aws

Guida alla gestione

Amazon EMR

Amazon EMR: Guida alla gestione

Table of Contents

Che cos'è Amazon EMR?

Amazon EMR (precedentemente Amazon Elastic MapReduce) è una piattaforma di cluster gestita che semplifica l'esecuzione di framework di Big Data, come ad esempio Apache Hadoop e Apache Spark, su AWS per elaborare e analizzare grandi quantità di dati. Utilizzando questi framework e i relativi progetti open source, è possibile elaborare i dati per scopi di analisi e carichi di lavoro di business intelligence. Amazon EMR consente inoltre di trasformare e spostare grandi quantità di dati all'interno e all'esterno di altri datastore e database AWS, come Amazon Simple Storage Service (Amazon S3) e Amazon DynamoDB.

Se si usa Amazon EMR per la prima volta, è consigliabile iniziare leggendo anche le seguenti sezioni:

- Amazon EMR: questo servizio fornisce i punti salienti di Amazon EMR, i dettagli del prodotto e informazioni sui prezzi.
- Tutorial: Nozioni di base su Amazon EMR: questo tutorial ti consente di iniziare a utilizzare rapidamente Amazon EMR.

In questa sezione

Panoramica di Amazon EMR

Questo argomento fornisce una panoramica dei cluster Amazon EMR e include come inviare il lavoro a un cluster, come vengono elaborati i dati e le varie condizioni del cluster durante l'elaborazione.

In questo argomento

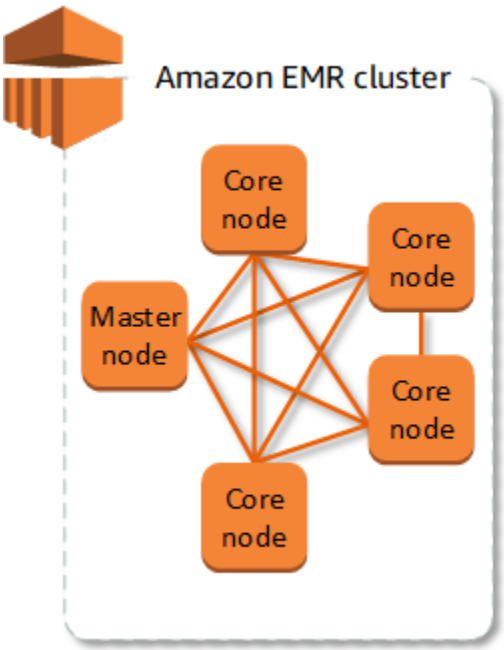
Comprensione di cluster e nodi

Il componente centrale di Amazon EMR è il cluster. Un cluster è una raccolta di istanze Amazon Elastic Compute Cloud (Amazon EC2). Ogni istanza nel cluster è denominata nodo. Ogni nodo ha un ruolo all'interno del cluster, indicato come tipo di nodo. Amazon EMR installa anche diversi componenti software su ogni tipo di nodo, dando a ogni nodo un ruolo in un'applicazione distribuita come Apache Hadoop.

I tipi di nodo in Amazon EMR sono i seguenti:

- Nodo primario: un nodo che gestisce il cluster eseguendo componenti software per coordinare la distribuzione di dati e attività tra altri nodi per l'elaborazione. Il nodo primario tiene traccia dello stato delle attività e monitora lo stato del cluster. Ogni cluster dispone di un nodo primario ed è possibile creare un singolo nodo cluster con solo il nodo primario.
- Nodo principale: un nodo con componenti software che eseguono attività e archiviano dati nel File system distribuito Hadoop (HDFS) sul cluster. I cluster multi-nodo dispongono di almeno un nodo principale.
- Nodo attività: un nodo con componenti software che esegue solo le attività e non archivia i dati in HDFS. I nodi attività sono facoltativi.

Il seguente diagramma rappresenta un cluster con un singolo nodo primario e quattro nodi principali.



Invio di lavori a un cluster

Quando si esegue un cluster su Amazon EMR, si hanno diverse opzioni circa il modo di specificare il lavoro che deve essere svolto.

- Fornisci l'intera definizione lavoro da svolgere in funzioni che specifichi come fasi al momento della creazione di un cluster. Questo viene generalmente fatto per i cluster che elaborano una determinata quantità di dati e poi terminano quando l'elaborazione è completa.
- Crea un cluster di lunga durata e utilizza la console Amazon EMR, l'API di Amazon EMR o la AWS CLI per inviare fasi che possono contenere uno o più processi. Per ulteriori informazioni, consulta Invio di lavoro a un cluster.
- Crea un cluster, connettiti al nodo primario e ad altri nodi in base alle esigenze attraverso SSH e utilizza le interfacce che le applicazioni installate forniscono per eseguire le attività e inviare le query, con script o in modo interattivo. Per ulteriori informazioni, consulta la Guida ai rilasci di Amazon EMR.

Elaborazione di dati

Quando avvii il cluster, devi scegliere i framework e le applicazioni da installare per le tue esigenze di elaborazione dati. Per elaborare i dati nel cluster Amazon EMR, è possibile inviare processi o query direttamente alle applicazioni installate, oppure è possibile eseguire fasi nel cluster.

Invio diretto di processi alle applicazioni

È possibile inviare processi e interagire direttamente con il software installato nel cluster Amazon EMR. A tale scopo, in genere, ci si connette al nodo primario attraverso una connessione sicura e si accede alle interfacce e agli strumenti disponibili per il software che è in esecuzione direttamente sul cluster. Per ulteriori informazioni, consulta Connessione a un cluster.

Esecuzione di fasi per elaborare i dati

È possibile inviare una o più fasi ordinate a un cluster Amazon EMR. Ogni fase è un'unità di lavoro che contiene le istruzioni per manipolare i dati per l'elaborazione da parte di software installato sul cluster.

Di seguito è riportato un processo di esempio che si articola in quattro fasi:

1. Invio di un set di dati di input per l'elaborazione.

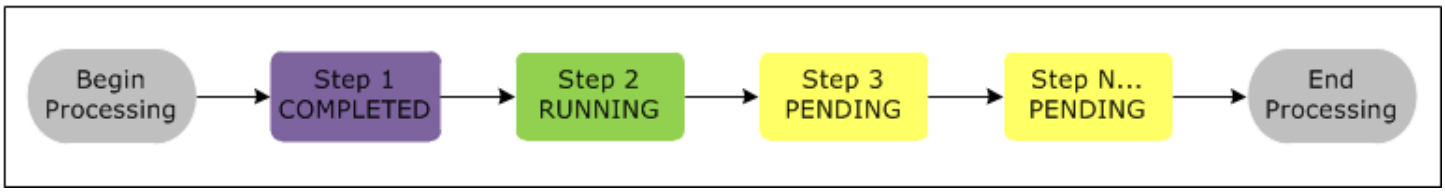
2. Elaborazione dell'output della prima fase con un programma Pig.
3. Elaborazione di un secondo set di dati di input con un programma Hive.
4. Scrittura di un set di dati di output.

Generalmente, quando si elaborano dati in Amazon EMR, l'input è costituito da dati memorizzati come file nel file system sottostante scelto, ad esempio Amazon S3 o HDFS. Questi dati passano da una fase all'altra della sequenza di elaborazione. La fase finale scrive i dati di output in un percorso specificato, ad esempio un bucket Amazon S3.

Le fasi vengono eseguite nella sequenza seguente:

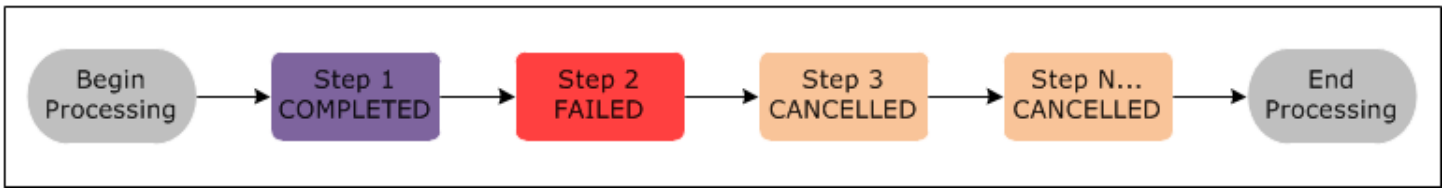
1. Una richiesta viene inviata per iniziare le fasi di elaborazione.
2. Lo stato di tutte le fasi è impostato su PENDING (IN SOSPESO).
3. Quando la prima fase nella sequenza inizia, il relativo stato diventa RUNNING (IN ESECUZIONE). Le altre fasi rimangono nello stato PENDING (IN SOSPESO).
4. Dopo il completamento della prima fase, il suo stato diventa COMPLETED (COMPLETATO).
5. La fase successiva nella sequenza inizia e il relativo stato diventa RUNNING (IN ESECUZIONE). Quando viene completata, il relativo stato diventa COMPLETED (COMPLETATO).
6. Questo schema si ripete per ogni fase fino al loro completamento e alla fine dell'elaborazione.

Il seguente diagramma rappresenta la sequenza delle fasi e il cambiamento di stato delle stesse durante la loro elaborazione.



Se una fase ha esito negativo durante l'elaborazione, lo stato diventa FAILED (NON RIUSCITO). È possibile stabilire il passaggio successivo per ogni fase. Per impostazione predefinita, le restanti fasi nella sequenza sono impostate sullo stato CANCELLED (ANNULLATO) e non eseguite in caso di insuccesso di una fase precedente. È inoltre possibile scegliere di ignorare l'errore e consentire alle fasi rimanenti di proseguire o di terminare il cluster immediatamente.

Il diagramma seguente rappresenta la sequenza delle fasi e il cambiamento di stato predefinito quando una fase ha esito negativo durante l'elaborazione.



Comprensione del ciclo di vita del cluster

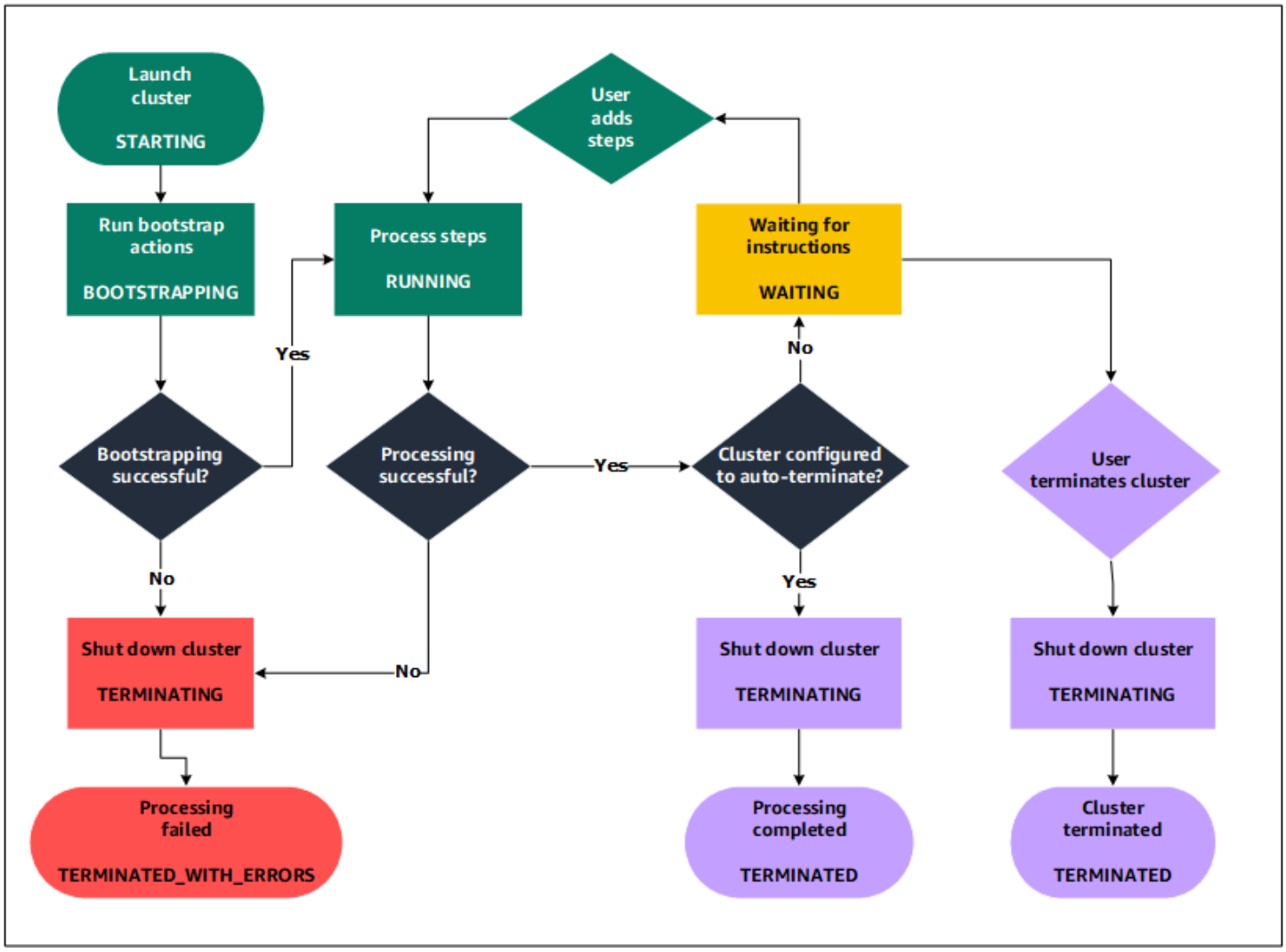
Un cluster Amazon EMR andato a buon fine segue questa procedura:

1. Innanzitutto, Amazon EMR esegue il provisioning di istanze EC2 nel cluster per ogni istanza in base alle specifiche. Per ulteriori informazioni, consulta Configurazione di hardware e reti cluster. Per tutte le istanze, Amazon EMR usa l'AMI predefinita di Amazon EMR o un'AMI Amazon Linux personalizzata specificata dall'utente. Per ulteriori informazioni, consulta Utilizzo di un'AMI personalizzata. Durante questa fase, lo stato del cluster è `STARTING`.
2. Amazon EMR esegue le operazioni di bootstrap specificate per ogni istanza. È possibile utilizzare le operazioni di bootstrap per installare applicazioni personalizzate ed eseguire le personalizzazioni necessarie. Per ulteriori informazioni, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo. Durante questa fase, lo stato del cluster è `BOOTSTRAPPING`.
3. Amazon EMR installa le applicazioni native specificate al momento della creazione del cluster, ad esempio Hive, Spark, Hadoop e così via.
4. Una volta completate le operazioni di bootstrap e installate le applicazioni native, lo stato del cluster è `RUNNING`. A questo punto, puoi connetterti alle istanze del cluster e il cluster eseguirà in sequenza tutte le fasi specificate durante la creazione del cluster. È possibile aggiungere ulteriori fasi, che vengono eseguite dopo il completamento di tutte le fasi precedenti. Per ulteriori informazioni, consulta Invio di lavoro a un cluster.
5. Dopo l'esecuzione corretta delle fasi, il cluster entra nello stato `WAITING`. Se un cluster è configurato per terminare in automatico al completamento dell'ultima fase, passa allo stato `TERMINATING` e poi allo stato `TERMINATED`. Se il cluster è configurato per attendere, è necessario terminarlo manualmente quando non è più necessario. Quando viene terminato manualmente, il cluster passa allo stato `TERMINATING` e poi allo stato `TERMINATED`.

Un errore durante il ciclo di vita del cluster porta Amazon EMR a terminare il cluster e tutte le istanze correlate, a meno che non sia stata abilitata la protezione da cessazione. Se un cluster viene terminato a causa di un errore, i dati archiviati al suo interno vengono eliminati e lo stato del cluster viene impostato su `TERMINATED_WITH_ERRORS`. Se hai attivato la protezione dall'arresto,

è possibile recuperare i dati dal cluster e quindi rimuovere la protezione e terminare il cluster. Per ulteriori informazioni, consulta Utilizzo della protezione da cessazione.

Il diagramma seguente rappresenta il ciclo di vita di un cluster e come ogni fase del ciclo di vita viene associata a un particolare stato del cluster.



Vantaggi dell'utilizzo di Amazon EMR

L'uso di Amazon EMR offre molti vantaggi. Questa sezione ne presenta una panoramica e fornisce link a ulteriori informazioni per approfondire l'argomento.

Argomenti

- Risparmio sui costi
- Integrazione di AWS

- Implementazione
- Scalabilità e flessibilità
- Affidabilità
- Sicurezza
- Monitoraggio
- Interfacce di gestione

Risparmio sui costi

Il prezzo di Amazon EMR dipende dal tipo di istanza, dal numero di istanze Amazon EC2 distribuite e dalla Regione in cui si avvia il cluster. I prezzi su richiesta offrono tariffe basse, ma è possibile ridurre ulteriormente i costi acquistando Istanze riservate o Istanze spot. In alcuni casi, le Istanze spot possono offrire risparmi significativi fino a un decimo dei prezzi su richiesta.

> **Note**
>
> Se utilizzi Amazon S3, Amazon Kinesis o DynamoDB con il cluster EMR, sono previsti costi aggiuntivi che vengono fatturati separatamente rispetto all'utilizzo di Amazon EMR.

> **Note**
>
> Quando si configura un cluster Amazon EMR in una sottorete privata, si consiglia di configurare anche gli endpoint VPC per Simple Storage Service (Amazon S3). Se il cluster EMR si trova in una sottorete privata senza endpoint VPC per Simple Storage Service (Amazon S3), verranno addebitati costi aggiuntivi del gateway NAT associati al traffico S3 perché il traffico tra il cluster EMR e S3 non rimarrà all'interno del VPC.

Per ulteriori informazioni su opzioni di prezzo e dettagli, consulta Prezzi di Amazon EMR.

Integrazione di AWS

Amazon EMR si integra con altri servizi AWS per fornire capacità e funzionalità relative a reti, archiviazione, sicurezza e altro per il cluster. Di seguito sono elencati diversi esempi di questa integrazione:

- Amazon EC2 per le istanze che costituiscono i nodi del cluster
- Amazon Virtual Private Cloud (Amazon VPC) per configurare la rete virtuale in cui è possibile avviare le istanze
- Amazon S3 per archiviare i dati di input e output
- Amazon CloudWatch per monitorare le prestazioni del cluster e configurare allarmi
- AWS Identity and Access Management (IAM) per configurare le autorizzazioni
- AWS CloudTrail per le richieste di audit effettuate al servizio
- AWS Data Pipeline per pianificare e avviare i cluster
- AWS Lake Formation per individuare, catalogare e proteggere i dati in un data lake Amazon S3

Implementazione

Il cluster EMR è costituito da istanze EC2 che eseguono il lavoro inviato al cluster. Quando si avvia il cluster, Amazon EMR configura le istanze con le applicazioni scelte, ad esempio Apache Hadoop o Spark. Scegli la dimensione e il tipo di istanza più adatti alle esigenze di elaborazione del cluster: elaborazione in batch, query a bassa latenza, streaming di dati o archiviazione di grandi quantità di dati. Per ulteriori informazioni sui tipi di istanza disponibili per Amazon EMR, consulta Configurazione di hardware e reti cluster.

Amazon EMR offre svariati modi per configurare software sul cluster. Ad esempio, è possibile installare una versione di Amazon EMR con un set scelto di applicazioni che possono includere framework versatili, come Hadoop, e applicazioni come Hive, Pig o Spark. È anche possibile installare una delle diverse distribuzioni di MapR. Amazon EMR usa Amazon Linux, che ti consente di installare il software sul tuo cluster manualmente sfruttando il gestore dei pacchetti yum o direttamente dalla fonte. Per ulteriori informazioni, consulta Configurazione del software cluster.

