente ejemplo se utiliza una referencia a un archivo JSON, *`instancegroupconfig.json`*, con la configuración de grupo de instancias.

```
aws emr add-instance-groups --cluster-id j-1EKZ3TYEVF1S2 --instance-groups file://your/
path/to/instancegroupconfig.json
```

Aplicar una política de escalado automático a un grupo de instancias existente o modificar una política aplicada

Utilice el comando `aws emr put-auto-scaling-policy` para aplicar una política de escalado automático a un grupo de instancias existente. El grupo de instancias debe formar parte de un clúster que utilice el rol de IAM; de escalado automático. En el siguiente ejemplo se utiliza una referencia a un archivo JSON, *`autoscaleconfig.json`*, que especifica la configuración de política de escalado.

```
aws emr put-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07 --auto-scaling-policy file://your/path/to/autoscaleconfig.json
```

El contenido del archivo `autoscaleconfig.json`, que define la misma regla de escalado ascendente, tal y como se muestra en el ejemplo anterior, se muestra a continuación.

```
{
          "Constraints": {
```

```
                "MaxCapacity": 10,
                "MinCapacity": 2
        },
        "Rules": [{
                "Action": {
                        "SimpleScalingPolicyConfiguration": {
                                "AdjustmentType": "CHANGE_IN_CAPACITY",
                                "CoolDown": 300,
                                "ScalingAdjustment": 1
                        }
                },
                "Description": "Replicates the default scale-out rule in the console
for YARN memory",
                "Name": "Default-scale-out",
                "Trigger": {
                        "CloudWatchAlarmDefinition": {
                                "ComparisonOperator": "LESS_THAN",
                                "Dimensions": [{
                                        "Key": "JobFlowID",
                                        "Value": "${emr.clusterID}"
                                }],
                                "EvaluationPeriods": 1,
                                "MetricName": "YARNMemoryAvailablePercentage",
                                "Namespace": "AWS/ElasticMapReduce",
                                "Period": 300,
                                "Statistic": "AVERAGE",
                                "Threshold": 15,
                                "Unit": "PERCENT"
                        }
                }
        }]
 }
```

Eliminación de una política de escalado automático desde un grupo de instancias

```
aws emr remove-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07
```

Recuperación de una configuración de política de escalado automático

El comando `describe-cluster` recupera la configuración de la política en el bloque InstanceGroup. Por ejemplo, el comando siguiente recupera la configuración para el clúster con un ID de clúster de `j-1CWOHP4PI30VJ`.

```
aws emr describe-cluster --cluster-id j-1CWOHP4PI30VJ
```

El comando produce el siguiente resultado de ejemplo.

```
{
    "Cluster": {
        "Configurations": [],
        "Id": "j-1CWOHP4PI30VJ",
        "NormalizedInstanceHours": 48,
        "Name": "Auto Scaling Cluster",
        "ReleaseLabel": "emr-5.2.0",
        "ServiceRole": "EMR_DefaultRole",
        "AutoTerminate": false,
        "TerminationProtected": true,
        "MasterPublicDnsName": "ec2-54-167-31-38.compute-1.amazonaws.com",
        "LogUri": "s3n://aws-logs-232939870606-us-east-1/elasticmapreduce/",
        "Ec2InstanceAttributes": {
            "Ec2KeyName": "performance",
            "AdditionalMasterSecurityGroups": [],
            "AdditionalSlaveSecurityGroups": [],
            "EmrManagedSlaveSecurityGroup": "sg-09fc9362",
            "Ec2AvailabilityZone": "us-east-1d",
            "EmrManagedMasterSecurityGroup": "sg-0bfc9360",
            "IamInstanceProfile": "EMR_EC2_DefaultRole"
        },
        "Applications": [
            {
                "Name": "Hadoop",
                "Version": "2.7.3"
            }
        ],
        "InstanceGroups": [
            {
                "AutoScalingPolicy": {
                    "Status": {
                        "State": "ATTACHED",
                        "StateChangeReason": {
                            "Message": ""
                        }
                    },
```

```
                    "Constraints": {
                        "MaxCapacity": 10,
                        "MinCapacity": 2
                    },
                    "Rules": [
                        {
                            "Name": "Default-scale-out",
                            "Trigger": {
                                "CloudWatchAlarmDefinition": {
                                    "MetricName": "YARNMemoryAvailablePercentage",
                                    "Unit": "PERCENT",
                                    "Namespace": "AWS/ElasticMapReduce",
                                    "Threshold": 15,
                                    "Dimensions": [
                                        {
                                            "Key": "JobFlowId",
                                            "Value": "j-1CWOHP4PI30VJ"
                                        }
                                    ],
                                    "EvaluationPeriods": 1,
                                    "Period": 300,
                                    "ComparisonOperator": "LESS_THAN",
                                    "Statistic": "AVERAGE"
                                }
                            },
                            "Description": "",
                            "Action": {
                                "SimpleScalingPolicyConfiguration": {
                                    "CoolDown": 300,
                                    "AdjustmentType": "CHANGE_IN_CAPACITY",
                                    "ScalingAdjustment": 1
                                }
                            }
                        },
                        {
                            "Name": "Default-scale-in",
                            "Trigger": {
                                "CloudWatchAlarmDefinition": {
                                    "MetricName": "YARNMemoryAvailablePercentage",
                                    "Unit": "PERCENT",
                                    "Namespace": "AWS/ElasticMapReduce",
                                    "Threshold": 75,
                                    "Dimensions": [
                                        {
```

```
                                    "Key": "JobFlowId",
                                    "Value": "j-1CWOHP4PI30VJ"
                                }
                            ],
                            "EvaluationPeriods": 1,
                            "Period": 300,
                            "ComparisonOperator": "GREATER_THAN",
                            "Statistic": "AVERAGE"
                        }
                    },
                    "Description": "",
                    "Action": {
                        "SimpleScalingPolicyConfiguration": {
                            "CoolDown": 300,
                            "AdjustmentType": "CHANGE_IN_CAPACITY",
                            "ScalingAdjustment": -1
                        }
                    }
                }
            ]
        },
        "Configurations": [],
        "InstanceType": "m5.xlarge",
        "Market": "ON_DEMAND",
        "Name": "Core - 2",
        "ShrinkPolicy": {},
        "Status": {
            "Timeline": {
                "CreationDateTime": 1479413437.342,
                "ReadyDateTime": 1479413864.615
            },
            "State": "RUNNING",
            "StateChangeReason": {
                "Message": ""
            }
        },
        "RunningInstanceCount": 2,
        "Id": "ig-3M16XBE8C3PH1",
        "InstanceGroupType": "CORE",
        "RequestedInstanceCount": 2,
        "EbsBlockDevices": []
    },
    {
        "Configurations": [],
```

```
                "Id": "ig-OP62I28NSE8M",
                "InstanceGroupType": "MASTER",
                "InstanceType": "m5.xlarge",
                "Market": "ON_DEMAND",
                "Name": "Master - 1",
                "ShrinkPolicy": {},
                "EbsBlockDevices": [],
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "CreationDateTime": 1479413437.342,
                        "ReadyDateTime": 1479413752.088
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "RunningInstanceCount": 1
            }
        ],
        "AutoScalingRole": "EMR_AutoScaling_DefaultRole",
        "Tags": [],
        "BootstrapActions": [],
        "Status": {
            "Timeline": {
                "CreationDateTime": 1479413437.339,
                "ReadyDateTime": 1479413863.666
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Cluster ready after last step completed."
            }
        }
    }
}
```

Cambiar manualmente el tamaño de un clúster en ejecución

Puede agregar y eliminar instancias en los grupos de instancias principales y de tarea, así como en las flotas de instancias de un clúster en ejecución mediante la AWS Management Console, la AWS CLI o la API de Amazon EMR. Si un clúster utiliza grupos de instancias, es necesario cambiar de

forma explícita el recuento de instancias. Si el clúster utiliza flotas de instancias, puede cambiar las unidades de destino en las instancias bajo demanda y las instancias de spot. A continuación, la flota de instancias añadirá y eliminará instancias para satisfacer los requisitos del nuevo destino. Para obtener más información, consulte Opciones de flota de instancias. Las aplicaciones pueden utilizar las instancias de Amazon EC2 recién aprovisionadas para alojar nodos en cuanto las instancias estén disponibles. Cuando se eliminan instancias, Amazon EMR finaliza las tareas de forma que no se interrumpan los trabajos y se protejan frente a la pérdida de datos. Para obtener más información, consulte Terminación al completar las tareas.

Cambiar el tamaño de un clúster con la consola

Puede utilizar la consola de Amazon EMR para cambiar el tamaño de un clúster en ejecución.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para cambiar el recuento de instancias de un clúster existente con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2, en el panel de navegación izquierdo, elija Clústeres y seleccione el clúster que desee actualizar. El clúster debe estar en ejecución; no puede cambiar el tamaño de un clúster que se esté aprovisionando o se haya terminado.
3. En la pestaña Instancias de la página de detalles del clúster, observe el panel Grupos de instancias.
4. Para cambiar el tamaño de un grupo de instancias existente, selecciona el botón de opción situado junto al grupo de instancias principales o de tarea cuyo tamaño desee cambiar y, a continuación, seleccione Cambiar el tamaño del grupo de instancias. Especifique el nuevo número de instancias para el grupo de instancias y, a continuación, seleccione Cambiar tamaño.

> **Note**
>
> Si decide reducir el tamaño de un grupo de instancias en ejecución, Amazon EMR seleccionará de forma inteligente las instancias que desee eliminar del grupo para reducir al mínimo la pérdida de datos. Para tener un control más detallado de la acción de cambio de tamaño, puede seleccionar el ID del grupo de instancias, seleccionar las instancias que desea eliminar y, a continuación, utilizar la opción Terminar. Para obtener más información sobre el comportamiento inteligente de reducción vertical, consulte Reducción vertical del clúster.

5. Si desea cancelar la acción de cambio de tamaño, puede seleccionar el botón de opción de un grupo de instancias con el estado Cambiando el tamaño y, a continuación, seleccionar Detener el cambio de tamaño en la lista de acciones.
6. Para agregar uno o más grupos de instancias de tarea al clúster en respuesta al aumento de la carga de trabajo, seleccione Agregar grupo de instancias de tareas en la lista de acciones. Seleccione el tipo de instancia de Amazon EC2, introduzca el número de instancias del grupo de tareas y, a continuación, seleccione Agregar grupo de instancias de tareas para volver al panel Grupos de instancias de su clúster.

Old console

Para cambiar el recuento de instancias de un clúster existente con la consola antigua

1. En la página Cluster List (Lista de clústeres), elija un clúster para cambiarlo de tamaño.
2. En la página Cluster Details (Detalles del clúster), elija Hardware.
3. Si el clúster utiliza grupos de instancias, elija Resize (Cambiar tamaño) en la columna Instance count (Recuento de instancias) para el grupo de instancias cuyo tamaño desea cambiar, escriba un nuevo recuento de instancias y, a continuación, seleccione la marca de verificación verde.

   -O BIEN-

   Si el clúster utiliza flotas de instancias, seleccione Cambiar tamaño en la columna Capacidad aprovisionada, escriba valores nuevos para Unidades bajo demanda y Unidades de spot y, a continuación, seleccione Cambiar tamaño.

Cuando se realiza un cambio en el número de nodos, se actualiza el Status (Estado) de las actualizaciones del grupo de instancias. Cuando finaliza el cambio solicitado, el Status (Estado) es Running (En ejecución).

Cambiar el tamaño de un clúster con la AWS CLI

Puede utilizar la AWS CLI para cambiar el tamaño de un clúster en ejecución. Puede aumentar o reducir el número de nodos de tareas y puede aumentar el número de nodos secundarios en un clúster en ejecución. También es posible finalizar una instancia en el grupo de instancias principales con la AWS CLI o la API. Esto debe hacerse con precaución. Apagar una instancia en el grupo de instancias secundarias conlleva un riesgo de pérdida de datos y la instancia no se sustituye automáticamente.

Además de cambiar el tamaño de los grupos se instancias secundarias y de tareas, también puede agregar uno o varios grupos de instancias de tareas a un clúster en ejecución con la AWS CLI.

Para cambiar el tamaño de un clúster cambiando el recuento de instancias con la AWS CLI

Puede agregar instancias al grupo principal o al grupo de tareas, así como eliminar instancias del grupo de tareas, utilizando el subcomando `modify-instance-groups` de la AWS CLI con el parámetro `InstanceCount`. Para agregar instancias a los grupos de instancias principales o de tareas, aumente el `InstanceCount`. Para reducir el número de instancias en el grupo de tareas, reduzca el `InstanceCount`. Al cambiar el recuento de instancias del grupo de tareas a 0 se eliminan todas las instancias, pero no el grupo de instancias.

- Para aumentar el número de instancias del grupo de instancias de tareas de 3 a 4, escriba el siguiente comando y sustituya *ig-31JXXXXXXBT0* por el ID del grupo de instancias.

  ```
  aws emr modify-instance-groups --instance-groups
   InstanceGroupId=ig-31JXXXXXXBT0,InstanceCount=4
  ```

  Para recuperar el `InstanceGroupId`, utilice el subcomando `describe-cluster`. La salida es un objeto JSON denominado `Cluster` que contiene el ID de cada grupo de instancias. Para utilizar este comando, necesita el ID del clúster (que puede recuperar con el comando `aws emr list-clusters` o la consola). Para recuperar el ID del grupo de instancias, escriba el siguiente comando y sustituya *j-2AXXXXXXGAPLF* por el ID del clúster.

  ```
  aws emr describe-cluster --cluster-id j-2AXXXXXXGAPLF
  ```

Con la AWS CLI, también es posible terminar una instancia en el grupo de instancias principales con el subcomando `--modify-instance-groups`.

> **Warning**
>
> Debe prestarse especial atención cuando se especifique `EC2InstanceIdsToTerminate`. Las instancias se terminan de forma inmediata, independientemente del estado de las aplicaciones que se ejecutan en ellas y de que la instancia no se sustituya automáticamente. Esto ocurre independientemente de la configuración de Scale down behavior (Comportamiento de escalado descendente) del clúster. Terminar una instancia de esta forma conlleva el riesgo de pérdida de datos y de un comportamiento del clúster imprevisible.

Para terminar una instancia específica, necesita el ID del grupo de instancias (devuelto por el subcomando `aws emr describe-cluster --cluster-id`) y el ID de la instancia (devuelto por el subcomando `aws emr list-instances --cluster-id`); una vez que consiga dichos ID, escriba el comando siguiente, sustituya *`ig-6RXXXXXX07SA`* por el ID del grupo de instancias y sustituya *`i-f9XXXXf2`* por el ID de la instancia.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-6RXXXXXX07SA,EC2InstanceIdsToTerminate=i-f9XXXXf2
```

Para más información sobre el uso de comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Para cambiar el tamaño de un clúster agregando grupos de instancias de tareas con la AWS CLI

La AWS CLI le permite agregar entre 1 y 48 grupos de instancias de tareas a un clúster con el subcomando `--add-instance-groups`. Los grupos de instancias de tareas solo se pueden agregar a un clúster que contenga un grupo de instancias principales y un grupo de instancias de núcleo. Con la AWS CLI, puede agregar hasta cinco grupos de instancias de tareas cada vez que utilice el subcomando `--add-instance-groups`.

1. Para añadir un único grupo de instancias de tareas a un clúster, escriba el siguiente comando y sustituya *`j-JXBXXXXXX37R`* por el ID del clúster.

```
aws emr add-instance-groups --cluster-id j-JXBXXXXXX37R --instance-groups
 InstanceCount=6,InstanceGroupType=task,InstanceType=m5.xlarge
```

2. Para añadir varios grupos de instancias de tareas a un clúster, escriba el siguiente comando y sustituya *j-JXBXXXXXX37R* por el ID del clúster. Puede añadir hasta cinco grupos de instancias de tareas en un solo comando.

```
aws emr add-instance-groups --cluster-id j-JXBXXXXXX37R --instance-
groups InstanceCount=6,InstanceGroupType=task,InstanceType=m5.xlarge
 InstanceCount=10,InstanceGroupType=task,InstanceType=m5.xlarge
```

   Para más información sobre el uso de comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Interrupción de un cambio de tamaño

Si utiliza la versión 4.1.0 o posterior de Amazon EMR, puede realizar un cambio de tamaño en medio de una operación de cambio de tamaño existente. Además, puede parar una solicitud de cambio de tamaño enviada previamente o enviar una nueva solicitud para anular una solicitud anterior sin necesidad de esperar a que finalice. También puede detener un cambio de tamaño existente desde la consola o a través de la llamada a la API `ModifyInstanceGroups` con el recuento actual como el recuento de destino del clúster.

La siguiente captura de pantalla muestra un grupo de instancias de tareas que se está cambiando de tamaño, pero que puede interrumpirse eligiendo Stop (Detener).

Para interrumpir un cambio de tamaño con la AWS CLI

Puede utilizar la AWS CLI para detener un cambio de tamaño con el subcomando `modify-instance-groups`. Supongamos que tiene seis instancias en un grupo de instancias y que desea aumentarlo a 10. Más tarde decide que desea cancelar esta solicitud:

- La solicitud inicial:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=10
```

La segunda solicitud para detener la primera solicitud:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=6
```

> **Note**
>
> Dado que este proceso es asíncrono, es posible que vea que los recuentos de instancias cambian con respecto las solicitudes de la API previas antes de que se realicen las solicitudes siguientes. En el caso de disminución, es posible que si tiene el trabajo en ejecución en los nodos, el grupo de instancias podría no reducirse hasta que los nodos hayan completado su trabajo.

Estado de suspensión

Un grupo de instancias pasa a estado suspendido si se encuentra demasiados errores al intentar iniciar los nuevos nodos del clúster. Por ejemplo, si los nodos nuevos devuelven error repetidamente al llevar a cabo acciones de arranque, el grupo de instancias pasa al estado SUSPENDED, en lugar de aprovisionar de forma continua nuevos nodos. Después de haber resuelto el error subyacente, restablezca el número deseado de nodos en el grupo de instancias del clúster y, a continuación, el grupo de instancias reanuda la asignación de nodos. La modificación de una instancia grupo indica a Amazon EMR que vuelva a intentar aprovisionar los nodos. Ningún nodo en ejecución se reinicia ni se termina.

En la AWS CLI, el subcomando `list-instances` devuelve todas las instancias y sus estados como hace el subcomando `describe-cluster`. Si Amazon EMR detecta un error con un grupo de instancias, cambia el estado del grupo a `SUSPENDED`.

Para restablecer un clúster en estado SUSPENDED con la AWS CLI

Escriba el subcomando `describe-cluster` con el parámetro `--cluster-id` para ver el estado de las instancias en el clúster.

- Para ver información sobre todas las instancias y los grupos de instancias de un clúster, escriba el siguiente comando y sustituya *j-3KVXXXXXXY7UG* por el ID del clúster.

```
aws emr describe-cluster --cluster-id j-3KVXXXXXXY7UG
```

La salida muestra información sobre los grupos de la instancia y el estado de las instancias:

```
{
    "Cluster": {
        "Status": {
            "Timeline": {
                "ReadyDateTime": 1413187781.245,
                "CreationDateTime": 1413187405.356
            },
            "State": "WAITING",
            "StateChangeReason": {
                "Message": "Waiting after step completed"
            }
        },
        "Ec2InstanceAttributes": {
            "Ec2AvailabilityZone": "us-west-2b"
        },
        "Name": "Development Cluster",
        "Tags": [],
        "TerminationProtected": false,
        "RunningAmiVersion": "3.2.1",
        "NormalizedInstanceHours": 16,
        "InstanceGroups": [
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1413187775.749,
                        "CreationDateTime": 1413187405.357
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "MASTER",
                "InstanceGroupType": "MASTER",
                "InstanceType": "m5.xlarge",
                "Id": "ig-3ETXXXXXXFYV8",
                "Market": "ON_DEMAND",
```

```
                "RunningInstanceCount": 1
            },
            {
                "RequestedInstanceCount": 1,
                "Status": {
                    "Timeline": {
                        "ReadyDateTime": 1413187781.301,
                        "CreationDateTime": 1413187405.357
                    },
                    "State": "RUNNING",
                    "StateChangeReason": {
                        "Message": ""
                    }
                },
                "Name": "CORE",
                "InstanceGroupType": "CORE",
                "InstanceType": "m5.xlarge",
                "Id": "ig-3SUXXXXXQ9ZM",
                "Market": "ON_DEMAND",
                "RunningInstanceCount": 1
            }
...
}
```

Para ver información sobre un grupo de instancias determinado, escriba el subcomando `list-instances` con los parámetros `--cluster-id` e `--instance-group-types`. Puede consultar información para los grupos principales, de núcleo o de tarea.

```
aws emr list-instances --cluster-id j-3KVXXXXXXY7UG --instance-group-types "CORE"
```

Utilice el subcomando `modify-instance-groups` con el parámetro `--instance-groups` para restablecer un clúster en el estado SUSPENDED. El subcomando `describe-cluster` devuelve el id del grupo de instancias.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-3SUXXXXXQ9ZM,InstanceCount=3
```

Consideraciones a la hora de reducir el tamaño del clúster

Si decide reducir el tamaño de un clúster en ejecución, tenga en cuenta el comportamiento y las prácticas recomendadas de Amazon EMR siguientes:

- Para reducir el impacto en los trabajos en curso, Amazon EMR selecciona de forma inteligente las instancias que se van a eliminar. Para más información sobre el comportamiento de reducción vertical del clúster, consulte Terminación al completar las tareas en la Guía de administración de Amazon EMR.
- Al reducir verticalmente el tamaño de un clúster, Amazon EMR copia los datos de las instancias que elimina a las instancias restantes. Asegúrese de que haya suficiente capacidad de almacenamiento para estos datos en las instancias que permanecen en el grupo.
- Amazon EMR intenta retirar HDFS en las instancias del grupo. Antes de reducir el tamaño de un clúster, le recomendamos que minimice las E/S de escritura en HDFS.
- Para tener un control más detallado a la hora de reducir el tamaño de un clúster, puede ver el clúster en la consola y acceder a la pestaña Instancias. Seleccione el ID del grupo de instancias cuyo tamaño desee cambiar. A continuación, use la opción Terminar para las instancias específicas que desee eliminar.

Configurar los tiempos de espera de la capacidad de aprovisionamiento

Cuando usa flotas de instancias, puede configurar los tiempos de espera de aprovisionamiento. Un tiempo de espera de aprovisionamiento indica a Amazon EMR que deje de aprovisionar la capacidad de la instancia si el clúster supera un umbral de tiempo especificado durante el lanzamiento del clúster o las operaciones de escalado del clúster. En los siguientes temas, se explica cómo configurar un tiempo de espera de aprovisionamiento para el lanzamiento del clúster y las operaciones de escalado vertical del clúster.

Temas

- Configurar los tiempos de espera de aprovisionamiento para el lanzamiento del clúster en Amazon EMR
- Personalizar un periodo de tiempo de espera de aprovisionamiento para el cambio de tamaño del clúster en Amazon EMR

Configurar los tiempos de espera de aprovisionamiento para el lanzamiento del clúster en Amazon EMR

Puede definir un periodo de tiempo de espera para aprovisionar instancias de spot para cada flota del clúster. Si Amazon EMR no puede aprovisionar la capacidad de spot, puede elegir terminar el clúster o, en su lugar, aprovisionar capacidad bajo demanda. Si el periodo de tiempo de espera finaliza durante el proceso de cambio de tamaño del clúster, Amazon EMR cancela las solicitudes de spot no aprovisionadas. Las instancias de spot no aprovisionadas no se transfieren a la capacidad bajo demanda.