Scalabilità e flessibilità

Amazon EMR fornisce flessibilità per dimensionare il cluster verso l'alto o verso il basso al variare delle esigenze di computing. È possibile ridimensionare il cluster per aggiungere istanze per i carichi di lavoro di picco e rimuovere le istanze per controllare i costi quando tali carichi di lavoro si riducono. Per ulteriori informazioni, consulta Ridimensionamento manuale di un cluster in esecuzione.

Amazon EMR fornisce anche la possibilità di eseguire più gruppi di istanze in modo da poter utilizzare le Istanze on demand in un gruppo per garantire la potenza di elaborazione insieme alle Istanze spot in un altro gruppo e completare i processi più velocemente e a costi inferiori. È anche possibile

mescolare diversi tipi di istanza per sfruttare i prezzi migliori per un tipo di Istanza spot rispetto a un'altra. Per ulteriori informazioni, consulta Quando occorre utilizzare le istanze Spot?.

Inoltre, Amazon EMR offre la flessibilità di utilizzare diversi file system per i dati di input, output e intermedi. Ad esempio, puoi scegliere il File system distribuito Hadoop (HDFS) che viene eseguito sui nodi primario e principali del cluster per elaborare i dati che devi archiviare oltre il ciclo di vita del cluster. È possibile scegliere il File system EMR (EMR File System, EMRFS) per utilizzare Amazon S3 come livello di dati per le applicazioni in esecuzione sul cluster in modo da poter separare il calcolo e l'archiviazione e mantenere i dati al di fuori del ciclo di vita del cluster. Come ulteriore vantaggio, EMRFS offre la possibilità di dimensionare verso l'alto o verso il basso le esigenze di calcolo e archiviazione in modo indipendente. È possibile scalare le esigenze di calcolo ridimensionando il cluster e le esigenze di archiviazione con Amazon S3. Per ulteriori informazioni, consulta Utilizzo di archiviazione e file system.

Affidabilità

Amazon EMR monitora i nodi del cluster e termina e sostituisce in automatico un'istanza in caso di esito negativo.

Amazon EMR fornisce opzioni di configurazione che controllano la modalità di terminazione del cluster (automatica o manuale). Se configuri la terminazione automatica del cluster, questa viene terminata una volta completate tutte le fasi. Si tratta di un cluster transitorio. Tuttavia, è possibile configurare il cluster in modo che continui a funzionare anche dopo il completamento dell'elaborazione, in modo da poter scegliere di terminarlo manualmente quando non è più necessario. In alternativa, è possibile creare un cluster, interagire direttamente con le applicazioni installate e quindi terminare manualmente il cluster quando non è più necessario. I cluster in questi esempi vengono definiti cluster di lunga durata.

Inoltre, è possibile configurare la protezione di terminazione per evitare che le istanze principali del cluster vengano terminate a causa di errori o problemi durante l'elaborazione. Quando la protezione di terminazione è abilitata, è possibile ripristinare i dati dalle istanze prima della terminazione. Le impostazioni predefinite di queste opzioni differiscono a seconda che si avvii il cluster utilizzando la console, la CLI o l'API. Per ulteriori informazioni, consulta Utilizzo della protezione da cessazione.

Sicurezza

Amazon EMR sfrutta altri servizi AWS, come IAM e Amazon VPC, e funzionalità come le coppie di chiavi Amazon EC2, per proteggere i cluster e i dati.

IAM

Amazon EMR si integra con IAM per gestire le autorizzazioni. L'utente definisce le autorizzazioni utilizzando policy IAM da collegare a utenti o gruppi IAM. Le autorizzazioni definite nella policy determinano le azioni che gli utenti o i membri del gruppo possono eseguire e le risorse a cui possono accedere. Per ulteriori informazioni, consulta Funzionamento di Amazon EMR con IAM.

Inoltre, Amazon EMR usa ruoli IAM per il servizio Amazon EMR stesso e il profilo dell'istanza EC2 per le istanze. Questi ruoli concedono autorizzazioni per far accedere il servizio e le istanze ad altri servizi AWS per conto dell'utente. Sono disponibili un ruolo predefinito per il servizio Amazon EMR e un ruolo predefinito per il profilo dell'istanza EC2. I ruoli predefiniti utilizzano policy gestite da AWS, che vengono create in automatico la prima volta che avvii un cluster EMR dalla console e scegli le autorizzazioni predefinite. È anche possibile creare i ruoli IAM predefiniti dalla AWS CLI. Se si desidera gestire le autorizzazioni invece di AWS, è possibile scegliere ruoli personalizzati per il servizio e il profilo dell'istanza. Per ulteriori informazioni, consulta Configurazione dei ruoli di servizio IAM per le autorizzazioni di Amazon EMR per i servizi e risorse AWS.

Gruppi di sicurezza

Amazon EMR utilizza gruppi di sicurezza per controllare il traffico in entrata e in uscita verso le tue istanze EC2. Quando si avvia il cluster, Amazon EMR utilizza un gruppo di sicurezza per l'istanza primaria e un gruppo di sicurezza condiviso dalle istanze principali/attività. Amazon EMR consente di configurare le regole dei gruppi di sicurezza per garantire la comunicazione tra le istanze del cluster. Facoltativamente, è possibile configurare gruppi di sicurezza aggiuntivi e assegnarli alle istanze primarie e principali/attività per ottenere regole più avanzate. Per ulteriori informazioni, consulta Controllo del traffico di rete con gruppi di sicurezza.

Crittografia

Amazon EMR supporta la crittografia lato server e lato client Amazon S3 facoltativa con EMRFS per proteggere i dati archiviati in Amazon S3. Con la crittografia lato server, Amazon S3 crittografa i dati dopo il caricamento.

Con la crittografia lato client, i processi di crittografia e decrittografia avvengono nel client EMRFS sul tuo cluster EMR. Puoi gestire la chiave root per la crittografia lato client utilizzando AWS Key Management Service (AWS KMS) o il tuo sistema di gestione delle chiavi personale.

Per ulteriori informazioni, consulta Configurazione della crittografia Amazon S3 con le proprietà EMRFS.

Amazon VPC

Amazon EMR supporta l'avvio di cluster in un cloud privato virtuale (Virtual Private Cloud, VPC) in Amazon VPC. Un VPC è una rete virtuale isolata in AWS che offre la possibilità di controllare aspetti avanzati della configurazione di rete e dell'accesso. Per ulteriori informazioni, consulta Configurazione delle reti.

AWS CloudTrail

Amazon EMR si integra con CloudTrail per registrare le informazioni sulle richieste provenienti da o per conto del tuo account AWS. Con queste informazioni, puoi tenere traccia di chi e quando sta accedendo al cluster e dell'indirizzo IP da cui è stata effettuata la richiesta. Per ulteriori informazioni, consulta Registrazione delle chiamate API di Amazon EMR in AWS CloudTrail.

Coppie di chiavi Amazon EC2

È possibile monitorare e interagire con il cluster creando una connessione sicura tra il computer remoto e il nodo primario. Per questa connessione dovrai utilizzare il protocollo di rete Secure Shell (SSH) oppure Kerberos per l'autenticazione. Se si utilizza SSH, è necessaria una coppia di chiavi Amazon EC2. Per ulteriori informazioni, consulta Utilizzo di una coppia di chiavi EC2 per le credenziali SSH.

Monitoraggio

Puoi utilizzare le interfacce di gestione e i file di log di Amazon EMR per risolvere problemi del cluster, come esiti negativi o errori. Amazon EMR consente di archiviare i file di log in Amazon S3 in modo da poter archiviare i log e risolvere eventuali problemi anche dopo la terminazione del cluster. Amazon EMR fornisce anche uno strumento opzionale per il debug nella console Amazon EMR per sfogliare i file di log in base a fasi, processi e attività. Per ulteriori informazioni, consulta Configurazione della registrazione e del debug di cluster.

Amazon EMR si integra con CloudWatch per monitorare i parametri delle prestazioni del cluster e i processi all'interno del cluster. Puoi configurare gli allarmi in base a diversi parametri, ad esempio se il cluster è inattivo o la percentuale di spazio di archiviazione utilizzata. Per ulteriori informazioni, consulta Monitoraggio di parametri di Amazon EMR con CloudWatch.

Interfacce di gestione

Esistono vari modi per interagire con Amazon EMR:

- Console: un'interfaccia utente grafica che consente di avviare e gestire i cluster. Attraverso la console si compilano i moduli Web per specificare i dettagli dei cluster da avviare, visualizzare i dettagli dei cluster esistenti, eseguire il debug e terminare i cluster. L'uso della console è il modo più semplice per iniziare a familiarizzare con Amazon EMR: infatti, non richiede competenze in termini di programmazione. La console è disponibile online all'indirizzo https://console.aws.amazon.com/elasticmapreduce/home.
- AWS Command Line Interface (AWS CLI): un'applicazione client in esecuzione sul computer locale per connettersi ad Amazon EMR e creare e gestire i cluster. La AWS CLI contiene una serie di comandi specifici per Amazon EMR. Consente di scrivere script che automatizzano il processo di avvio e gestione dei cluster. Se preferisci lavorare attraverso una riga di comando, AWS CLI rappresenta la soluzione migliore. Per ulteriori informazioni, consulta Amazon EMR nella Guida di riferimento ai comandi della AWS CLI.
- Software Development Kit (SDK): gli SDK contengono funzioni che invocano Amazon EMR per creare e gestire i cluster. Permettono di scrivere applicazioni che automatizzano il processo di creazione e gestione dei cluster. Utilizzare gli SDK è l'opzione migliore per ampliare o personalizzare la funzionalità di Amazon EMR. Amazon EMR è attualmente disponibile nei seguenti SDK: Go, Java, .NET (C# e VB.NET), Node.js, PHP, Python e Ruby. Per ulteriori informazioni sugli SDK, consulta Strumenti per AWS e Codice di esempio e librerie di Amazon EMR.
- Web Service API: un'interfaccia di basso livello che è possibile utilizzare per chiamare il servizio Web direttamente, utilizzando JSON. Utilizzare l'API è l'opzione migliore per creare un SDK personalizzato che invochi Amazon EMR. Per ulteriori informazioni, consulta la Guida di riferimento alle API di Amazon EMR.

Panoramica dell'architettura di Amazon EMR

L'architettura di servizio di Amazon EMR è composta da diversi livelli, ognuno dei quali fornisce determinate capacità e funzionalità al cluster. Questa sezione fornisce una panoramica dei livelli e i componenti di ciascuno di essi.

In questo argomento

Archiviazione

Il livello di archiviazione include diversi file system utilizzati con il cluster. Esistono diversi tipi di opzioni di archiviazione, come indicato di seguito.

File system distribuito Hadoop (HDFS)

File system distribuito Hadoop (HDFS) è un file di sistema distribuito e scalabile per Hadoop. HDFS distribuisce i dati che archivia tra le istanze del cluster, archiviando più copie dei dati su istanze diverse per garantire che non vadano persi in caso di guasto di una singola istanza. HDFS è uno spazio di archiviazione temporaneo che viene recuperato quando si termina un cluster. HDFS è utile per memorizzare nella cache i risultati intermedi durante l'elaborazione di MapReduce o per i carichi di lavoro con I/O casuali significativi.

Per ulteriori informazioni, consulta Archiviazione dell'istanza oppure la HDFS User Guide (Guida per l'utente di HDFS) sul sito Web di Apache Hadoop.

File system EMR (EMRFS)

Utilizzando il file system EMR (EMRFS), Amazon EMR estende Hadoop per aggiungere la possibilità di accedere direttamente ai dati memorizzati in Amazon S3 come se si trattasse di un file system del tipo di HDFS. È possibile utilizzare HDFS o Amazon S3 come file system nel cluster. Amazon S3 viene utilizzato principalmente per memorizzare i dati di input e di output, mentre i risultati intermedi sono memorizzati in HDFS.

File system locale

Il file system locale fa riferimento a un disco con connessione locale. Alla creazione di un cluster Hadoop, ogni nodo viene creato da un'istanza Amazon EC2 fornita con un blocco preconfigurato di archiviazione su disco precollegato, chiamato archivio istanza. I dati in qualsiasi volume di archivio istanza sono persistenti solo durante il ciclo di vita della relativa istanza Amazon EC2.

Gestione delle risorse del cluster

Il livello di gestione delle risorse è responsabile della gestione delle risorse del cluster e della pianificazione dei processi per l'elaborazione dei dati.

Per impostazione predefinita, Amazon EMR usa YARN (Yet Another Resource Negotiator), che è un componente introdotto in Apache Hadoop 2.0 per gestire centralmente le risorse del cluster per

più framework di elaborazione dati. Tuttavia, esistono altri framework e applicazioni offerti in Amazon EMR che non utilizzano YARN come gestore di risorse. Ogni nodo di Amazon EMR può inoltre contare su un agente che gestisce i componenti YARN, preserva l'integrità del cluster e comunica con Amazon EMR.

Poiché le Istanze Spot vengono spesso utilizzate per eseguire nodi attività, Amazon EMR dispone delle funzionalità predefinite per la pianificazione dei processi YARN in modo che i processi in esecuzione non abbiano esito negativo quando i nodi attività in esecuzione sulle Istanze Spot vengono terminati. Amazon EMR esegue questa operazione consentendo ai processi master delle applicazioni di funzionare solo sui nodi principali. Il processo master dell'applicazione controlla i processi in esecuzione e deve rimanere attivo per tutta la durata del processo.

Amazon EMR rilascio 5.19.0 e successivi utilizzano la caratteristica integrata etichette nodo YARN per questo scopo. (Le versioni precedenti utilizzavano una patch di codice). Le proprietà nelle classificazioni di configurazione `yarn-site` e `capacity-scheduler` sono configurate per impostazione predefinita in modo che capacity-scheduler e fair-scheduler YARN sfruttino le etichette dei nodi. Amazon EMR etichetta in automatico i nodi principali con l'etichetta `CORE` e imposta le proprietà in modo che i master dell'applicazione siano pianificati solo sui nodi con l'etichetta CORE. La modifica manuale delle proprietà correlate nelle classificazioni di configurazione del sito di YARN e del pianificatore di capacità o direttamente nei file XML associati potrebbe interrompere o alterare questa funzionalità.

Framework di elaborazione dati

Il livello di framework di elaborazione dati è il motore utilizzato per elaborare e analizzare i dati. Sono disponibili molti framework che funzionano su YARN o che hanno una propria gestione delle risorse. Sono disponibili diversi framework per diversi tipi di esigenze di elaborazione, come batch, interattivi, in-memory, streaming e così via. Il framework scelto varia a seconda del caso d'uso. Questo ha un impatto sulle lingue e sulle interfacce disponibili dal livello dell'applicazione, che è il livello utilizzato per interagire con i dati che si desidera elaborare. I principali framework di elaborazione disponibili per Amazon EMR sono MapReduce e Hadoop Spark.

Hadoop MapReduce

Hadoop MapReduce è un modello di programmazione open source per l'elaborazione distribuita. Questo modello semplifica il processo di scrittura di applicazioni distribuite in parallelo gestendo tutta la logica, mentre le funzioni Map e Reduce vengono fornite dall'utente. La funzione Map mappa i dati in gruppi di coppie chiave-valore denominati risultati intermedi. La funzione Reduce combina i risultati

intermedi, applica algoritmi aggiuntivi e genera l'output finale. Per MapReduce sono disponibili più framework, ad esempio Hive, che genera in automatico programmi Map e Reduce.

Per ulteriori informazioni, consulta How map and reduce operations are actually carried out (Come vengono eseguite le operazioni Map e Reduce) sul sito Web del wiki di Apache Hadoop.

Apache Spark

Spark è un framework di cluster e un modello di programmazione per l'elaborazione di carichi di lavoro di big data. Come Hadoop MapReduce, Spark è un sistema di elaborazione distribuito open source, ma utilizza grafici aciclici diretti per i piani di esecuzione e la cache in-memory per i set di dati. Quando si esegue Spark su Amazon EMR, è possibile utilizzare EMRFS per accedere direttamente ai dati in Amazon S3. Spark supporta più moduli di query interattivi, come SparkSQL.

Per ulteriori informazioni, consulta Apache Spark su cluster Amazon EMR nella Guida ai rilasci di Amazon EMR.

Applicazioni e programmi

Amazon EMR supporta numerose applicazioni come Hive, Pig e la libreria Spark Streaming per fornire funzionalità quali l'utilizzo di linguaggi di livello superiore per creare carichi di lavoro di elaborazione, l'utilizzo di algoritmi di machine learning, la creazione di applicazioni di elaborazione del flusso di lavoro e la creazione di data warehouse. Inoltre, Amazon EMR supporta anche i progetti open source che sfruttano le proprie funzionalità di gestione dei cluster anziché utilizzare YARN.

È possibile interagire con le applicazioni eseguite in Amazon EMR attraverso varie librerie e lingue. Ad esempio, puoi utilizzare Java, Hive o Pig con MapReduce o Spark Streaming, Spark SQL, MLlib e GraphX con Spark.

Per ulteriori informazioni, consulta la Guida ai rilasci di Amazon EMR.

Configurazione di Amazon EMR

Completa le attività in questa sezione prima di avviare un cluster Amazon EMR per la prima volta:

Prima di utilizzare Amazon EMR per la prima volta, è necessario completare le seguenti operazioni:

Registrati per creare un Account AWS

Se non disponi di un Account AWS, completa la procedura seguente per crearne uno.

Come registrarsi a un Account AWS

1. Apri la pagina all'indirizzo https://portal.aws.amazon.com/billing/signup.
2. Seguire le istruzioni online.

   Nel corso della procedura di registrazione riceverai una telefonata, durante la quale sarà necessario inserire un codice di verifica attraverso la tastiera del telefono.

   Durante la registrazione di un Account AWS, viene creato un Utente root dell'account AWS. L'utente root dispone dell'accesso a tutte le risorse e tutti i Servizi AWS nell'account. Come best practice di sicurezza, assegna l'accesso amministrativo a un utente amministrativo e utilizza solo l'utente root per eseguire attività che richiedono l'accesso di un utente root.

Al termine del processo di registrazione, riceverai un'e-mail di conferma da AWS. È possibile visualizzare l'attività corrente dell'account e gestire l'account in qualsiasi momento accedendo all'indirizzo https://aws.amazon.com/ e selezionando Il mio account.

Creazione di un utente amministratore

Dopo aver effettuato l'accesso a un Account AWS, crea un utente amministratore in modo da non utilizzare l'utente root per le attività quotidiane.

Protezione dell'Utente root dell'account AWS

1. Accedi alla AWS Management Console come proprietario dell'account scegliendo Utente root e immettendo l'indirizzo email dell'Account AWS. Nella pagina successiva, inserisci la password.

Per informazioni sull'accesso utilizzando un utente root, consulta la pagina Accesso come utente root della Guida per l'utente di Accedi ad AWS.

2. Abilita l'autenticazione a più fattori (MFA) per l'utente root.

   Per ricevere istruzioni, consulta Abilitare un dispositivo MFA virtuale per l'utente root dell'Account AWS (console) nella Guida per l'utente IAM.

Creazione di un utente amministratore

- Per le tue attività amministrative quotidiane, assegna l'accesso amministrativo a un utente amministratore in AWS IAM Identity Center.

  Per ricevere istruzioni, consulta Nozioni di base nella Guida per l'utente di AWS IAM Identity Center.

Accesso come utente amministratore

- Per accedere con l'utente IAM Identity Center, utilizza l'URL di accesso che è stato inviato al tuo indirizzo e-mail quando hai creato l'utente IAM Identity Center.

  Per ricevere assistenza nell'accesso mediante un utente IAM Identity Center, consulta Accedere al portale di accesso AWS nella Guida per l'utente Accedi ad AWS.

Creazione di una coppia di chiavi Amazon EC2 per SSH

> **Note**
>
> Con Amazon EMR versione 5.10.0 o versioni successive, puoi configurare Kerberos per autenticare utenti e connessioni SSH su un cluster. Per ulteriori informazioni, consulta Utilizzo di Kerberos per l'autenticazione con Amazon EMR.

Per eseguire l'autenticazione e la connessione ai nodi in un cluster mediante un canale sicuro utilizzando il protocollo SSH (Secure Shell), crea una coppia di chiavi Amazon Elastic Compute Cloud (Amazon EC2) prima di avviare il cluster. Puoi anche creare un cluster senza una coppia di chiavi.

Questa operazione viene di solito effettuata con cluster transitori che si avviano, eseguono fasi e quindi terminano in automatico.

Se...	Allora...
Disponi già di una coppia di chiavi Amazon EC2 che desideri utilizzare o non hai bisogno di autenticarti nel cluster.	Salta questo passaggio.
Occorre creare una coppia di chiavi.	Consulta Creazione di una coppia di chiavi mediante Amazon EC2.

Fasi successive

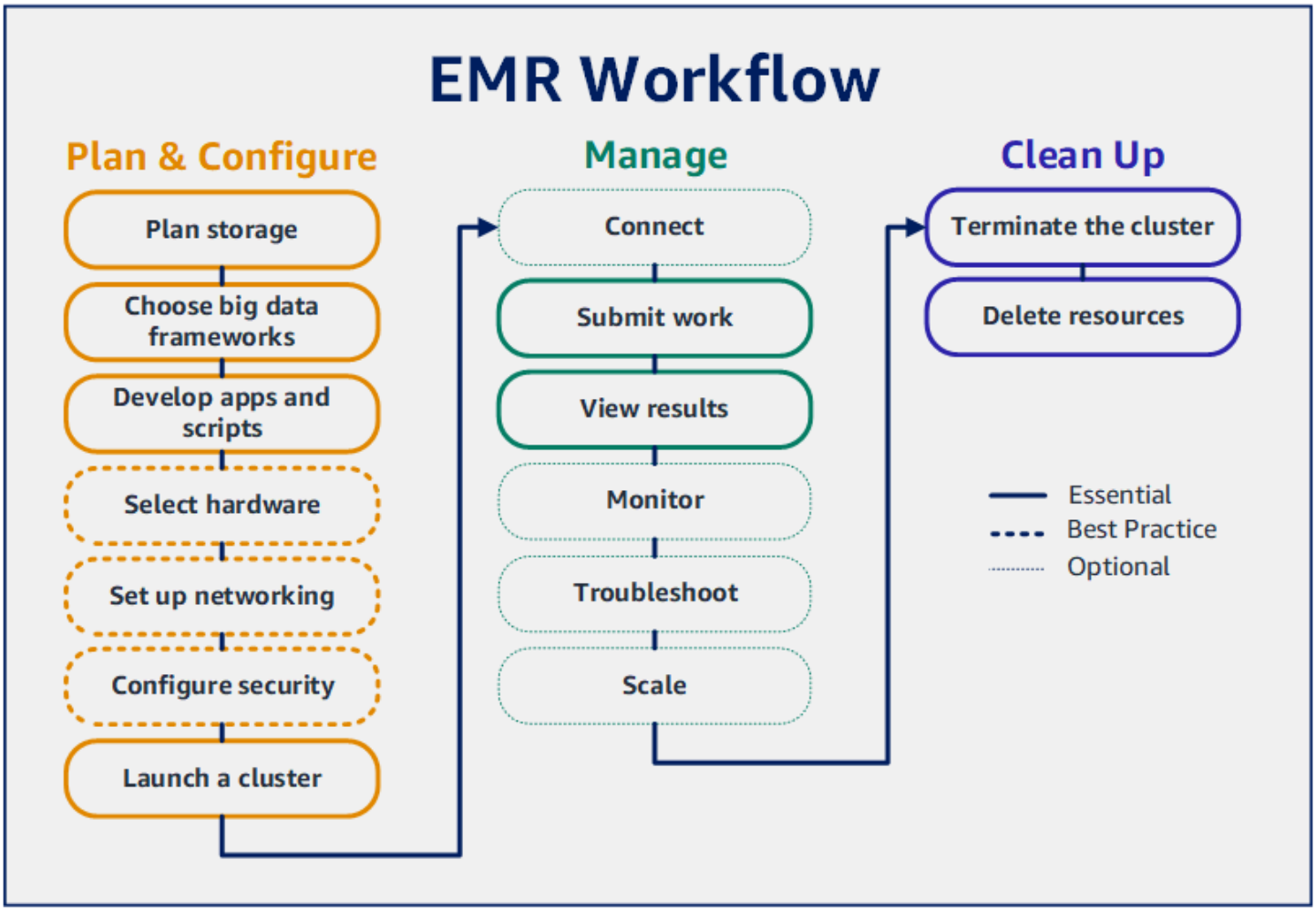
- Per ulteriori informazioni sulla creazione di un cluster di esempio, consulta Tutorial: Nozioni di base su Amazon EMR.
- Per ulteriori informazioni su come configurare un cluster personalizzato e controllare l'accesso a quest'ultimo, consulta Pianificazione e configurazione di cluster e Sicurezza in Amazon EMR.

Tutorial: Nozioni di base su Amazon EMR

Panoramica

Con Amazon EMR è possibile configurare un cluster per elaborare e analizzare i dati con framework di big data in pochi minuti. Questo tutorial mostra come avviare un cluster di esempio utilizzando Spark e come eseguire un semplice script PySpark archiviato in un bucket Amazon S3. Copre le attività essenziali di Amazon EMR in tre categorie principali del flusso di lavoro: pianificazione e configurazione, gestione e pulizia.

Nel corso del tutorial sono disponibili collegamenti ad argomenti più dettagliati e idee per ulteriori fasi nella sezione Fasi successive. In caso di domande o problemi, contatta il team Amazon EMR sul nostro Forum di discussione.



Prerequisiti

- Prima di avviare un cluster Amazon EMR, assicurati di completare le attività descritte in Configurazione di Amazon EMR.

Costo

- Il cluster di esempio che crei viene eseguito in un ambiente reale. Il cluster accumula costi minimi. Per evitare costi aggiuntivi, assicurati di completare le processi di pulizia nell'ultima fase di questo tutorial. I costi maturano alla tariffa al secondo in base ai prezzi di Amazon EMR. I costi variano anche in base alla Regione. Per ulteriori informazioni, consulta Prezzi di Amazon EMR.
- Potrebbero esserci degli costi minimi per i file di piccole dimensioni che archivi su Amazon S3. Alcuni o tutti i costi per Amazon S3 potrebbero essere annullati se non superi i limiti di utilizzo del Piano gratuito di AWS. Per ulteriori informazioni, consulta Prezzi di Amazon S3 e Piano gratuito di AWS.

Fase 1: pianificazione e configurazione di un cluster Amazon EMR

Preparare la archiviazione per Amazon EMR

Quando utilizzi Amazon EMR, puoi scegliere tra una varietà di file system per archiviare dati di input, dati di output e file di log. In questo tutorial, utilizzi EMRFS per archiviare i dati in un bucket S3. EMRFS è un'implementazione del file system Hadoop che permette di leggere e scrivere file standard su Amazon S3. Per ulteriori informazioni, consulta Utilizzo di archiviazione e file system.

Per creare un bucket per questo tutorial, segui le istruzioni in Come creare un bucket S3? nella Guida per l'utente della console Amazon Simple Storage Service. Crea il bucket nella stessa Regione AWS in cui intendi avviare il cluster Amazon EMR. Ad esempio, US West (Oregon) us-west-2 (Stati Uniti occidentali (Oregon) us-west-2).

I bucket e le cartelle utilizzati con Amazon EMR presentano le seguenti limitazioni:

- I nomi devono includere lettere minuscole, numeri, punti (.) e trattini (-).
- I nomi non devono terminare con numeri.
- Il nome del bucket deve essere univoco per tutti gli account AWS.
- Una cartella di output deve essere vuota.

Preparazione di un'applicazione con i dati di input per Amazon EMR

Il modo più comune per preparare un'applicazione per Amazon EMR è caricare l'applicazione e i relativi dati di input su Amazon S3. Poi, quando invii il lavoro al cluster, specificare le posizioni Amazon S3 per i vostri script e dati.

Questa fase prevede il caricamento di uno script PySpark di esempio nel bucket Amazon S3. Ti abbiamo fornito uno script PySpark perché tu possa utilizzarlo. Lo script elabora i dati di ispezione degli stabilimenti alimentari e restituisce un file dei risultati nel bucket S3. Il file di risultati elenca i primi dieci stabilimenti alimentari con il maggior numero di violazioni di tipo "Red (Rosso)".

Puoi anche caricare dati di input di esempio su Amazon S3 per l'elaborazione da parte dello script PySpark. I dati di input sono una versione modificata dei risultati delle ispezioni condotte dal 2006 al 2020 dal Dipartimento di sanità pubblica della Contea di King, Washington. Per ulteriori informazioni, consulta Dati pubblici della Contea di King: dati di ispezione sugli stabilimenti alimentari. Di seguito sono riportare righe di esempio del set di dati:

```
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
```

Preparazione dello script PySpark di esempio per EMR

1. Copia il codice di esempio fornito di seguito in un nuovo file nell'editor che hai scelto.

```
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. Salva il file con nome `health_violations.py`.
3. Carica `health_violations.py` su Amazon S3 nel bucket che hai creato per questo tutorial. Per istruzioni, consulta Caricamento di un oggetto in un bucket nella Guida alle operazioni di base di Amazon Simple Storage Service.

Preparazione dei dati di input di esempio per EMR

1. Scarica il file zip food_establishment_data.zip.

2. Decomprimere e salvare `food_establishment_data.zip` come `food_establishment_data.csv` sulla tua macchina.
3. Carica il file CSV nel bucket S3 creato per questo tutorial. Per istruzioni, consulta Caricamento di un oggetto in un bucket nella Guida alle operazioni di base di Amazon Simple Storage Service.

Per ulteriori informazioni sulla configurazione di dati per EMR, consulta Preparazione dei dati di input.

Avvio di un cluster Amazon EMR

Dopo aver preparato una posizione di archiviazione e l'applicazione, è possibile avviare un cluster Amazon EMR di esempio. In questa fase, avvii un cluster Apache Spark utilizzando l'ultima versione di rilascio di Amazon EMR.

New console

Avvio di un cluster con Spark installato con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. Nella pagina Crea Cluster, prendi nota dei valori predefiniti per Versione, Tipo di istanza, Numero di istanze e Autorizzazioni. Questi campi si compilano in automatico con i valori che funzionano per i cluster per uso generico.
4. Nel campo Nome cluster, inserisci un nome univoco che ti aiuti a identificare il cluster, ad esempio *`Il mio primo cluster`*.
5. In Applicazioni, seleziona l'opzione Spark per installare Spark nel cluster.

> **Note**
>
> Scegli le applicazioni che desideri nel cluster Amazon EMR prima di avviarlo. Non è possibile aggiungere o rimuovere applicazioni da un cluster dopo l'avvio.

6. In Registri del cluster, seleziona la casella di spunta Pubblica log specifici del cluster su Amazon S3. Sostituisci il valore di Posizione di Amazon S3 con il bucket Amazon S3 che hai creato seguito da **`/logs`**. Ad esempio, **`s3://DOC-EXAMPLE-BUCKET/logs`**. Aggiungendo

/logs si crea una nuova cartella chiamata "logs" nel tuo bucket, nella quale Amazon EMR copierà i file di log del cluster.

7. In Configurazione e autorizzazioni di sicurezza, scegli la tua coppia di chiavi EC2. Nella stessa sezione, seleziona il menu a discesa Service role for Amazon EMR (Ruolo di servizio per Amazon EMR) e scegli EMR_DefaultRole. Quindi, seleziona il menu a discesa IAM role for instance profile (Ruolo IAM per il profilo dell'istanza) e scegli EMR_EC2_DefaultRole.
8. Scegli Crea cluster per avviare il cluster e apri la pagina dei dettagli del cluster.
9. Cerca l'indicazione Stato accanto al nome del cluster. Lo stato cambia da Avvio in corso a In esecuzione a In attesa mentre Amazon EMR effettua il provisioning del cluster. Potrebbe essere necessario scegliere l'icona di aggiornamento a destra o aggiornare il browser per visualizzare gli aggiornamenti di stato.

Quando lo stato del cluster passa a In attesa, il cluster è attivo, in esecuzione e pronto per accettare lavoro. Per ulteriori informazioni sulla lettura del riepilogo di un cluster, consulta Visualizzazione dello stato e dei dettagli di un cluster. Per ulteriori informazioni sullo stato del cluster, consulta Comprensione del ciclo di vita del cluster.

Old console

Avvio di un cluster con Spark installato con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster per aprire la procedura guidata Opzioni rapide.
3. Prendi nota dei valori predefiniti per Versione, Tipo di istanza, Numero di istanze e Autorizzazioni nella pagina Crea cluster - Opzioni rapide. Questi campi si compilano in automatico con i valori che funzionano per i cluster generici.
4. Inserisci un Nome cluster per agevolare l'identificazione del cluster. Ad esempio, *Il mio primo cluster*.
5. Lascia l'opzione Registrazione di log abilitata ma sostituisci la Cartella S3 con il bucket Amazon S3 che hai creato, seguito da **/logs**. Ad esempio, **s3://DOC-EXAMPLE-BUCKET/logs**. Aggiungendo **/logs** si crea una nuova cartella chiamata "logs" (registri) nel tuo bucket, nella quale EMR copierà i file di log del tuo cluster.
6. In Applications (Applicazioni), seleziona l'opzione Spark per installare Spark nel cluster.

> **Note**
>
> Scegliere le applicazioni che si desidera nel cluster Amazon EMR prima di avviare il cluster. Dopo l'avvio, non è possibile aggiungere o rimuovere applicazioni da un cluster.

7. Scegliere il tuo EC2 key pair (Coppia di chiavi EC2) in Security and access (Sicurezza e accesso).
8. Scegli Crea cluster per avviare il cluster e apri la pagina di stato del cluster.
9. Cerca l'indicazione Status (Stato) accanto al nome del cluster. Lo stato cambia da Avvio in corso a In esecuzione a In attesa mentre Amazon EMR effettua il provisioning del cluster. Potrebbe essere necessario scegliere l'icona di aggiornamento a destra o aggiornare il browser per visualizzare gli aggiornamenti di stato.

Quando lo stato del cluster passa a In attesa, il cluster è attivo, in esecuzione e pronto per accettare lavoro. Per ulteriori informazioni sulla lettura del riepilogo di un cluster, consulta Visualizzazione dello stato e dei dettagli di un cluster. Per ulteriori informazioni sullo stato del cluster, consulta Comprensione del ciclo di vita del cluster.

CLI

Avvio di un cluster con Spark installato con la AWS CLI

1. Crea ruoli predefiniti IAM che puoi utilizzare per creare il cluster utilizzando il comando seguente.

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

Per ulteriori informazioni su come `create-default-roles`, consulta la Guida di riferimento ai comandi della AWS CLI.

2. Crea un cluster Spark con il comando seguente. Inserisci un nome per il cluster attraverso l'opzione `--name` e specifica il nome della coppia di chiavi EC2 con l'opzione `--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
```

Prendi nota degli altri valori richiesti per `--instance-type`, `--instance-count` e `--use-default-roles`. Questi valori sono stati scelti per i cluster generici. Per ulteriori informazioni su come `create-cluster`, consulta la Guida di riferimento ai comandi della AWS CLI.

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

L'output restituito dovrebbe essere simile al seguente. L'output mostra il `ClusterId` e `ClusterArn` del nuovo cluster. Nota il tuo `ClusterId`. Utilizza il `ClusterId` per verificare lo stato del cluster e inviare il lavoro.

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

3. Verifica lo stato del cluster con il comando seguente.

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

Verrà visualizzato un output come il seguente con l'oggetto `Status` per il nuovo cluster.

```
{
    "Cluster": {
        "Id": "myClusterId",
        "Name": "My First EMR Cluster",
        "Status": {
            "State": "STARTING",
            "StateChangeReason": {
```

```
                "Message": "Configuring cluster software"
            }
        }
    }
}
```

La valore `State` cambia da `STARTING` a `RUNNING` a `WAITING` mentre Amazon EMR effettua il provisioning del cluster.

Lo stato del cluster passa a **`WAITING`** quando il cluster è attivo, in esecuzione e pronto per accettare lavoro. Per ulteriori informazioni sullo stato del cluster, consulta Comprensione del ciclo di vita del cluster.

Fase 2: gestione del cluster Amazon EMR

Invio di lavoro ad Amazon EMR

Dopo aver avviato un cluster, è possibile inviare il lavoro al cluster in esecuzione per elaborare e analizzare i dati. Invia lavoro a un cluster Amazon EMR come una fase. Una fase è un'unità di lavoro costituita da uno o più operazioni. Ad esempio, potresti inviare una fase per calcolare valori o per trasferire ed elaborare dati. Puoi inviare fasi quando crei un cluster o a un cluster in esecuzione. In questa parte del tutorial, invii `health_violations.py` come fase al cluster in esecuzione. Per ulteriori informazioni sulle fasi, consulta Invio di lavoro a un cluster.

New console

Invio di un'applicazione Spark come fase con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona il cluster a cui desideri inviare il lavoro. Lo stato del cluster deve essere In attesa.
3. Scegli la scheda Steps (Fasi), quindi scegli Add step (Aggiungi fase).
4. Configura la fase in base alle seguenti linee guida:

- Per Type (Tipo), scegli Spark application (Applicazione Spark). Dovresti visualizzare altri campi per Deploy mode (Modalità di implementazione), Application location (Percorso dell'applicazione) e Spark-submit options (Opzioni Spark-submit).
- In Name (Nome), inserisci un nuovo nome. Se in un cluster sono presenti più fasi, l'assegnazione di un nome a ciascuna fase aiuta a tenerne traccia.
- Per Deploy mode (Modalità di implementazione), lascia il valore predefinito Cluster mode (Modalità cluster). Per ulteriori informazioni sulle modalità di implementazione di Spark, consulta la sezione Cluster mode overview (Panoramica della modalità cluster) nella documentazione di Apache Spark.
- Per Application location (Percorso dell'applicazione), inserisci il percorso dello script `health_violations.py` in Amazon S3, ad esempio *`s3://DOC-EXAMPLE-BUCKET/health_violations.py`*.
- Lascia vuoto il campo Spark-submit options (Opzioni Spark-submit). Per ulteriori informazioni sulle opzioni di `spark-submit`, consulta la sezione Launching applications with spark-submit (Avvio di applicazioni con spark-submit).
- Nel campo Argomenti, inserisci i seguenti argomenti e valori:

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

  Sostituisci *`s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv`* con l'URI del bucket S3 dei dati di input che hai preparato in Preparazione di un'applicazione con i dati di input per Amazon EMR.

  Sostituisci *`DOC-EXAMPLE-BUCKET`* con il nome del bucket creato per questo tutorial e *`myOutputFolder`* con un nome per la cartella di output del cluster.

- Per Action on failure (Operazione in caso di errore), accetta l'opzione predefinita, ossia Continue (Continua). In questo modo, se la fase riscontra un errore, il cluster continua a funzionare.