> **Note**
>
> No se puede personalizar el periodo de espera del aprovisionamiento en la consola antigua. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

Siga los pasos que se describen a continuación para personalizar un periodo de tiempo de espera de aprovisionamiento para el lanzamiento del clúster con la consola de Amazon EMR.

New console

Para configurar el tiempo de espera de aprovisionamiento al crear un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En la página Crear clúster, vaya a Configuración del clúster y seleccione Flotas de instancias.
4. En Opción de escalado y aprovisionamiento del clúster, especifique el tamaño de spot de las flotas principales y de tarea.
5. En Configuración del tiempo de espera de spot, seleccione Terminar el clúster tras finalizar el tiempo de espera de spot o Cambiar a bajo demanda tras finalizar el tiempo de espera de spot. A continuación, especifique el periodo de tiempo de espera para el aprovisionamiento de las instancias de spot. El valor predeterminado es de 1 hora.
6. Seleccione cualquier otra opción que se aplique a su clúster.

7. Para lanzar el clúster con el tiempo de espera configurado, seleccione Crear clúster.

AWS CLI

Para especificar un tiempo de espera de aprovisionamiento con el comando **create-cluster**

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"LaunchSp
{"OnDemandSpecification":{"AllocationStrategy":"lowest-
price"}},"InstanceTypeConfigs":[{"WeightedCapacity":1,"EbsConfiguration":
{"EbsBlockDeviceConfigs":[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemand
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecifi
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification"
{"AllocationStrategy":"lowest-price"}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemand
 - 2"}]'
```

Personalizar un periodo de tiempo de espera de aprovisionamiento para el cambio de tamaño del clúster en Amazon EMR

Puede definir un periodo de tiempo de espera para aprovisionar instancias de spot para cada flota del clúster. Si Amazon EMR no puede aprovisionar la capacidad de spot, cancela la solicitud de cambio de tamaño y detiene los intentos de aprovisionar capacidad de spot adicional. Al crear un clúster, puede configurar el tiempo de espera. Para un clúster en ejecución, puede agregar o actualizar un tiempo de espera.

Cuando vence el periodo de tiempo de espera, Amazon EMR envía automáticamente los eventos a una transmisión de Eventos de Amazon CloudWatch. Con CloudWatch, puede crear reglas que asignen eventos de acuerdo con un patrón especificado y, a continuación, dirijan los eventos a destinos para tomar medidas. Por ejemplo, puede configurar una regla para enviar una notificación

por correo electrónico. Para obtener más información sobre cómo crear reglas, consulte Creación de reglas para los eventos de Amazon EMR con CloudWatch. Para obtener más información sobre los diferentes detalles de eventos, consulte Eventos de cambio de estado de la flota de instancias.

Ejemplos de tiempos de espera de aprovisionamiento agotados para cambiar el tamaño del clúster

Especificar un tiempo de espera de aprovisionamiento para el cambio de tamaño con la AWS CLI

En el siguiente ejemplo, se utiliza el comando `create-cluster` para agregar un tiempo de espera de aprovisionamiento para cambiar el tamaño.

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"InstanceTyp
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecificat
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification":
{"AllocationStrategy":"lowest-price"}},"ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":20},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":25}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 2"}]'
```

En el siguiente ejemplo, se utiliza el comando `modify-instance-fleet` para agregar un tiempo de espera de aprovisionamiento para cambiar el tamaño.

```
aws emr modify-instance-fleet \
--cluster-id j-XXXXXXXXXXXXX \
--instance-fleet '{"InstanceFleetId":"if-XXXXXXXXXXXX","ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":30},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":60}}}' \
--region us-east-1
```

En el siguiente ejemplo, se utiliza `add-instance-fleet-command` para agregar un tiempo de espera de aprovisionamiento para cambiar el tamaño.

```
aws emr add-instance-fleet \
--cluster-id j-XXXXXXXXXXXX \
--instance-fleet
 '{"InstanceFleetType":"TASK","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"InstanceTypeCo
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
{"SpotResizeSpecification":{"TimeoutDurationMinutes":30},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":35}}}' \
--region us-east-1
```

Especificar un tiempo de espera de aprovisionamiento para el cambio de tamaño y el lanzamiento con la AWS CLI

En el siguiente ejemplo, se utiliza el comando `create-cluster` para agregar un tiempo de espera de aprovisionamiento para cambiar el tamaño y realizar el lanzamiento.

```
aws emr create-cluster \
--release-label emr-5.35.0 \
--service-role EMR_DefaultRole \
--ec2-attributes '{"InstanceProfile":"EMR_EC2_DefaultRole","SubnetIds":["subnet-
XXXXX"]}' \
--instance-fleets
 '[{"InstanceFleetType":"MASTER","TargetOnDemandCapacity":1,"TargetSpotCapacity":0,"LaunchSpeci
{"OnDemandSpecification":{"AllocationStrategy":"lowest-price"}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 1"},
{"InstanceFleetType":"CORE","TargetOnDemandCapacity":1,"TargetSpotCapacity":1,"LaunchSpecificat
{"SpotSpecification":
{"TimeoutDurationMinutes":120,"TimeoutAction":"SWITCH_TO_ON_DEMAND"},"OnDemandSpecification":
{"AllocationStrategy":"lowest-price"}},"ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":20},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":25}},"InstanceTypeConfigs":
[{"WeightedCapacity":1,"EbsConfiguration":{"EbsBlockDeviceConfigs":
[{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp2"},"VolumesPerInstance":2}]},"BidPriceAsPercentageOfOnDemandPri
 - 2"}]'
```

Consideraciones para los tiempos de espera de aprovisionamiento para el cambio de tamaño

Al configurar los tiempos de espera de aprovisionamiento del clúster para las flotas de instancias, tenga en cuenta los siguientes comportamientos.

- Puede configurar tiempos de espera de aprovisionado para instancias de spot y bajo demanda. El tiempo de espera mínimo de aprovisionamiento es de 5 minutos. El tiempo de espera máximo de aprovisionamiento es de 7 días.
- Solo puede configurar los tiempos de espera de aprovisionamiento para un clúster de EMR que utilice flotas de instancias. Debe configurar cada núcleo y cada flota de tareas por separado.
- Al crear un clúster, puede configurar los tiempos de espera de aprovisionamiento. Puede agregar un tiempo de espera o actualizar uno existente para un clúster en ejecución.
- Si envía varias operaciones de cambio de tamaño, Amazon EMR realiza un seguimiento de los tiempos de espera de aprovisionamiento para cada operación de cambio de tamaño. Por ejemplo, establezca el tiempo de espera de aprovisionamiento en un clúster en *60* minutos. A continuación, envíe una operación de cambio de tamaño *R1* a las *T1*. Envíe una segunda operación de cambio de tamaño *R2* a las *T2*. El tiempo de espera de aprovisionamiento de R1 vence a las *T1 + 60 minutos*. El tiempo de espera de aprovisionamiento de R2 vence a las *T2 + 60 minutos*.
- Si envía una nueva operación de aumento de tamaño de escalado vertical antes de que se agote el tiempo de espera, Amazon EMR continúa intentando aprovisionar capacidad para su clúster de EMR.

Reducción vertical del clúster

> **Note**
>
> Las opciones de comportamiento de reducción vertical ya no se admiten desde la versión 5.10.0 de Amazon EMR. Debido a la introducción de la facturación por segundo en Amazon EC2, el comportamiento predeterminado de reducción vertical para los clústeres de Amazon EMR ahora es terminarse al completar las tareas.

Con las versiones de 5.1.0 a 5.9.1 de Amazon EMR, existen dos opciones para el comportamiento de reducción vertical: terminarse en el límite de instancia por hora para la facturación de Amazon EC2 o terminarse al completar las tareas. A partir de la versión 5.10.0 de Amazon EMR, la configuración para la terminación en el límite por instancia y hora ha quedado obsoleta debido

a la introducción de la facturación por segundo de Amazon EC2. No recomendamos especificar la terminación en el límite por instancia y hora en las versiones en las que la opción está disponible.

> **Warning**
>
> Si utiliza la AWS CLI para emitir un comando `modify-instance-groups` con `EC2InstanceIdsToTerminate`, estas instancias se terminan de forma inmediata, sin tener en cuenta estos ajustes e independientemente del estado de las aplicaciones que se ejecutan en ellas. Terminar una instancia de esta forma conlleva el riesgo de pérdida de datos y de un comportamiento del clúster imprevisible.

Cuando se especifica la terminación al completar la tarea, Amazon EMR envía a la lista negra y vacía las tareas de los nodos antes de terminar las instancias de Amazon EC2. Sea cual sea el comportamiento especificado, Amazon EMR no termina las instancias de Amazon EC2 de los grupos de instancias principales si ello puede dar lugar a que se dañen los datos de HDFS.

Terminación al completar las tareas

Amazon EMR le permite reducir verticalmente el clúster sin que ello afecte a su carga de trabajo. Amazon EMR retira con gracia YARN, HDFS y otros daemons en nodos principales y de tarea durante una operación de reducción de tamaño sin perder datos o interrumpir trabajos. Amazon EMR solo reduce el tamaño de los grupos de instancias si el trabajo asignados a los grupos se ha completado y están inactivos. Para la retirada con gracia de NodeManager de YARN, puede ajustar manualmente el tiempo que un nodo espera para la retirada.

Este tiempo se establece mediante una propiedad de la clasificación de configuración `YARN-site`. Si utiliza la versión 5.12.0 o una posterior de Amazon EMR, especifique la propiedad `YARN.resourcemanager.nodemanager-graceful-decommission-timeout-secs`. Si utiliza versiones anteriores de Amazon EMR, especifique la propiedad `YARN.resourcemanager.decommissioning.timeout`.

Si sigue habiendo contenedores en ejecución en aplicaciones YARN cuando transcurre el tiempo de espera de retirada, se obliga la retirada del nodo y las reprogramaciones de YARN afectan a contenedores de otros nodos. El valor predeterminado es de 3 600 segundos (una hora). Puede definir este tiempo de espera con un valor arbitrariamente alto para obligar a que la reducción con gracia espere más. Para obtener más información, consulte Graceful Decommission of YARN Nodes en la documentación de Apache Hadoop.

Grupos de nodos de tarea

Amazon EMR selecciona de manera inteligente las instancias que no tienen tareas ejecutándose en ningún paso o aplicación y las elimina de un clúster en primer lugar. Si se están utilizando todas las instancias del clúster, Amazon EMR espera a que las tareas se completen en una instancia antes de eliminarla del clúster. El tiempo de espera predeterminado es de 1 hora. Este valor se puede cambiar con la configuración `YARN.resourcemanager.decommissioning.timeout`. Amazon EMR utiliza de forma dinámica la nueva configuración. Puede establecerlo en un número arbitrariamente grande para garantizar que Amazon EMR no termine ninguna tarea y, al mismo tiempo, reduzca el tamaño del clúster.

Grupos de nodos principales

En los nodos principales, tanto los daemons NodeManager de YARN y DataNode de HDFS se deben retirar para que el grupo de instancias se reduzca. Para YARN, la reducción con gracia garantiza que un nodo marcado para retirada solo pase al estado `DECOMMISSIONED` si no hay contenedores o aplicaciones pendientes o incompletos. La retirada finaliza de inmediato si no hay contenedores en ejecución en el nodo al principio de la retirada.

Para HDFS, la reducción con gracia garantiza que la capacidad de destino de HDFS sea lo suficientemente grande como para adaptarse a todos los bloques existentes. Si la capacidad de destino no es lo suficientemente grande, solo una cantidad parcial de las instancias secundarios se retiran de forma que los nodos restantes pueden gestionar los datos actuales residentes en HDFS. Debe garantizar una capacidad de HDFS adicional para permitir una retirada adicional. También debes intentar minimizar las E/S de escritura antes de intentar reducir los grupos de instancias. Un exceso de E/S de escritura puede retrasar la terminación de la operación de cambio de tamaño.

Otro límite es el factor de replicación predeterminado, `dfs.replication` dentro de `/etc/hadoop/conf/hdfs-site`. Cuando crea un clúster, Amazon EMR configura el valor en función del número de instancias en el clúster: 1 con 1-3 instancias, 2 para clústeres con 4-9 instancias y 3 para clústeres con 10 o más instancias.

⚠ Warning

1. Establecer `dfs.replication` en 1 en clústeres con menos de cuatro nodos puede conllevar la pérdida de datos del HDFS si un solo nodo deja de funcionar. Se recomienda que utilice un clúster con al menos cuatro nodos principales para las cargas de trabajo de producción.

2. Amazon EMR no permitirá que los clústeres escalen los nodos principales por debajo de `dfs.replication`. Por ejemplo, si `dfs.replication = 2`, el número mínimo de nodos principales es 2.
3. Cuando utiliza el escalado administrado, el escalado automático o decide cambiar el tamaño del clúster manualmente, se recomienda que establezca `dfs.replication` en 2 o más.

La reducción con gracia no permite reducir los nodos principales por debajo del factor de replicación de HDFS. Esto permite que HDFS cierre archivos debido a la falta de réplicas. Para evitar este límite, reduzca el factor de replicación y reinicie el daemon NameNode.

Configurar el comportamiento de reducción vertical de Amazon EMR

> **Note**
>
> La opción de comportamiento de reducción vertical con terminación de la instancia después de la hora ya no es compatible con la versión 5.10.0 y posteriores de Amazon EMR. Las siguientes opciones de comportamiento de reducción vertical solo aparecen en la consola de Amazon EMR para las versiones de 5.1.0 a 5.9.1.

Puede utilizar la AWS Management Console, la AWS CLI o la API de Amazon EMR para configurar el comportamiento de reducción vertical al crear un clúster.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Consulte Novedades de la consola para obtener información sobre las diferencias entre la consola antigua y la nueva.

New console

Para configurar el comportamiento de reducción vertical con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.

2. En EMR en EC2 situado en el panel de navegación izquierdo, elija Clústeres y, a continuación, elija Crear clúster.
3. En la sección Opción de escalado y aprovisionamiento del clúster, busque Terminación del clúster y seleccione terminar el clúster manualmente o pida que Amazon EMR lo termine después de un periodo de tiempo de inactividad específico. Si lo desea, active la protección de la terminación contra errores.
4. Elija cualquier otra opción que se aplique a su clúster.
5. Para lanzar el clúster, elija Crear clúster.

Old console

Para configurar el comportamiento de reducción vertical con la consola antigua

1. Abra la consola de Amazon EMR en https://console.aws.amazon.com/elasticmapreduce.
2. Elija Create cluster. Vaya a Opciones avanzadas y seleccione los ajustes de configuración en Paso 1: Software y pasos y Paso 2: Hardware.
3. En Paso 3: Configuración general del clúster, seleccione el comportamiento de reducción vertical que desee. Complete las configuraciones restantes y cree el clúster.

AWS CLI

Para configurar el comportamiento de reducción vertical con la AWS CLI

- Use la opción `--scale-down-behavior` para especificar `TERMINATE_AT_INSTANCE_HOUR` o `TERMINATE_AT_TASK_COMPLETION`.

Terminar un clúster

En esta sección se describen los métodos de terminación de un clúster. Para obtener información sobre cómo habilitar la protección de terminación y la terminación automática de clústeres, consulte Control de la terminación de los clústeres. Puede terminar clústeres en los estados `STARTING`, `RUNNING`o `WAITING`. Un clúster en el estado `WAITING` se debe terminar o se ejecuta de forma indefinida, generando cargos en su cuenta. Puede terminar un clúster que no sea capaz de salir del estado `STARTING` o no pueda completar un paso.

Si está terminando un clúster que tiene definida la protección de terminación, primero debe anular la protección para de poder terminar el clúster. Los clústeres se pueden terminar mediante la consola, la AWS CLI o mediante programación utilizando la API `TerminateJobFlows`.

En función de la configuración del clúster, este puede tardar entre 5 y 20 minutos en terminarse por completo y liberar los recursos asignados, tales como instancias EC2.

> **Note**
>
> No puede reiniciar un clúster terminado, pero puede clonar un clúster terminado para reutilizar su configuración en un clúster nuevo. Para obtener más información, consulte Clonación de un clúster con la consola.

> **Important**
>
> Amazon EMR utiliza el rol de servicio de Amazon EMR y el rol `AWSServiceRoleForEMRCleanup` para limpiar los recursos del clúster que están en su cuenta y que ya no utiliza, como las instancias de Amazon EC2. Debe incluir acciones para que las políticas de rol eliminen o terminen los recursos. De lo contrario, Amazon EMR no podrá realizar estas acciones de limpieza y podría incurrir en costos por los recursos no utilizados que permanecen en el clúster.

Terminar un clúster con la consola

Puede terminar uno o varios clústeres mediante la consola de Amazon EMR. Los pasos para terminar un clúster en la consola varían en función de si la protección de terminación está activada o desactivada. Para terminar un clúster protegido, primero debe deshabilitar la protección de terminación.

New console

Para terminar un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. Seleccione Clústeres y, a continuación, elija el clúster que desea terminar.

3. En el menú desplegable Acciones, seleccione Terminar clúster para abrir la ventana Terminar clúster.
4. Cuando se le solicite, seleccione Terminar. Según la configuración del clúster, la terminación puede tardar de 5 a 10 minutos. Para obtener más información sobre cómo terminar clústeres en Amazon EMR, consulte Terminar un clúster.

Old console

Para terminar un clúster con la protección de finalización desactivada mediante la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Seleccione el clúster que va a terminar. Puede seleccionar varios clústeres y terminarlos al mismo tiempo.
3. Elija Terminate (Terminar).
4. Cuando se le pregunte, elija Terminate (Finalizar).

   Amazon EMR termina las instancias del clúster y deja de guardar los datos de registro.

Para terminar un clúster con la protección de terminación activada mediante la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. En la página Cluster List (Lista de clústeres), seleccione el clúster que desea terminar. Puede seleccionar varios clústeres y terminarlos al mismo tiempo.
3. Elija Terminate (Terminar).
4. Cuando se le solicite, elija Change (Cambiar) para desactivar la protección de terminación. Si ha seleccionado varios clústeres, elija Turn off all (Desactivar todo) para deshabilitar la protección de terminación de todos los clústeres a la vez.
5. En el cuadro de diálogo Terminate clusters (Terminar clústeres), para Termination Protection (Protección de terminación), elija Off (Desactivada) y, a continuación, haga clic en la marca de verificación para confirmar.
6. Haga clic en Terminate (Terminar).

Amazon EMR termina las instancias del clúster y deja de guardar los datos de registro.

Terminar un clúster con la AWS CLI

Para terminar un clúster sin protección utilizando la AWS CLI

Para terminar un clúster sin protección utilizando la AWS CLI, utilice el subcomando `terminate-clusters` con el parámetro --cluster-ids.

- Escriba el comando siguiente para terminar un único clúster y sustituya *j-3KVXXXXXXX7UG* por el ID del clúster.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
  ```

  Para terminar varios clústeres, escriba el comando siguiente y sustituya *j-3KVXXXXXXX7UG* y *j-WJ2XXXXXX8EU* por los ID de los clústeres.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
  ```

  Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Para terminar un clúster con protección utilizando la AWS CLI

Para terminar un clúster protegido utilizando la AWS CLI, en primer lugar deshabilite la protección de terminación utilizando el subcomando `modify-cluster-attributes` con el parámetro `--no-termination-protected`. A continuación, utilice el subcomando `terminate-clusters` con el parámetro `--cluster-ids` para terminarlo.

1. Escriba el comando siguiente para deshabilitar la protección de terminación y sustituya *j-3KVTXXXXXX7UG* por el ID del clúster.

   ```
   aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
   ```

2. Para terminar el clúster, escriba el comando siguiente y sustituya *j-3KVXXXXXXX7UG* por el ID del clúster.

```
aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
```

Para terminar varios clústeres, escriba el comando siguiente y sustituya *j-3KVXXXXXXX7UG* y *j-WJ2XXXXXX8EU* por los ID de los clústeres.

```
aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
```

Para obtener más información sobre cómo usar los comandos de Amazon EMR en la AWS CLI, consulte https://docs.aws.amazon.com/cli/latest/reference/emr.

Terminar un clúster con la API

La operación `TerminateJobFlows` finaliza el procesamiento de pasos, carga los datos de registro desde Amazon EC2 en Amazon S3 (si está configurado) y termina el clúster de Hadoop. Un clúster también termina automáticamente si establece `KeepJobAliveWhenNoSteps` en `False` en una solicitud `RunJobFlows`.

Puede utilizar esta acción para terminar un único clúster o una lista de clústeres por sus ID de clúster.

Para obtener más información sobre los parámetros de entrada únicos para `TerminateJobFlows`, consulte TerminateJobFlows. Para obtener más información acerca de los parámetros genéricos en la solicitud, consulte Parámetros de solicitud comunes.

Clonación de un clúster con la consola

Puede utilizar la consola de Amazon EMR para clonar un clúster, lo que realiza una copia de la configuración del clúster original para utilizarla como base de un nuevo clúster.

> **Note**
>
> Hemos rediseñado la consola de Amazon EMR para que sea más fácil de utilizar. Puede clonar clústeres que utilizan el escalado automático en la nueva consola, pero solo puede crear nuevos clústeres si desea escalarlos manualmente o utilizar el escalado administrado. Consulte Novedades de la consola para obtener más información sobre las diferencias entre la consola antigua y la nueva.

New console

Para clonar un clúster con la nueva consola

1. Inicie sesión en la AWS Management Console y abra la consola de Amazon EMR en https://console.aws.amazon.com/emr.
2. En EMR en EC2 en el panel de navegación izquierdo, seleccione Clústeres.
3. Para clonar un clúster desde la lista de clústeres

   a. Utilice las opciones de búsqueda y filtro para encontrar el clúster que desea clonar en la vista de lista.

   b. Seleccione la casilla de verificación a la izquierda de la fila del clúster que desee clonar.

   c. La opción Clonar ahora estará disponible en la parte superior de la vista de lista. Seleccione Clonar para iniciar el proceso de clonación. Si el clúster tiene pasos configurados, seleccione Incluir pasos y Continuar si desea clonar los pasos junto con las demás configuraciones del clúster.

   d. Revise la configuración del nuevo clúster que se ha copiado del clúster clonado. Ajuste la configuración si es necesario. Cuando esté satisfecho con la configuración del nuevo clúster, seleccione Crear clúster para lanzar el nuevo clúster.
4. Para clonar un clúster desde la página de detalles de un clúster

   a. Para ir a la página de detalles del clúster que desea clonar, seleccione su ID de clúster en la vista de lista de clústeres.

   b. En la parte superior de la página de detalles del clúster, seleccione Clonar el clúster en el menú Acciones para iniciar el proceso de clonación. Si el clúster tiene pasos configurados, seleccione Incluir pasos y Continuar si desea clonar los pasos junto con las demás configuraciones del clúster.

   c. Revise la configuración del nuevo clúster que se ha copiado del clúster clonado. Ajuste la configuración si es necesario. Cuando esté satisfecho con la configuración del nuevo clúster, seleccione Crear clúster para lanzar el nuevo clúster.

Old console

Para clonar un clúster utilizando la consola antigua

1. Vaya hasta la nueva consola de Amazon EMR y seleccione Ir a la consola antigua en el panel de navegación lateral. Para más información sobre lo que puede esperar al cambiar a la consola antigua, consulte Uso de la consola antigua.
2. Elija Create cluster.
3. En la página Cluster List (Lista de clústeres), haga clic en el clúster que desea clonar.
4. En la parte superior de la página Cluster Details (Detalles del clúster), haga clic en Clone (Clonar).

   En el cuadro de diálogo, elija Yes (Sí) para incluir los pasos del clúster original en el clúster de clonado. Elija No para clonar la configuración del clúster original sin incluir ninguno de los pasos.