5. Seleziona Aggiungi per inviare la fase. La fase viene visualizzata nella console con lo stato In attesa.
6. Monitora lo stato della fase. Dovrebbe passare da Pending (In attesa) a Running (In esecuzione) a Completed (Completata). Per aggiornare lo stato nella console, scegli l'icona di aggiornamento a destra di Filter (Filtro). Lo script richiede circa un minuto per l'esecuzione.

Quando lo stato diventa Completed (Completata), significa che la fase è stata completata correttamente.

Old console

Invio di un'applicazione Spark come fase con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona il nome del cluster in Elenco cluster. Lo stato del cluster deve essere Waiting (In attesa).
3. Scegli Fasi e quindi Aggiungi fase.
4. Configura la fase in base alle seguenti linee guida:
   - Per Tipo di fase, scegli Applicazione Spark. Visualizzerai altri campi per Deploy Mode (Modalità di implementazione), Spark-submit options (Opzioni Spark-submit) e Application location (Percorso dell'applicazione).
   - Per Name (Nome), è possibile lasciare l'impostazione predefinita o digitare un nuovo nome. Se in un cluster sono presenti più fasi, assegnare un nome a ogni fase aiuta a tenerne traccia.
   - Per Deploy Mode (Modalità di implementazione), lascia il valore predefinito Cluster. Per ulteriori informazioni sulle modalità di implementazione Spark, consulta Cluster mode overview (Panoramica della modalità cluster) nella documentazione di Apache Spark.
   - Lascia il campo Spark-submit options (Opzioni Spark-submit) vuoto. Per ulteriori informazioni sulle opzioni `spark-submit`, consulta Avvio di applicazioni con spark-submit.
   - Per Application location (Percorso dell'applicazione), inserisci il percorso dello script `health_violations.py` in Amazon S3. Ad esempio: *`s3://DOC-EXAMPLE-BUCKET/health_violations.py`*.
   - Nel campo Arguments (Argomenti), immetti i seguenti argomenti e valori:

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

Sostituisci *s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv* con l'URI S3 dei dati di input che hai preparato in Preparazione di un'applicazione con i dati di input per Amazon EMR.

Sostituisci *DOC-EXAMPLE-BUCKET* con il nome del bucket creato per questo tutorial, e *myOutputFolder* con un nome per la cartella di output del cluster.

- Per Action on failure (Operazione in caso di errore), accetta l'opzione predefinita Continue (Continua) per assicurare che il cluster continui a funzionare qualora la fase avesse esito negativo.

5. Seleziona Add (Aggiungi) per inviare la fase. La fase viene visualizzata nella console con lo stato In attesa.
6. Verifica lo stato della fase da cui cambiare Pending (In attesa) a Running (In esecuzione) a Completed (Completato). Per aggiornare lo stato nella console, scegli l'icona di aggiornamento a destra di Filter (Filtro). Lo script richiede circa un minuto per l'esecuzione.

Saprai che la fase è terminata correttamente quando il relativo stato passa a Completata.

CLI

Invio di un'applicazione Spark come fase con la AWS CLI

1. Accertati di disporre del `ClusterId` del cluster che hai avviato in Avvio di un cluster Amazon EMR. In alternativa, puoi recuperare l'ID del cluster con il comando seguente.

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

2. Invia `health_violations.py` come fase con il comando `add-steps` e il `ClusterId`.

- È possibile specificare un nome per la fase al posto di *"La mia applicazione Spark"*. Nell'array `Args`, sostituisci *s3://DOC-EXAMPLE-BUCKET/health_violations.py* con il percorso della tua applicazione `health_violations.py`.
- Sostituisci *s3://DOC-EXAMPLE-BUCKET/food_establishment_data.csv* con il percorso S3 del set di dati `food_establishment_data.csv`.
- Sostituisci *s3://DOC-EXAMPLE-BUCKET/MyOutputFolder* con il percorso S3 del bucket designato e un nome per la cartella di output del cluster.

- `ActionOnFailure=CONTINUE` indica che il cluster continua a funzionare qualora la fase avesse esito negativo.

```
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>]
```

Per ulteriori informazioni sull'invio di fasi mediante la CLI, consulta la Guida di riferimento ai comandi della AWS CLI.

Dopo aver inviato la fase, dovresti visualizzare l'output simile al seguente con un elenco di `StepIds`. Dal momento che hai inviato una sola fase, vedrai un solo ID nell'elenco. Copia il tuo ID della fase. Utilizzare l'ID della fase per verificare lo stato della fase.

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

3. Esegui una query sullo stato della fase con il comando `describe-step`.

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

L'output restituito dovrebbe essere simile al seguente con le informazioni sul fase.

```
{
    "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"
        }
    }
}
```

Lo `State` della fase cambia da `PENDING` a `RUNNING` a `COMPLETED` durante la sua esecuzione. La fase richiede circa un minuto per essere eseguita, quindi potrebbe essere necessario controllarne lo stato più volte.

Saprai che la fase è terminata correttamente quando il relativo `State` passa a **`COMPLETED`**.

Per ulteriori informazioni sul ciclo di vita delle fasi, consulta Esecuzione di fasi per elaborare i dati.

Visualizzazione dei risultati

Se l'esecuzione della fase è stata riuscita, puoi visualizzarne i risultati nella cartella output di Amazon S3.

Visualizzazione dei risultati di **`health_violations.py`**

1. Apri la console Amazon S3 all'indirizzo https://console.aws.amazon.com/s3/.
2. Scegli il Bucket name (Nome bucket) e la cartella di output che hai specificato all'invio della fase. Ad esempio, *`DOC-EXAMPLE-BUCKET`* e quindi *`myOutputFolder`*.
3. Verifica che nella cartella di output siano presenti i seguenti elementi:
   - Un oggetto di piccole dimensioni denominato `_SUCCESS`.
   - Un file CSV che inizia con il prefisso `part-` che contiene i risultati.
4. Scegli l'oggetto con i risultati, quindi scegli Download per salvare i risultati nel file system locale.

5. Apri i risultati nell'editor che preferisci. Il file di output elenca i primi dieci stabilimenti alimentari con il maggior numero di violazioni di tipo "red (rosso)". Il file di output mostra anche il numero totale di violazioni di tipo "red (rosso)" per ogni stabilimento.