   > **Note**
   >
   > Para clústeres creados utilizando AMI 3.1.1 y versiones posteriores (Hadoop 2.x) o AMI 2.4.8 y versiones posteriores (Hadoop 1.x), si clona un clúster e incluye pasos, todos los pasos del sistema (como la configuración de Hive) se clonan con pasos enviados por el usuario, hasta 1 000 en total. Los pasos más antiguos que ya no aparezcan en el historial de pasos de la consola no se pueden clonar. Para AMI anteriores, solo se pueden clonar 256 pasos (incluidos los pasos del sistema). Para obtener más información, consulte Enviar trabajo a un clúster.
5. La página Create Cluster (Crear clúster) aparece con una copia de la configuración del clúster original. Revise la configuración, realice todos los cambios necesarios y, a continuación, haga clic en Create Cluster (Crear clúster).

Automatizar clústeres periódicos con AWS Data Pipeline

AWS Data Pipeline es un servicio que automatiza el movimiento y la transformación de datos. Puede utilizarlo para programar la transferencia de datos de entrada en Amazon S3 y para programar el lanzamiento de clústeres para procesar dichos datos. Por ejemplo, considere el caso en que tiene un servidor web que graba registros de tráfico. Si desea ejecutar un clúster semanal para analizar los datos de tráfico, puede utilizar AWS Data Pipeline para programar dichos clústeres. AWS Data

Pipeline es un flujo de trabajo basado en datos, de modo que una tarea (lanzamiento del clúster) puede depender de otra tarea (trasladar los datos de entrada a Amazon S3). También tiene la funcionalidad de reintento eficaz.

Para obtener más información sobre AWS Data Pipeline, consulte la Guía para desarrolladores de AWS Data Pipeline, especialmente los tutoriales referentes a Amazon EMR:

- Tutorial: lanzamiento de un flujo de trabajo de Amazon EMR
- Introducción: procesar registros web con AWS Data Pipeline, Amazon EMR y Hive
- Tutorial: importación y exportación de Amazon DynamoDB utilizando AWS Data Pipeline

Solucionar problemas de clústeres

Un clúster de EMR se ejecuta en un ecosistema complejo que comprende software de código abierto, código de aplicación personalizado y Servicios de AWS. Si se produce un problema con alguna de estas partes, es posible que el clúster falle o tarde más de lo esperado en completarse. Los temas siguientes le ayudarán a identificar problemas del clúster y a solucionarlos.

Temas

- ¿Qué herramientas hay disponibles para la resolución de problemas?
- Ver y reiniciar Amazon EMR y procesos de aplicaciones (daemons)
- Errores comunes en Amazon EMR
- Solucionar problemas de un clúster con errores
- Solucionar problemas de un clúster lento
- Solucionar problemas de un clúster de Lake Formation

Al desarrollar una nueva aplicación de Hadoop, le recomendamos que habilite la depuración y procese un subconjunto pequeño, pero representativo, de sus datos para probar la aplicación. También puede ejecutar la aplicación paso a paso para probar cada paso por separado. Para obtener más información, consulte Configurar el registro y la depuración de un clúster y Paso 5: comprobar el clúster paso a paso.

¿Qué herramientas hay disponibles para la resolución de problemas?

Para identificar y corregir los errores del clúster, puede utilizar las herramientas que se describen en esta página. Es posible que tenga que inicializar algunas de las herramientas al lanzar el clúster. De forma predeterminada, hay otras herramientas disponibles para cada clúster.

Temas

- Ver detalles del clúster de EMR
- Ver detalles de errores del clúster de EMR
- Ejecutar scripts y configurar procesos de Amazon EMR
- Ver archivos de registro de

- Supervisar el rendimiento del clúster de EMR

Ver detalles del clúster de EMR

Puede utilizar la AWS Management Console, la AWS CLI o la API de EMR para recuperar información detallada sobre un clúster de EMR y sobre la ejecución de trabajos. Para obtener más información sobre cómo usar la AWS Management Console y la AWS CLI, consulte Ver el estado y los detalles del clúster.

Panel de detalles de la consola de Amazon EMR

En la lista Clústeres de la consola de Amazon EMR puede ver información de alto nivel sobre el estado de cada clúster de su cuenta y su Región de AWS. En la lista, se muestran todos los clústeres activos y terminados que ha lanzado en los dos últimos meses. En la lista Clusters (Clústeres), puede seleccionar el Name (Nombre) de un clúster para ver los detalles del clúster. Esta información está organizada en distintas categorías para poder consultarla más fácilmente.

La opción Interfaces de usuario de aplicaciones, disponible en la página de detalles del clúster, puede ser especialmente útil para la resolución de problemas. Proporciona el estado de las aplicaciones de YARN y, en algunos casos, como en las aplicaciones de Spark, puede explorar las diferentes métricas y facetas, como trabajos, etapas y ejecutores. Para obtener más información, consulte Ver el historial de aplicaciones. Esta característica solo está disponible para las versiones 5.8.0 y posteriores de Amazon EMR.

Interfaz de línea de comandos de Amazon EMR

Puede encontrar detalles sobre un clúster desde la AWS CLI con el argumento `--describe`.

API de Amazon EMR

Puede encontrar detalles sobre un clúster desde la API utilizando la acción `DescribeJobFlows`.

Ver detalles de errores del clúster de EMR

Cuando un clúster de EMR termina con un error, las API `DescribeCluster` y `ListClusters` devuelven un código de error y un mensaje de error. En el caso de determinados errores del clúster, la matriz de datos `ErrorDetail` puede ayudarle a solucionar el error.

Para obtener una lista de códigos de error que incluyen datos de `ErrorDetail`, consulte Códigos de error con información ErrorDetail.

> **Note**
>
> Mejoramos continuamente nuestros mensajes de error para que reciba la información más reciente y pertinente. No recomendamos analizar el texto desde `ErrorMessage` porque está sujeto a cambios.

Ejecutar scripts y configurar procesos de Amazon EMR

Como parte del proceso de resolución de problemas, puede resultarle útil ejecutar scripts personalizados en el clúster o ver y configurar los procesos del clúster.

Ver y reiniciar los procesos de la aplicación

Puede resultar útil ver los procesos en ejecución en el clúster para diagnosticar posibles problemas. Para detener y reiniciar los procesos del clúster, puede conectarse al nodo maestro del clúster. Para obtener más información, consulte Ver y reiniciar Amazon EMR y procesos de aplicaciones (daemons).

Ejecutar comandos y scripts sin una conexión SSH

Para ejecutar un comando o un script en el clúster como paso, puede usar las herramientas `command-runner.jar` o `script-runner.jar` sin establecer una conexión SSH con el nodo maestro. Para obtener más información, consulte Ejecutar comandos y scripts en un clúster de Amazon EMR.

Ver archivos de registro de

Amazon EMR y Hadoop generan archivos de registro cuando se ejecuta el clúster. Puede acceder a estos archivos de registro desde diversas herramientas, en función de la configuración que haya especificado al lanzar el clúster. Para obtener más información, consulte Configurar el registro y la depuración de un clúster.

Archivos de registro en el nodo maestro

Cada clúster publica archivos de registro en el directorio /mnt/var/log/ en el nodo principal. Estos archivos de registro solo están disponibles mientras se ejecuta el clúster.

Archivos de registro archivados en Amazon S3

Si lanza el clúster y especifica una ruta de registro de Amazon S3, el clúster copia los archivos de registro almacenados en /mnt/var/log/ en el nodo maestro a Amazon S3, en intervalos de 5 minutos. Esto garantiza que tenga acceso a los archivos de registro incluso después de que el clúster se termine. Dado que los archivos están archivados en intervalos de 5 minutos, los últimos minutos de un clúster terminado de forma repentina podrían no estar disponibles.

Supervisar el rendimiento del clúster de EMR

Amazon EMR proporciona varias herramientas para supervisar el rendimiento del clúster.

Interfaces web de Hadoop

Cada clúster publica una serie de interfaces web en el nodo principal que contienen información sobre el clúster. Puede acceder a estas páginas web mediante un túnel SSH para conectarlas en el nodo principal. Para obtener más información, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

Métricas de CloudWatch

Cada clúster notifica las métricas a CloudWatch. CloudWatch es un servicio web que hace un seguimiento de las métricas y que se puede utilizar para establecer alarmas en dichas métricas. Para obtener más información, consulte Monitorización de métricas de Amazon EMR con CloudWatch.

Ver y reiniciar Amazon EMR y procesos de aplicaciones (daemons)

Cuando realice la solución de problemas en un clúster, conviene que cree una lista de los procesos en ejecución. Es posible que también desee detener o reiniciar procesos. Por ejemplo, puede reiniciar un proceso después de cambiar una configuración o detectar un problema con un proceso determinado tras analizar los archivos de registro y los mensajes de error.

Existen dos tipos de procesos que se ejecutan en un clúster: procesos de Amazon EMR (por ejemplo, instance-controller y Log Pusher) y procesos asociados a las aplicaciones instaladas en el clúster (por ejemplo, hadoop-hdfs-namenode y hadoop-yarn-resourcemanager).

Para trabajar con procesos directamente en un clúster, antes debe conectarse al nodo maestro. Para obtener más información, consulte Conectarse a un clúster.

Ver procesos en ejecución

El método que utilice para ver los procesos en ejecución en un clúster varía según la versión de Amazon EMR que utilice.

EMR 5.30.0 and 6.0.0 and later

Example : enumerar todos los procesos en ejecución

En el siguiente ejemplo, se utiliza `systemctl` y se especifica `--type` para ver todos los procesos.

```
systemctl --type=service
```

Example : enumerar procesos específicos

En el siguiente ejemplo, se muestra una lista de todos los procesos con nombres que contengan `hadoop`.

```
systemctl --type=service | grep -i hadoop
```

Ejemplo de resultados:

```
hadoop-hdfs-namenode.service           loaded active running Hadoop namenode
hadoop-httpfs.service                  loaded active running Hadoop httpfs
hadoop-kms.service                     loaded active running Hadoop kms
hadoop-mapreduce-historyserver.service loaded active running Hadoop historyserver
hadoop-state-pusher.service            loaded active running Daemon process that
processes and serves EMR metrics data.
hadoop-yarn-proxyserver.service        loaded active running Hadoop proxyserver
hadoop-yarn-resourcemanager.service    loaded active running Hadoop resourcemanager
hadoop-yarn-timelineserver.service     loaded active running Hadoop timelineserver
```

Example : ver un informe de estado detallado de un proceso específico

En el siguiente ejemplo, se muestra un informe de estado detallado del servicio `hadoop-hdfs-namenode`.

```
sudo systemctl status hadoop-hdfs-namenode
```

Ejemplo de resultados:

```
hadoop-hdfs-namenode.service - Hadoop namenode
   Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
   Active: active (running) since Wed 2021-08-18 21:01:46 UTC; 26min ago
 Main PID: 9733 (java)
    Tasks: 0
   Memory: 1.1M
   CGroup: /system.slice/hadoop-hdfs-namenode.service
           # 9733 /etc/alternatives/jre/bin/java -Dproc_namenode -Xmx1843m -server -
XX:OnOutOfMemoryError=kill -9 %p ...

Aug 18 21:01:37 ip-172-31-20-123 systemd[1]: Starting Hadoop namenode...
Aug 18 21:01:37 ip-172-31-20-123 su[9715]: (to hdfs) root on none
Aug 18 21:01:37 ip-172-31-20-123 hadoop-hdfs-namenode[9683]: starting namenode,
 logging to /var/log/hadoop-hdfs/ha...out
Aug 18 21:01:46 ip-172-31-20-123 hadoop-hdfs-namenode[9683]: Started Hadoop
 namenode:[  OK  ]
Aug 18 21:01:46 ip-172-31-20-123 systemd[1]: Started Hadoop namenode.
Hint: Some lines were ellipsized, use -l to show in full.
```

EMR 4.x - 5.29.0

Example : enumerar todos los procesos en ejecución

En el ejemplo siguiente, se muestra una lista de todos los procesos en ejecución.

```
initctl list
```

EMR 2.x - 3.x

Example : enumerar todos los procesos en ejecución

En el ejemplo siguiente, se muestra una lista de todos los procesos en ejecución.

```
ls /etc/init.d/
```

Detener y reiniciar procesos

Después de determinar qué procesos se están ejecutando, puede detenerlos y, a continuación, reiniciarlos si es necesario.

EMR 5.30.0 and 6.0.0 and later

Example : detener un proceso

En el siguiente ejemplo, se detiene el proceso `hadoop-hdfs-namenode`.

```
sudo systemctl stop hadoop-hdfs-namenode
```

Puede consultar el `status` para comprobar que el proceso se ha detenido.

```
sudo systemctl status hadoop-hdfs-namenode
```

Ejemplo de resultados:

```
hadoop-hdfs-namenode.service - Hadoop namenode
  Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
  Active: failed (Result: exit-code) since Wed 2021-08-18 21:37:50 UTC; 8s ago
Main PID: 9733 (code=exited, status=143)
```

Example : iniciar un proceso

En el siguiente ejemplo, se inicia el proceso `hadoop-hdfs-namenode`.

```
sudo systemctl start hadoop-hdfs-namenode
```

Puede consultar el estado para comprobar que el proceso se está ejecutando.

```
sudo systemctl status hadoop-hdfs-namenode
```

Ejemplo de resultados:

```
hadoop-hdfs-namenode.service - Hadoop namenode
   Loaded: loaded (/etc/systemd/system/hadoop-hdfs-namenode.service; enabled; vendor
 preset: disabled)
   Active: active (running) since Wed 2021-08-18 21:38:24 UTC; 2s ago
  Process: 13748 ExecStart=/etc/init.d/hadoop-hdfs-namenode start (code=exited,
 status=0/SUCCESS)
 Main PID: 13800 (java)
    Tasks: 0
   Memory: 1.1M
```

```
   CGroup: /system.slice/hadoop-hdfs-namenode.service
           # 13800 /etc/alternatives/jre/bin/java -Dproc_namenode -Xmx1843m -server
 -XX:OnOutOfMemoryError=kill -9 %p...
```

EMR 4.x - 5.29.0

Example : detener un proceso en ejecución

En el siguiente ejemplo, se detiene el servicio `hadoop-hdfs-namenode`.

```
sudo stop hadoop-hdfs-namenode
```

Example : reiniciar un proceso detenido

En el siguiente ejemplo, se reinicia el servicio `hadoop-hdfs-namenode`. Debe usar el comando `start` y no `restart`.

```
sudo start hadoop-hdfs-namenode
```

Example : comprobar el estado del proceso

A continuación, se obtiene el estado de `hadoop-hdfs-namenode`. Puede utilizar el comando `status` para comprobar que el proceso se ha detenido o se ha iniciado.

```
sudo status hadoop-hdfs-namenode
```

EMR 2.x - 3.x

Example : detener el proceso de una aplicación

En el siguiente ejemplo, se detiene el servicio `hadoop-hdfs-namenode`, que está asociado a la versión de Amazon EMR instalada en el clúster.

```
sudo /etc/init.d/hadoop-hdfs-namenode stop
```

Example : reiniciar el proceso de una aplicación

El siguiente comando de ejemplo reinicia el proceso `hadoop-hdfs-namenode`:

```
sudo /etc/init.d/hadoop-hdfs-namenode start
```

Example : detener un proceso de Amazon EMR

En el siguiente ejemplo, se detiene un proceso, como instance-controller, que no está asociado a la versión de Amazon EMR del clúster.

```
sudo /sbin/stop instance-controller
```

Example : reiniciar un proceso de Amazon EMR

En el siguiente ejemplo, se reinicia un proceso, como instance-controller, que no está asociado a la versión de Amazon EMR del clúster.

```
sudo /sbin/start instance-controller
```

> **Note**
>
> Los comandos `/sbin/start, stop` y `restart` son symlinks a `/sbin/intictl`. Para obtener más información acerca de `initctl`, consulte la página del manual de initctl escribiendo `man initctl` en el símbolo del sistema.

Errores comunes en Amazon EMR

A veces, los clústeres fallan o procesan los datos con lentitud. En las siguientes secciones, se enumeran algunos problemas comunes de los clústeres con sugerencias sobre cómo solucionarlos.

Temas

- Códigos de error con información ErrorDetail
- Errores de recursos
- Errores de entrada y salida
- Errores de permisos
- Errores de clúster de Hive
- Errores de VPC
- Errores de clúster de streaming
- Errores de clúster JAR personalizados
- Errores de AWS GovCloud (Oeste de EE. UU.)

- Buscar un clúster que falta

Códigos de error con información ErrorDetail

Cuando un clúster de EMR termina con un error, las API `DescribeCluster` y `ListClusters` devuelven un código de error y un mensaje de error. En el caso de algunos errores de clústeres, la matriz de datos `ErrorDetail` puede ayudarle a solucionar el error.

Los errores que incluyen una matriz `ErrorDetail` proporcionan los siguientes detalles:

`ErrorCode`

Un código de error único que se puede utilizar para el acceso mediante programación.

`ErrorData`

Una lista de identificadores en pares clave-valor que puede utilizar para la programación o la búsqueda manual. Para obtener una descripción de los valores `ErrorData` que incluye un código de error, consulte la página de resolución de problemas del código de error.

`ErrorMessage`

Descripción del error con un enlace a más información en la documentación de Amazon EMR.

> **Note**
>
> No recomendamos analizar el texto desde `ErrorMessage` porque está sujeto a cambios.

Códigos de error por categoría

- Códigos de errores de arranque
- Códigos de errores internos
- Códigos de errores de validación

Códigos de errores de arranque

En las siguientes secciones, se proporciona información sobre la resolución de problemas relacionados con los códigos de errores de arranque.

Temas

- BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE
- BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY
- BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY

BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE

Información general

Cuando un clúster termina con un error `BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE`, se produce un error en una acción de arranque en la instancia principal. Para obtener más información sobre las acciones de arranque, consulte Crear acciones de arranque para instalar software adicional.

Resolución

Para resolver este error, revise los detalles que se muestran en el error de la API, modifique el script de acción de arranque y cree un clúster nuevo con la acción de arranque actualizada.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`primary-instance-id`

El ID de la instancia principal en la que se produjo un error en la acción de arranque.

`bootstrap-action`

El número ordinal de la acción de arranque que falló. Un script con un valor de `bootstrap-action` de `1` es la primera acción de arranque que se ejecuta en la instancia.

`return-code`

El código de retorno de la acción de arranque que falló.

`amazon-s3-path`

La ubicación en Amazon S3 de la acción de arranque que falló.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir la causa raíz del error de la acción de arranque. A continuación, lance un clúster nuevo.

1. Revise los archivos de registro de acciones de arranque en Amazon S3 para identificar la causa raíz del error. Para obtener más información sobre cómo ver los registros de Amazon EMR, consulte Ver archivos de registro de .
2. Si activó los registros del clúster al crear la instancia, consulte el registro `stdout` para obtener más información. Puede encontrar el registro `stdout` de la acción de arranque en esta ubicación de Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

   Para obtener más información sobre registros del clúster, consulte Configurar el registro y la depuración de un clúster.
3. Para determinar el error de la acción de arranque, revise las excepciones en los registros `stdout` y el valor `return-code` en `ErrorData`.
4. Utilice los resultados del paso anterior para revisar la acción de arranque de forma que evite las excepciones o pueda gestionarlas correctamente cuando se produzcan.
5. Lance un clúster nuevo con la acción de arranque actualizada.

BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY

Información general

Un clúster termina con el error `BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY` cuando la instancia principal no puede descargar un script de acción de arranque desde la ubicación de Amazon S3 que especifique. Las posibles causas son las siguientes:

- El archivo del script de acción de arranque no se encuentra en la ubicación de Amazon S3.
- El rol de servicio de las instancias de Amazon EC2 del clúster (también denominado perfil de instancia de EC2 para Amazon EMR) no tiene permisos para acceder al bucket de Amazon S3 en el que reside el script de acción de arranque. Para obtener más información acerca de los roles de servicio, consulte Rol de servicio para instancias de EC2 del clúster (perfil de instancia de EC2).

Para obtener más información sobre las acciones de arranque, consulte Crear acciones de arranque para instalar software adicional.

Resolución

Para resolver este error, asegúrese de que la instancia principal tenga el acceso adecuado al script de acción de arranque.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`primary-instance-id`

El ID de la instancia principal en la que se produjo un error en la acción de arranque.

`bootstrap-action`

El número ordinal de la acción de arranque que falló. Un script con un valor de `bootstrap-action` de `1` es la primera acción de arranque que se ejecuta en la instancia.

`amazon-s3-path`

La ubicación en Amazon S3 de la acción de arranque que falló.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir la causa raíz del error de la acción de arranque. A continuación, lance un clúster nuevo.

Pasos para la solución de problemas

1. Utilice el valor `amazon-s3-path` de la matriz `ErrorData` para buscar el script de acción de arranque correspondiente en Amazon S3.
2. Si activó los registros del clúster al crear la instancia, consulte el registro `stdout` para obtener más información. Puede encontrar el registro `stdout` de la acción de arranque en esta ubicación de Amazon S3:

```
s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-
actions/Failed_Bootstrap_Action_Number/stdout.gz
```

Para obtener más información sobre registros del clúster, consulte Configurar el registro y la depuración de un clúster.

3. Para determinar el error de la acción de arranque, revise las excepciones en los registros `stdout` y el valor `return-code` en `ErrorData`.
4. Utilice los resultados del paso anterior para revisar la acción de arranque de forma que evite las excepciones o pueda gestionarlas correctamente cuando se produzcan.
5. Lance un clúster nuevo con la acción de arranque actualizada.

BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY

Información general

El error `BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY` indica que la instancia principal no encuentra el script de acción de arranque que la instancia acaba de descargar del bucket de Amazon S3 especificado.

Resolución

Para resolver este error, confirme que la instancia principal tenga el acceso adecuado al script de acción de arranque.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`primary-instance-id`

El ID de la instancia principal en la que se produjo un error en la acción de arranque.

`bootstrap-action`

El número ordinal de la acción de arranque que falló. Un script con un valor de `bootstrap-action` de 1 es la primera acción de arranque que se ejecuta en la instancia.

`amazon-s3-path`

La ubicación en Amazon S3 de la acción de arranque que falló.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir la causa raíz del error de la acción de arranque. A continuación, lance un clúster nuevo.

1. Para encontrar el script de acción de arranque correspondiente en Amazon S3, utilice el valor `amazon-s3-path` de la matriz `ErrorData`.
2. Revise los archivos de registro de acciones de arranque en Amazon S3 para identificar la causa raíz del error. Para obtener más información sobre cómo ver los registros de Amazon EMR, consulte Ver archivos de registro de .

   > **Note**
   >
   > Si no activó los registros del clúster, debe crear un clúster nuevo con las mismas configuraciones y acciones de arranque. Para asegurarse de que los registros del clúster estén activados, consulte Configurar el registro y la depuración de un clúster.

3. Revise el registro `stdout` de las acciones de arranque y confirme que no haya procesos personalizados que eliminen los archivos de la carpeta `/emr/instance-controller/lib/bootstrap-actions` de sus instancias principales. Puede encontrar el registro `stdout` de la acción de arranque en esta ubicación de Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

4. Lance un clúster nuevo con la acción de arranque actualizada.

Códigos de errores internos

En las siguientes secciones, se proporciona información sobre la resolución de problemas relacionados con los códigos de errores internos.

Temas

- INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ
- INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY
- INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY

INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ

Información general

Un clúster termina con un error `INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ` cuando la zona de disponibilidad seleccionada no tiene capacidad suficiente para cumplir con la solicitud de tipo de instancia de Amazon EC2. La subred que seleccione para un clúster determina la zona de disponibilidad. Para obtener más información sobre las subredes para Amazon EMR, consulte Configurar redes.

Resolución

Para resolver este error, modifique las configuraciones del tipo de instancia y cree un clúster nuevo con la solicitud actualizada.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`instance-type`

El tipo de instancia que no tiene capacidad.

`availability-zone`

La zona de disponibilidad en la que se resuelve la subred.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir la causa raíz del error de configuración del clúster:

- Revise las prácticas recomendadas para la configuración del clúster. Consulte Prácticas recomendadas para la configuración del clúster en la Guía de administración de Amazon EMR.
- Solucione los problemas de lanzamiento y revise la configuración. Consulte Solucionar problemas de lanzamiento de instancias en la Guía del usuario de Amazon EC2 de instancias de Linux.
- Lance un clúster nuevo con la configuración del clúster actualizada.

INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY

Información general

Un clúster termina con un error `INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY` cuando Amazon EMR no puede gestionar la solicitud de instancia de spot para el nodo principal porque las instancias no están disponibles por el precio máximo de spot o uno inferior. Para obtener más información, consulte Instancias de spot en la Guía del usuario de Amazon EC2 para instancias de Linux.

Resolución

Para resolver este error, especifique los tipos de instancia para su clúster que estén dentro de su precio objetivo o aumente el límite de precio para el mismo tipo de instancia.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`primary-instance-id`

El ID de la instancia principal del clúster que falló.

`instance-type`

El tipo de instancia que no tiene capacidad.

`availability-zone`

La zona de disponibilidad en la que reside la subred.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para solucionar los problemas de su estrategia de configuración del clúster y, a continuación, lance un clúster nuevo:

1. Revise las prácticas recomendadas para las instancias de spot de Amazon EC2 y la estrategia de configuración del clúster. Para obtener más información, consulte Prácticas recomendadas para instancias de spot de EC2 en la Guía del usuario de Amazon EC2 de instancias de Linux y Prácticas recomendadas para la configuración del clúster.
2. Modifique las configuraciones del tipo de instancia o la zona de disponibilidad y cree un clúster nuevo con la solicitud actualizada.
3. Si el problema persiste, use la capacidad bajo demanda para la instancia principal.

INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY

Información general

Un clúster termina con un error `INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY` cuando no hay suficiente capacidad para gestionar una solicitud de instancia de spot para su nodo principal. Para obtener más información, consulte Instancias de spot en la Guía del usuario de Amazon EC2 para instancias de Linux.

Resolución

Para resolver este error, especifique los tipos de instancia para su clúster que estén dentro de su precio objetivo o aumente el límite de precio para el mismo tipo de instancia.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`primary-instance-id`

El ID de la instancia principal del clúster que falló.

`instance-type`

El tipo de instancia que no tiene capacidad.

availability-zone

La zona de disponibilidad en la que se resuelve la subred.

public-doc

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para solucionar los problemas de su estrategia de configuración del clúster y, a continuación, lance un clúster nuevo:

1. Revise las prácticas recomendadas para las instancias de spot de Amazon EC2 y la estrategia de configuración del clúster. Para obtener más información, consulte Prácticas recomendadas para instancias de spot de EC2 en la Guía del usuario de Amazon EC2 de instancias de Linux y Prácticas recomendadas para la configuración del clúster.
2. Modifique las configuraciones del tipo de instancia y cree un clúster nuevo con la solicitud actualizada.
3. Si el problema persiste, use la capacidad bajo demanda para la instancia principal.

Códigos de errores de validación

En las siguientes secciones, se proporciona información sobre la resolución de problemas relacionados con los códigos de errores de validación.

Temas

- VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC
- VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC
- VALIDATION_ERROR_INVALID_SSH_KEY_NAME
- VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED

VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC

Información general

Cuando el clúster y las subredes a las que hace referencia pertenecen a distintas nubes privadas virtuales (VPC), el clúster termina con un error

`VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC`. Puede lanzar clústeres con Amazon EMR con la configuración de las flotas de instancias en las subredes de una VPC. Para obtener más información sobre las flotas de instancias, consulte Configurar flotas de instancias en la Guía de administración de Amazon EMR.

Resolución

Para resolver este error, utilice subredes que pertenezcan a la misma VPC que el clúster.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`vpc`

Para cada par subred:VPC, el ID de la VPC a la que pertenece la subred.

`subnet`

Para cada par subred:VPC, el ID de la subred.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir el error:

1. Revise los ID de subred que aparecen en la matriz `ErrorData` y confirme que pertenezcan a la VPC en la que quiere lanzar el clúster de EMR.
2. Modifique las configuraciones de la subred. Puede usar uno de los siguientes métodos para buscar todas las subredes públicas y privadas disponibles en una VPC.

   - Vaya a la consola de Amazon VPC. Elija Subredes y vea todas las subredes que residen en la Región de AWS del clúster. Para buscar solo subredes públicas o privadas, aplique el filtro Asignar automáticamente una dirección IPv4 pública. Para buscar y seleccionar subredes en la VPC que utiliza el clúster, utilice la opción Filtrar por VPC. Para obtener más información sobre cómo crear subredes, consulte Creación de una subred en la Guía del usuario de Amazon Virtual Private Cloud.

- Use la AWS CLI para buscar todas las subredes públicas y privadas disponibles en la VPC que usa su clúster. Para obtener más información, consulte la API describe-subnets. Para crear subredes en una VPC, consulta la API create-subnet.

3. Lance un clúster nuevo con subredes desde la misma VPC que el clúster.

VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC

Información general

Cuando el clúster y los grupos de seguridad que asigna al clúster pertenecen a distintas nubes privadas virtuales (VPC), el clúster termina con un error `VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC`. Para obtener más información sobre los grupos de seguridad, consulte Especificación de los grupos de seguridad adicionales administrados por Amazon EMR y Control del tráfico de red con grupos de seguridad.

Resolución

Para resolver este error, utilice grupos de seguridad que pertenezcan a la misma VPC que el clúster.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`vpc`

Para cada par security-group:VPC, el ID de la VPC a la que pertenece el grupo de seguridad.

`security-group`

Para cada par security-group:VPC, el ID del grupo de seguridad.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir el error:

1. Revise los ID de grupo de seguridad que aparecen en la matriz `ErrorData` y confirme que pertenezcan a la VPC en la que quiere lanzar el clúster de EMR.

2. Vaya a la consola de Amazon VPC. Elija Grupos de seguridad para enumerar todos los grupos de seguridad de la región que seleccione. Busque los grupos de seguridad de la misma VPC que el clúster y, a continuación, modifique la configuración del grupo de seguridad.
3. Lance un clúster nuevo con grupos de seguridad desde la misma VPC que el clúster.

VALIDATION_ERROR_INVALID_SSH_KEY_NAME

Información general

Un clúster termina con un error `VALIDATION_ERROR_INVALID_SSH_KEY_NAME` cuando usa un par de claves de Amazon EC2 que no es válido para SSH en la instancia principal. Es posible que el nombre del par de claves sea incorrecto o que el par de claves no exista en la Región de AWS solicitada. Para obtener más información sobre pares de claves, consulte Pares de claves de Amazon EC2 e instancias de Linux en la Guía del usuario de Amazon EC2 de instancias de Linux.

Resolución

Para resolver este error, cree un clúster nuevo con un nombre de par de claves SSH válido.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`ssh-key`

El nombre del par de claves SSH que proporcionó al crear el clúster.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir el error:

1. Compruebe el archivo *`keypair`*.pem y confirme que coincida con el nombre de la clave SSH que ve en la consola de Amazon EMR.
2. Vaya a la consola de Amazon EC2. Compruebe que el nombre de la clave SSH que utilizó esté disponible en la Región de AWS que utiliza el clúster. Encontrará la Región de AWS junto al ID de cuenta en la parte superior de la AWS Management Console.

3. Lance un clúster nuevo con un nombre de clave SSH válido.

VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED

Información general

Un clúster termina con un error `VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED` cuando la Región de AWS y las zonas de disponibilidad del clúster no admiten el tipo de instancia especificado para uno o más grupos de instancias. Amazon EMR puede admitir un tipo de instancia en una zona de disponibilidad de una región, pero no en otra. La subred que seleccione para un clúster determina la zona de disponibilidad dentro de la región. Para ver una lista de tipos de instancia y regiones compatibles con Amazon EMR, consulte Tipos de instancias admitidas.

Resolución

Para resolver este error, especifique los tipos de instancia para el clúster que Amazon EMR admite en la región y la zona de disponibilidad en la que solicita el clúster.

Para solucionar los problemas del clúster de EMR con errores, consulte la información `ErrorDetail` que devuelven las API `DescribeCluster` y `ListClusters`. Para obtener más información, consulte Códigos de error con información ErrorDetail. La matriz `ErrorData` de `ErrorDetail` devuelve la siguiente información para este código de error:

`instance-types`

La lista de tipos de instancia no compatibles.

`availability-zones`

La lista de zonas de disponibilidad en las que se resuelve la subred.

`public-doc`

La URL pública de la documentación del código de error.

Pasos que completar

Siga estos pasos para identificar y corregir el error:

1. Utilice la AWS CLI para recuperar los tipos de instancia disponibles en una zona de disponibilidad. Para ello, puede usar el comando `ec2 describe-instance-type-`

`offerings` para filtrar los tipos de instancia disponibles por ubicación (Región de AWS o zona de disponibilidad). Por ejemplo, el siguiente comando devuelve los tipos de instancias que se ofrecen en la zona de disponibilidad especificada, *us-east-2a*.

```
aws ec2 describe-instance-type-offerings --location-type "availability-
zone" --filters Name=location,Values=us-east-2a --region us-east-2 --query
 "InstanceTypeOfferings[*].[InstanceType]" --output text | sort
```

Para obtener más información sobre cómo descubrir los tipos de instancia disponibles, consulte Buscar un tipo de instancia de Amazon EC2.

2. Tras determinar los tipos de instancia que están disponibles en la misma región y zona de disponibilidad que el clúster, elija una de las siguientes soluciones para continuar:

   a. Cree un clúster nuevo y elija una subred para el clúster que esté en una zona de disponibilidad en la que el tipo de instancia que ha seleccionado esté disponible y sea compatible con Amazon EMR.

   b. Cree un clúster nuevo en la misma región y subred de Amazon EC2 que el clúster en el que se produjo el error, pero con un tipo de instancia que Amazon EMR admita en esa ubicación.

Para ver una lista de tipos de instancia y regiones compatibles con Amazon EMR, consulte Tipos de instancias admitidas. Para comparar las capacidades de los tipos de instancia, consulte Tipos de instancia de Amazon EC2.

Errores de recursos

Los siguientes errores suelen ser causados habitualmente por recursos limitados en el clúster.

Temas

- El clúster termina con NO_SLAVE_LEFT y los nodos principales con FAILED_BY_MASTER
- Cannot replicate block, only managed to replicate to zero nodes.
- EC2 QUOTA EXCEEDED
- Too many fetch-failures
- File could only be replicated to 0 nodes instead of 1
- Nodos incluidos en la lista de denegados
- Errores de limitación

- Tipo de instancia no compatible
- EC2 no tiene capacidad

El clúster termina con NO_SLAVE_LEFT y los nodos principales con FAILED_BY_MASTER

Normalmente, esto ocurre porque la protección de terminación está deshabilitada y todos los nodos secundarios superan la capacidad de almacenamiento en disco especificada por el umbral de utilización máximo en la configuración de clasificación `yarn-site`, que corresponde al archivo `yarn-site.xml`. Este valor es el 90 % de forma predeterminada. Cuando la utilización del disco en un nodo secundario supera el umbral de utilización, el servicio de estado NodeManager de YARN informa de que el nodo se encuentra en el estado `UNHEALTHY`. Mientras esté en este estado, Amazon EMR lo incluye en la lista de denegados y no le asigna contenedores YARN. Si el nodo sigue en mal estado transcurridos 45 minutos, Amazon EMR marca la instancia de Amazon EC2 asociada para su terminación como `FAILED_BY_MASTER`. Cuando todas las instancias de Amazon EC2 asociadas con nodos principales se marcan para su terminación, el clúster termina con el estado `NO_SLAVE_LEFT` porque no hay recursos para ejecutar trabajos.

Sobrepasar la utilización del disco en un nodo secundario podría causar una reacción en cadena. Si un único nodo supera el umbral de utilización del disco debido a HDFS, es posible que otros nodos estén también cerca del umbral. El primer nodo supera el umbral de uso del disco, por lo que Amazon EMR lo agrega a la lista de denegados. Esto aumenta la carga de uso del disco en los nodos restantes, ya que estos comienzan a replicar entre ellos los datos HDFS que perdieron en el nodo incluido en la lista de denegados. Uno por uno, los nodos van adoptando el estado `UNHEALTHY` de la misma manera, y el clúster finalmente termina.

Prácticas recomendadas y recomendaciones

Configurar el hardware del clúster con almacenamiento suficiente

Al crear un clúster, asegúrese de que haya suficientes nodos secundarios y de que cada uno tenga un almacén de instancias y volúmenes de almacenamiento de EBS para HDFS apropiados. Para obtener más información, consulte Cálculo de la capacidad de HDFS requerida de un clúster. También puede añadir instancias secundarias a grupos de instancias existentes de forma manual o mediante el escalado automático. Las instancias nuevas tienen la misma configuración de almacenamiento que el resto de las instancias del grupo. Para obtener más información, consulte Usar el escalado de clústeres.

Habilitar la protección contra la terminación

Habilite la protección de terminación. De esta forma, si un nodo principal se incluye en la lista de denegados, es posible conectarse a la instancia de Amazon EC2 asociada mediante SSH para solucionar el problema y recuperar los datos. Si habilita la protección de terminación, tenga en cuenta que Amazon EMR no sustituye la instancia de Amazon EC2 por una nueva. Para obtener más información, consulte Uso de la protección de terminación.

Crear una alarma para la métrica de CloudWatch MRUnhealthyNodes

Esta métrica indica el número de nodos que tienen el estado `UNHEALTHY`. Es equivalente a la métrica `mapred.resourcemanager.NoOfUnhealthyNodes` de YARN. Puede configurar una notificación para esta alarma que le avise de los nodos en mal estado 45 minutos antes de que se agote el tiempo de espera. Para obtener más información, consulte Monitorización de métricas de Amazon EMR con CloudWatch.

Retocar la configuración mediante yarn-site

Las opciones mostradas a continuación se pueden ajustar de acuerdo con los requisitos de la aplicación. Por ejemplo, es posible que desee aumentar el umbral de utilización del disco si un nodo adopta el estado `UNHEALTHY` aumentando el valor de `yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`.

Puede establecer estos valores al crear un clúster mediante la clasificación de configuración `yarn-site`. Para obtener más información, consulte Configuración de aplicaciones en la Guía de publicación de Amazon EMR. También puede conectarse mediante SSH a las instancias de Amazon EC2 asociadas con los nodos principales y, a continuación, agregar los valores de `/etc/hadoop/conf.empty/yarn-site.xml` con un editor de texto. Después de realizar los cambios, debe reiniciar hadoop-yarn-nodemanager tal y como se muestra a continuación.

> **Important**
>
> Al reiniciar el servicio NodeManager, los contenedores YARN activos se cancelan a menos que `yarn.nodemanager.recovery.enabled` se establezca en `true` mediante la clasificación de configuración `yarn-site` al crear el clúster. Asimismo, debe especificar el directorio en el que se va a almacenar el estado del contenedor mediante la propiedad `yarn.nodemanager.recovery.dir`.

```
sudo /sbin/stop hadoop-yarn-nodemanager
sudo /sbin/start hadoop-yarn-nodemanager
```

Para obtener más información sobre las propiedades `yarn-site` actuales y sus valores predeterminados, consulte página relacionada con la configuración predeterminada de YARN en la documentación de Apache Hadoop.

Propiedad	Valor predeterminado	Descripción
yarn.nodemanager.disk-health-checker.interval-ms	120 000	La frecuencia (en segundos) con la que se ejecuta el comprobador de estado del disco.
yarn.nodemanager.disk-health-checker.min-healthy-disks	0,25	La fracción mínima del número de discos que deben estar en buen estado para que NodeManager lance contenedores nuevos. Esto se corresponde con yarn.node manager.local-dirs (de forma predeterminada, `/mnt/yarn` en Amazon EMR) y yarn.node manager.log-dirs (de forma predeterminada, `/var/log/hadoop-yarn/containers` , que está vinculado mediante symlinks a `mnt/var/log/hadoop-yarn/containers` en Amazon EMR).
`yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`	90,0	El porcentaje máximo de utilización del espacio en disco permitido después del cual un disco se marca como dañado. Los valores están

Propiedad	Valor predeterminado	Descripción
		comprendidos entre 0,0 y 100,0. Si el valor es superior o igual a 100, NodeManager comprueba si el disco está lleno. Esto se aplica a `yarn-nodemanager.local-dirs` y `yarn.nodemanager.log-dirs` .
`yarn.nodemanager.disk-health-checker.min-free-space-per-disk-mb`	0	El espacio mínimo que debe estar disponible en un disco para que se pueda utilizar. Esto se aplica a `yarn-nodemanager.local-dirs` y `yarn.nodemanager.log-dirs` .

Cannot replicate block, only managed to replicate to zero nodes.

El error “No se puede replicar el bloque, solo se pudo replicar a cero nodos” suele ocurrir cuando un clúster no tiene suficiente espacio de almacenamiento de HDFS. Este error se produce cuando se generan más datos en el clúster de los que pueden almacenarse en HDFS. Ve este error solo mientras se está ejecutando el clúster, porque cuando el trabajo termina se libera el espacio de HDFS que se estaba utilizando.

La cantidad de espacio de HDFS disponible para un clúster depende del número y del tipo de instancias de Amazon EC2 que se utilizan como nodos principales. Los nodos de tarea no se utilizan para almacenamiento de HDFS. Todo el espacio en disco en cada instancia de Amazon EC2, incluidos los volúmenes de almacenamiento de EBS adjuntos, está disponible para HDFS. Para obtener más información sobre la cantidad de almacenamiento local para cada tipo de instancia de EC2, consulte Familias y tipos de instancia en la Guía del usuario de Amazon EC2 de instancias de Linux.

El otro factor que puede influir en la cantidad de espacio de HDFS disponible es el factor de replicación, que es el número de copias de cada bloque de datos que se almacena en HDFS para redundancia. El factor de replicación aumenta con el número de nodos del clúster: existen tres copias

de cada bloque de datos para un clúster con 10 o más nodos, 2 copias de cada bloque por un clúster con 4 a 9 nodos y 1 copia (sin redundancia) para clústeres con 3 o menos nodos. El espacio de HDFS total disponible se divide por el factor de replicación. En algunos casos, como, por ejemplo, aumentando el número de nodos de 9 a 10, el aumento del factor de replicación pueden en realidad hacer que la cantidad de espacio de HDFS disponible se reduzca.

Por ejemplo, un clúster con diez nodos secundarios de tipo m1.large tendría 2 833 GB de espacio disponible para HDFS ((10 nodos X 850 GB por nodo)/factor de replicación de 3).

Si el clúster supera la cantidad de espacio disponible para HDFS, puede añadir más nodos secundarios a su clúster o utilizar la compresión de datos para crear más espacio de HDFS. Si el clúster se puede detener y reiniciar, podría plantearse el uso de nodos principales de un tipo de instancia de Amazon EC2 mayor. También puede plantearse la posibilidad de ajustar el factor de replicación. Tenga en cuenta que la disminución del factor de replicación reduce la redundancia de datos de HDFS y la capacidad de su clúster recuperarse frente a bloques de HDFS perdidos o dañados.

EC2 QUOTA EXCEEDED

Si recibe un mensaje `EC2 QUOTA EXCEEDED`, se puede deber a varias causas. En función de las diferencias de configuración, los clústeres anteriores pueden tardar entre 5 y 20 minutos en terminar y liberar los recursos asignados. Si aparece un error `EC2 QUOTA EXCEEDED` al intentar lanzar un clúster, puede deberse a que aún no se hayan liberado los recursos de un clúster terminado recientemente. Este mensaje también puede deberse al cambio de tamaño de un grupo o flota de instancias a un tamaño de destino mayor que la cuota de instancias actual para la cuenta. Esto puede ocurrir de forma manual o automática a través de escalado automático.

Tenga en cuenta las opciones siguientes para resolver el problema:

- Siga las instrucciones de AWS Service Quotas en Referencia general de Amazon Web Services para solicitar un aumento del límite de servicio. Para algunas API, configurar un evento de CloudWatch puede ser una mejor opción que aumentar los límites. Para obtener más información, consulte Cuándo configurar los eventos de EMR en CloudWatch.
- Si uno o varios clústeres en ejecución no se ejecutan según la capacidad, cambie de tamaño los grupos de instancia o reduzca las capacidades de destino en flotas de instancia para clústeres en ejecución.
- Cree clústeres con menos instancias EC2 o una capacidad de destino reducida.

Too many fetch-failures

La presencia de mensajes de error Too many fetch-failures (Demasiados errores de recuperación) o Error reading task output (Error al leer la salida de la tarea) en los registros de intentos de tareas o de pasos indican que la tarea en ejecución depende de la salida de otra tarea. Esto suele ocurrir cuando una tarea de reducción se pone en cola para ejecutarse y requiere la salida de una o más tareas de asignación y la salida no está disponible aún.

Hay varias razones por las que la salida podría no estar disponible:

- La tarea requisito previo aún se está procesando. Suele ser una tarea de asignación.
- Los datos podrían no estar disponibles debido a una mala conectividad de red si los datos se encuentran en otra instancia.
- Si HDFS se utiliza para recuperar la salida, es posible que exista un problema con HDFS.

La causa más frecuente de este error es que la tarea anterior sigue en procesamiento. Esto es especialmente probable si los errores se producen cuando las tareas de reducción son las primeras que se intentan ejecutar. Puede comprobar si este es el caso revisando el registro syslog para el paso de clúster que devuelve el error. Si el syslog muestra las tareas de asignación y reducción progresando, esto indica que la fase de reducción ha comenzado mientras hay tareas de asignación que no se han completado aún.

Una cosa que hay que buscar en los registros es un porcentaje de progreso de asignación que pasa al 100% y, a continuación, disminuye hasta un valor inferior. Cuando el porcentaje de asignación está al 100%, eso no significa que todas las tareas de asignación se han completado. Simplemente significa que Hadoop está ejecutando todas las tareas de asignación. Si este valor vuelve a disminuir por debajo del 100%, significa que una tarea de asignación ha devuelto un error y, en función de la configuración, Hadoop podría intentar volver a programar la tarea. Si el porcentaje de asignación permanece en el 100 % en los registros, consulte las métricas de CloudWatch, en concreto `RunningMapTasks`, para comprobar si la tarea de asignación sigue procesándose. También puede encontrar esta información a través de la interfaz web de Hadoop en el nodo principal.

Si está viendo este problema, hay varias cosas que puede probar:

- Indique a la fase de reducción que espere más tiempo antes de empezar. Puede hacerlo modificando el ajuste de configuración de Hadoop mapred.reduce.slowstart.completed.maps a un tiempo superior. Para obtener más información, consulte Crear acciones de arranque para instalar software adicional.