   Di seguito è riportato un esempio di risultati `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
```

Per ulteriori informazioni sull'output dei cluster Amazon EMR, consulta Configurazione di un percorso di output.

(Facoltativo) Eseguire la connessione al cluster Amazon EMR in esecuzione

Quando utilizzi Amazon EMR, potresti voler connetterti a un cluster in esecuzione per leggere i file di log, eseguire il debug del cluster o utilizzare strumenti CLI come la shell Spark. Amazon EMR consente di connetterti a un cluster utilizzando il protocollo Secure Shell (SSH). Questa sezione illustra come configurare SSH, connettersi al cluster e visualizzare i file di log per Spark. Per ulteriori informazioni sulla connessione a un cluster, consulta Autenticazione nei nodi cluster Amazon EMR.

Autorizza le connessioni SSH al cluster

Prima di connetterti al cluster, devi modificare i gruppi di sicurezza del cluster per autorizzare le connessioni SSH in entrata. I gruppi di sicurezza Amazon EC2 agiscono come firewall virtuali per controllare il traffico in entrata e uscita del cluster. Quando hai creato il cluster per questo tutorial, Amazon EMR ha creato i seguenti gruppi di sicurezza per tuo conto:

ElasticMapReduce-master

Il gruppo di sicurezza gestito da Amazon EMR predefinito associato al nodo primario. In un cluster Amazon EMR, il nodo primario è un'istanza Amazon EC2 che gestisce il cluster.

ElasticMapReduce-slave

Il gruppo di sicurezza predefinito associato ai nodi principali e attività.

New console

Concessione dell'accesso SSH alle origini attendibili per il gruppo di sicurezza primario con la nuova console

Per modificare i gruppi di sicurezza, devi disporre dell'autorizzazione per gestire i gruppi di sicurezza per il VPC in cui si trova il cluster. Per ulteriori informazioni, consulta Modifica delle autorizzazioni per un utente e la Policy di esempio che consente di gestire i gruppi di sicurezza EC2 nella Guida per l'utente IAM.

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona il cluster da aggiornare. Si apre la pagina dei dettagli del cluster. In questa pagina dovrebbe essere preselezionata la scheda Properties (Proprietà) .
3. In Networking (Reti) nella scheda Properties (Proprietà), seleziona la freccia accanto a EC2 security groups (firewall) (Gruppi di sicurezza EC2 [firewall]) per espandere questa sezione. In Primary node (Nodo primario), seleziona il collegamento al gruppo di sicurezza. Dopo avere completato i seguenti passaggi, puoi facoltativamente tornare a questo passaggio, scegliere Core and task nodes (Nodi core e attività) e ripetere i passaggi seguenti per consentire al client SSH di accedere ai nodi core e attività.
4. Si apre la console EC2. Scegli la scheda Regole in entrata e quindi Modifica le regole in entrata.
5. Verifica la presenza di una regola in entrata che consenta l'accesso pubblico con le seguenti impostazioni. Se esiste, scegli Elimina per rimuoverla.

   - Tipo

     SSH
   - Porta

     22
   - Origine

Personalizzata 0.0.0.0/0

> **Warning**
>
> Prima di dicembre 2020, il gruppo di sicurezza ElasticMapReduce-master aveva una regola preconfigurata per consentire il traffico in entrata sulla porta 22 da tutte le origini. Questa regola è stata creata per semplificare le connessioni SSH iniziali al nodo principale. Ti consigliamo di rimuovere questa regola in entrata e limitare il traffico a origini affidabili.

6. Scorri fino alla fine dell'elenco di regole e seleziona Aggiungi regola.
7. Per Tipo, seleziona SSH. Selezionando SSH si inserisce in automatico TCP per Protocollo e 22 per Intervallo porta.
8. Per origine, seleziona Il mio IP per aggiungere in automatico il tuo indirizzo IP come indirizzo di origine. Puoi anche aggiungere un intervallo di indirizzi IP affidabili del client Custom (Personalizzato), o creare regole aggiuntive per altri client. In molti ambienti di rete, gli indirizzi IP vengono allocati in modo dinamico, perciò, nel futuro, potrebbe essere necessario aggiornare gli indirizzi IP per client affidabili.
9. Seleziona Salva.
10. Facoltativamente, scegli Nodi principali e di attività dall'elenco e ripeti le fasi descritte in precedenza per consentire l'accesso SSH dei client ai nodi principali e attività.

Old console

Concessione dell'accesso SSH alle origini attendibili per il gruppo di sicurezza primario con la vecchia console

Per modificare i gruppi di sicurezza, devi disporre dell'autorizzazione per gestire i gruppi di sicurezza per il VPC in cui si trova il cluster. Per ulteriori informazioni, consulta Modifica delle autorizzazioni per un utente e la Policy di esempio che consente di gestire i gruppi di sicurezza EC2 nella Guida per l'utente IAM.

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.

2. Scegli Cluster. Scegli il Nome del cluster che desideri modificare.
3. Scegli il link Gruppi di sicurezza per master in Sicurezza e accesso.
4. Scegli ElasticMapReduce-master dall'elenco.
5. Scegliere la scheda Inbound rules (Regole in entrata), quindi scegliere Edit inbound rules (Modifica le regole in entrata).
6. Verifica la presenza di una regola in entrata che consenta l'accesso pubblico con le seguenti impostazioni. Se esiste, scegli Elimina per rimuoverla.

   - Tipo

     SSH
   - Porta

     22
   - Origine

     Personalizzata 0.0.0.0/0

   > **Warning**
   >
   > Prima di dicembre 2020, il gruppo di sicurezza ElasticMapReduce-master aveva una regola preconfigurata per consentire il traffico in entrata sulla porta 22 da tutte le origini. Questa regola è stata creata per semplificare le connessioni SSH iniziali al nodo primario. Consigliamo vivamente di rimuovere questa regola in entrata e limitare il traffico alle origini affidabili.

7. Scorri fino alla fine dell'elenco di regole e seleziona Aggiungi regola.
8. Per Tipo, seleziona SSH.

   Selezionando SSH si inserisce in automatico TCP per Protocollo e 22 per Intervallo porta.
9. Per origine, seleziona Il mio IP per aggiungere in automatico il tuo indirizzo IP come indirizzo di origine. Puoi anche aggiungere un intervallo di indirizzi IP affidabili del client Custom (Personalizzato), o creare regole aggiuntive per altri client. In molti ambienti di rete, gli indirizzi IP vengono allocati in modo dinamico, perciò, nel futuro, potrebbe essere necessario aggiornare gli indirizzi IP per client affidabili.
10. Seleziona Salva.

11. Facoltativamente, scegli ElasticMapReduce-slave dall'elenco e ripeti le fasi descritte in precedenza per consentire l'accesso SSH dei client ai nodi principali e nodi attività.

Connettiti al cluster tramite AWS CLI

A prescindere dal sistema operativo, è possibile creare una connessione SSH al cluster utilizzando la AWS CLI.

Connessione al cluster e visualizzazione dei file di log con la AWS CLI

1. Utilizza il comando seguente per aprire una connessione SSH al cluster. Sostituisci *<mykeypair.key>* con il percorso completo e il nome di file della coppia di chiavi. Ad esempio, `C:\Users\<username>\.ssh\mykeypair.pem`.

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

2. Navigare a `/mnt/var/log/spark` per accedere ai registri Spark sul nodo principale del cluster. Quindi visualizza i file in quella posizione. Per un elenco di file di log aggiuntivi sul nodo principale, consulta Visualizzazione di file di log sul nodo primario.

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

Fase 3: pulizia delle risorse Amazon EMR

Terminazione di un cluster

Ora che hai inviato lavoro al cluster e hai visualizzato i risultati dell'applicazione PySpark, è possibile terminare il cluster. La terminazione di un cluster interrompe tutti i costi Amazon EMR e le istanze Amazon EC2 associati.

Amazon EMR mantiene gratuitamente i metadati relativi al cluster per due mesi dopo la terminazione del cluster. I metadati archiviati ti aiutano clonare il cluster per un nuovo lavoro o rivedere la configurazione del cluster a scopo di riferimento. I metadati non includono i dati che il cluster scrive in S3 o che sono archiviati in HDFS sul cluster.

> **Note**
>
> La console di Amazon EMR non consente di eliminare un cluster dalla visualizzazione elenco in seguito alla terminazione del cluster. Un cluster terminato scompare dalla console quando Amazon EMR ne cancella i metadati.

New console

Terminazione del cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. Seleziona Cluster, quindi scegli il cluster da terminare.
3. Nel menu a discesa Operazioni, scegli Termina cluster.
4. Nella finestra di dialogo, scegli Termina. A seconda della configurazione del cluster, la terminazione potrebbe richiedere da 5 a 10 minuti. Per ulteriori informazioni sui cluster Amazon EMR, consulta la sezione Terminazione di un cluster.

Old console

Terminazione del cluster con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Cluster, quindi scegli il cluster da terminare. Ad esempio, *`Il mio primo cluster EMR`*.
3. Scegli Termina per aprire il prompt Termina il cluster.
4. Sceglie Termina nel prompt aperto. A seconda della configurazione del cluster, la terminazione potrebbe richiedere da 5 a 10 minuti. Per ulteriori informazioni sulla terminazione dei cluster Amazon EMR, consulta Terminazione di un cluster.

> **Note**
>
> Se hai seguito con attenzione il tutorial, la protezione da terminazione dovrebbe essere disabilitata. La protezione da terminazione del cluster impedisce le

terminazioni accidentali. Se la protezione da terminazione è attiva, verrà richiesto di modificare l'impostazione prima di terminare il cluster. Scegli Modifica e quindi Disattivata.

CLI

Terminazione del cluster con la AWS CLI

1. Avvia il processo di terminazione del cluster con il comando seguente. Sostituisci *<myClusterId>* con l'ID del cluster di esempio. Il comando non restituisce output.

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

2. Per verificare che il processo di terminazione del cluster è in corso di esecuzione, controlla lo stato del cluster con il comando seguente.

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

Di seguito è riportato un esempio di output in formato JSON. Lo `Status` del cluster dovrebbe passare da **`TERMINATING`** a **`TERMINATED`**. La terminazione potrebbe richiedere da 5 a 10 minuti a seconda della configurazione del cluster. Per ulteriori informazioni sulla terminazione di un cluster Amazon EMR, consulta Terminazione di un cluster.

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

Eliminazione di risorse S3

Per evitare costi aggiuntivi, devi eliminare il bucket Amazon S3. L'eliminazione del bucket rimuove tutte le risorse di Amazon S3 per questo tutorial. Il bucket deve contenere:

- Lo script PySpark
- Il set di dati di input
- La cartella dei risultati di output
- La cartella dei file di log

Potrebbe essere necessario eseguire fasi aggiuntive per eliminare i file memorizzati se hai salvato lo script PySpark o l'output in una posizione diversa.

> **Note**
>
> Il cluster deve essere terminato prima di eliminare il bucket. In caso contrario, potrebbe non essere consentito svuotare il bucket.

Per eliminare il bucket, segui le istruzioni in Come eliminare un bucket S3? nella Guida per l'utente di Amazon Simple Storage Service.

Fasi successive

Ora hai avviato il tuo primo cluster Amazon EMR dall'inizio alla fine. Hai inoltre completato processi EMR essenziali come la preparazione e l'invio di applicazioni di Big Data, la visualizzazione dei risultati e la terminazione di un cluster.

Utilizza gli argomenti seguenti per ulteriori informazioni su come personalizzare il flusso di lavoro Amazon EMR.

Valutazione delle applicazioni di big data per Amazon EMR

Scopri e confronta le applicazioni di big data che si possono installare in un cluster nella Guida ai rilasci di Amazon EMR. La Guida ai rilasci descrive in dettaglio ogni versione di EMR e include suggerimenti per l'utilizzo di framework come Spark e Hadoop su Amazon EMR.

Pianificazione dell'hardware, della rete e della sicurezza del cluster

In questo tutorial hai creato un cluster EMR semplice senza configurare opzioni avanzate. Le opzioni avanzate consentono di specificare i tipi di istanza Amazon EC2, la rete cluster e la sicurezza del cluster. Per ulteriori informazioni sulla pianificazione e sull'avvio di un cluster che soddisfi i tuoi requisiti, consulta Pianificazione e configurazione di cluster e Sicurezza in Amazon EMR.

Gestione di cluster

Immergiti nel lavoro con i cluster in esecuzione Gestione di cluster. Per gestire un cluster, è possibile connettersi al cluster, eseguire il debug delle fasi e tenere traccia delle processi e l'integrità del cluster. Inoltre, puoi regolare le risorse del cluster in risposta alle richieste dei carichi di lavoro con Dimensionamento gestito da EMR.

Uso di un'interfaccia diversa

Oltre alla console di Amazon EMR, puoi gestire Amazon EMR utilizzando la AWS Command Line Interface, l'API del servizio Web o uno dei tanti SDK AWS supportati. Per ulteriori informazioni, consulta Interfacce di gestione.

Inoltre, puoi interagire in molti modi con le applicazioni installate nei cluster Amazon EMR. Alcune applicazioni come Apache Hadoop pubblicano interfacce Web visualizzabili. Per ulteriori informazioni, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Consultazione del blog tecnico su EMR

Per esempi dettagliati e discussioni tecniche approfondite sulle nuove caratteristiche di Amazon EMR, consulta il Blog sui Big Data AWS.

Novità della console

Amazon EMR è passato a una nuova esperienza. La nuova console offre un'interfaccia aggiornata che assicura un modo intuitivo per gestire l'ambiente Amazon EMR e consente di accedere facilmente alla documentazione, alle informazioni sui prodotti e ad altre risorse. Questa pagina descrive le differenze salienti tra le esperienze della vecchia e della nuova AWS Management Console di Amazon EMR.

In quale console mi trovo?

Per determinare la console Amazon EMR che utilizzi attualmente, visualizza l'URL della pagina della console nel tuo browser:

- URL della nuova console: https://console.aws.amazon.com/emr
- URL della vecchia console: https://console.aws.amazon.com/elasticmapreduce

> **Note**
>
> Amazon EMR ha una nuova esperienza di console. La console precedente è stata deprecata e non è più disponibile.

La funzionalità della console Amazon EMR sta migrando gradualmente verso la nuova esperienza. La tabella seguente elenca i principali componenti della console Amazon EMR e il relativo stato di migrazione della console.

Componente della console Amazon EMR	Nuova console	Vecchia console
EMR Studio[1]	✓	✓
Creazione e gestione dei cluster	✓	✓
Blocco dell'accesso pubblico	✓	✓

Componente della console Amazon EMR	Nuova console	Vecchia console
Monitoraggio con Eventi Amazon CloudWatch	✓	✓
Configurazioni di sicurezza	✓	✓
Cluster virtuali (Amazon EMR su EKS)	✓	✓
Visualizzazione e gestione delle sottoreti di Amazon Virtual Private Cloud[2]	✓	✓
Notebook[3]	✓	✓

[1] EMR Studio utilizza la nuova esperienza di interfaccia sia nella nuova sia nella vecchia console.

[2] Nella nuova console, è possibile visualizzare e gestire le sottoreti Amazon VPC nella sezione Networking (Reti) quando crei un cluster. Nella vecchia console, utilizza il collegamento nella barra di navigazione a sinistra per accedere all'elenco delle sottoreti di Amazon VPC.

[3] I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Uso della vecchia console

Amazon EMR ha una nuova esperienza di console. La console precedente è stata deprecata e non è più disponibile.

Riepilogo delle differenze

Questa sezione descrive le differenze tra le esperienze della nuova console Amazon EMR e della vecchia console Amazon EMR. Le differenze rientrano nelle seguenti categorie:

Compatibilità dei cluster tra la vecchia e la nuova console

In alcuni casi, un cluster creato nella vecchia console Amazon EMR potrebbe non essere compatibile con la nuova console. L'elenco seguente descrive i requisiti di compatibilità per la nuova console Amazon EMR.

- La nuova console supporta i cluster creati con Amazon EMR versione 5.20.1 e successive.
- È possibile clonare cluster che utilizzano il dimensionamento automatico nella nuova console, ma è possibile creare nuovi cluster solo se si desidera utilizzare il dimensionamento manuale o gestito.

Per creare e lavorare con cluster non compatibili con la nuova console, puoi utilizzare l'AWS Command Line Interface (AWS CLI), l'SDK AWS o la vecchia console.

Differenze durante la creazione di cluster

La tabella seguente evidenzia le differenze che potrai riscontrare durante la creazione di cluster tra le esperienze della nuova console Amazon EMR e della vecchia console Amazon EMR.

Funzionalità	Nuova console	Vecchia console
Terminologia: tipi di nodi del cluster Amazon EMR	Primario (primary), core (core), attività (task)	Principale (master), core (core), attività (task)

Funzionalità	Nuova console	Vecchia console
Versioni supportate da Amazon EMR [1]	Amazon EMR rilascio 5.20.1 e successivi	Tutte le versioni di Amazon EMR
Avvio rapido di un cluster	Con il pulsante Crea cluster nel pannello Riepilogo	Nella pagina Crea cluster - Opzioni rapide
Configurazione di un timeout di provisioning Spot	Definire un periodo di timeout per il provisioning delle istanze per ogni parco del cluster.	Non puoi personalizzare un timeout di provisioning quando crei un cluster.
Ruoli di servizio e ruolo di profilo dell'istanza Amazon EC2	La nuova console non crea ruoli predefiniti; è necessario creare ruoli con la console IAM oppure selezionare un ruolo IAM creato in precedenza	Supporta la creazione di ruoli predefiniti con le policy v1 e v2 oppure la selezione di un ruolo IAM creato in precedenza
Visibilità del cluster	Dall'interno della console Amazon EMR non è possibile rendere visibile un cluster a tutti gli utenti; è la policy IAM a determina re l'accesso al cluster	Dall'interno della console Amazon EMR puoi rendere visibile un cluster a tutti gli utenti se utilizzi le policy obsolete di creazione dei ruoli v1
Reti: configurazione di sottoreti private	Gli endpoint Amazon S3 e i gateway NAT devono essere configura ti dalle rispettive console Amazon S3 e Amazon VPC	È possibile configurare gli endpoint Amazon S3 e i gateway NAT direttamente dal flusso di lavoro Crea cluster nella vecchia console

Funzionalità	Nuova console	Vecchia console
Visualizzazione coerente del file system EMR (EMRFS CV)	Con il rilascio della funzionalità di solida coerenza di lettura dopo scrittura di Amazon S3 il 1° dicembre 2020, non è più necessario utilizzare la visualizzazione coerente EMRFS (EMRFS CV) con i cluster EMR	La funzionalità EMRFS CV è abilitata, ma è possibile disattivare EMRFS CV ed eliminare il database Amazon DynamoDB che utilizza; consulta la sezione Consistent view (Visualizzazione coerente) per ulteriori informazioni.
Debug	È possibile eseguire il debug dei processi utilizzando l'interfaccia utente dell'applicazione nella pagina dei dettagli del cluster	È possibile utilizzare uno strumento di debug, passaggio 3 nella sezione Advanced options (Opzioni avanzate), per eseguire il debug dei processi per i cluster eseguiti sulle versioni da 4.1.0 a 5.27.0 di Amazon EMR

[1] Non è possibile creare o modificare i cluster utilizzando versioni precedenti ad Amazon EMR 5.20.1 nella nuova console, ma tutti i cluster esistenti creati utilizzando versioni precedenti alla 5.20.1 continueranno a funzionare. Per creare e modificare cluster con versioni di Amazon EMR precedenti alla 5.20.1, utilizza l'API o la CLI o torna alla vecchia console.

Differenze durante l'elenco e la ricerca di cluster

La tabella seguente evidenzia le differenze che potrai riscontrare durante la visualizzazione e la ricerca di cluster nella vista elenco tra la esperienze della nuova console Amazon EMR e della vecchia console Amazon EMR.

> **Note**
>
> Sia per la vecchia che per la nuova console, quando si applica un filtro dati all'elenco dei cluster, viene eseguita una query sull'intero database. Tuttavia, quando si inserisce una

stringa di testo nella casella di ricerca, la ricerca si applica solo ai risultati che l'elenco ha caricato su lato client.

Funzionalità	Nuova console	Vecchia console
Visualizzazione dei dettagli del cluster	Puoi selezionare il Cluster ID (ID cluster) per visualizzare i dettagli esaustivi del cluster come opzioni di configurazione, interfacce utente persistenti delle applicazioni e log.	Puoi espandere e comprimere ogni riga del cluster per visualizzare le diverse informazioni, come i dettagli di configurazione, e accedere ai collegamenti per il monitoraggio e i log del cluster.
Ricerca di cluster	Utilizza un unico campo di ricerca per inserire query di ricerca di testo e per creare e applicare filtri di dati come "Status = Any active status" (Stato = Qualsiasi stato attivo).	Utilizza un menu a discesa per definire lo stato dei cluster (Active, Terminated, Failed - Attivo, Terminato, Non riuscito) e un campo separato per inserire una query di ricerca di testo.
Individuazione di cluster falliti	Per cercare i cluster con errori, applica il filtro Status (Stato) = Terminated with errors (Terminato con errori).	Per cercare i cluster con errori, applica il filtro Cluster falliti.

Differenze nella visualizzazione o modifica dei dettagli del cluster

La tabella seguente evidenzia le differenze che potrai riscontrare durante la visualizzazione e la modifica dei dettagli di un cluster esistente tra le esperienze della nuova e della vecchia console Amazon EMR.

Funzionalità	Nuova console	Vecchia console
Visualizzazione delle istanze presenti nei gruppi di istanze e nei parchi istanze, oltre alle opzioni di dimensionamento, provisioning, ridimensionamento e terminazione	Visualizza le opzioni e i dettagli delle istanze nella scheda Istanze. Visualizza le opzioni di terminazione nella scheda Proprietà.	Visualizza le opzioni di configurazione e terminazione dell'istanza nella scheda Hardware.
Visualizzazione delle interfacce utente, dei log e delle configurazioni delle applicazioni (Interfaccia utente Apache Spark, servizio di cronologia Spark, interfaccia utente Apache Tez, server della cronologia YARN)	Visualizza le configurazioni del cluster nella scheda Configurazioni. Avvia un'interfaccia utente di applicazione attiva e persistente per visualizzare i log di un'applicazione dalla scheda Applicazioni.	Visualizza le configurazioni del cluster nella scheda Configurazioni. Avvia un'interfaccia utente di applicazione attiva e persistente per visualizzare i log di un'applicazione dalla scheda Interfacce utente delle applicazioni. A partire da gennaio 2023, la cronologia delle applicazioni di livello superiore non è più disponibile.
Esportazione di un cluster nella CLI	Opzione disponibile nei menu dei dettagli del cluster e di visualizzazione dell'elenco delle operazioni come "View command for cloning cluster" (Visualizza comando per la clonazione del cluster)	Opzione disponibile nei menu di visualizzazione dell'elenco delle operazioni del cluster come "Esportazione AWS CLI"

Differenze nell'utilizzo delle configurazioni di sicurezza

La tabella seguente evidenzia le differenze che potrai riscontrare durante la configurazione delle opzioni di sicurezza tra le esperienze della nuova console Amazon EMR e della vecchia console Amazon EMR.

Funzionalità	Nuova console	Vecchia console
Configurazioni di sicurezza della clonazione	✓	
Governance federata con Trino e Apache Ranger	✓	
Utilizzo di un ruolo di runtime per inviare il lavoro a un cluster[1]	✓	
Autorizzazione per l'accesso ai dati del file system EMR (EMRFS)	Punti di accesso Amazon S3	Ruoli AWS Identity and Access Management (IAM)
Controlli sugli accessi a AWS Lake Formation	Ruoli di runtime	Federazione SAML

[1] Per trasmettere un ruolo durante l'invio delle fasi, il cluster deve utilizzare una configurazione di sicurezza collegata a una policy di autorizzazione IAM in modo che l'utente IAM possa trasmettere solo i ruoli approvati e i processi possano accedere alle risorse Amazon EMR. Per ulteriori informazioni, consulta Ruoli di runtime per le fasi di Amazon EMR.

Amazon EMR Studio

Amazon EMR Studio è un ambiente di sviluppo integrato (IDE) basato sul Web per notebook Jupyter completamente gestiti che vengono eseguiti su cluster Amazon EMR. Puoi configurare un EMR Studio per il tuo team per sviluppare, visualizzare ed eseguire il debug di applicazioni scritte in R, Python, Scala e PySpark. EMR Studio è integrato con AWS Identity and Access Management (IAM) e IAM Identity Center per consentire agli utenti di accedere utilizzando le proprie credenziali aziendali.

È possibile creare un EMR Studio in modo totalmente gratuito. Quando si utilizza EMR Studio, si applicano i costi applicabili per l'archiviazione Amazon S3 e per i cluster Amazon EMR. Per i dettagli e le caratteristiche principali del prodotto, consulta la pagina del servizio per Amazon EMR Studio.

Funzionalità principali di EMR Studio

Amazon EMR Studio offre le seguenti funzionalità:

- Autentica gli utenti con AWS Identity and Access Management (IAM) o AWS IAM Identity Center (IAM Identity Center) e il tuo provider di identità aziendale.
- Accedi e avvia i cluster Amazon EMR su richiesta per eseguire i processi di notebook Jupyter.
- Connetti ai cluster Amazon EMR su EKS per inviare il lavoro durante l'esecuzione del processo.
- Esplora e salva notebook di esempio. Per ulteriori informazioni sui notebook di esempio, consulta il repository GitHub dei notebook EMR Studio di esempio.
- Analizza i dati utilizzando Python, PySpark, Spark Scala, Spark R o SparkSQL e installa kernel e librerie personalizzati.
- Collabora in tempo reale con altri utenti nello stesso Workspace. Per ulteriori informazioni, consulta Configurazione della collaborazione di WorkSpace.
- Utilizza SQL Explorer di EMR Studio per sfogliare il catalogo dati, eseguire query SQL e scaricare i risultati prima di lavorare con i dati in un notebook.
- Esegui notebook parametrizzati nell'ambito dei flussi di lavoro pianificati con uno strumento di orchestrazione come Apache Airflow o Amazon Managed Workflows per Apache Airflow. Per ulteriori informazioni, consulta Orchestrazione dei processi di analisi su Notebook EMR con MWAA nel blog AWS Big Data.
- Repository del codice di collegamento come GitHub e BitBucket.
- Monitora ed esegui il debug dei processi utilizzando Spark History Server, l'interfaccia utente Tez o il server della cronologia YARN.

EMR Studio è inoltre idoneo a HIPAA ed è certificato secondo HITRUST CSF e SOC 2. Per ulteriori informazioni sulla conformità HIPAA per i servizi AWS, consulta https://aws.amazon.com/compliance/hipaa-compliance/. Per ulteriori informazioni sulla conformità HITRUST CSF per i servizi AWS, consulta https://aws.amazon.com/compliance/hitrust/. Per ulteriori informazioni su altri programmi di conformità per i servizi AWS, consulta Servizi AWS coperti dal programma di conformità.

Cronologia delle funzionalità di Amazon EMR Studio

Questa tabella elenca gli aggiornamenti della funzionalità di dimensionamento gestito da Amazon EMR.

Data di rilascio	Funzionalità
26 ottobre 2023	Aggiunta la possibilità di creare un'applicazione EMR Serverless con funzionalità interattive.
28 febbraio 2023	Aggiunto supporto per chiavi AWS KMS gestite dal cliente per l'archiviazione dei log delle applicazioni per le applicazioni EMR serverless.
23 febbraio 2023	Aggiunta creazione di ruoli IAM con un solo clic per l'invio di processi EMR serverless. Aggiunta ricerca ECR per quando si seleziona un'immagine personalizzata per le applicazioni EMR serverless.
27 gennaio 2023	I notebook di esecuzione headless possono tracciare l'avanzamento dell'esecuzione di ogni cella con `%execute_notebook` magic.
23 gennaio 2023	Le applicazioni persistenti sono state ottimizzate per consentire tempi di avvio più rapidi.

Come funziona Amazon EMR Studio

Amazon EMR Studio è una risorsa Amazon EMR creata per un team di utenti. Ogni Studio è un ambiente di sviluppo integrato autonomo basato sul Web per notebook Jupyter che vengono eseguiti su cluster Amazon EMR. Gli utenti registrano ad Studio utilizzando le proprie credenziali aziendali.

Ogni EMR Studio creato utilizza le seguenti risorse AWS:

- Amazon Virtual Private Cloud (VPC) con sottoreti: gli utenti eseguono kernel e applicazioni Studio su Amazon EMR e Amazon EMR su cluster EKS nel VPC specificato. EMR Studio può connettersi a qualsiasi cluster nelle sottoreti specificate al momento della creazione dello Studio.
- Ruoli IAM e policy di autorizzazione: per gestire le autorizzazioni utente, è possibile creare policy di autorizzazione IAM collegate all'identità IAM di un utente o a un ruolo utente. EMR Studio utilizza anche un ruolo di servizio IAM e gruppi di sicurezza per interagire con altri servizi AWS. Per ulteriori informazioni, consulta Controllo accessi e Definizione di gruppi di sicurezza per controllare il traffico di rete EMR Studio.
- Gruppi di sicurezza: EMR Studio utilizza gruppi di sicurezza per stabilire un canale di rete sicuro tra Studio e un cluster EMR.
- Una posizione di backup di Amazon S3: EMR Studio salva il lavoro del notebook in una posizione Amazon S3.

Nelle fasi seguenti viene descritto come creare e amministrare un EMR Studio:

1. Crea uno Studio nel tuo Account AWS con l'autenticazione IAM o IAM Identity Center. Per istruzioni, consultare Configurazione di un Amazon EMR Studio.
2. Assegnazione di utenti e gruppi a Studio Utilizza le policy di autorizzazione per impostare autorizzazioni granulari per ogni utente. Per ulteriori informazioni, consulta l'argomento Assegnazione e gestione degli utenti di EMR Studio.
3. Inizia a monitorare le azioni di EMR Studio con eventi AWS CloudTrail. Per ulteriori informazioni, consulta Monitoraggio delle operazioni di Amazon EMR Studio.
4. Fornisci più opzioni di cluster agli utenti di Studio con modelli cluster e Amazon EMR su endpoint gestiti EKS.

Autenticazione e accesso utente

Amazon EMR Studio supporta due modalità di autenticazione: la modalità di autenticazione IAM e la modalità di autenticazione IAM Identity Center. La modalità IAM utilizza AWS Identity and Access Management (IAM), mentre la modalità IAM Identity Center utilizza AWS IAM Identity Center. Quando crei un EMR Studio, scegli la modalità di autenticazione per tutti gli utenti di tale Studio.

Modalità di autenticazione IAM

Con la modalità di autenticazione IAM, è possibile utilizzare l'autenticazione IAM o la federazione IAM.

L'autenticazione IAM consente di gestire le identità IAM come utenti, gruppi e ruoli in IAM. Concedi agli utenti l'accesso a uno Studio con policy di autorizzazione IAM e controllo dell'accesso basato su attributi (ABAC).

La federazione IAM consente di stabilire la fiducia tra un gestore dell'identità digitale (IdP) di terze parti eAWS in modo da poter gestire le identità utente tramite il tuo IdP.

Modalità di autenticazione IAM Identity Center

La modalità di autenticazione IAM Identity Center consente agli utenti federati di accedere a un EMR Studio. È possibile utilizzare IAM Identity Center per l'autenticazione di utenti e gruppi della directory IAM Identity Center, della directory aziendale esistente o di un gestore dell'identità digitale (IdP) esterno come Azure Active Directory (AD). Quindi gestisci gli utenti con il provider di identità (IdP).

EMR Studio supporta l'uso dei seguenti gestori dell'identità digitale per IAM Identity Center:

- Active Directory AWS Managed Microsoft AD e autogestita: per ulteriori informazioni, consulta Connessione alla directory Microsoft AD.
- Provider basati su SAML: per un elenco completo, consulta Provider di identità supportati.
- La directory di IAM Identity Center: per ulteriori informazioni, consulta la sezione Gestione delle identità in IAM Identity Center.

In che modo l'autenticazione influisce sull'accesso e sull'assegnazione dell'utente

La modalità di autenticazione scelta per Amazon EMR Studio influisce sul modo in cui gli utenti accedono a uno Studio, sul modo in cui assegni un utente a uno Studio e sul modo in cui l'utente autorizza gli utenti (concede loro autorizzazioni) per eseguire operazioni come la creazione di nuovi cluster Amazon EMR.

Nella tabella seguente vengono riepilogati i metodi di accesso per EMR Studio in base alla modalità di autenticazione.

Opzioni di accesso a EMR Studio secondo la modalità di autenticazione

Modalità di autenticazione	Metodo di accesso	Descrizione
• IAM (autenticazione e federazione) • IAM Identity Center	URL di EMR Studio	Gli utenti accedono a uno Studio utilizzando l'URL di accesso allo Studio. Ad esempio, `https://xxxxxxxxxxxxxxxxxxxxxx.emrstudio-prod.us-east-1.amazonaws.com` . Gli utenti immettono le credenziali IAM quando si utilizza l'autenticazione IAM. Quando utilizzi la federazione IAM o IAM Identity Center, EMR Studio reindirizza gli utenti all'URL di accesso del gestore dell'identità digitale per inserire le credenziali. Nel contesto della federazione delle identità, questa opzione di accesso è chiamata l'accesso avviato dal Service Provider (SP).
• IAM (federazione) • IAM Identity Center	Portale del provider di identità (IdP)	Gli utenti accedono al portale del provider di identità, ad esempio il portale di Azure e avviano la console Amazon EMR. Dopo aver avviato la console Amazon EMR, gli utenti selezionano e aprono uno Studio dal Elenco degli Studio. È inoltre possibile configurare EMR Studio come applicazione SAML in modo che gli utenti possano accedere a uno Studio specifico dal portale del provider di identità. Per istruzioni, consulta Per configurare un EMR Studio come applicazione SAML nel portale IdP. Nel contesto della federazione delle identità, questa opzione di accesso è chiamata accesso avviato dal provider di identità (IdP).

Modalità di autenticazione	Metodo di accesso	Descrizione
• IAM (autenticazione)	AWS Management Console	Gli utenti accedono alla AWS Management Console utilizzando le credenziali IAM e apri uno Studio dal Elenco degli Studio nella console Amazon EMR.

La tabella seguente illustra l'assegnazione e l'autorizzazione dell'utente per EMR Studio in base alla modalità di autenticazione.

Assegnazione e autorizzazione utente di EMR Studio secondo la modalità di autenticazione

Modalità di autenticazione	Assegnazione dell'utente	Autorizzazione dell'utente
IAM (autenticazione e federazione)	Consenti l'operazione `CreateStudioPresignedUrl` in una policy di autorizzazione IAM collegata a un'identità IAM (utente, gruppo o ruolo). Per gli utenti federati, consenti l'operazione `CreateStudioPresignedUrl` in un IAM nel policy di autorizzazione configurata per il ruolo IAM utilizzato per la federazione. Utilizzare il controllo di accesso basato su attributi (ABAC) per specificare lo Studio o gli studi a cui l'utente può accedere. Per istruzioni, consultare Assegnare un utente o un gruppo a un EMR Studio.	Definire le policy di autorizzazione IAM che consentono determinate azioni di EMR Studio. Per gli utenti nativi, collega la policy di autorizzazione IAM a un'identità IAM (utente, gruppo o ruolo). Per gli utenti federati, consenti le operazioni di Studio nel policy di autorizzazione configurata per il ruolo IAM utilizzato per la federazione. Per ulteriori informazioni, consulta Configurazione delle autorizzazioni utente di EMR Studio per Amazon EC2 o Amazon EKS.

Modalità di autenticazione	Assegnazione dell'utente	Autorizzazione dell'utente
IAM Identity Center	Assegnare un utente a uno Studio mappando l'utente a uno Studio con una policy di sessione specificata. Per istruzioni, consultare Assegnare un utente o un gruppo a un EMR Studio.	Definire le policy di sessione IAM che consentono alcune operazioni di EMR Studio. Mappare una policy di sessione a un utente quando si assegna l'utente a un Studio. Per ulteriori informazioni, consulta Autorizzazioni utente per la modalità di autenticazione IAM Identity Center.

Controllo accessi

In Amazon EMR Studio, configuri l'autorizzazione utente (autorizzazioni) con policy basate su identitàAWS Identity and Access Management (IAM). In queste policy basate su identità, specificare quali operazioni e risorse sono consentite, nonché le condizioni in base alle quali le operazioni sono consentite.

Autorizzazioni utente per la modalità di autenticazione IAM

Per impostare le autorizzazioni utente quando si utilizza l'autenticazione IAM per EMR Studio, è possibile consentire operazioni come `elasticmapreduce:RunJobFlow` in una policy di autorizzazione IAM. Puoi creare una o più policy di autorizzazione da utilizzare. Ad esempio, è possibile creare una policy di base che non consente a un utente di creare nuovi cluster Amazon EMR e un'altra policy che consente la creazione di cluster. Per un elenco di tutte le operazioni di Studio, consulta Autorizzazioni AWS Identity and Access Management per gli utenti di EMR Studio.

Autorizzazioni utente per la modalità di autenticazione IAM Identity Center

Quando utilizzi l'autenticazione IAM Identity Center, crei un singolo ruolo utente di EMR Studio. Il ruolo utente è un ruolo IAM dedicato che uno Studio assume quando un utente accede.

Aallegare le policy di sessione IAM al ruolo utente di EMR Studio. Un policy di sessione è un tipo speciale di policy di autorizzazione IAM che limita ciò che un utente federato può fare durante una sessione di accesso a Studio. Le policy di sessione consentono di impostare autorizzazioni specifiche per un utente o un gruppo senza la necessità di creare più ruoli utente per EMR Studio.

Durante l'assegnazione di utenti e gruppi allo Studio, potrai mappare una policy di sessione a tale utente o gruppo per applicare le autorizzazione granulari. Potrai inoltre aggiornare la policy di sessione utente o un gruppo in qualsiasi momento. Amazon EMR archivia ogni mappatura delle policy di sessione che hai creato.

Per ulteriori informazioni sulle policy di sessione, consulta Policy e autorizzazioni nella Guida per l'utente di AWS Identity and Access Management.

WorkSpace

I WorkSpace sono gli elementi costitutivi principali di Amazon EMR Studio. Per organizzare i notebook, gli utenti creano uno o più istanze WorkSpace in uno Studio. Per ulteriori informazioni, consulta Informazioni sulle nozioni di base di WorkSpace.

Simile alle istanze WorkSpace in JupyterLab, una istanza WorkSpace preserva lo stato del lavoro del notebook. Tuttavia, l'interfaccia utente dell'istanza WorkSpace amplia l'interfaccia open source JupyterLab con strumenti aggiuntivi per aiutare di creare e collegare cluster EMR, eseguire processi, esplorare notebook di esempio e collegare repository Git.

L'elenco seguente include le caratteristiche principali di EMR Studio WorkSpaces:

- La visibilità del WorkSpace è basata sullo Studio. Le istanze WorkSpace creati in uno Studio non sono visibili in altri Studio.
- Per impostazione predefinita, un WorkSpace è condiviso e può essere visualizzato da tutti gli utenti di Studio. Tuttavia, solo un utente alla volta può aprire e lavorare in un WorkSpace. Lavorare contemporaneamente con altri utenti è possibile con Configurazione della collaborazione di WorkSpace
- È possibile collaborare contemporaneamente con altri utenti in un WorkSpace quando si abilita la collaborazione di WorkSpace. Per ulteriori informazioni, consulta Configurazione della collaborazione di WorkSpace.
- I notebook in un WorkSpace condividono lo stesso cluster EMR per eseguire i comandi. Puoi collegare un'istanza WorkSpace a un cluster Amazon EMR in esecuzione su Amazon EC2 o a un cluster virtuale Amazon EMR su EKS e endpoint gestito.
- I WorkSpace possono passare a un'altra zona di disponibilità associata alle sottoreti di uno Studio. È possibile arrestare e riavviare un WorkSpace per richiedere il processo di failover. Quando si riavvia un WorkSpace, EMR Studio avvia il WorkSpace in una zona di disponibilità diversa nel VPC di Studio quando Studio è configurato con l'accesso a più zone di disponibilità. Se Studio dispone

di una sola zona di disponibilità, EMR Studio tenta di avviare il workspace in una sottorete diversa. Per ulteriori informazioni, consulta Risolvi i problemi di connettività di WorkSpace.

- Un workspace può connettersi a cluster in una qualsiasi delle sottoreti associate a uno Studio.

Per ulteriori informazioni sulla creazione e configurazione di WorkSpace EMR Studio, consulta Informazioni sulle nozioni di base di WorkSpace.

Archiviazione di notebook in Amazon EMR Studio

Quando utilizzi un'istanza WorkSpace, EMR Studio salva automaticamente le celle nei file notebook ad una cadenza regolare al posizione Amazon S3 associato al tuo Studio. Questo processo di backup mantiene il lavoro tra le sessioni in modo da poter tornare in un secondo momento senza commettere modifiche a un repository Git. Per ulteriori informazioni, consulta Salvataggio del contenuto del WorkSpace.

Quando elimini un file notebook da un WorkSpace, EMR Studio elimina automaticamente la versione di backup da Amazon S3. Tuttavia, se elimini un'istanza WorkSpace senza prima eliminare i file notebook, i file notebook rimangono in Amazon S3 e continuano ad addebitare costi di archiviazione. Per ulteriori informazioni, consulta Eliminazione dei file di un'istanza WorkSpace e di un notebook.

Considerazioni su EMR Studio

Considerazioni

Quando lavori con EMR Studio, tieni in considerazione i seguenti aspetti:

- EMR Studio è disponibile nelle seguenti Regioni AWS: Stati Uniti orientali (Virginia settentrionale, Ohio), Stati Uniti occidentali (California settentrionale, Oregon), Asia Pacifico (Mumbai, Seoul, Singapore, Sydney, Tokyo), Canada (Centrale), Europa (Francoforte, Irlanda, Londra, Parigi, Stoccolma) e Sud America (San Paolo).
- Per consentire agli utenti di effettuare il provisioning di nuovi cluster EMR in esecuzione su Amazon EC2 per un Workspace, puoi associare un EMR Studio a un set di modelli di cluster. Gli amministratori possono definire modelli di cluster con Service Catalog e scegliere se un utente o un gruppo può accedere ai modelli o a nessuno dei modelli all'interno di uno Studio.
- Quando definisci le autorizzazioni di accesso ai file notebook archiviati in Amazon S3 o leggi segreti da AWS Secrets Manager, utilizza il ruolo di servizio Amazon EMR. Le politicy di sessione non sono supportate con queste autorizzazioni.

- Puoi creare più EMR Studio per controllare l'accesso ai cluster EMR in VPC diversi.
- Utilizza la AWS CLI per configurare cluster Amazon EMR su EKS. È quindi possibile utilizzare l'interfaccia Studio per collegare cluster ai Workspace con un endpoint gestito per eseguire processi notebook.
- EMR Studio non supporta i seguenti comandi magic Python:
  - `%alias`
  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - `%%js`
  - `%%javascript`
  - Modifica di `proxy_user` mediante `%configure`
  - Modifica di `KERNEL_USERNAME` mediante `%env` o `%set_env`
- I cluster Amazon EMR su EKS non supportano i comandi SparkMagic per EMR Studio.
- Per scrivere istruzioni Scala a più righe nelle celle del notebook, assicurarsi che tutte le righe tranne l'ultima finiscano con un punto. Nell'esempio seguente viene utilizzata la sintassi corretta per le istruzioni Scala a più righe.

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

Problemi noti

- Assicurati di disattivare gli strumenti di gestione proxy come FoxyProxy o SwitchyOmega nel browser prima di creare uno Studio. I proxy attivi possono causare errori quando scegli Create Studio (Crea Studio) e tradursi in un messaggio di errore Network Failure (Errore di rete).
- I kernel che vengono eseguiti su cluster Amazon EMR su EKS possono non avviarsi a causa di problemi di timeout. Se si verifica un errore o un problema durante l'avvio del kernel, è necessario chiudere il file notebook, arrestare il kernel e in seguito riaprire il file notebook.
- L'operazione Restart kernel (Riavvia kernel) non funziona come previsto quando si usa un cluster Amazon EMR su EKS. Dopo aver selezionato Restart kernel (Riavvia kernel), aggiorna il Workspace affinché il riavvio abbia effetto.

- Se un Workspace non è collegato a un cluster, viene visualizzato un messaggio di errore quando un utente dello Studio apre un file notebook e tenta di selezionare un kernel. Puoi ignorare questo messaggio di errore scegliendo Ok, ma è necessario collegare il Workspace a un cluster e selezionare un kernel prima di poter eseguire il codice del notebook.
- Quando utilizzi Amazon EMR 6.2.0 con una configurazione di sicurezza per impostare la protezione del cluster, l'interfaccia del Workspace appare vuota e non funziona come previsto. Se desideri configurare la crittografia dei dati o l'autorizzazione Amazon S3 per EMRFS per un cluster, consigliamo di utilizzare un'altra versione di Amazon EMR supportata. EMR Studio funziona con Amazon EMR versione 5.32.0 (Amazon EMR serie 5.x) o 6.2.0 (Amazon EMR serie 6.x) e versioni successive.
- Quando avvii l'interfaccia utente Spark sul cluster da un file notebook, puoi visualizzare le informazioni su un processo dopo l'esecuzione del codice del notebook. Tuttavia, quando avvii il server della cronologia Spark dall'elenco Cluster dello Studio, il processo potrebbe non essere visualizzato per un massimo di due minuti.
- Quando Esecuzione del debug di Amazon EMR su processi Amazon EC2, i collegamenti all'interfaccia utente Spark sul cluster potrebbero non funzionare o non essere visualizzati. Per rigenerare i collegamenti, crea una nuova cella del notebook ed esegui il comando `%%info`.
- Jupyter Enterprise Gateway non elimina i kernel inattivi sul nodo primario di un cluster nelle seguenti versioni Amazon EMR: 5.32.0, 5.33.0, 6.2.0 e 6.3.0. I kernel inattivi consumano risorse di elaborazione e possono causa l'interruzione dei cluster a esecuzione prolungata. È possibile configurare l'eliminazione del kernel inattivo per Jupyter Enterprise Gateway utilizzando il seguente script di esempio. Puoi Connessione al nodo primario tramite SSH oppure inviare lo script come fase. Per ulteriori informazioni, consulta Esecuzione di comandi e script su un cluster 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
```