- Asigne el recuento de reductores a la capacidad de reductor total del clúster. Esto se hace ajustando la opción de configuración de Hadoop mapred.reduce.tasks para el trabajo.
- Utilice un código de clase de combinador para minimizar el número de salidas que se tienen que recuperar.
- Compruebe que no haya ningún problema con el servicio de Amazon EC2 que esté afectando al rendimiento de red del clúster. Puede hacerlo utilizando el Panel de estado del servicio.
- Revise los recursos de CPU y de memoria de las instancias en su clúster para asegurarse de que su procesamiento de datos no esté desbordando los recursos de los nodos. Para obtener más información, consulte Configuración del hardware y las redes de los clústeres.
- Compruebe la versión de la imagen de máquina de Amazon (AMI) utilizada en el clúster de Amazon EMR. Si la versión es de la 2.3.0 a la 2.4.4 incluida, actualice a una versión posterior. Las versiones de AMI en el rango especificado utilizan una versión de Jetty que podría fallar a la hora de entregar la salida desde la fase de asignación. El error de recuperación se produce cuando los reductores no pueden obtener la salida desde la fase de asignación.

  Jetty es un servidor HTTP de código abierto que se utiliza para comunicaciones de equipo a equipo dentro de un clúster de Hadoop.

File could only be replicated to 0 nodes instead of 1

Cuando un archivo se escribe en HDFS, se replica a varios nodos secundarios. Cuando vea este error, significa que el daemon NameNode no tiene ninguna instancia de DataNode disponible para escribir datos en HDFS. En otras palabras, la replicación de bloques no se está produciendo. Este error puede deberse a una serie de problemas:

- El sistema de archivos HDFS podría haberse quedado sin espacio. Esta es la causa más probable.
- Las instancias DataNode podrían no haber estado disponibles cuando se ejecutó el trabajo.
- La comunicación de las instancias DataNode podría haber estado bloqueada con el nodo principal.
- Las instancias del grupo de instancias secundarias podrían no estar disponibles.
- Es posible que falten permisos. Por ejemplo, el daemon JobTracker podría no tener permisos para crear la información de seguimiento del trabajo.
- El ajuste de espacio reservado para una instancia DataNode podría ser insuficiente. Compruebe si este es el caso comprobando la opción de configuración dfs.datanode.du.reserved.

Para comprobar si este problema se debe a que HDFS se ha quedado sin espacio en disco, examine la métrica `HDFSUtilization` en CloudWatch. Si este valor es demasiado alto, puede añadir nodos secundarios adicionales en el clúster. Si tiene un clúster que cree que podría quedarse sin espacio en disco de HDFS, puede configurar una alarma en CloudWatch que le avise si el valor de `HDFSUtilization` supera un determinado nivel. Para obtener más información, consulte Cambiar manualmente el tamaño de un clúster en ejecución y Monitorización de métricas de Amazon EMR con CloudWatch.

Si el problema no es que HDFS se quede sin espacio, compruebe los registros DataNode, los registros de NameNode y la conectividad de red para los demás problemas que podrían haber evitado que HDFS replicara los datos. Para obtener más información, consulte Ver archivos de registro de .

Nodos incluidos en la lista de denegados

El daemon NodeManager es responsable de lanzar y administrar los contenedores en los nodos secundarios y de tareas. El daemon ResourceManager que se ejecuta en el nodo maestro asigna los contenedores al daemon NodeManager. El ResourceManager monitorea el nodo NodeManager a través de un latido.

Hay un par de situaciones en las que el daemon ResourceManager envía a la lista de denegados un NodeManager y lo elimina del grupo de nodos disponible para procesar tareas:

- Si NodeManager no ha enviado un latido al daemon ResourceManager en los últimos 10 minutos (600 000 milisegundos). Este periodo de tiempo puede configurarse mediante la opción de configuración `yarn.nm.liveness-monitor.expiry-interval-ms`. Para obtener más información sobre cómo cambiar la configuración de Yarn, consulte Configuración de aplicaciones en la Guía de publicación de Amazon EMR.
- NodeManager comprueba el estado de los discos determinado por `yarn.nodemanager.local-dirs` y `yarn.nodemanager.log-dirs`. Las comprobaciones incluyen permisos y espacio libre en disco (< 90 %). Si un disco no supera la comprobación, NodeManager deja de utilizarlo pero sigue notificando que el estado del nodo es correcto. Si varios discos no superan la compilación, el estado del nodo se notifica como no correcto a ResourceManager y los nuevos contenedores no se asignan al nodo.

El nodo principal de la aplicación también puede incluir en la lista de denegados un nodo NodeManager si tiene más de tres tareas con errores. Puede cambiar esto a un valor superior utilizando el parámetro de configuración `mapreduce.job.maxtaskfailures.per.tracker`.

Otras opciones de configuración que podría cambiar controlan cuántas veces se intenta una tarea antes de marcarla como errónea: `mapreduce.map.max.attempts` para tareas de asignación y `mapreduce.reduce.maxattempts` para tareas de reducción. Para obtener más información sobre cómo cambiar la configuración, consulte Configuración de aplicaciones en la Guía de publicación de Amazon EMR.

Errores de limitación

Los errores “Limitado desde *Amazon EC2* al lanzar el clúster” y “No se pudieron aprovisionar las instancias debido a la limitación de *Amazon EC2*” se producen cuando Amazon EMR no puede completar una solicitud porque otro servicio ha limitado la actividad. Amazon EC2 es el origen más habitual de errores de limitación, pero otros servicios pueden ser la causa de estos errores. Los límites de servicio de AWS se aplican por región para mejorar el rendimiento, y un error de limitación indica que ha superado el límite de servicio de su cuenta en esa región.

Causas posibles

El origen más habitual de errores de limitación de Amazon EC2 es que se está lanzando un gran número de instancias de clúster, de modo que se supera el límite de servicio de instancias de EC2. Las instancias de clúster pueden lanzarse por las siguientes razones:

- Se han creado nuevos clústeres.
- Se ha cambiado manualmente el tamaño de los clústeres. Para obtener más información, consulte Cambiar manualmente el tamaño de un clúster en ejecución.
- Los grupos de instancias en un clúster añaden instancias (escalado) como resultado de una regla de escalado automático. Para obtener más información, consulte Descripción de las reglas de escalado automático.
- Las flotas de instancia de un clúster añaden instancias para satisfacer una mayor capacidad de destino. Para obtener más información, consulte Configurar flotas de instancias.

También es posible que la frecuencia o el tipo de solicitud de API hecha a Amazon EC2 provoque errores de limitación. Para obtener más información sobre cómo limita Amazon EC2 las solicitudes de API, consulte Tasa de solicitudes de la API de consulta en la Referencia de la API de Amazon EC2.

Soluciones

Tenga en cuenta las soluciones siguientes:

- Siga las instrucciones de AWS Service Quotas en Referencia general de Amazon Web Services para solicitar un aumento del límite de servicio. Para algunas API, configurar un evento de CloudWatch puede ser una mejor opción que aumentar los límites. Para obtener más información, consulte Cuándo configurar los eventos de EMR en CloudWatch.
- Si tiene clústeres que se lanzan según una misma programación (por ejemplo, al final de una hora), considere la posibilidad de escalonar las horas de inicio.
- Si tiene clústeres que están dimensionados para picos de demanda y periódicamente tiene problemas de capacidad de la instancia, considere la posibilidad de especificar el escalado automático para añadir y eliminar instancias bajo demanda. De esta forma, las instancias se utilizan de manera más eficiente y, en función del perfil de la demanda, es posible que se soliciten menos instancias en un momento dado en una cuenta. Para obtener más información, consulte Uso del escalado automático con una política personalizada para grupos de instancias .

Tipo de instancia no compatible

Si crea un clúster y este falla y muestra el mensaje de error “El tipo de instancia solicitado *InstanceType* no se admite en la zona de disponibilidad solicitada”, significa que creó el clúster y especificó un tipo de instancia para uno o más grupos de instancias que Amazon EMR no acepta en la región y la zona de disponibilidad en las que se creó el clúster. Amazon EMR puede admitir un tipo de instancia en una zona de disponibilidad de una región y no en otra. La subred que seleccione para un clúster determina la zona de disponibilidad dentro de la región.

Solución

Determinar los tipos de instancia disponibles en una zona de disponibilidad mediante la AWS CLI

- Utilice el comando `ec2 run-instances` con la opción `--dry-run`. En el ejemplo siguiente, sustituya *m5.xlarge* por el tipo de instancia que desee utilizar, *ami-035be7bafff33b6b6* por la AMI asociada con ese tipo de instancia y *subnet-12ab3c45* por una subred en la zona de disponibilidad que desea consultar.

  ```
  aws ec2 run-instances --instance-type m5.xlarge --dry-run --image-
  id ami-035be7bafff33b6b6 --subnet-id subnet-12ab3c45
  ```

  Para obtener instrucciones sobre cómo encontrar un ID de AMI, consulte Buscar una AMI de Linux. Para encontrar un ID de subred, puede usar el comando describe-subnets.

Para obtener más información sobre cómo descubrir los tipos de instancia disponibles, consulte Buscar un tipo de instancia de Amazon EC2.

Después de determinar los tipos de instancia disponibles, puede hacer lo siguiente:

- Cree el clúster en la misma región y subred EC2 y elija un tipo de instancia diferente con capacidades similares a la elección inicial. Para ver una lista de los tipos de instancias admitidos, consulte Tipos de instancias admitidas. Para comparar las capacidades de los tipos de instancia de EC2, consulte Tipos de instancia de Amazon EC2.
- Elija una subred para el clúster en una zona de disponibilidad en la que el tipo de instancia esté disponible y sea compatible con Amazon EMR.

EC2 no tiene capacidad

Se produce el error “EC2 no tiene capacidad para *InstanceType*” cuando intenta crear un clúster o agregar instancias a un clúster en una zona de disponibilidad que ya no tiene el tipo de instancia de EC2 especificado. La subred que seleccione para un clúster determina la zona de disponibilidad.

Para crear un clúster, realice alguna de las siguientes operaciones:

- Especificar un tipo de instancia diferente con capacidades similares
- Crear el clúster en una región diferente
- Seleccione una subred en una zona de disponibilidad en la que pueda estar disponible el tipo de instancia que desee.

Para agregar instancias en un clúster en ejecución, haga una de estas acciones:

- Modifique las configuraciones de grupo de instancias o las configuraciones de flota de instancias para agregar tipos de instancia disponibles con capacidades similares. Para ver una lista de los tipos de instancias admitidos, consulte Tipos de instancias admitidas. Para comparar las capacidades de los tipos de instancia de EC2, consulte Tipos de instancia de Amazon EC2.
- Termine el clúster y vuelva a crearlo en una región y zona de disponibilidad en la que el tipo de instancia esté disponible.

Errores de entrada y salida

Los errores siguientes son habituales en las operaciones de entrada y salida de clúster.

Temas

- ¿La ruta hacia Amazon Simple Storage Service (Amazon S3) tiene por lo menos tres barras?
- ¿Está intentando atravesar de forma recursiva directorios de entrada?
- ¿Ya existe el directorio de salida?
- ¿Está intentando especificar un recurso mediante una URL HTTP?
- ¿Está haciendo referencia a un bucket de Amazon S3; con un formato de nombre no válido?
- ¿Tiene problemas para cargar datos hacia o desde Amazon S3?

¿La ruta hacia Amazon Simple Storage Service (Amazon S3) tiene por lo menos tres barras?

Cuando especifique un bucket de Amazon S3, deberá incluir una barra de terminación al final de la URL. Por ejemplo, en lugar de hacer referencia a un bucket como “s3n://*DOC-EXAMPLE-BUCKET1*”, debería utilizar “s3n://*DOC-EXAMPLE-BUCKET1*/”; de lo contrario, Hadoop genera un error en el clúster en la mayoría de los casos.

¿Está intentando atravesar de forma recursiva directorios de entrada?

Hadoop no busca directorios de entrada de forma recursiva para archivos. Si tiene una estructura de directorios como /corpus/01/01.txt, /corpus/01/02.txt, /corpus/02/01.txt, etc. y especifica /corpus/ como parámetro de entrada para el clúster, Hadoop no encuentra los archivos de entrada, ya que el directorio /corpus/ está vacío y Hadoop no comprueba el contenido de los subdirectorios. Del mismo modo, Hadoop no comprueba recursivamente los subdirectorios de buckets de Amazon S3.

Los archivos de entrada deben estar directamente en el directorio de entrada o bucket de Amazon S3 que especifique, no en subdirectorios.

¿Ya existe el directorio de salida?

Si especifica una ruta de salida que ya existe, Hadoop generará un error en el clúster en la mayoría de los casos. Esto significa que si ejecuta un clúster una vez y, a continuación, lo vuelve a ejecutar con los mismos parámetros, probablemente funcionará la primera vez y, a continuación, nunca más; después de la primera ejecución, existe la ruta de salida y, por lo tanto, hace que todas las ejecuciones sucesivas generen un error.

¿Está intentando especificar un recurso mediante una URL HTTP?

Hadoop no acepta las ubicaciones de los recursos especificados mediante el prefijo http://. No puede hacer referencia a un recurso con una dirección URL HTTP. Por ejemplo, transferir http://mysite/myjar.jar como parámetro JAR provoca que el clúster devuelva un error.

¿Está haciendo referencia a un bucket de Amazon S3; con un formato de nombre no válido?

Si intenta utilizar un nombre de bucket como "*DOC-EXAMPLE-BUCKET1*.1" con Amazon EMR, el clúster devolverá un error porque Amazon EMR requiere que los nombres de bucket sean nombres de host RFC 2396 válidos; el nombre no puede terminar por un número. Además, debido a los requisitos de Hadoop, los nombres de bucket de Amazon S3 que se utilizan con Amazon EMR solo puede contener letras en minúsculas, números, puntos (.) y guiones (-). Para obtener más información acerca de las convenciones de nomenclatura de buckets de Amazon S3, consulte Restricciones y limitaciones de los buckets en la Guía del usuario de Amazon Simple Storage Service.

¿Tiene problemas para cargar datos hacia o desde Amazon S3?

Amazon S3 es el origen de entrada y salida más popular para Amazon EMR. Un error común consiste en tratar Amazon S3 como si fuera un sistema de archivos habitual. Existen diferencias entre Amazon S3 y un sistema de archivos que debe tener en cuenta a la hora de ejecutar el clúster.

- Si se produce un error interno en Amazon S3, la aplicación tiene que gestionarlo correctamente y volver a intentar la operación.
- Si las llamadas a Amazon S3 tardan demasiado tiempo en devolverse, es posible que la aplicación tenga que reducir la frecuencia con la que llama a Amazon S3.
- Enumerar todos los objetos en un bucket de Amazon S3 es una llamada costosa. La aplicación debe minimizar el número de veces que lo hace.

Existen varias formas de mejorar la forma en la que el clúster interactúa con Amazon S3.

- Inicie el clúster con la versión de lanzamiento más reciente de Amazon EMR.
- Utilice S3DistCp para introducir y sacar objetos de Amazon S3. S3DistCp implementa la gestión de errores, los reintentos y las interrupciones para ajustarse a los requisitos de Amazon S3. Para obtener más información, consulte Copia distribuida mediante S3DistCp.

- Diseñe la aplicación teniendo en cuenta la consistencia final. Utilice HDFS para el almacenamiento de datos intermedios mientras que el clúster se está ejecutando y Amazon S3 únicamente para entrada de los datos iniciales y salida de los resultados finales.
- Si los clústeres confirmarán 200 o más transacciones por segundo en Amazon S3, contacte con la asistencia técnica para preparar el bucket para que haga más transacciones por segundo y plantéese el uso de las estrategias de partición de claves que se describen en Consejos y trucos de rendimiento de Amazon S3.
- Defina el ajuste de configuración de Hadoop io.file.buffer.size en 65 536. Esto hace que Hadoop dedique menos tiempo a buscar a través de objetos de Amazon S3.
- Plantéese deshabilitar la característica de ejecución especulativa de Hadoop si su clúster experimenta problemas de simultaneidad de Amazon S3. Esto también resulta útil cuando se solucionan problemas de un clúster lento. Para ello, establezca las propiedades `mapreduce.reduce.speculative` `mapreduce.map.speculative` y en `false`. Cuando lance un clúster, podrá establecer estos valores mediante la clasificación de configuración `mapred-env`. Para obtener más información, consulte Configuración de aplicaciones en la Guía de publicación de Amazon EMR.
- Si ejecuta un clúster de Hive, consulte ¿Tiene problemas para cargar datos hacia o desde Amazon S3 en Hive?.

Para obtener información adicional, consulte Prácticas recomendadas de errores de Amazon S3 en la Guía del usuario de Amazon Simple Storage Service.

Errores de permisos

Los siguientes errores son comunes cuando se utilizan permisos o credenciales.

Temas

- ¿Está transfiriendo las credenciales correctas en SSH?
- Si está utilizando IAM, ¿tiene definido el conjunto de políticas de Amazon EC2 correcto?

¿Está transfiriendo las credenciales correctas en SSH?

Si no puede utilizar SSH para conectarse al nodo principal, se trata muy probablemente de un problema con sus credenciales de seguridad.

En primer lugar, compruebe que el archivo .pem que contiene su clave SSH disponga de los permisos adecuados. Puede utilizar chmod para cambiar los permisos de su archivo .pem tal y como se muestra en el siguiente ejemplo, donde debería sustituir mykey.pem por el nombre de su propio archivo .pem.

```
chmod og-rwx mykey.pem
```

La segunda posibilidad es que no se esté utilizando el par de claves que especificó al crear el clúster. Esto es fácil de hacer si ha creado varios pares de claves. Consulte los detalles del clúster en la consola de Amazon EMR (o utilice la opción `--describe` de la CLI) para el nombre del par de claves que se especificó cuando se creó el clúster.

Una vez que haya verificado que está utilizando el par de claves correcto y que los permisos se han definido correctamente en el archivo .pem, puede utilizar el siguiente comando para utilizar SSH para conectarse al nodo maestro, donde debería sustituir mykey.pem por el nombre de su archivo .pem y `hadoop@ec2-01-001-001-1.compute-1.amazonaws.com` por el nombre de DNS público del nodo principal (disponible a través de la opción `--describe` en la CLI o a través de la consola de Amazon EMR).

> **Important**
>
> Debe utilizar el nombre de inicio de sesión `hadoop` cuando se conecte a un nodo de clúster de Amazon EMR; de lo contrario, es posible que se produzca un error similar a `Server refused our key`.

```
ssh -i mykey.pem hadoop@ec2-01-001-001-1.compute-1.amazonaws.com
```

Para obtener más información, consulte Conectarse al nodo principal mediante SSH.

Si está utilizando IAM, ¿tiene definido el conjunto de políticas de Amazon EC2 correcto?

Dado que Amazon EMR utiliza instancias de EC2 como nodos, los usuarios de Amazon EMR también deben tener definidas determinadas políticas de Amazon EC2 para que Amazon EMR pueda

administrar dichas instancias en nombre del usuario. Si no tiene definidos los permisos necesarios, Amazon EMR devuelve el error: “La cuenta no tiene autorización para llamar a EC2”.

Para obtener más información sobre las políticas de Amazon EC2 que la cuenta de IAM tiene que definir para ejecutar Amazon EMR, consulte Cómo funciona Amazon EMR con IAM.

Errores de clúster de Hive

Normalmente, puede encontrar la causa de un error de Hive en el archivo `syslog`, para el que tiene un enlace en el panel Steps (Pasos). Si no puede determinar el problema allí, consulte el mensaje de error de intento de tareas de Hadoop. Encontrará un enlace al mismo en el panel Task Attempts (Intentos de tareas).

Los siguientes errores son comunes en los clústeres de Hive.

Temas

- ¿Está utilizando la última versión de Hive?
- ¿Ha detectado un error de sintaxis en el script de Hive?
- ¿Ha devuelto error un trabajo al ejecutarlo de forma interactiva?
- ¿Tiene problemas para cargar datos hacia o desde Amazon S3 en Hive?

¿Está utilizando la última versión de Hive?

La última versión de Hive presenta todas las revisiones actuales y correcciones de errores y podría resolver el problema.

¿Ha detectado un error de sintaxis en el script de Hive?

Si un paso devuelve un error, examine el archivo `stdout` de los registros para el paso que se ejecutó en el script de Hive. Si el error no se encuentra allí, examine el archivo `syslog` de los registros del intento de tarea que ha devuelto error. Para obtener más información, consulte Ver archivos de registro de .

¿Ha devuelto error un trabajo al ejecutarlo de forma interactiva?

Si ejecuta Hive de forma interactiva en el nodo principal y el clúster ha fallado, vea las entradas `syslog` en el registro de intento de tarea para el intento de tarea fallido. Para obtener más información, consulte Ver archivos de registro de .

¿Tiene problemas para cargar datos hacia o desde Amazon S3 en Hive?

Si tiene problemas para tener acceso a los datos en Amazon S3, compruebe antes las causas posibles incluidas en ¿Tiene problemas para cargar datos hacia o desde Amazon S3?. Si ninguno de estos problemas es la causa, tenga en cuenta las siguientes opciones específicas de Hive.

- Asegúrese de utilizar la última versión de Hive que presenta todas las revisiones actuales y correcciones de errores que podría resolver el problema. Para obtener más información, consulte Apache Hive.
- El uso de `INSERT OVERWRITE` requiere mostrar el contenido del bucket o carpeta de Amazon S3. Se trata de una operación costosa. Si es posible, elimine manualmente la ruta en lugar de que Hive enumere y elimine los objetos existentes.
- Si utiliza versiones de Amazon EMR anteriores a la 5.0, puede utilizar el comando siguiente de HiveQL para guardar previamente en caché los resultados de una operación de listado de Amazon S3 localmente en el clúster:

```
set hive.optimize.s3.query=true;
```

- Utilice las particiones estáticas donde sea posible.
- En algunas versiones de Hive y Amazon EMR, es posible que el uso de ALTER TABLES devuelva un error debido a que la tabla se almacena en una ubicación distinta a la esperada por Hive. La solución consiste en añadir o actualizar lo siguiente en `/home/hadoop/conf/core-site.xml`:

```
<property>
    <name>fs.s3n.endpoint</name>
    <value>s3.amazonaws.com</value>
</property>
```

Errores de VPC

Los siguientes errores son comunes a la configuración de VPC en Amazon EMR.

Temas

- Configuración de subredes no válida
- Falta el conjunto de opciones de DHCP
- Errores de permisos

- Errores que dan lugar a START_FAILED
- Clúster Terminated with errors y NameNode no se inicia

Configuración de subredes no válida

En la página Cluster Details (Detalles del clúster), en el campo Status (Estado), ve un error similar al siguiente:

```
The subnet configuration was invalid: Cannot find route to InternetGateway
in main RouteTable rtb-id for vpc vpc-id.
```

Para solucionar este problema, debe crear una gateway de Internet y asociarla a la VPC. Para obtener más información, consulte Adding an internet gateway to your VPC (Cómo añadir una gateway de Internet a la VPC).

De forma alternativa, compruebe que ha configurado la VPC con las opciones Enable DNS resolution (Habilitar resolución de DNS) y Enable DNS hostname support (Habilitar el soporte de nombres de host DNS) habilitadas. Para obtener más información, consulte Utilización de DNS con su VPC.

Falta el conjunto de opciones de DHCP

Puede ver un error de paso en el registro del sistema (syslog) del clúster con un error similar al siguiente:

```
ERROR org.apache.hadoop.security.UserGroupInformation
(main): PriviledgedActionException as:hadoop (auth:SIMPLE)
cause:java.io.IOException:
org.apache.hadoop.yarn.exceptions.ApplicationNotFoundException: Application
with id 'application_id' doesn't exist in RM.
```

o bien

```
ERROR org.apache.hadoop.streaming.StreamJob (main): Error Launching job :
org.apache.hadoop.yarn.exceptions.ApplicationNotFoundException: Application
with id 'application_id' doesn't exist in RM.
```

Para solucionar este problema, debe configurar una VPC que incluya un conjunto de opciones de DHCP cuyos parámetros se hayan definido en los siguientes valores:

> **Note**
>
> Si utiliza la región AWS GovCloud (Oeste de EE. UU.), defina domain-name como **`us-gov-west-1.compute.internal`** en lugar del valor utilizado en el ejemplo siguiente.

- domain-name = **`ec2.internal`**

  Use **`ec2.internal`** si su región es Este de EE. UU. (Norte de Virginia). Para las demás regiones, utilice *`nombre-región`***`.compute.internal`**. Por ejemplo en us-west-2, utilice domain-name=**`us-west-2.compute.internal`**.
- domain-name-servers = **`AmazonProvidedDNS`**

Para obtener más información, consulte Conjuntos de opciones de DHCP.

Errores de permisos

Un error en el registro `stderr` de un paso indica que un recurso de Amazon S3 no tiene los permisos adecuados. Se trata de un error 403 y su aspecto es:

```
Exception in thread "main" com.amazonaws.services.s3.model.AmazonS3Exception: Access
 Denied (Service: Amazon S3; Status Code: 403; Error Code: AccessDenied; Request
 ID: REQUEST_ID
```

Si ActionOnFailure está definido en `TERMINATE_JOB_FLOW`, esto daría lugar a que el clúster termine con el estado, `SHUTDOWN_COMPLETED_WITH_ERRORS`.

Algunas formas de solucionar este problema son:

- Si está utilizando una política de bucket de Amazon S3 dentro de una VPC, asegúrese de proporcionar acceso a todos los buckets. Para ello, cree un punto de conexión de VPC y seleccione Permitir todo en la opción Política al crear el punto de conexión.
- Asegúrese de que las políticas asociadas con recursos de S3 incluyan la VPC en la que lanzar el clúster.
- Pruebe a ejecutar el siguiente comando desde el clúster para verificar que puede acceder al bucket

  ```
  hadoop fs -copyToLocal s3://path-to-bucket /tmp/
  ```

- Puede obtener información más específica sobre la depuración definiendo el parámetro `log4j.logger.org.apache.http.wire` en `DEBUG` en el archivo `/home/hadoop/conf/log4j.properties` en el clúster. Puede comprobar el archivo de registro `stderr` después de intentar acceder al bucket desde el clúster. El archivo de registro proporcionará información más detallada:

```
Access denied for getting the prefix for bucket - us-west-2.elasticmapreduce with
 path samples/wordcount/input/
15/03/25 23:46:20 DEBUG http.wire: >> "GET /?prefix=samples%2Fwordcount%2Finput
%2F&delimiter=%2F&max-keys=1 HTTP/1.1[\r][\n]"
15/03/25 23:46:20 DEBUG http.wire: >> "Host: us-
west-2.elasticmapreduce.s3.amazonaws.com[\r][\n]"
```

Errores que dan lugar a **`START_FAILED`**

Antes de la AMI 3.7.0, para las VPC donde se especifica un nombre de host, Amazon EMR asigna los nombres de host internos de la subred a direcciones de dominio personalizadas del siguiente modo: `ip-`*`X.X.X.X.customdomain.com`*`.tld`. Por ejemplo, si el nombre de host fuera `ip-10.0.0.10` y la VPC tuviera la opción de nombre de dominio definida en customdomain.com, el nombre de host resultante asignado por Amazon EMR sería `ip-10.0.1.0.customdomain.com`. Se añade una entrada en `/etc/hosts` para resolver el nombre de host a 10.0.0.10. Este comportamiento se cambia con la AMI 3.7.0 y ahora Amazon EMR respeta completamente la configuración de DHCP de la VPC. Anteriormente, los clientes también podrían utilizar una acción de arranque para especificar un mapeo de nombre de host.

Si desea conservar este comportamiento, debe proporcionar la DNS y reenviar la configuración de resolución que necesita para el dominio personalizado.

Clúster **`Terminated with errors`** y NameNode no se inicia

Al lanzar un clúster de EMR en una VPC que hace uso de un nombre de dominio de DNS personalizado, el clúster podría devolver el siguiente mensaje de error en la consola:

```
Terminated with errors  On the master instance(instance-id), bootstrap action 1
 returned a  non-zero return code
```

El error se debe a que NameNode no se puede iniciar. Esto dará lugar al siguiente error en los registros de NameNode, cuyo URI de Amazon S3 tiene el formato:

s3://*mybucket*/*logs*/*cluster-id*/daemons/*master instance-id*/hadoop-hadoop-namenode-*master node hostname*.log.gz:

```
2015-07-23 20:17:06,266 WARN
      org.apache.hadoop.hdfs.server.namenode.FSNamesystem (main): Encountered
 exception
      loading fsimage  java.io.IOException: NameNode is not formatted.
      at

 org.apache.hadoop.hdfs.server.namenode.FSImage.recoverTransitionRead(FSImage.java:212)
           at

 org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFSImage(FSNamesystem.java:1020)
           at

 org.apache.hadoop.hdfs.server.namenode.FSNamesystem.loadFromDisk(FSNamesystem.java:739)
           at
      org.apache.hadoop.hdfs.server.namenode.NameNode.loadNamesystem(NameNode.java:537)
           at
      org.apache.hadoop.hdfs.server.namenode.NameNode.initialize(NameNode.java:596)

      at  org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:765)
           at
      org.apache.hadoop.hdfs.server.namenode.NameNode.<init>(NameNode.java:749)
      at

 org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1441)
           at
      org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1507)
```

Esto es debido a un posible problema en una instancia de EC2 que puede tener varios conjuntos de nombres de dominio completos al lanzar clústeres de EMR en una VPC, que hace uso de un servidor de DNS proporcionado por AWS y un servidor de DNS proporcionado por el usuario personalizado. Si el servidor de DNS proporcionado por el usuario no ofrece ningún registro de puntero (PTR) para ningún registro A utilizado para designar nodos en un clúster de EMR, los clústeres devolverán un error al iniciarse cuando se configuran de esta manera. La solución consiste en añadir un registro PTR por cada registro A que se crea cuando se lanza una instancia EC2 en cualquiera de las subredes de la VPC.

Errores de clúster de streaming

Normalmente, puede encontrar la causa de un error de streaming en un archivo `syslog`. Encontrará un enlace al mismo en el panel Steps (Pasos).

Los siguientes errores son comunes a los clústeres de streaming.

Temas

- ¿Los datos que se envían al mapeador están en formato equivocado?
- ¿Se agota el tiempo de espera del script?
- ¿Está transfiriendo un argumento de streaming no válido?
- ¿El script se cierra con un error?

¿Los datos que se envían al mapeador están en formato equivocado?

Para comprobar si este es el caso, busque un mensaje de error en el archivo `syslog` un intento de tarea con error en los registros de intento de tareas. Para obtener más información, consulte Ver archivos de registro de .

¿Se agota el tiempo de espera del script?

El tiempo de espera predeterminado para un script de mapeador o reductor es 600 segundos. Si el script tarda más tiempo, el intento de tarea devolverá un error. Puede comprobar si es así comprobando el archivo `syslog` de un intento de tarea con error en los registros de intento de tareas. Para obtener más información, consulte Ver archivos de registro de .

Puede cambiar el límite de tiempo estableciendo un nuevo valor para el ajuste de configuración de `mapred.task.timeout`. Esta configuración especifica el número de milisegundos tras el que Amazon EMR terminará una tarea que no tiene entrada de lectura, salida de escritura o ha actualizado su cadena de estado. Puede actualizar este valor transfiriendo un argumento de streaming adicional `-jobconf mapred.task.timeout=800000`.

¿Está transfiriendo un argumento de streaming no válido?

Hadoop Streaming admite únicamente los siguientes argumentos. Si transfiere argumentos distintos de los que se indican a continuación, el clúster devolverá un error.

```
-blockAutoGenerateCacheFiles
```

```
-cacheArchive
-cacheFile
-cmdenv
-combiner
-debug
-input
-inputformat
-inputreader
-jobconf
-mapper
-numReduceTasks
-output
-outputformat
-partitioner
-reducer
-verbose
```

Además, Hadoop Streaming solo reconoce argumentos transferidos mediante sintaxis de Java; es decir, precedidos de un único guion. Si transfiere argumentos precedidos de un guion doble, el clúster fallará.

¿El script se cierra con un error?

Si su script de mapeador o reductor termina con un error, puede localizar el error en el archivo `stderr` de los registros de intento de tarea del intento de tarea que ha devuelto error. Para obtener más información, consulte Ver archivos de registro de .

Errores de clúster JAR personalizados

Los siguientes errores son comunes en los clústeres JAR personalizados.

Temas

- ¿Su JAR lanza una excepción antes de crear un trabajo?
- ¿Su JAR lanza un error dentro de una tarea de asignación?

¿Su JAR lanza una excepción antes de crear un trabajo?

Si el programa principal de su JAR personalizado arroja una excepción al crear el trabajo de Hadoop, el mejor lugar consiste donde buscar es el archivo `syslog` de los registros de pasos. Para obtener más información, consulte Ver archivos de registro de .

¿Su JAR lanza un error dentro de una tarea de asignación?

Si su JAR personalizado y mapeador lanzan una excepción al procesar los datos de entrada, el mejor lugar donde buscar es el archivo `syslog` de los registros de intento de tarea. Para obtener más información, consulte Ver archivos de registro de .

Errores de AWS GovCloud (Oeste de EE. UU.)

La región AWS GovCloud (Oeste de EE. UU.) difiere de otras regiones en la seguridad, configuración y ajustes predeterminados. Por tanto, utilice la siguiente lista de comprobación para resolver errores de Amazon EMR que sean específicos de la región AWS GovCloud (Oeste de EE. UU.) antes de utilizar recomendaciones más generales de resolución de problemas.

- Compruebe que sus roles de IAM estén correctamente configurados. Para obtener más información, consulte Configuración de los roles de servicio de IAM de los permisos de Amazon EMR para los servicios y recursos de AWS.
- Asegúrese de que la configuración de VPC haya configurado correctamente el soporte de la resolución DNS/nombre de host, la gateway de Internet y los parámetros del conjunto de opciones de DHCP. Para obtener más información, consulte Errores de VPC.

Si estos pasos no resuelven el problema, continúe con los pasos para la resolución de los errores de Amazon EMR comunes. Para obtener más información, consulte Errores comunes en Amazon EMR.

Buscar un clúster que falta

Si el clúster no aparece en la lista de consolas o en la API `ListClusters`, compruebe lo siguiente:

- Confirme que la antigüedad del clúster desde el momento de su finalización es inferior a dos meses. Amazon EMR conserva la información de metadatos sobre clústeres completados, gratuitamente, durante dos meses. No puede eliminar los clústeres completados de la consola; en su lugar, Amazon EMR purga los clústeres completados automáticamente después de dos meses.
- Confirme que tiene los permisos de rol para ver el clúster.
- Confirme que está viendo la misma Región de AWS donde reside el clúster.

Solucionar problemas de un clúster con errores

Esta sección le muestra el proceso de resolución de problemas de un clúster que ha generado un error. Esto significa que el clúster terminó con un código de error.

> **Note**
>
> Cuando un clúster de EMR termina con un error, las API `DescribeCluster` y `ListClusters` devuelven un código de error y un mensaje de error. En el caso de algunos errores del clúster, la matriz de datos `ErrorDetail` también puede ayudarle a solucionar el error. Para obtener más información, consulte Códigos de error con información ErrorDetail.

Si el clúster se ejecuta, pero tarda mucho en devolver resultados, consulte Solucionar problemas de un clúster lento.

Temas

- Paso 1: recopilar datos sobre el problema
- Paso 2: comprobar el entorno
- Paso 3: comprobar el último cambio de estado
- Paso 4: examinar los archivos de registro
- Paso 5: comprobar el clúster paso a paso

Paso 1: recopilar datos sobre el problema

El primer paso para solucionar problemas de un clúster es recopilar información sobre lo que ha fallado y el estado y la configuración actuales del clúster. Esta información se utilizará en los siguientes pasos para confirmar o descartar las posibles causas del problema.

Definir el problema

Una definición clara del problema es el primer punto de partida. Algunas preguntas que debe hacerse:

- ¿Qué esperaba que sucediera? ¿Qué pasó en su lugar?
- ¿Cuándo se produjo este problema por primera vez? ¿Con qué frecuencia ha ocurrido desde entonces?

- ¿Ha cambiado algo en la forma en que configuro o ejecuto mi clúster?

Detalles de clúster

Los siguientes detalles del clúster son útiles para ayudar a detectar problemas. Para obtener más información sobre cómo recopilar esta información, consulte Ver el estado y los detalles del clúster.

- El identificador del clúster. (También se denomina identificador de flujo de tareas).
- La Región de AWS y la zona de disponibilidad en la que se lanzó el clúster.
- El estado del clúster, incluidos los detalles del último cambio de estado.
- Tipo y número de instancias de EC2 especificados para los nodos maestro, principal y de tarea.

Paso 2: comprobar el entorno

Amazon EMR opera como parte de un ecosistema de servicios web y software de código abierto. Lo que afecta a dichas dependencias pueden influir en el rendimiento de Amazon EMR.

Temas

- Comprobar las interrupciones de servicio
- Comprobar los límites de uso
- Comprobar la versión
- Comprobar la configuración de subredes de Amazon VPC

Comprobar las interrupciones de servicio

Amazon EMR utiliza varios servicios de Amazon Web Services internamente. Ejecuta servidores virtuales en Amazon EC2, almacena datos y scripts en Amazon S3 e informa de las métricas a CloudWatch. Los eventos que interrumpen estos servicios son poco frecuentes, pero, cuando se producen, pueden provocar problemas en Amazon EMR.

Antes de continuar, compruebe el Panel de estado del servicio. Compruebe la región en la que lanzó el clúster para ver si hay interrupciones en alguno de estos servicios.

Comprobar los límites de uso

Si va a lanzar un clúster grande, ha lanzado varios clústeres simultáneamente o es un usuario que comparte una Cuenta de AWS con otros usuarios, es posible que el clúster haya fallado porque ha superado el límite de servicio de AWS.

Amazon EC2 limita el número de instancias de servidores virtuales que se ejecutan en una sola región AWS a 20 instancias reservadas o bajo demanda. Si lanza un clúster con más de 20 nodos o lanza un clúster que hace que el número total de instancias de EC2 activas en la Cuenta de AWS supere las 20, el clúster no podrá lanzar todas las instancias de EC2 que necesita y podría fallar. Cuando esto ocurre, Amazon EMR devuelve un error `EC2 QUOTA EXCEEDED`. Puede solicitar que AWS aumente el número de instancias de EC2 que puede ejecutar en su cuenta. Para ello, solicite un aumento de límite de instancias de Amazon EC2.

Otro factor que puede provocar que supere los límites de uso es el retraso que transcurre entre la finalización de un clúster y el momento en que libera todos sus recursos. En función de las diferencias de configuración, un clúster puede tardar entre 5 y 20 minutos terminar por completo y liberar los recursos asignados. Si aparece un error `EC2 QUOTA EXCEEDED` al intentar lanzar un clúster, puede deberse a que aún no se hayan liberado los recursos de un clúster terminado recientemente. En este caso, puede solicitar un aumento de la cuota de Amazon EC2 o puede esperar 20 minutos y volver a lanzar el clúster.

Amazon S3 limita el número de buckets creados en una cuenta a 100. Si el clúster crea un bucket nuevo que supera este límite, se producirá un error en la creación del bucket y es posible que el clúster falle.

Comprobar la versión

Compare la etiqueta de versión que ha usado para lanzar el clúster con la última versión de Amazon EMR. Cada versión de Amazon EMR incorpora mejoras, como nuevas aplicaciones, características, parches y errores corregidos. Puede que el problema que afecta a su clúster ya se haya solucionado en la última versión. Si es posible, vuelva a ejecutar el clúster con la versión más reciente.

Comprobar la configuración de subredes de Amazon VPC

Si el clúster se lanzó en una subred de Amazon VPC, la subred debe configurarse como se describe en Configurar redes. Además, compruebe que la subred en la que lanza el clúster tenga suficientes direcciones IP elásticas libres para asignar una a cada nodo del clúster.

Paso 3: comprobar el último cambio de estado

El último cambio de estado ofrece información sobre qué ocurrió la última vez que el clúster cambió de estado. Esto a menudo contiene información capaz de indicar lo que ha funcionado mal cuando el clúster cambia su estado a `FAILED`. Por ejemplo, si lanza un clúster de streaming y especifica una ubicación de salida que ya existe en Amazon S3, se producirá el error con un último cambio de estado “El directorio de salida de streaming ya existe” en el clúster.

Puede localizar el valor del último cambio de estado desde la consola consultando el panel de detalles del clúster, desde la CLI utilizando los argumentos `list-steps` o `describe-cluster` o desde el API utilizando las acciones `DescribeCluster` y `ListSteps`. Para obtener más información, consulte Ver el estado y los detalles del clúster.

Paso 4: examinar los archivos de registro

El siguiente paso consiste en examinar los archivos de registro para encontrar un código de error u otro indicio del problema que ha sufrido el clúster. Para obtener información sobre los archivos de registro disponibles, dónde encontrarlos y cómo verlos, consulte Ver archivos de registro de .

Es posible que sea necesario un poco de trabajo de investigación para determinar qué pasó. Hadoop ejecuta los trabajos en los intentos de tareas en varios nodos del clúster. Amazon EMR puede iniciar intentos de tareas especulativos y terminar los demás intentos de tareas que no se completen primero. Esto genera una actividad importante que se registra en los archivos de registro del controlador, stderr y syslog a medida que se produce. Además, se ejecutan varios intentos de tareas simultáneamente, pero un archivo de registro solo puede mostrar los resultados de forma lineal.

Para comenzar, compruebe los registros de acciones de arranque para ver si hay errores o cambios de configuración inesperados durante el lanzamiento del clúster. A partir de ahí, consulte los registros de pasos para identificar los trabajos de Hadoop lanzados como parte de un paso con errores. Examine los registros de trabajos de Hadoop para identificar los intentos fallidos de tareas. El registro de intentos de tarea contendrá detalles sobre la causa del error de un intento de tarea.

En las siguientes secciones, se describe cómo utilizar los distintos archivos de registro para identificar errores en el clúster.

Comprobar los registros de acción de arranque

Las acciones de arranque ejecutan scripts en el clúster a medida que se lanza. Por lo general, se utilizan para instalar software adicional en el clúster o para modificar los valores predeterminados de

los valores de configuración. La comprobación de estos registros puede proporcionar información sobre los errores que se produjeron durante la configuración del clúster, así como sobre los cambios en los ajustes de configuración que podrían afectar al rendimiento.

Comprobar los registros de pasos

Hay cuatro tipos de registros de pasos.

- controlador: contiene archivos generados por Amazon EMR (Amazon EMR) que se deben a errores encontrados al intentar ejecutar el paso. Si se produce un error en el paso durante la carga, puede encontrar el registro de seguimiento de la pila en este registro. Aquí se describen con frecuencia los errores al cargar la aplicación o al acceder a ella, así como los errores que faltan en el archivo de asignación.
- stderr: contiene los mensajes de error que se produjeron al procesar el paso. Los errores de carga de la aplicación se describen a menudo aquí. En ocasiones, este registro contiene un seguimiento de pila.
- stdout: contiene el estado generado por los ejecutables de asignación y reducción. Los errores de carga de la aplicación se describen a menudo aquí. En ocasiones, este registro contiene mensajes de error de la aplicación.
- syslog: contiene registros de software ajeno a Amazon, como Apache y Hadoop. Los errores de streaming suelen describirse aquí.

Compruebe si hay errores obvios en stderr. Si stderr muestra una lista corta de errores, el paso se detuvo rápidamente y se produjo un error. En la mayoría de los casos, esto se debe a un error en las aplicaciones de asignación y reducción que se ejecutan en el clúster.

Examine las últimas líneas del controlador y de syslog en busca de avisos de errores. Siga cualquier aviso sobre tareas con errores, especialmente si dice “Trabajo con errores”.

Comprobar los registros de intento de tarea

Si el análisis anterior de los registros de pasos reveló una o más tareas fallidas, investigue los registros de los intentos de tareas correspondientes para obtener información de error más detallada.

Paso 5: comprobar el clúster paso a paso

Una técnica útil cuando se intenta localizar el origen de un error consiste en reiniciar el clúster y enviar los pasos al mismo uno a uno. Esto le permite comprobar los resultados de cada paso antes

de procesar el siguiente y le ofrece la oportunidad de corregir y volver a ejecutar un paso que ha fallado. Esto también tiene la ventaja de cargar los datos de entrada solo una vez.

Para probar un clúster paso a paso

1. Lance un clúster nuevo, con keep alive y la protección de terminación habilitados. Keep alive mantiene el clúster en ejecución después de que haya procesado todos los pasos pendientes. La protección de terminación evita que un clúster se cierre en caso de que se produzca un error. Para obtener más información, consulte Configuración de un clúster para que continúe o termine después de la ejecución de pasos y Uso de la protección de terminación.
2. Envíe un paso al clúster. Para obtener más información, consulte Enviar trabajo a un clúster.
3. Cuando el paso completa el procesamiento, compruebe los errores en los archivos de registro del paso. Para obtener más información, consulte Paso 4: examinar los archivos de registro. La manera más rápida de localizar estos archivos de registro consiste en conectar al nodo maestro y visualizar los archivos de registro allí. Los archivos de registro del paso no aparecen hasta que el paso se ejecuta durante algún tiempo, finaliza o genera un error.
4. Si el paso se realiza correctamente sin error, ejecutar el siguiente paso. Si se produjeron errores, investigar el error en los archivos de registro. Si se trató de un error en el código, corríjalo y vuelva a ejecutar el paso. Continúe hasta que todos los pasos se ejecuten sin errores.
5. Cuando haya terminado la depuración del clúster y desea terminarlo, tendrá que terminarlo manualmente. Esto es necesario porque el clúster se lanzó con la protección de terminación habilitada. Para obtener más información, consulte Uso de la protección de terminación.

Solucionar problemas de un clúster lento

Esta sección describe el proceso de solución de problemas de un clúster que sigue en ejecución, pero que tarda mucho en devolver los resultados. Para obtener más información sobre qué hacer si el clúster ha terminado con un código de error, consulte Solucionar problemas de un clúster con errores

Amazon EMR le permite especificar el número y el tipo de instancias en el clúster. Estas especificaciones son los medios principales que afectan a la velocidad con que la que se completa el procesamiento de datos. Una cosa que debería tener en cuenta es volver a ejecutar el clúster, esta vez especificando instancias EC2 con más recursos o especificar un número mayor de instancias en el clúster. Para obtener más información, consulte Configuración del hardware y las redes de los clústeres.

Los siguientes temas le guiarán a través del proceso de identificación de causas alternativas de un clúster lento.

Temas

- Paso 1: recopilar datos sobre el problema
- Paso 2: comprobar el entorno
- Paso 3: examinar los archivos de registro
- Paso 4: comprobar el clúster y el estado de la instancia
- Paso 5: comprobar si hay grupos suspendidos
- Paso 6: revisar los ajustes de configuración
- Paso 7: examinar los datos de entrada

Paso 1: recopilar datos sobre el problema

El primer paso para solucionar problemas de un clúster es recopilar información sobre lo que ha fallado y el estado y la configuración actuales del clúster. Esta información se utilizará en los siguientes pasos para confirmar o descartar las posibles causas del problema.

Definir el problema

Una definición clara del problema es el primer punto de partida. Algunas preguntas que debe hacerse:

- ¿Qué esperaba que sucediera? ¿Qué pasó en su lugar?
- ¿Cuándo se produjo este problema por primera vez? ¿Con qué frecuencia ha ocurrido desde entonces?
- ¿Ha cambiado algo en la forma en que configuro o ejecuto mi clúster?