- Quando utilizzi una policy di terminazione automatica con Amazon EMR versioni 5.32.0, 5.33.0, 6.2.0 o 6.3.0, Amazon EMR contrassegna un cluster come inattivo e potrebbe terminarlo in automatico anche se disponi di un kernel Python3 attivo. Questo perché l'esecuzione di un kernel

Python3 non invia un processo Spark sul cluster. Per utilizzare la terminazione automatica con un kernel Python3, consigliamo di utilizzare Amazon EMR versione 6.4.0 o successive. Per ulteriori informazioni sulla terminazione automatica, consulta Utilizzo di una policy di terminazione automatica.

- Quando si utilizza `%%display` per visualizzare un DataFrame Spark in una tabella, le tabelle molto ampie potrebbero essere troncate. È possibile fare clic con il pulsante destro del mouse sull'output e selezionare Creare nuova vista per l'output per ottenere una schermata scorrevole dell'output.
- L'avvio di un kernel basato su Spark, come PysPark, Spark o SparkR, avvia una sessione Spark e l'esecuzione di una cella in un notebook mette in coda i processi Spark in quella sessione. Quando interrompi una cella in esecuzione, il processo Spark continua a essere eseguito. Per interrompere il processo Spark, è necessario utilizzare l'interfaccia utente Spark sul cluster. Per istruzioni sulla modalità di connessione all'interfaccia utente di Spark, consulta Debug di applicazioni e processi con EMR Studio.

Limitazioni delle caratteristiche

Amazon EMR Studio non supporta le seguenti caratteristiche di Amazon EMR:

- Collegamento ed esecuzione di processi su cluster EMR con una configurazione di sicurezza che specifica l'autenticazione Kerberos
- Cluster con più nodi primari
- Cluster che utilizzano istanze Amazon EC2 basate su AWS Graviton2 per versioni di Amazon EMR 6.x precedenti alla 6.9.0 e versioni 5.x precedenti alla 5.36.1

Limiti del servizio per EMR Studio

La tabella seguente mostra i limiti del servizio per EMR Studio.

Elemento	Limite
EMR Studio	Un massimo di 100 per ciascun account AWS
Sottoreti	Un massimo di 5 associate a ciascun EMR Studio
Gruppi IAM Identity Center	Un massimo di 5 assegnati a ciascun EMR Studio

Elemento	Limite
Utenti IAM Identity Center	Un massimo di 100 assegnati a ciascun EMR Studio

Best practice per VPC e sottoreti

Utilizzare le seguenti best practice per configurare un Amazon Virtual Private Cloud (Amazon VPC) con sottoreti per EMR Studio:

- È possibile specificare un massimo di cinque sottoreti in un VPC da associare allo Studio. Si consiglia di fornire più sottoreti in diverse zone di disponibilità per supportare la disponibilità di Workspace e consentire agli utenti di Studio l'accesso ai cluster in diverse zone di disponibilità. Per ulteriori informazioni sul funzionamento di VPC, sottoreti e zone di disponibilità, consulta VPC e sottoreti nella Amazon Virtual Private CloudGuida per l'utente.
- Le sottoreti specificate dovrebbero essere in grado di comunicare tra loro.
- Per consentire agli utenti di collegare un Workspace a repository Git ospitati pubblicamente, è necessario specificare solo sottoreti private che hanno accesso a Internet tramite Network Address Translation (NAT). Per ulteriori informazioni sulla configurazione di una sottorete privata per Amazon EMR, consulta Sottoreti private.
- Quando si utilizza Amazon EMR su EKS con EMR Studio, deve esserci almeno una sottorete in comune tra Studio e il cluster Amazon EKS utilizzato per registrare il cluster virtuale. In caso contrario, l'endpoint gestito non verrà visualizzato come opzione nei Workspace Studio. Puoi creare un cluster Amazon EKS e associarlo a una sotterete appartenente allo Studio o creare uno Studio e specificare le sottoreti del tuo cluster EKS.
- Se intendi utilizzare Amazon EMR su EKS con EMR Studio, scegli lo stesso VPC dei nodi worker del cluster Amazon EKS.

Requisiti del cluster per Amazon EMR Studio

Cluster Amazon EMR in esecuzione su Amazon EC2

Tutti i cluster Amazon EMR in esecuzione su Amazon EC2 creati per un Workspace EMR Studio devono soddisfare i seguenti requisiti. I cluster creati utilizzando l'interfaccia di EMR Studio soddisfano in automatico questi requisiti.

- Il cluster deve utilizzare Amazon EMR versione 5.32.0 (Amazon EMR serie 5.x) o 6.2.0 (Amazon EMR serie 6.x) o versioni successive. Puoi creare un cluster utilizzando la console Amazon EMR, la AWS Command Line Interface o l'SDK, per poi collegarlo a un Workspace EMR Studio. Gli utenti dello Studio possono anche creare e allegare un cluster durante la creazione o l'utilizzo di un Workspace Amazon EMR. Per ulteriori informazioni, consulta Collegamento di un calcolo a un WorkSpace EMR Studio.
- Il cluster deve essere all'interno di Amazon Virtual Private Cloud. La piattaforma EC2-Classic non è supportata.
- Sul cluster devono essere installati Spark, Livy e Jupyter Enterprise Gateway. Se prevedi di utilizzare il cluster per SQL Explorer, devi installare sia Presto che Spark.
- Per utilizzare SQL Explorer, il cluster deve utilizzare Amazon EMR versione 5.34.0 o successive oppure versione 6.4.0 e disporre di Presto installato. Se desideri specificare AWS Glue Data Catalog come metastore Hive per Presto, devi configurarlo sul cluster. Per ulteriori informazioni, consulta Utilizzo di Presto con AWS Glue Data Catalog.
- Il cluster deve trovarsi in una sottorete privata con Network Address Translation (NAT) per utilizzare repository Git in hosting pubblico con EMR Studio.

Quando si lavora con EMR Studio, si consigliano le seguenti configurazioni del cluster.