Detalles de clúster

Los siguientes detalles del clúster son útiles para ayudar a detectar problemas. Para obtener más información sobre cómo recopilar esta información, consulte Ver el estado y los detalles del clúster.

- El identificador del clúster. (También se denomina identificador de flujo de tareas).
- La Región de AWS y la zona de disponibilidad en la que se lanzó el clúster.
- El estado del clúster, incluidos los detalles del último cambio de estado.

- Tipo y número de instancias de EC2 especificados para los nodos maestro, principal y de tarea.

Paso 2: comprobar el entorno

Temas

- Comprobar las interrupciones de servicio
- Comprobar los límites de uso
- Comprobar la configuración de subredes de Amazon VPC
- Reiniciar el clúster

Comprobar las interrupciones de servicio

Amazon EMR utiliza varios servicios de Amazon Web Services internamente. Ejecuta servidores virtuales en Amazon EC2, almacena datos y scripts en Amazon S3 e informa de las métricas a CloudWatch. Los eventos que interrumpen estos servicios son poco frecuentes, pero, cuando se producen, pueden provocar problemas en Amazon EMR.

Antes de continuar, compruebe el Panel de estado del servicio. Compruebe la región en la que lanzó el clúster para ver si hay interrupciones en alguno de estos servicios.

Comprobar los límites de uso

Si va a lanzar un clúster grande, ha lanzado varios clústeres simultáneamente o es un usuario que comparte una Cuenta de AWS con otros usuarios, es posible que el clúster haya fallado porque ha superado el límite de servicio de AWS.

Amazon EC2 limita el número de instancias de servidores virtuales que se ejecutan en una sola región AWS a 20 instancias reservadas o bajo demanda. Si lanza un clúster con más de 20 nodos o lanza un clúster que hace que el número total de instancias de EC2 activas en la Cuenta de AWS supere las 20, el clúster no podrá lanzar todas las instancias de EC2 que necesita y podría fallar. Cuando esto ocurre, Amazon EMR devuelve un error `EC2 QUOTA EXCEEDED`. Puede solicitar que AWS aumente el número de instancias de EC2 que puede ejecutar en su cuenta. Para ello, solicite un aumento de límite de instancias de Amazon EC2.

Otro factor que puede provocar que supere los límites de uso es el retraso que transcurre entre la finalización de un clúster y el momento en que libera todos sus recursos. En función de las diferencias de configuración, un clúster puede tardar entre 5 y 20 minutos terminar por completo

y liberar los recursos asignados. Si aparece un error `EC2 QUOTA EXCEEDED` al intentar lanzar un clúster, puede deberse a que aún no se hayan liberado los recursos de un clúster terminado recientemente. En este caso, puede solicitar un aumento de la cuota de Amazon EC2 o puede esperar 20 minutos y volver a lanzar el clúster.

Amazon S3 limita el número de buckets creados en una cuenta a 100. Si el clúster crea un bucket nuevo que supera este límite, se producirá un error en la creación del bucket y es posible que el clúster falle.

Comprobar la configuración de subredes de Amazon VPC

Si el clúster se lanzó en una subred de Amazon VPC, la subred debe configurarse como se describe en Configurar redes. Además, compruebe que la subred en la que lanza el clúster tenga suficientes direcciones IP elásticas libres para asignar una a cada nodo del clúster.

Reiniciar el clúster

La ralentización de procesamiento puede deberse a una condición transitoria. Plantéese terminar y reiniciar el clúster para ver si el rendimiento mejora.

Paso 3: examinar los archivos de registro

El siguiente paso consiste en examinar los archivos de registro para encontrar un código de error u otro indicio del problema que ha sufrido el clúster. Para obtener información sobre los archivos de registro disponibles, dónde encontrarlos y cómo verlos, consulte Ver archivos de registro de .

Es posible que sea necesario un poco de trabajo de investigación para determinar qué pasó. Hadoop ejecuta los trabajos en los intentos de tareas en varios nodos del clúster. Amazon EMR puede iniciar intentos de tareas especulativos y terminar los demás intentos de tareas que no se completen primero. Esto genera una actividad importante que se registra en los archivos de registro del controlador, stderr y syslog a medida que se produce. Además, se ejecutan varios intentos de tareas simultáneamente, pero un archivo de registro solo puede mostrar los resultados de forma lineal.

Para comenzar, compruebe los registros de acciones de arranque para ver si hay errores o cambios de configuración inesperados durante el lanzamiento del clúster. A partir de ahí, consulte los registros de pasos para identificar los trabajos de Hadoop lanzados como parte de un paso con errores. Examine los registros de trabajos de Hadoop para identificar los intentos fallidos de tareas. El registro de intentos de tarea contendrá detalles sobre la causa del error de un intento de tarea.

En las siguientes secciones, se describe cómo utilizar los distintos archivos de registro para identificar errores en el clúster.

Comprobar los registros de acción de arranque

Las acciones de arranque ejecutan scripts en el clúster a medida que se lanza. Por lo general, se utilizan para instalar software adicional en el clúster o para modificar los valores predeterminados de los valores de configuración. La comprobación de estos registros puede proporcionar información sobre los errores que se produjeron durante la configuración del clúster, así como sobre los cambios en los ajustes de configuración que podrían afectar al rendimiento.

Comprobar los registros de pasos

Hay cuatro tipos de registros de pasos.

- controlador: contiene archivos generados por Amazon EMR (Amazon EMR) que se deben a errores encontrados al intentar ejecutar el paso. Si se produce un error en el paso durante la carga, puede encontrar el registro de seguimiento de la pila en este registro. Aquí se describen con frecuencia los errores al cargar la aplicación o al acceder a ella, así como los errores que faltan en el archivo de asignación.
- stderr: contiene los mensajes de error que se produjeron al procesar el paso. Los errores de carga de la aplicación se describen a menudo aquí. En ocasiones, este registro contiene un seguimiento de pila.
- stdout: contiene el estado generado por los ejecutables de asignación y reducción. Los errores de carga de la aplicación se describen a menudo aquí. En ocasiones, este registro contiene mensajes de error de la aplicación.
- syslog: contiene registros de software ajeno a Amazon, como Apache y Hadoop. Los errores de streaming suelen describirse aquí.

Compruebe si hay errores obvios en stderr. Si stderr muestra una lista corta de errores, el paso se detuvo rápidamente y se produjo un error. En la mayoría de los casos, esto se debe a un error en las aplicaciones de asignación y reducción que se ejecutan en el clúster.

Examine las últimas líneas del controlador y de syslog en busca de avisos de errores. Siga cualquier aviso sobre tareas con errores, especialmente si dice “Trabajo con errores”.

Comprobar los registros de intento de tarea

Si el análisis anterior de los registros de pasos reveló una o más tareas fallidas, investigue los registros de los intentos de tareas correspondientes para obtener información de error más detallada.

Comprobar los registros de daemon de Hadoop

En raras ocasiones, Hadoop podría fallar. Para comprobar si ese es el caso, consulte los registros de Hadoop. Están ubicados en `/var/log/hadoop/` en cada nodo.

Puede utilizar los registros de JobTracker para asignar un intento de tarea fallido al nodo en el que se ejecutó. Una vez que conozca el nodo asociado al intento de tarea, puede comprobar el estado de la instancia de EC2 que aloja ese nodo para comprobar si se ha producido algún problema, por ejemplo, si se ha quedado sin CPU o memoria.

Paso 4: comprobar el clúster y el estado de la instancia

Un clúster de Amazon EMR se compone de nodos que se ejecutan en instancias de Amazon EC2. Si dichas instancias se ven limitadas por los recursos (como, por ejemplo, quedarse sin CPU o memoria), tienen problemas de conectividad de red o se terminan, la velocidad de procesamiento del clúster se resiente.

Existen hasta tres tipos de nodos en un clúster:

- nodo maestro: administra el clúster. Si experimenta problemas de rendimiento, se ve afectado todo el clúster.
- nodos principales: procesan tareas de map-reduce y mantienen Hadoop Distributed FileSystem (HDFS). Si uno de estos nodos experimenta un problema de rendimiento, puede ralentizar las operaciones de HDFS, así como el procesamiento de MapReduce. Puede añadir más nodos secundarios a un clúster para mejorar el rendimiento, pero no puede eliminar nodos secundarios. Para obtener más información, consulte Cambiar manualmente el tamaño de un clúster en ejecución.
- nodos de tareas: procesan tareas map-reduce. Se trata exclusivamente de recursos informáticos y no almacenan datos. Puede añadir nodos de tareas a un clúster para acelerar el rendimiento o eliminar los nodos de tareas que no sean necesarios. Para obtener más información, consulte Cambiar manualmente el tamaño de un clúster en ejecución.

Al examinar el estado de un clúster, debe examinar tanto el rendimiento global del clúster, así como el rendimiento de instancias concretas. Existen varias herramientas que puede utilizar:

Comprobar el estado del clúster con CloudWatch

Cada clúster de Amazon EMR notifica las métricas a CloudWatch. Estas métricas proporcionan información sobre el rendimiento de resumen acerca del clúster, como la carga total, utilización de HDFS, ejecución de tareas, tareas restante, bloques corruptos, etc. Al consultar las métricas de CloudWatch, obtiene el panorama general de lo que ocurre en el clúster y puede obtener información sobre los motivos por los que se ralentiza su procesamiento. Además de utilizar CloudWatch para analizar un problema de rendimiento existente, puede definir alarmas que hagan que CloudWatch avise en caso de que se produzca un problema de rendimiento. Para obtener más información, consulte Monitorización de métricas de Amazon EMR con CloudWatch.

Comprobar el estado del trabajo y de HDFS

Utilice la pestaña Interfaces de usuario de aplicaciones de la página de detalles del clúster para ver los detalles de las aplicaciones de YARN. Para determinadas aplicaciones, puede consultar información adicional y tener acceso a los logs directamente. Esto resulta especialmente útil para las aplicaciones Spark. Para obtener más información, consulte Ver el historial de aplicaciones.

Hadoop proporciona una serie de interfaces web que puede utilizar para ver información. Para obtener más información sobre cómo acceder a estas interfaces web, consulte Ver las interfaces web alojadas en clústeres de Amazon EMR.

- JobTracker: proporciona información sobre el progreso del trabajo que procesa el clúster. Puede utilizar esta interfaz para identificar si se ha bloqueado un trabajo.
- NameNode de HDFS: proporciona información sobre el porcentaje de uso de HDFS y el espacio disponible en cada nodo. Puede utilizar esta interfaz para identificar cuando HDFS se ve limitado por los recursos y requiere capacidad adicional.
- TaskTracker: proporciona información sobre las tareas del trabajo que procesa el clúster. Puede utilizar esta interfaz para identificar cuando se ha bloqueado una tarea.

Comprobar el estado de la instancia con Amazon EC2

Otra forma de buscar información sobre el estado de las instancias en el clúster consiste en utilizar la consola de Amazon EC2. Dado que cada nodo del clúster se ejecuta en una instancia de EC2, puede utilizar las herramientas proporcionadas por Amazon EC2 para comprobar su estado. Para obtener más información, consulte Ver instancias del clúster en Amazon EC2.

Paso 5: comprobar si hay grupos suspendidos

Un grupo de instancias queda suspendido cuando encuentra demasiados errores al intentar lanzar nodos. Por ejemplo, si nodos nuevos devuelven error repetidamente al llevar a cabo acciones de arranque, el grupo de instancias, después de un tiempo, pasará al estado `SUSPENDED` en lugar intentar de forma continua aprovisionar nuevos nodos.

Un nodo podría no cargarse si:

- Hadoop o el clúster están estropeados por algún motivo y no aceptan un nuevo nodo en el clúster
- Una acción de arranque falla en el nuevo nodo
- El nodo no funciona correctamente y no puede iniciar sesión con Hadoop

Si un grupo de instancias está en estado `SUSPENDED` y el clúster está en estado `WAITING`, puede añadir un paso de clúster para restablecer el número deseado de nodos secundarios y de tareas. Al añadir el paso se reanuda el procesamiento del clúster y coloca el grupo de instancias de nuevo en estado `RUNNING`.

Para obtener más información sobre cómo restablecer un clúster en estado suspendido, consulte Estado de suspensión.

Paso 6: revisar los ajustes de configuración

Los ajustes de configuración especifican detalles acerca de cómo se ejecuta un clúster, como cuántas veces se vuelve a intentar una tarea y la cantidad de memoria que hay disponible para clasificación. Al lanzar un clúster con Amazon EMR, existen ajustes específicos de Amazon EMR además de los ajustes de configuración estándar de Hadoop. Los ajustes de configuración se almacenan en el nodo principal del clúster. Puede comprobar los ajustes de configuración para asegurarse de que su clúster tenga los recursos que necesita para ejecutarse de forma eficaz.

Amazon EMR define los ajustes de configuración de Hadoop predeterminados que utiliza para lanzar un clúster. Los valores se basan en la AMI y el tipo de instancia que especifique para el clúster. Puede modificar los ajustes de configuración a partir de los valores predeterminados mediante una acción de arranque o especificando nuevos valores en parámetros de ejecución de trabajo. Para obtener más información, consulte Crear acciones de arranque para instalar software adicional. Para determinar si una acción de arranque ha cambiado los ajustes de configuración, compruebe los registros de la acción de arranque.

Amazon EMR registra los ajustes de Hadoop utilizados para ejecutar cada trabajo. Los datos de registro se almacenan en un archivo denominado `job_job-id_conf.xml` en el directorio `/mnt/var/log/hadoop/history/` del nodo principal, donde *job-id* se sustituye por el identificador del trabajo. Si ha habilitado el archivado de registros, estos datos se copian a Amazon S3 en la carpeta `logs/date/jobflow-id/jobs`, donde *fecha* es la fecha en que se ejecutó el trabajo y *jobflow-id* es el identificador del clúster.

Los siguientes ajustes de configuración de trabajo de Hadoop son especialmente útiles para investigar los problemas de rendimiento. Para obtener más información acerca de los ajustes de configuración de Hadoop y cómo afectan al comportamiento de Hadoop, visite http://hadoop.apache.org/docs/.

Warning

1. Establecer `dfs.replication` en 1 en clústeres con menos de cuatro nodos puede conllevar la pérdida de datos del HDFS si un solo nodo deja de funcionar. Se recomienda que utilice un clúster con al menos cuatro nodos principales para las cargas de trabajo de producción.
2. Amazon EMR no permitirá que los clústeres escalen los nodos principales por debajo de `dfs.replication`. Por ejemplo, si `dfs.replication = 2`, el número mínimo de nodos principales es 2.
3. Cuando utiliza el escalado administrado, el escalado automático o decide cambiar el tamaño del clúster manualmente, se recomienda que establezca `dfs.replication` en 2 o más.

Opción de configuración	Descripción
dfs.replication	El número de nodos de HDFS en los que un bloque único (como el bloque de disco duro) se copian para producir un entorno similar a RAID. Determina el número de nodos de HDFS que contienen una copia del bloque.
io.sort.mb	Memoria total disponible para clasificación. Este valor debería ser 10x io.sort.factor. Este ajuste también puede utilizarse para calcular la memoria total utilizada por el

Opción de configuración	Descripción
	nodo de tareas calculando io.sort.mb multiplicado por mapred.tasktracker.ap.tasks.maximum.
io.sort.spill.percent	Utilizado durante la clasificación, momento en que el disco empezará a utilizarse porque la memoria de clasificación asignada se está llenando.
mapred.child.java.opts	Obsoleto. Utilice mapred.map.child.java.opts y mapred.reduce.child.java.opts en su lugar. Las opciones de Java TaskTracker se utiliza al ejecutar una JVM para una tarea que se ejecuta dentro. Un parámetro común es “-Xmx” para configurar el tamaño de memoria máximo.
mapred.map.child.java.opts	Las opciones de Java TaskTracker se utiliza al ejecutar una JVM para una tarea Map que se ejecuta dentro. Un parámetro común es “-Xmx” para configurar el tamaño de montón de memoria máximo.
mapred.map.tasks.speculative.execution	Determina si los intentos de tarea Map de la misma tarea se pueden lanzar en paralelo.
mapred.reduce.tasks.speculative.execution	Determina si los intentos de tarea Reduce de la misma tarea se pueden lanzar en paralelo.
mapred.map.max.attempts	El número máximo de veces que se puede intentar una tarea Map. Si todos fallan, entonces la tarea Map se marca como error.
mapred.reduce.child.java.opts	Las opciones de Java TaskTracker se utiliza al ejecutar una JVM para una tarea Reduce que se ejecuta dentro. Un parámetro común es “-Xmx” para configurar el tamaño de montón de memoria máximo.
mapred.reduce.max.attempts	El número máximo de veces que se puede intentar una tarea Reduce. Si todos fallan, entonces la tarea Map se marca como error.

Opción de configuración	Descripción
mapred.reduce.slowstart.completed.maps	La cantidad de tareas Map que deben completar antes de intentar las tareas Reduce. Si no se espera lo suficiente, se podrían devolver errores “Demasiados errores de recuperación” en los intentos.
mapred.reuse.jvm.num.tasks	Una tarea se ejecuta dentro de una única JVM. Especifica cuántas tareas podría reutilizar la misma JVM.
mapred.tasktracker.map.tasks.maximum	La cantidad máxima de tareas que se pueden ejecutar en paralelo por nodo de tareas durante el mapeo.
mapred.tasktracker.reduce.tasks.maximum	La cantidad máxima de tareas que se pueden ejecutar en paralelo por nodo de tareas durante la reducción.

Si las tareas de clúster utilizan mucha memoria, puede mejorar el rendimiento utilizando menos tareas por nodo secundario y reduciendo el tamaño de montón de rastreador de trabajos.

Paso 7: examinar los datos de entrada

Compruebe los datos de entrada. ¿Están distribuidos de manera uniforme entre los valores clave? Si los datos están muy sesgados hacia uno o varios valores clave, la carga de procesamiento podría estar asignada a un pequeño número de nodos, mientras que los demás nodos están inactivos. Esta distribución desequilibrada de trabajo puede dar lugar a tiempos de procesamiento más lentos.

Un ejemplo de conjunto de datos desequilibrado sería la ejecución de un clúster para alfabetizar palabras, pero disponer de un conjunto de datos que contenga solo palabras que comienzan con la letra "a". Cuando el trabajo se ha planificado, los valores de procesamiento del nodo que comienzan por "a" serían abrumadores, mientras que los nodos que procesan palabras que comienzan por otras letras estarían inactivos.

Solucionar problemas de un clúster de Lake Formation

Esta sección le guía por el proceso de resolución de problemas comunes al utilizar Amazon EMR con AWS Lake Formation.

No se permite el acceso al lago de datos

Debe optar explícitamente por el filtrado de datos en los clústeres de Amazon EMR antes de poder analizar y procesar los datos de su lago de datos. Cuando se produzca un error en el acceso a los datos, verá un mensaje `Access is not allowed` genérico en el resultado de las entradas de su bloc de notas.

Para activar y permitir el filtrado de datos en Amazon EMR, consulte Permitir el filtrado de datos en Amazon EMR en la Guía para desarrolladores de AWS Lake Formation para obtener instrucciones.

Vencimiento de la sesión

El tiempo de espera para EMR Notebooks y Zeppelin se controla mediante el rol de IAM para la configuración `Maximum CLI/API session duration` de Lake Formation. El valor predeterminado para esta configuración es de una hora. Cuando se agote el tiempo de espera de una sesión, verá el siguiente mensaje en el resultado de las entradas de su bloc de notas al intentar ejecutar los comandos de Spark SQL.

```
Error 401    HTTP ERROR: 401 Problem accessing /sessions/2/statements.
Reason:  JWT token included in request failed validation.
Powered by Jetty:// 9.3.24.v20180605
 org.springframework.web.client.HttpClientErrorException: 401 JWT token included in
 request failed validation…
```

Para validar su sesión actualice la página. Se le pedirá que vuelva a autenticarse con su IdP y se le redirigirá de vuelta al bloc de notas. Puede continuar ejecutando consultas después de la segunda autenticación.

No hay permisos para el usuario en la tabla solicitada

Al intentar acceder a una tabla a la que no tiene acceso, verá la siguiente excepción en el resultado de las entradas de su bloc de notas al intentar ejecutar comandos de Spark SQL.

```
org.apache.spark.sql.AnalysisException:
 org.apache.hadoop.hive.ql.metadata.HiveException: Unable to fetch table table.
Resource does not exist or requester is not authorized to access requested
 permissions.
(Service: AWSGlue; Status Code: 400; Error Code: AccessDeniedException; Request ID: …
```

Para obtener acceso a la tabla, debe conceder acceso al usuario. Para ello, actualice los permisos asociados a esta tabla en Lake Formation.

Consulta de datos entre cuentas compartidos con Lake Formation

Al usar Amazon EMR para acceder a los datos compartidos desde otra cuenta, algunas bibliotecas de Spark intentarán llamar a la operación de la API `Glue:GetUserDefinedFunctions`. Como las versiones 1 y 2 de los permisos administrados de AWS RAM no admiten esta acción, obtiene el siguiente mensaje de error:

```
"ERROR: User: arn:aws:sts::012345678901:assumed-role/my-
spark-role/i-06ab8c2b59299508a is not authorized to perform:
glue:GetUserDefinedFunctions on resource: arn:exampleCatalogResource
because no resource-based policy allows the glue:GetUserDefinedFunctions
action"
```

Para resolver este error, el administrador del lago de datos que creó el recurso compartido debe actualizar los permisos administrados de AWS RAM asociados al recurso compartido. La versión 3 de los permisos administrados de AWS RAM permite a los directores llevar a cabo la acción `glue:GetUserDefinedFunctions`.

Si crea un nuevo recurso compartido, Lake Formation aplica la última versión del permiso administrado por AWS RAM de forma predeterminada y no es necesario que haga nada. Para habilitar el acceso a los datos entre cuentas para los recursos compartidos existentes, debe actualizar los permisos administrados de AWS RAM a la versión 3.

Puede ver los permisos de AWS RAM asignados a los recursos compartidos en AWS RAM. Los siguientes permisos se incluyen en la versión 3:

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Para actualizar la versión de los permisos administrados de AWS RAM de los recursos compartidos existentes

Como administrador del lago de datos, puede actualizar los permisos administrados de AWS RAM a una versión más reciente. Para ello, siga las instrucciones de la Guía del usuario de AWS RAM o revoque todos los permisos existentes para el tipo de recurso y vuelva a concederlos. Si revoca los permisos, AWS RAM elimina el recurso compartido de AWS RAM asociado al tipo de recurso. Al volver a conceder los permisos, AWS RAM adjunta la versión más reciente de los permisos administrados de AWS RAM para crear nuevos recursos compartidos.

Inserción, creación y alteración de tablas

No se admite la inserción, creación ni alteración de tablas en las bases de datos protegidas por políticas de Lake Formation. Al realizar estas operaciones, verá la siguiente excepción en el resultado de las entradas de su bloc de notas al intentar ejecutar los comandos de Spark SQL:

```
java.io.IOException:
 com.amazon.ws.emr.hadoop.fs.shaded.com.amazonaws.services.s3.model.AmazonS3Exception:
           Access Denied (Service: Amazon S3; Status Code: 403; Error Code:
 AccessDenied; Request ID: …
```

Para obtener más información, consulte Limitaciones de la integración de Amazon EMR con AWS Lake Formation.

Escribir aplicaciones que lanzan y administran clústeres

Temas

- Muestra completa de código fuente de Java de Amazon EMR
- Conceptos comunes para las llamadas a la API
- Utilizar los SDK para llamar a las API de Amazon EMR
- Administrar las Service Quotas de Amazon EMR

Puede acceder a la funcionalidad proporcionada por la API de Amazon EMR llamando a las funciones contenedoras de uno de los SDK de AWS. Los SDK de AWS proporcionan funciones específicas del lenguaje que contienen la API del servicios web y simplifican la conexión al servicio web, gestionando muchos de los detalles de conexión para usted. Para obtener más información sobre cómo llamar a Amazon EMR; mediante uno de los SDK, consulte Utilizar los SDK para llamar a las API de Amazon EMR.