- Imposta la modalità di implementazione per le sessioni Spark nella modalità cluster. La modalità cluster posiziona i processi principali dell'applicazione sui nodi principali e non sul nodo primario di un cluster. Così facendo, si alleggerisce il nodo primario delle potenziali pressioni in termini di memoria. Per ulteriori informazioni, consulta Panoramica della modalità cluster nella documentazione di Apache Spark.
- Modificare il timeout Livy dall'impostazione predefinita di un'ora a sei ore come nella seguente configurazione di esempio.

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

- Creare flotte di istanze diverse con un massimo di 30 istanze e selezionare più tipi di istanza nel parco istanze Spot. Ad esempio, è possibile specificare i seguenti tipi di istanza ottimizzati per la

memoria per i carichi di lavoro Spark: r5.2x, r5.4x, r5.8x, r5.12x, r5.16x, r4.2x, r4.4x, r4.8x, r4.12, ecc. Per ulteriori informazioni, consulta Configurazione di parchi istanze.

- Utilizzare la strategia di allocazione ottimizzata per la capacità per le istanze Spot per aiutare Amazon EMR a effettuare selezioni efficaci delle istanze basate su informazioni dettagliate sulla capacità in tempo reale di Amazon EC2. Per ulteriori informazioni, consulta Strategia di allocazione per parchi istanze.
- Abilita il dimensionamento gestito sul cluster. Impostare il parametro dei nodi principali massimi sulla capacità minima persistente che si prevede di utilizzare e configurare la scalabilità su un parco istanze di processi ben diversificato in esecuzione su istanze Spot per risparmiare sui costi. Per ulteriori informazioni, consulta Utilizzo del dimensionamento gestito in Amazon EMR.

Ti invitiamo a mantenere abilitato Amazon EMR Block Public Access e questo per limitare il traffico SSH in entrata a fonti attendibili. L'accesso in ingresso a un cluster consente agli utenti di eseguire notebook sul cluster. Per ulteriori informazioni, consulta Utilizzo del blocco dell'accesso pubblico di Amazon EMR e Controllo del traffico di rete con gruppi di sicurezza.

Cluster Amazon EMR su EKS

Oltre ai cluster EMR in esecuzione su Amazon EC2, puoi configurare e gestire cluster Amazon EMR su EKS per EMR Studio utilizzando la AWS CLI. Imposta Amazon EMR sui cluster EKS utilizzando le seguenti linee guida:

- Crea un endpoint HTTPS gestito per Amazon EMR su cluster EKS. Gli utenti collegano un Workspace a un endpoint gestito. Il cluster Amazon Elastic Kubernetes Service (EKS) utilizzato per registrare il cluster virtuale deve disporre di una sottorete privata per supportare gli endpoint gestiti.
- Utilizza un cluster Amazon EKS con almeno una sottorete privata e una Network Address Translation (NAT) quando desideri utilizzare i repository Git in hosting pubblico.
- Evita l'uso di AMI Arm Amazon Linux ottimizzate per Amazon EKS non supportate per gli endpoint gestiti di Amazon EMR su EKS.
- Evita l'uso di cluster Amazon EKS solo per AWS Fargate, che non sono supportati.

Configurazione di Amazon EMR Studio

Questa sezione è dedicata agli amministratori di EMR Studio. Illustra come configurare un EMR Studio per il tuo team e fornisce istruzioni su processi come l'assegnazione di utenti e gruppi, la configurazione di modelli di cluster e l'ottimizzazione di Apache Spark per EMR Studio.

Argomenti

Autorizzazioni di amministratore per creare e gestire un EMR Studio

Le autorizzazioni IAM descritte in questa pagina consentono di creare e gestire un EMR Studio. Per informazioni dettagliate su ciascuna autorizzazione necessaria, consulta Autorizzazioni necessarie per gestire un EMR Studio.

Autorizzazioni necessarie per gestire un EMR Studio

Nella tabella seguente sono elencate le operazioni relative alla creazione e alla gestione di un EMR Studio. Nella tabella vengono inoltre visualizzate le autorizzazioni necessarie per ciascuna operazione.

> **Note**
>
> Le operazioni di `SessionMapping` di IAM Identity Center e Studio ti occorrono soltanto quando utilizzi la modalità di autenticazione IAM Identity Center.

Autorizzazioni per creare e gestire un EMR Studio

Operazione	Autorizzazioni
Creazione di uno Studio	`"elasticmapreduce:CreateStudio", "sso:CreateManagedApplicationInstance", "iam:PassRole"`

Operazione	Autorizzazioni
Descrizione di uno Studio	`"elasticmapreduce:DescribeStudio", "sso:GetManagedApplicationInstance"`
Elencazione degli Studio	`"elasticmapreduce:ListStudios"`
Eliminazione di uno Studio	`"elasticmapreduce:DeleteStudio", "sso:DeleteManagedApplicationInstance"`
Additional permissions required when you use IAM Identity Center mode	
Assegnazione di utenti o gruppi a uno Studio	`"elasticmapreduce:CreateStudioSessionMapping", "sso:GetProfile", "sso:ListDirectoryAssociations", "sso:ListProfiles", "sso:AssociateProfile" "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup"`
Recupero dei dettagli delle assegnazioni Studio per un utente o un gruppo	`"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:GetManagedApplicationInstance", "elasticmapreduce:GetStudioSessionMapping"`
Elencazione di tutti gli utenti e i gruppi assegnati a uno Studio	`"elasticmapreduce:ListStudioSessionMappings"`

Operazione	Autorizzazioni
Aggiornamento della policy di sessione associata a un utente o a un gruppo assegnato a uno Studio	`"sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:GetManagedApplicationInstance", "elasticmapreduce:UpdateStudioSessionMapping"`
Rimozione di un utente o un gruppo da uno Studio	`"elasticmapreduce:DeleteStudioSessionMapping", "sso-directory:SearchUsers", "sso-directory:SearchGroups", "sso-directory:DescribeUser", "sso-directory:DescribeGroup", "sso:ListDirectoryAssociations", "sso:GetProfile", "sso:GetManagedApplicationInstance" "sso:ListProfiles", "sso:DisassociateProfile"`

Creare una policy con le autorizzazioni di amministrazione per EMR Studio

1. Segui le istruzioni in Creazione di policy IAM per creare una policy utilizzando uno dei seguenti esempi. Le autorizzazioni di cui hai bisogno dipendono dalla tua modalità di autenticazione per EMR Studio.

   Inserisci valori personalizzati per questi elementi:

   - Sostituisci *<your-resource-ARN>* per specificare il nome della risorsa Amazon (ARN) dell'oggetto o degli oggetti che l'istruzione copre per i tuoi casi d'uso.
   - Sostituisci *<region>* con il codice della Regione AWS in cui prevedi di creare lo Studio.
   - Sostituisci *<aws-account_id>* con l'ID dell'account AWS per lo Studio.
   - Sostituisci *<EMRStudio-Service-Role>* e *<EMRStudio-User-Role>* con i nomi del tuo ruolo di servizio EMR Studio e ruolo utente di EMR Studio.

Example Policy di esempio: autorizzazioni di amministrazione quando si utilizza la modalità di autenticazione 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 Policy di esempio: autorizzazioni di amministrazione quando si utilizza la modalità di autenticazione IAM Identity Center

> **Note**
>
> Le API IAM Identity Center e IAM Identity Center Directory non supportano la specifica di un ARN nell'elemento della risorsa di un'istruzione di policy IAM. Per consentire

l'accesso a IAM Identity Center e IAM Identity Center Directory, le seguenti autorizzazioni specificano tutte le risorse, "Resource":"*", per le operazioni di IAM Identity Center. Per ulteriori informazioni, consulta la sezione Operazioni, risorse e chiavi di condizione per la directory IAM Identity Center.

```
{
    "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. Collega la policy all'identità IAM (utente, ruolo o gruppo). Per ricevere istruzioni, consulta Aggiunta e rimozione di autorizzazioni per identità IAM.

Configurazione di un Amazon EMR Studio

Completa la procedura seguente per configurare un Amazon EMR Studio.

Prima di iniziare

> **Note**
>
> Se intendi utilizzare EMR Studio con Amazon EMR su EKS, ti consigliamo di configurare Amazon EMR su EKS per EMR Studio prima di configurare uno Studio.

Prima di configurare un EMR Studio, assicurati di disporre dei seguenti elementi:

- Un Account AWS. Per ricevere istruzioni, consulta Configurazione di Amazon EMR.
- Autorizzazioni per creare e gestire un EMR Studio. Per ulteriori informazioni, consulta the section called “Autorizzazioni di amministratore per creare un EMR Studio”.

- Un bucket Amazon S3 in cui EMR Studio può eseguire il backup dei Workspace e dei file notebook nel tuo Studio. Per le istruzioni, consulta Creazione di un bucket nella Amazon Simple Storage ServiceGuida per l'utente (S3).
- Se desideri collegarti a un cluster Amazon EMR su EC2 o Amazon EMR su EKS o utilizzare i repository Git, devi disporre di un Amazon Virtual Private Cloud (VPC) per Studio e un massimo di cinque sottoreti. Non è necessario un VPC per utilizzare EMR Studio con EMR Serverless. Per suggerimenti su come configurare la rete, consulta Best practice per VPC e sottoreti.

Configurazione di un EMR Studio

1. Scelta di una modalità di autenticazione per Amazon EMR Studio
2. Creare le seguenti risorse EMR Studio.
   - Creazione di un ruolo di servizio EMR Studio
   - Configurazione delle autorizzazioni utente di EMR Studio per Amazon EC2 o Amazon EKS
   - (Facoltativo) Definizione di gruppi di sicurezza per controllare il traffico di rete EMR Studio.
3. Creazione di un EMR Studio
4. Assegnare un utente o un gruppo a un EMR Studio

Una volta completate queste fasi di impostazione, puoi Utilizzare un Amazon EMR Studio.

Scelta di una modalità di autenticazione per Amazon EMR Studio

EMR Studio supporta due modalità di autenticazione: la modalità di autenticazione IAM e la modalità di autenticazione IAM Identity Center. La modalità IAM utilizza AWS Identity and Access Management (IAM), mentre la modalità IAM Identity Center utilizza AWS IAM Identity Center. Quando crei un EMR Studio, scegli la modalità di autenticazione per tutti gli utenti di tale Studio. Per ulteriori informazioni sulle differenti modalità di autenticazione, consulta Autenticazione e accesso utente.

Utilizzare la tabella seguente per scegliere una modalità di autenticazione per EMR Studio.

Se hai...	Consigliamo…
Già familiarità con o hai precedentemente configurato l'autenticazione o la federazione IAM	Modalità di autenticazione IAM, che offre i vantaggi seguenti:

Se hai...	Consigliamo…
	• Fornisce una configurazione rapida per EMR Studio se si già gestiscono identità come utenti e gruppi in IAM. • Funziona con i provider di identità compatibili con with OpenID Connect (OIDC) o Security Assertion Markup Language 2.0 (SAML 2.0). • Supporta l'utilizzo di più provider di identità con lo stesso Account AWS. • Disponibile in un gran numero di Regioni AWS. • Conforme a SOC 2.
Nuovo per AWS o per Amazon EMR	Modalità di autenticazione IAM Identity Center, che offre le seguenti caratteristiche: • Supporta una facile incarico di utenti e gruppi ad risorse AWS. • Funziona con Microsoft Active Directory e provider di identità SAML 2.0. • Facilita la configurazione della federazione multi-account in modo che non sia necessari o configurare separatamente la federazio ne per ciascun Account AWS nella tua organizzazione.

Configurare una modalità di autenticazione IAM per Amazon EMR Studio

Con la modalità di autenticazione IAM, è possibile utilizzare l'autenticazione IAM o la federazione IAM. L'autenticazione IAM consente di gestire le identità IAM come utenti, gruppi e ruoli in IAM. Concedi agli utenti l'accesso a uno Studio con policy di autorizzazione IAM e controllo dell'accesso basato su attributi (ABAC). La federazione IAM consente di stabilire la fiducia tra un gestore dell'identità digitale (IdP) di terze parti eAWS in modo da poter gestire le identità utente tramite il tuo IdP.

> **Note**
>
> Se usi già IAM per controllare l'accesso a risorse AWS o se hai già configurato il tuo gestore dell'identità digitale (IdP) per IAM, consulta Autorizzazioni utente per la modalità di autenticazione IAM per impostare le autorizzazioni utente quando si utilizza la modalità di autenticazione IAM per EMR Studio.

Uso della federazione IAM per Amazon EMR Studio

Per utilizzare la federazione IAM per EMR Studio, crei una relazione di attendibilità tra Account AWS e il gestore dell'identità digitale (IdP) e consenti agli utenti federati di accedere alla AWS Management Console. Le fasi che intraprendi per creare questa relazione di fiducia differiscono a seconda dello standard federativo del tuo IdP.

In generale, si completano le seguenti attività per configurare la federazione con un IdP esterno. Per istruzioni complete, consulta Abilitare gli utenti federati SAML 2.0 per accedere alla AWS Management Console e Abilitare il gestore identità personalizzato per accedere alla AWS Management Console nella Guida per l'utente di AWS Identity and Access Management.

1. Raccogli informazioni dal tuo gestore dell'identità digitale. Questo di solito significa generare un documento di metadati per convalidare le richieste di autenticazione SAML dal tuo gestore dell'identità digitale.
2. Crea un'entità IAM del provider di identità per archiviare informazioni sul tuo gestore dell'identità digitale. Per ricevere istruzioni, consulta Creazione di provider di identità IAM.
3. Creare uno o più ruoli IAM per il gestore dell'identità digitale. EMR Studio assegna un ruolo a un utente federato quando l'utente effettua l'accesso. Il ruolo consente all'IdP di richiedere credenziali di sicurezza provvisorie per l'accesso a AWS. Per ricevere istruzioni, consulta Creazione di un ruolo per un provider di identità di terza parte (federazione). Le policy di autorizzazione assegnate al ruolo determinano in cosa possono fare gli utenti federati in AWS e in uno Studio EMR. Per ulteriori informazioni, consulta Autorizzazioni utente per la modalità di autenticazione IAM.
4. (Per i provider SAML) Completare la attendibilità di SAML configurando il tuo IdP con informazioni su AWS e i ruoli che si desidera assumere gli utenti federati. Questo processo di configurazione crea una relazione di attendibilità tra IdP e AWS. Per ulteriori informazioni, consulta Configurazione del provider di identità SAML 2.0 con una relazione di attendibilità e aggiungendo attestazioni.

Per configurare un EMR Studio come applicazione SAML nel portale del provider di identità

È possibile configurare un particolare EMR Studio come applicazione SAML utilizzando un deep link allo Studio. In questo modo, gli utenti accedono al tuo portale del provider di identità e lanciano uno Studio specifico invece di navigare attraverso la console Amazon EMR.

- Utilizza il seguente formato per configurare un deep link a EMR Studio come URL di destinazione dopo la verifica dell'asserzione SAML.

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

Configurazione della modalità di autenticazione IAM Identity Center per Amazon EMR Studio

Per preparare AWS IAM Identity Center per EMR Studio, è necessario configurare l'origine dell'identità e allocare utenti e gruppi. L'allocazione è il processo che rende disponibili le informazioni su utenti e gruppi per l'utilizzo da parte di IAM Identity Center e delle applicazioni che utilizzano IAM Identity Center. Per ulteriori informazioni, consulta Provisioning di utenti e gruppi.

EMR Studio supporta l'uso dei seguenti gestori dell'identità digitale per IAM Identity Center:

- Active Directory AWS Managed Microsoft AD e autogestita: per ulteriori informazioni, consulta Connessione alla directory Microsoft AD.
- Provider basati su SAML: per un elenco completo, consulta Provider di identità supportati.
- La directory di IAM Identity Center: per ulteriori informazioni, consulta la sezione Gestione delle identità in IAM Identity Center.

Configurazione di IAM Identity Center per EMR Studio

1. Per configurare IAM Identity Center per EMR Studio, sono necessari i seguenti elementi:

   - Un account di gestione nella tua organizzazione AWS se utilizzi più account nell'organizzazione.

   > **Note**
   >
   > Per abilitare IAM Identity Center e allocare utenti e gruppi, dovresti utilizzare esclusivamente l'account di gestione. Dopo aver configurato IAM Identity Center, puoi

utilizzare un account membro per creare un EMR Studio e assegnare utenti e gruppi. Per ulteriori informazioni sulla terminologia di AWS, consulta Terminologia e concetti AWS Organizations.

- Se hai abilitato IAM Identity Center prima del 25 novembre 2019, potrebbe essere necessario abilitare le applicazioni che utilizzano IAM Identity Center per gli account nella tua organizzazione AWS. Per ulteriori informazioni, consulta la sezione Abilitazione di applicazioni integrate con IAM Identity Center negli account AWS.
- Accertati di soddisfare i prerequisiti elencati nella pagina Prerequisiti di IAM Identity Center.

2. Segui le istruzioni riportate nella sezione Abilitazione di IAM Identity Center per abilitare IAM Identity Center nella Regione AWS in cui desideri creare l'EMR Studio.
3. Collega IAM Identity Center al tuo gestore dell'identità digitale (IdP) e alloca gli utenti e i gruppi che desideri assegnare allo Studio.

Se utilizzi...	Esegui questa operazione…
Una directory Microsoft AD	1. Segui le istruzioni in Connessione alla directory Microsoft AD per connettere l'Active Directory autogestita o la directory AWS Managed Microsoft AD utilizzando AWS Directory Service. 2. Per allocare utenti e gruppi per IAM Identity Center, puoi sincronizzare i dati di identità dall'AD di origine a IAM Identity Center. Puoi sincronizzare le identità della tua AD di origine in molti modi. Un possibile modo è quello di assegnare utenti o gruppi AD a un account AWS nell'organizzazione. Per ricevere istruzioni, consulta Accesso single sign-on. La sincronizzazione può richiedere fino a due ore. Dopo avere completato questa fase, verranno visualizzati gli utenti e i gruppi sincronizzati nell'archivio identità.

Se utilizzi...	Esegui questa operazione…
	Note Gli utenti e i gruppi non vengono visualizzati nell'archivio identità fino a quando non sincronizzi le informazioni di utenti e gruppi o utilizzi l'allocazione utente just-in-time (JIT). Per ulteriori informazioni, consulta Provisioning quando gli utenti provengono da Active Directory. 3. (Facoltativo) Dopo aver sincronizzato gli utenti e i gruppi di AD, è possibile rimuoverne l'accesso all'account AWS configurato nella fase precedente. Per ricevere istruzioni, consulta Rimozione dell'accesso degli utenti.
Un provider di identità esterno	Segui le istruzioni in Connessione al provider di identità esterno.
La directory di IAM Identity Center	Quando crei utenti e gruppi in IAM Identity Center, l'allocazione è automatica. Per ulteriori informazioni, consulta la sezione Gestione delle identità in IAM Identity Center.

Ora puoi assegnare utenti e gruppi dall'archivio identità a un EMR Studio. Per istruzioni, consultare Assegnare un utente o un gruppo a un EMR Studio.

Creazione di un ruolo di servizio EMR Studio

Informazioni sul ruolo di servizio EMR Studio

Ogni EMR Studio utilizza un ruolo IAM con autorizzazioni che consentono allo Studio di interagire con altri servizi AWS. Questo ruolo di servizio deve includere autorizzazioni che consentano a EMR Studio di stabilire un canale di rete sicuro tra Workspace e cluster, per archiviare i file notebook in Amazon S3 Control e per accedere ad AWS Secrets Manager durante il collegamento di un Workspace a un repository Git.

Utilizza il ruolo di servizio Studio (anziché le policy di sessione) per definire tutte le autorizzazioni di accesso ad Amazon S3 per l'archiviazione dei file notebook e per definire autorizzazioni di accesso AWS Secrets Manager.

Come creare un ruolo di servizio per EMR Studio su Amazon EC2 o Amazon EKS

1. Segui le istruzioni in Creazione di un ruolo per delegare le autorizzazioni a un servizio AWS per creare il ruolo di servizio utilizzando le policy di affidabilità riportate di seguito.

> **Important**
>
> La policy di affidabilità riportata di seguito include le chiavi di condizione globale `aws:SourceArn` e `aws:SourceAccount` per limitare le autorizzazioni che concedi a EMR Studio a determinate risorse del tuo account. In questo modo puoi proteggerti dal problema del "confused deputy".