> **Important**
>
> La máxima velocidad de solicitudes para Amazon EMR; es una solicitud cada diez segundos.

Muestra completa de código fuente de Java de Amazon EMR

Los desarrolladores pueden llamar a la API de Amazon EMR mediante código Java personalizado para hacer lo mismo que con la consola de Amazon EMR o la CLI. Esta sección proporciona todos los pasos necesarios para instalar el AWS Toolkit for Eclipse y ejecutar un ejemplo de código fuente de Java totalmente funcional para agregar pasos a un clúster de Amazon EMR.

> **Note**
>
> Este ejemplo se centra en Java, pero Amazon EMR también admite varios lenguajes de programación a través de la colección de SDK de Amazon EMR. Para obtener más información, consulte Utilizar los SDK para llamar a las API de Amazon EMR.

Este código fuente de Java de ejemplo muestra cómo realizar las siguientes tareas a través de la API de Amazon EMR:

- Recuperar las credenciales de AWS y enviarlas a Amazon EMR para realizar llamadas a la API
- Configurar un nuevo paso personalizado y un nuevo paso predefinido
- Agregar nuevos pasos a un clúster de Amazon EMR existente
- Recuperar los ID de paso de clúster de un clúster en ejecución

> **Note**
>
> Este ejemplo muestra cómo añadir pasos a un clúster existente y requiere que tenga un clúster activo en su cuenta.

Antes de comenzar, instale una versión del Eclipse IDE for Java EE Developers (IDE de Eclipse para Java EE Developers) que coincida con su plataforma informática. Para obtener más información, consulte Eclipse Downloads.

A continuación, instale el complemento Database Development para Eclipse.

Para instalar el complemento Database Development para Eclipse

1. Abra el IDE de Eclipse.
2. Elija Help (Ayuda) e Install New Software (Instalar software nuevo).
3. En el campo Work with (Trabajar con), escriba **`http://download.eclipse.org/releases/kepler`** o la ruta que coincida con el número de versión de su IDE de Eclipse.
4. En la lista de elementos, elija Database Development (Desarrollo de bases de datos) y Finish (Finalizar).
5. Reinicie Eclipse cuando se le solicite.

A continuación, instale el Kit de herramientas para Eclipse para hacer que estén disponibles las útiles plantillas de proyecto de código fuente preconfiguradas.

Para instalar el Kit de herramientas para Eclipse

1. Abra el IDE de Eclipse.
2. Elija Help (Ayuda) e Install New Software (Instalar software nuevo).
3. En el campo Work with (Trabajar con), escriba **`https://aws.amazon.com/eclipse`**.

4. En la lista de elementos, seleccione AWS Toolkit for Eclipse y Terminar.
5. Reinicie Eclipse cuando se le solicite.

A continuación, cree un nuevo proyecto de Java de AWS y ejecute el código fuente de Java de ejemplo.

Para crear un nuevo proyecto de Java de AWS

1. Abra el IDE de Eclipse.
2. Elija File (Archivo), New (Nuevo) y Other (Otros).
3. En el cuadro de diálogo Seleccionar un asistente, seleccione Proyecto de Java de AWS y Siguiente.
4. En el cuadro de diálogo Nuevo proyecto de Java de AWS, en el campo **`Project name:`**, introduzca el nombre del proyecto nuevo, por ejemplo, **`EMR-sample-code`**.
5. Seleccione Configurar cuentas de AWS, introduzca sus claves de acceso públicas y privadas y seleccione Terminar. Para obtener más información sobre cómo crear las claves de acceso, consulte ¿Cómo obtengo credenciales de seguridad? en la Referencia general de Amazon Web Services.

   > **Note**
   >
   > No debe incrustar las claves de acceso directamente en el código. El SDK de Amazon EMR le permite colocar las claves de acceso en ubicaciones conocidas para que no tenga que mantenerlas en el código.

6. En el proyecto de Java nuevo, haga clic con el botón derecho en la carpeta src y, a continuación, elija New (Crear) y Class (Clase).
7. En el cuadro de diálogo Java Class (Clase de Java), en el campo Name (Nombre), introduzca un nombre para la clase nueva, por ejemplo **`main`**.
8. En la sección Which method stubs would you like to create? (¿Qué stubs de método le gustaría crear?), elija public static void main(String[] args) y Finish (Finalizar).
9. Escriba el código fuente de Java dentro de la clase nueva y añada las instrucciones import adecuadas para las clases y los métodos del ejemplo. Para su comodidad se muestra a continuación el código fuente completo.

> **Note**
>
> En el siguiente código de ejemplo, sustituya el ID de clúster de ejemplo (JobFlowId) *j-xxxxxxxxxxxx*, por un ID de clúster válido de su cuenta que encontrará en la AWS Management Console o utilizando el siguiente comando de la AWS CLI:
>
> ```
> aws emr list-clusters --active | grep "Id"
> ```
>
> Además, sustituya la ruta de Amazon S3 de ejemplo, *s3://path/to/my/jarfolder*, por la ruta válida a su archivo JAR. Por último, sustituya el nombre de clase de ejemplo *com.my.Main1* por el nombre correcto de la clase en su JAR, si procede.

```
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();

  // Run a bash script using a predefined step in the StepFactory helper class
     StepFactory stepFactory = new StepFactory();
     StepConfig runBashScript = new StepConfig()
       .withName("Run a bash script")
       .withHadoopJarStep(stepFactory.newScriptRunnerStep("s3://jeffgoll/emr-
scripts/create_users.sh"))
       .withActionOnFailure("CONTINUE");

     // Run a custom jar file as a step
     HadoopJarStepConfig hadoopConfig1 = new HadoopJarStepConfig()
        .withJar("s3://path/to/my/jarfolder") // replace with the location of the
 jar to run as a step
        .withMainClass("com.my.Main1") // optional main class, this can be omitted
 if jar above has a manifest
        .withArgs("--verbose"); // optional list of arguments to pass to the jar
     StepConfig myCustomJarStep = new StepConfig("RunHadoopJar", hadoopConfig1);

     AddJobFlowStepsResult result = emr.addJobFlowSteps(new
 AddJobFlowStepsRequest()
    .withJobFlowId("j-xxxxxxxxxxxx") // replace with cluster id to run the steps
    .withSteps(runBashScript,myCustomJarStep));

          System.out.println(result.getStepIds());

 }
}
```

10. Elija Run (Ejecutar), Run As (Ejecutar como) y Java Application (Aplicación de Java).
11. Si la muestra se ejecuta correctamente, aparece una lista de ID para los nuevos pasos aparecerá en la ventana de la consola Eclipse IDE. El resultado correcto es similar al siguiente:

```
[s-39BLQZRJB2E5E, s-1L6A4ZU2SAURC]
```

Conceptos comunes para las llamadas a la API

Temas

- Puntos de conexión para Amazon EMR
- Especificar parámetros de clúster en Amazon EMR

- Zonas de disponibilidad en Amazon EMR
- Cómo utilizar archivos y bibliotecas adicionales en clústeres de Amazon EMR

Al escribir una aplicación que llame a la API de Amazon EMR, existen varios conceptos que se aplican a la hora de llamar a una de las funciones contenedoras de un SDK.

Puntos de conexión para Amazon EMR

Un punto de enlace es una URL que es el punto de entrada de un servicio web. Cada solicitud de servicio web debe contener un punto de enlace. El punto de conexión especifica la región de AWS en la que se crean, se describen o se terminan los clústeres. Tiene el formulario `elasticmapreduce.`*`regionname`*`.amazonaws.com`. Si especifica el punto de conexión general (`elasticmapreduce.amazonaws.com`), Amazon EMR dirige la solicitud a un punto de conexión de la región predeterminada. Para las cuentas creadas el 8 de marzo de 2013 o después de esa fecha, la región predeterminada es us-west-2; en el caso de cuentas más antiguas, la región predeterminada es us-east-1.

Para obtener más información acerca de los puntos de conexión de Amazon EMR, consulte Regiones y puntos de conexión en la Referencia general de Amazon Web Services.

Especificar parámetros de clúster en Amazon EMR

Los parámetros `Instances` le permiten configurar el tipo y el número de instancias EC2 para crear nodos para procesar los datos. Hadoop reparte el procesamiento de los datos entre varios nodos del clúster. El nodo principal es responsable de realizar un seguimiento del estado de los nodos secundarios y de tareas y de sondear los nodos para conocer el estado del resultado de los trabajos. Los nodos secundarios y de tareas hacen el procesamiento real de los datos. Si tiene un clúster de un solo nodo, el nodo sirve tanto como nodo principal y nodo secundario.

El parámetro `KeepJobAlive` en una solicitud `RunJobFlow` determina si se debe terminar el clúster cuando se queda sin pasos de clúster que ejecutar. Defina este valor en `False` cuando sepa que el clúster se ejecuta según lo previsto. Al resolver problemas del flujo de trabajo y añadir pasos mientras la ejecución del clúster se suspende, defina el valor en `True`. Esto reduce la cantidad de tiempo y los gastos necesarios para cargar los resultados en Amazon Simple Storage Service (Amazon S3), solo para repetir el proceso después de modificar un paso para reiniciar el clúster.

Si `KeepJobAlive` es `true`, después de obtener correctamente el clúster para completar su trabajo, debe enviar una solicitud `TerminateJobFlows` o el clúster sigue en ejecución y genera cargos de AWS.

Para obtener más información sobre los parámetros que son únicos para `RunJobFlow`, consulte RunJobFlow. Para obtener más información acerca de los parámetros genéricos en la solicitud, consulte Parámetros de solicitud comunes.

Zonas de disponibilidad en Amazon EMR

Amazon EMR utiliza instancias de EC2 como nodos para procesar clústeres. Estas instancias de EC2 tienen ubicaciones compuestas de zonas de disponibilidad y regiones. Las regiones son ubicaciones dispersas emplazadas en zonas geográficas distintas. Las zonas de disponibilidad son ubicaciones diferentes dentro de una región aisladas en caso de error en otras zonas de disponibilidad. Cada zona de disponibilidad proporciona conectividad de red económica y de baja latencia con otras zonas de disponibilidad dentro de la misma región. Para ver una lista de las regiones y los puntos de conexión de Amazon EMR, consulte Regiones y puntos de conexión en la Referencia general de Amazon Web Services.

El parámetro `AvailabilityZone` especifica la ubicación del clúster. Este parámetro es opcional y, en general, no se aconseja su uso. Cuando no se especifica `AvailabilityZone`, Amazon EMR elige automáticamente el mejor valor de `AvailabilityZone` para el clúster. Puede encontrar este parámetro útil si desea coubicar sus instancias con otras instancias en ejecución existentes y su clúster necesita leer o escribir datos de dichas instancias. A fin de obtener más información, consulte la Guía del usuario de Amazon EC2 para instancias de Linux.

Cómo utilizar archivos y bibliotecas adicionales en clústeres de Amazon EMR

Hay ocasiones en las que le podría interesar utilizar archivos adicionales o bibliotecas personalizadas con las aplicaciones de mapeador o reductor. Por ejemplo, podría utilizar una biblioteca que convierta un archivo PDF en texto sin formato.

Para almacenar en caché un archivo que utilice el mapeador o s reductor al utilizar Hadoop Streaming

- En el campo `args` del JAR:, añada el siguiente argumento:

```
-cacheFile s3://bucket/path_to_executable#local_path
```

El archivo, `local_path`, está en el directorio de trabajo del mapeador, que podría hacer referencia al archivo.

Utilizar los SDK para llamar a las API de Amazon EMR

Temas

- Usar el AWS SDK for Java para crear un clúster de Amazon EMR

Los SDK de AWS ofrecen funciones que encapsulan la API y se encargan de muchos de los detalles de la conexión, tales como el cálculo de firmas, el control de reintentos de solicitud y el control de errores. Los SDK también contienen código de ejemplo, tutoriales y otros recursos para ayudarle a empezar a escribir aplicaciones que llaman a AWS. Llamar a las funciones contenedoras de un SDK puede simplificar en gran medida el proceso de escribir una aplicación de AWS.

Para obtener más información sobre cómo descargar y utilizar los SDK de AWS, consulte la sección SDK en Herramientas para Amazon Web Services.

Usar el AWS SDK for Java para crear un clúster de Amazon EMR

El AWS SDK for Java ofrece tres paquetes con las funcionalidades de Amazon EMR:

- com.amazonaws.services.elasticmapreduce
- com.amazonaws.services.elasticmapreduce.model
- com.amazonaws.services.elasticmapreduce.util

Para obtener más información sobre estos paquetes, consulte la referencia de la API de AWS SDK for Java.

En el siguiente ejemplo se ilustra cómo los SDK pueden simplificar la programación con Amazon EMR. El código de muestra siguiente utiliza el objeto `StepFactory`, una clase helper para crear los tipos de pasos de Amazon EMR comunes, para crear un clúster de Hive interactivo con la depuración habilitada.

```
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) {
  AWSCredentialsProvider profile = null;
  try {
   credentials_profile = new ProfileCredentialsProvider("default"); // specifies any
 named profile in .aws/credentials as the credentials provider
       } catch (Exception e) {
           throw new AmazonClientException(
                   "Cannot load credentials from .aws/credentials file. " +
                   "Make sure that the credentials file exists and that the profile
 name is defined within it.",
                   e);
       }

  // create an EMR client using the credentials and region specified in order to create
 the cluster
  AmazonElasticMapReduce emr = AmazonElasticMapReduceClientBuilder.standard()
   .withCredentials(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());

       // specify applications to be installed and configured when EMR creates the
 cluster
  Application hive = new Application().withName("Hive");
  Application spark = new Application().withName("Spark");
  Application ganglia = new Application().withName("Ganglia");
  Application zeppelin = new Application().withName("Zeppelin");

  // create the cluster
  RunJobFlowRequest request = new RunJobFlowRequest()
          .withName("MyClusterCreatedFromJava")
```

```
          .withReleaseLabel("emr-5.20.0") // specifies the EMR release version label,
 we recommend the latest release
          .withSteps(enabledebugging)
          .withApplications(hive,spark,ganglia,zeppelin)
          .withLogUri("s3://path/to/my/emr/logs") // a URI in S3 for log files is
 required when debugging is enabled
          .withServiceRole("EMR_DefaultRole") // replace the default with a custom IAM
 service role if one is used
          .withJobFlowRole("EMR_EC2_DefaultRole") // replace the default with a custom
 EMR role for the EC2 instance profile if one is used
          .withInstances(new JobFlowInstancesConfig()
              .withEc2SubnetId("subnet-12ab34c56")
              .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());

 }

}
```

Como mínimo debe transferir un rol de servicio y un rol de flujo de trabajo correspondiente a EMR_DefaultRole y EMR_EC2_DefaultRole, respectivamente. Para ello, invoque este comando de la AWS CLI para la misma cuenta. En primer lugar, vea si ya existen los roles:

```
aws iam list-roles | grep EMR
```

Tanto el perfil de instancia (EMR_EC2_DefaultRole) como el rol de servicio (EMR_DefaultRole) se mostrarán si existen:

```
"RoleName": "EMR_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_DefaultRole"
            "RoleName": "EMR_EC2_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_EC2_DefaultRole"
```

Si los roles predeterminados no existen, puede utilizar el siguiente comando para crearlos:

```
aws emr create-default-roles
```

Administrar las Service Quotas de Amazon EMR

Temas

- ¿Qué son las Service Quotas de Amazon EMR?
- Cómo administrar Service Quotas para Amazon EMR
- Cuándo configurar los eventos de EMR en CloudWatch

En los temas de esta sección, se describen las Service Quotas de EMR (anteriormente denominadas límites de servicio), cómo administrarlas en la AWS Management Console y cuándo es mejor utilizar los eventos de CloudWatch en lugar de las Service Quotas para supervisar los clústeres y activar acciones.

¿Qué son las Service Quotas de Amazon EMR?

Su cuenta de AWS tiene Service Quotas predeterminadas, también denominadas límites, para cada servicio de AWS. El servicio de EMR tiene dos tipos de límites:

- Límites de los recursos: puede utilizar EMR para crear recursos de EC2. Sin embargo, estos recursos de EC2 están sujetos a Service Quotas. Las limitaciones de recursos de esta categoría son las siguientes:
  - El número máximo de clústeres activos que se pueden ejecutar a la vez.
  - El número máximo de instancias activas por grupo de instancias.
- Límites de las API: cuando se utilizan las API de EMR, los dos tipos de limitaciones son los siguientes:
  - Límite de ráfaga: es el número máximo de llamadas a la API que se pueden realizar a la vez. Por ejemplo, el número máximo de solicitudes a la API AddInstanceFleet que se pueden realizar por segundo se establece de forma predeterminada en 5 llamadas por segundo. Esto implica que el límite de ráfaga de la API AddInstanceFleet es de 5 llamadas por segundo o que, en un momento dado, se puede realizar un máximo de 5 llamadas a la API AddInstanceFleet. Sin embargo, después de usar el límite de ráfaga, las llamadas subsiguientes estarán limitadas por el límite de frecuencia.
  - Límite de frecuencia: es la tasa de reabastecimiento de la capacidad de ampliación de la API. Por ejemplo, la tasa de reabastecimiento de las llamadas de AddInstanceFleet se establece

de forma predeterminada en 0,5 llamadas por segundo. Esto significa que cuando alcance el límite de ráfaga, tendrás que esperar al menos 2 segundos (0,5 llamadas/segundo x 2 segundos = 1 llamada) para realizar la llamada a la API. Si realiza una llamada antes, el servicio web de EMR le limitará. En cualquier momento, solo puede realizar tantas llamadas como tenga la capacidad de ampliación sin que se le limite. Cada segundo adicional que espere, la capacidad de ampliación aumentará en 0,5 llamadas hasta alcanzar el límite máximo de 5, que es el límite de ráfaga.

Cómo administrar Service Quotas para Amazon EMR

Service Quotas es una característica de AWS que le permite ver y administrar sus Service Quotas de Amazon EMR, o límites, desde una ubicación central mediante la AWS Management Console, la API o la CLI. Para obtener más información sobre la consulta de las cuotas y la solicitud de aumentos, consulte AWS Service Quotas en la Referencia general de Amazon Web Services.

Para algunas API, configurar un evento de CloudWatch puede ser una mejor opción que aumentar las Service Quotas. También puede ahorrar tiempo si utiliza CloudWatch para configurar alarmas y activar las solicitudes de aumento de forma proactiva, antes de alcanzar las Service Quotas. Para obtener más información, consulte Cuándo configurar los eventos de EMR en CloudWatch.

Cuándo configurar los eventos de EMR en CloudWatch

En algunas API de sondeo, como DescribeCluster, DescribeStep y ListClusters, configurar un evento de CloudWatch puede reducir el tiempo de respuesta a los cambios y liberar las Service Quotas. Por ejemplo, si tiene una función de Lambda configurada para ejecutarse cuando el estado de un clúster cambie, por ejemplo, cuando se complete un paso o se termine un clúster, puede usar ese activador para iniciar la siguiente acción del flujo de trabajo en lugar de esperar al siguiente sondeo. De lo contrario, si tiene instancias dedicadas de Amazon EC2 o funciones de Lambda que consultan constantemente la API de EMR para detectar cambios, no solo desperdiciará recursos de computación, sino que también podría alcanzar sus Service Quotas.

Los siguientes son algunos casos en los que podría beneficiarse de la transición a una arquitectura basada en eventos.

Caso 1: Sondeo de EMR mediante llamadas a la API DescribeCluster para completar pasos

Example Sondeo de EMR mediante llamadas a la API DescribeCluster para completar pasos

Un patrón común consiste en enviar un paso a un clúster en ejecución y sondear Amazon EMR para conocer el estado del paso, normalmente mediante las API DescribeCluster o DescribeStep. Esta tarea también se puede realizar con un retraso mínimo conectándose al evento Step Status Change de Amazon EMR.

Este evento contiene la siguiente información en su carga útil.

```
{
  "version": "0",
  "id": "999cccaa-eaaa-0000-1111-123456789012",
  "detail-type": "EMR Step Status Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T20:53:09Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "ERROR",
    "actionOnFailure": "CONTINUE",
    "stepId": "s-ZYXWVUTSRQPON",
    "name": "CustomJAR",
    "clusterId": "j-123456789ABCD",
    "state": "FAILED",
    "message": "Step s-ZYXWVUTSRQPON (CustomJAR) in Amazon EMR cluster j-123456789ABCD
 (Development Cluster) failed at 2016-12-16 20:53 UTC."
  }
}
```

En el mapa de detalles, una función de Lambda podría analizar “state”, “stepId” o “clusterId” para encontrar la información pertinente.

Caso 2: Sondeo de EMR en busca de clústeres disponibles para ejecutar flujos de trabajo

Example Sondeo de EMR en busca de clústeres disponibles para ejecutar flujos de trabajo

Un patrón de los clientes que ejecutan varios clústeres es ejecutar los flujos de trabajo en los clústeres tan pronto como estén disponibles. Si hay muchos clústeres en ejecución y es necesario realizar un flujo de trabajo en un clúster que está en espera, un patrón podría consistir en sondear EMR mediante llamadas a la API DescribeCluster o ListClusters para ver los clústeres disponibles. Otra forma de reducir el retraso a la hora de saber cuándo un clúster está listo para un paso sería procesar el evento State Change de Amazon EMR.

Este evento contiene la siguiente información en su carga útil.

```
{
  "version": "0",
  "id": "999cccaa-eaaa-0000-1111-123456789012",
  "detail-type": "EMR Cluster State Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T20:43:05Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "INFO",
    "stateChangeReason": "{\"code\":\"\"}",
    "name": "Development Cluster",
    "clusterId": "j-123456789ABCD",
    "state": "WAITING",
    "message": "Amazon EMR cluster j-123456789ABCD ..."
  }
}
```

Para este caso, se podría configurar una función de Lambda para enviar inmediatamente un flujo de trabajo en espera a un clúster en cuanto su estado cambie a EN ESPERA.

Caso 3: Sondeo de EMR para la terminación del clúster

Example Sondeo de EMR para la terminación del clúster

Un patrón común de clientes que ejecutan muchos clústeres de EMR consiste en sondear Amazon EMR en busca de los clústeres terminados para que ya no se les envíe trabajo. Puede

implementar este patrón con las llamadas a las API DescribeCluster y ListClusters o mediante el evento Cluster State Change de Amazon EMR.

Tras la terminación del clúster, el evento emitido se parece al siguiente ejemplo.

```
{
  "version": "0",
  "id": "1234abb0-f87e-1234-b7b6-000000123456",
  "detail-type": "EMR Cluster State Change",
  "source": "aws.emr",
  "account": "123456789012",
  "time": "2016-12-16T21:00:23Z",
  "region": "us-east-1",
  "resources": [],
  "detail": {
    "severity": "INFO",
    "stateChangeReason": "{\"code\":\"USER_REQUEST\",\"message\":\"Terminated by user
 request\"}",
    "name": "Development Cluster",
    "clusterId": "j-123456789ABCD",
    "state": "TERMINATED",
    "message": "Amazon EMR Cluster jj-123456789ABCD (Development Cluster) has
 terminated at 2016-12-16 21:00 UTC with a reason of USER_REQUEST."
  }
}
```

La sección “detalles” de la carga incluye el id. de clúster y el estado sobre los que se puede actuar.

Glosario de AWS

Para ver la terminología más reciente de AWS, consulte el Glosario de AWS en la Referencia de Glosario de AWS.