```
{
  "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. Rimuovere le autorizzazioni di ruolo predefinite. Quindi, includere le autorizzazioni dalla seguente policy di autorizzazione IAM di esempio. In alternativa, è possibile creare una policy personalizzata che utilizzi Autorizzazioni del ruolo di servizio EMR Studio.

   Se applicabile, modifica `"Resource":"*"` nella policy seguente per specificare il nome della risorsa Amazon (ARN) della risorsa o delle risorse che l'istruzione copre per i tuoi casi d'uso.

   Important

   - Accesso a per l'API `ModifyNetworkInterfaceAttribute` deve rimanere così com'è nella seguente policy a causa di limitazioni tecniche con il controllo degli accessi basato su tag di Amazon EC2 e il modo in cui EMR Studio utilizza `ModifyNetworkInterfaceAttribute`.
   - Affinché EMR Studio funzioni con il ruolo di servizio, le istruzioni seguenti devono rimanere invariate: `AllowAddingEMRTagsDuringDefaultSecurityGroupCreation` e `AllowAddingTagsDuringEC2ENICreation`.
   - Per utilizzare la policy di esempio, è necessario taggare le seguenti risorse con la chiave **`"for-use-with-amazon-emr-managed-policies"`** e il valore **`"true"`**.
     - Amazon Virtual Private Cloud (VPC) designato per EMR Studio.
     - Ogni sottorete che si desidera utilizzare con Studio.
     - Qualsiasi gruppo di sicurezza EMR Studio personalizzato. È necessario applicare tag a tutti i gruppi di sicurezza creati durante il periodo di anteprima di EMR Studio se si desidera continuare a utilizzarli.
     - Segreti gestiti in AWS Secrets Manager che vengono utilizzati da utenti di Studio per collegare i repository Git a un Workspace.

     Puoi applicare i tag alle risorse utilizzando la scheda Tag nella schermata delle risorse pertinente nella sezione 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. Offri al ruolo di servizio l'accesso in lettura e scrittura alla tua posizione Amazon S3 per EMR Studio. Utilizza il seguente set minimo di autorizzazioni. Per ulteriori informazioni, consulta l'esempio Amazon S3: consente l'accesso in lettura e scrittura agli oggetti di un bucket S3, in modo programmatico e nella console.

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

Se si crittografa il bucket Amazon S3, devi includere le seguenti autorizzazioni per AWS Key Management Service.

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

Ruolo di servizio minimo per EMR Serverless

Se desideri eseguire carichi di lavoro interattivi con EMR Serverless tramite notebook di EMR Studio, utilizza la stessa policy di attendibilità utilizzata per configurare EMR Studio nella sezione precedente, Come creare un ruolo di servizio per EMR Studio su Amazon EC2 o Amazon EKS.

Per la tua policy IAM, la policy minima valida prevede le seguenti autorizzazioni. Aggiorna *bucket-name* con il nome del bucket che intendi utilizzare quando configuri EMR Studio e WorkSpace. EMR Studio utilizza il bucket per il backup delle istanze WorkSpaces e dei file notebook nel Studio.

```
{
    "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"]
        }
    ]
}
```

Se intendi utilizzare un bucket Amazon S3 crittografato, aggiungi le seguenti autorizzazioni al policy:

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

Autorizzazioni del ruolo di servizio EMR Studio

Nella tabella seguente sono elencate le operazioni eseguite da EMR Studio utilizzando il ruolo di servizio, insieme alle operazioni IAM necessarie per ogni operazione.

Operazione	Azioni

Stabilisci un canale di rete protetto tra un Workspace e un cluster EMR ed esegui le azioni di eliminazione necessarie.	`"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"`

Utilizza le credenziali Git archiviate in AWS Secrets Manager per collegare i repository Git a un Workspace.	`"secretsmanager:GetSecretValue"`
Applica tag AWS all'interfaccia di rete e ai gruppi di sicurezza predefiniti creati da EMR Studio durante la configurazione del canale di rete sicuro. Per ulteriori informazioni, consulta Assegnazione di tag alle risorse AWS.	`"ec2:CreateTags"`

Accedi o carica i file notebook e i metadati in Amazon S3.	`"s3:PutObject",`
	`"s3:GetObject",`
	`"s3:GetEncryptionConfiguration",`
	`"s3:ListBucket",`
	`"s3:DeleteObject"`

Operazione	Azioni

	Se utilizzi un bucket Amazon S3 crittografato, devi includere le seguenti autorizzazioni.
	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Abilita e configura la collaborazione di Workspace.	`"iam:GetUser",`
	`"iam:GetRole",`
	`"iam:ListUsers",`
	`"iam:ListRoles",`
	`"sso:GetManagedApplicationInstance",`
	`"sso-directory:SearchUsers"`

Configurazione delle autorizzazioni utente di EMR Studio per Amazon EC2 o Amazon EKS

È necessario configurare le policy di autorizzazione utente per Amazon EMR Studio in modo da poter impostare autorizzazioni granulari per utenti e gruppi. Per informazioni sul funzionamento delle autorizzazioni utente in EMR Studio, consulta Controllo accessi in Come funziona Amazon EMR Studio.

> **Note**
>
> Le autorizzazioni descritte in questa sezione non applicano il controllo dell'accesso ai dati. Per gestire l'accesso ai set di dati di input, è necessario configurare le autorizzazioni per i cluster utilizzati da Studio. Per ulteriori informazioni, consulta Sicurezza in Amazon EMR.

Creazione di un ruolo utente di EMR Studio per la modalità di autenticazione IAM Identity Center

Quando utilizzi la modalità di autenticazione IAM Identity Center per EMR Studio, devi creare un ruolo utente.

Creazione di un ruolo utente per EMR Studio

1. Segui le istruzioni in Creazione di un ruolo per delegare le autorizzazioni a un servizio AWS nella Guida per l'utente di AWS Identity and Access Management per creare il ruolo utente.

   Utilizza la seguente policy di affidabilità quando crei il ruolo.

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

2. Rimuovere le autorizzazioni e le policy di ruolo predefinite.
3. Allegare tutte le policy di sessione di EMR Studio al ruolo utente prima di assegnare utenti e gruppi a uno Studio. Per ricevere istruzioni su come creare policy di sessione, consulta Creazione di policy di autorizzazione per gli utenti di EMR Studio.

Creazione di policy di autorizzazione per gli utenti di EMR Studio

Completa la seguente procedura per creare policy di autorizzazione per EMR Studio.

> **Note**
>
> Per impostare le autorizzazioni di accesso di Amazon S3 per la memorizzazione dei file notebook e le autorizzazioni di accesso di AWS Secrets Manager per leggere i segreti durante il collegamento dei Workspace ai repository Git, utilizza il ruolo di servizio EMR Studio.

1. Crea una o più policy di autorizzazione IAM che specificano le operazioni che possono essere compiute da un utente in Studio. Ad esempio, utilizzando le policy di esempio in questa pagina è possibile creare tre policy separate per utenti di Studio di base, intermedi e avanzati.

La tabella Autorizzazioni AWS Identity and Access Management per gli utenti di EMR Studio suddivide ogni operazione dello Studio che un utente può eseguire ed elenca le operazioni IAM minime necessarie per eseguirla. Per ricevere istruzioni, consulta Creare policy IAM.

La policy di autorizzazione deve includere le istruzioni seguenti.

```
{
            "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"
}
```

Imposta la proprietà per la collaborazione di Workspace

La collaborazione di Workspace consente a più utenti di lavorare contemporaneamente nello stesso Workspace e può essere configurata con il pannello Collaboration (Collaborazione) nell'interfaccia utente di Workspace. Per visualizzare e utilizzare il pannello Collaboration (Collaborazione), un utente deve disporre delle seguenti autorizzazioni. Tutti gli utenti con queste autorizzazioni possono visualizzare e utilizzare il pannello Collaboration (Collaborazione).

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

Per limitare l'accesso al pannello Collaboration (Collaborazione), puoi utilizzare il controllo degli accessi basato su tag. Quando un utente crea un Workspace, EMR Studio applica un tag di default con una chiave `creatorUserId` il cui valore è l'ID dell'utente che crea il Workspace.

> **Note**
>
> EMR Studio non ha aggiunto il tag `creatorUserId` a Workspace creati prima del 16 novembre 2021. Per limitare chi può configurare la collaborazione, si consiglia di aggiungere manualmente il tag `creatorUserId` al Workspace e di utilizzare il controllo degli accessi basato su tag nei criteri di autorizzazione utente.

La seguente istruzione di esempio consente a un utente di configurare la collaborazione per tutti i Workspace con la chiave di tag `creatorUserId` il cui valore corrisponde all'ID dell'utente (indicato dalla variabile `aws:userId` della policy). In altre parole, l'istruzione consente a un utente di configurare la collaborazione per Workspace creati. Per ulteriori informazioni sulle variabili della policy, consulta Elementi delle policy IAM: variabili e tag.

```
{
    "Sid": "UserRolePermissionsForCollaboration",
    "Action": [
        "elasticmapreduce:UpdateEditor",
        "elasticmapreduce:PutWorkspaceAccess",
        "elasticmapreduce:DeleteWorkspaceAccess",
        "elasticmapreduce:ListWorkspaceAccessIdentities"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Condition": {
        "StringEquals": {
            "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
        }
    }
}
```

2. Collega la policy di autorizzazione alla tua identità IAM.

   La tabella seguente riepiloga l'identità IAM a cui si collega una policy di autorizzazione, a seconda della modalità di autenticazione di EMR Studio. Per ricevere istruzioni su come allegare una policy, consulta Aggiunta e rimozione di autorizzazioni per identità IAM.

Se utilizzi...	Collega la policy a…
Autenticazione IAM	Le identità IAM (utenti, gruppi di utenti o ruoli). Ad esempio, prova a collegare una policy di autorizzazione a un utente nel Account AWS.
Federazione IAM con un provider di identità esterno (IdP)	Il ruolo o i ruoli IAM creati per il tuo IdP esterno. Per esempio, una federazione IAM per SAML 2.0. EMR Studio utilizza le autorizzazioni collegate ai ruoli IAM per l'utente o gli utenti con accesso federato a uno Studio.
IAM Identity Center	Il tuo ruolo utente di Amazon EMR Studio.

Esempi di policy utente

La seguente policy utente di base consente la maggior parte delle operazioni di EMR Studio, ma non consente a un utente di creare nuovi cluster Amazon EMR.

Policy di base

> **Important**
>
> La policy di esempio non include l'autorizzazione `CreateStudioPresignedUrl`, che è necessario concedere a un utente quando si utilizza la modalità di autenticazione IAM. Per ulteriori informazioni, consulta Assegnare un utente o un gruppo a un EMR Studio.

La policy di esempio include elementi `Condition` per applicare il controllo degli accessi basato su tag (TBAC) in modo da poter utilizzare la policy con il ruolo di servizio di esempio per EMR Studio. Per ulteriori informazioni, consulta Creazione di un ruolo di servizio 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 seguente policy utente intermedia consente la maggior parte delle operazioni di EMR Studio e consente a un utente di creare nuovi cluster Amazon EMR utilizzando un modello di cluster.

Policy intermedia

> **Important**
>
> La policy di esempio non include l'autorizzazione `CreateStudioPresignedUrl`, che è necessario concedere a un utente quando si utilizza la modalità di autenticazione IAM. Per ulteriori informazioni, consulta Assegnare un utente o un gruppo a un EMR Studio.

La policy di esempio include elementi `Condition` per applicare il controllo degli accessi basato su tag (TBAC) in modo da poter utilizzare la policy con il ruolo di servizio di esempio per EMR Studio. Per ulteriori informazioni, consulta Creazione di un ruolo di servizio 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 seguente policy utente avanzata consente tutte le operazioni di EMR Studio e consente a un utente di creare nuovi cluster Amazon EMR utilizzando un modello di cluster o fornendo una configurazione cluster.

Policy avanzata

> **Important**
>
> La policy di esempio non include l'autorizzazione `CreateStudioPresignedUrl`, che è necessario concedere a un utente quando si utilizza la modalità di autenticazione IAM. Per ulteriori informazioni, consulta Assegnare un utente o un gruppo a un EMR Studio.

La policy di esempio include elementi `Condition` per applicare il controllo degli accessi basato su tag (TBAC) in modo da poter utilizzare la policy con il ruolo di servizio di esempio per EMR Studio. Per ulteriori informazioni, consulta Creazione di un ruolo di servizio 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 seguente policy utente contiene le autorizzazioni utente minime necessarie per utilizzare un'applicazione interattiva EMR Serverless con EMR Studio WorkSpaces.

Policy EMR Serverless interattiva

In questo esempio di policy che prevede le autorizzazioni utente per le applicazioni interattive EMR Serverless con EMR Studio, sostituisci i segnaposto per *serverless-runtime-role* ed *emr-studio-service-role* con il ruolo di servizio EMR Studio e il ruolo runtime EMR Serverless corretti.

```
{
    "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": "*"
        }
    ]
}
```

Autorizzazioni AWS Identity and Access Management per gli utenti di EMR Studio

La tabella seguente include ogni operazione di Amazon EMR Studio che un utente potrebbe eseguire ed elenca le azioni IAM minime necessarie per eseguire tale operazione. Consenti queste operazioni nelle policy di autorizzazione IAM (quando utilizzi l'autenticazione IAM) o nelle policy di sessione (quando utilizzi l'autenticazione IAM Identity Center) per EMR Studio.

La tabella riporta inoltre le operazioni consentite in ciascuna delle policy di autorizzazione di esempio per EMR Studio. Per ulteriori informazioni su policy di autorizzazione di esempio, consulta Creazione di policy di autorizzazione per gli utenti di EMR Studio.

Azione	Base	Intermedia	Avanzata	Azioni associate
Creazione ed eliminazione di Workspace	Sì	Sì	Sì	`"elasticmapreduce:CreateEditor", "elasticmapreduce:DescribeEditor", "elasticmapreduce:ListEditors", "elasticmapreduce:DeleteEditor"`
Visualizza il pannello Collaborazione, abilita	Sì	Sì	Sì	`"elasticmapreduce:UpdateEditor",`

Azione	Base	Intermedia	Avanzata	Azioni associate
la collaborazione di WorkSpace e aggiungi collaboratori. Per ulteriori informazioni, consulta Imposta la proprietà per la collaborazione di Workspace.				`"elasticmapreduce:PutWorkspaceAccess", "elasticmapreduce:DeleteWorkspaceAccess", "elasticmapreduce:ListWorkspaceAccessIdentities"`
Visualizza un elenco di bucket di archiviazione Amazon S3 Control nello stesso account di Studio quando crei un nuovo cluster EMR e accedi ai log del container quando utilizzi un'interfaccia utente Web per eseguire il debug delle applicazioni	Sì	Sì	Sì	`"s3:ListAllMyBuckets", "s3:ListBucket", "s3:GetBucketLocation", "s3:GetObject"`
Accesso ai Workspace	Sì	Sì	Sì	`"elasticmapreduce:DescribeEditor", "elasticmapreduce:ListEditors", "elasticmapreduce:StartEditor", "elasticmapreduce:StopEditor", "elasticmapreduce:OpenEditorInConsole"`

Azione	Base	Intermedia	Avanzata	Azioni associate
Collegamento o scollegamento di cluster Amazon EMR esistenti associati al Workspace	Sì	Sì	Sì	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster", "elasticmapreduce:ListInstanceGroups", "elasticmapreduce:ListBootstrapActions"`
Collegamento o scollegamento di cluster Amazon EMR su EKS	Sì	Sì	Sì	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListManagedEndpoints", "emr-containers:DescribeManagedEndpoint", "emr-containers:GetManagedEndpointSessionCredentials"`

Azione	Base	Intermedia	Avanzata	Azioni associate
Collegamento o scollegamento delle applicazioni EMR Serverless associate al Workspace	No	Sì	Sì	`"elasticmapreduce:AttachEditor", "elasticmapreduce:DetachEditor", "emr-serverless:GetApplication", "emr-serverless:StartApplication", "emr-serverless:ListApplications", "emr-serverless:GetDashboardForJobRun", "emr-serverless:AccessInteractiveEndpoints", "iam:PassRole"` L'autorizzazione `PassRole` è necessaria per passare il ruolo di runtime del processo EMR Serverless. Per ulteriori informazioni, consulta la sezione Ruoli di runtime del processo nella Guida per l'utente di Amazon EMR serverless.

Azione	Base	Intermedia	Avanzata	Azioni associate
Esecuzione del debug dei processi Amazon EMR su EC2 con interfacce utente dell'applicazione persistente	Sì	Sì	Sì	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "elasticmapreduce:ListClusters", "elasticmapreduce:ListSteps", "elasticmapreduce:DescribeCluster", "s3:ListBucket", "s3:GetObject"`
Esecuzione del debug dei processi Amazon EMR su EC2 con interfacce utente dell'applicazione su cluster	Sì	Sì	Sì	`"elasticmapreduce:GetOnClusterAppUIPresignedURL"`

Azione	Base	Intermedia	Avanzata	Azioni associate
Esecuzione del debug delle esecuzioni di processo Amazon EMR su EKS utilizzando Spark History Server	Sì	Sì	Sì	`"elasticmapreduce:CreatePersistentAppUI", "elasticmapreduce:DescribePersistentAppUI", "elasticmapreduce:GetPersistentAppUIPresignedURL", "emr-containers:ListVirtualClusters", "emr-containers:DescribeVirtualCluster", "emr-containers:ListJobRuns", "emr-containers:DescribeJobRun", "s3:ListBucket", "s3:GetObject"`
Creazione ed eliminazione di repository Git	Sì	Sì	Sì	`"elasticmapreduce:CreateRepository", "elasticmapreduce:DeleteRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository", "secretsmanager:CreateSecret", "secretsmanager:ListSecrets", "secretsmanager:TagResource"`

Azione	Base	Intermedia	Avanzata	Azioni associate
Collegamento e scollegamento di repository Git	Sì	Sì	Sì	`"elasticmapreduce:LinkRepository", "elasticmapreduce:UnlinkRepository", "elasticmapreduce:ListRepositories", "elasticmapreduce:DescribeRepository"`
Creazione di nuovi cluster da modelli di cluster predefiniti	No	Sì	Sì	`"servicecatalog:SearchProducts", "servicecatalog:DescribeProduct", "servicecatalog:DescribeProductView", "servicecatalog:DescribeProvisioningParameters", "servicecatalog:ProvisionProduct", "servicecatalog:UpdateProvisionedProduct", "servicecatalog:ListProvisioningArtifacts", "servicecatalog:DescribeRecord", "servicecatalog:ListLaunchPaths", "cloudformation:DescribeStackResources", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`

Azione	Base	Intermedia	Avanzata	Azioni associate
Creazione di nuovi cluster fornendo una configurazione cluster	No	No	Sì	`"elasticmapreduce:RunJobFlow", "iam:PassRole", "elasticmapreduce:ListClusters", "elasticmapreduce:DescribeCluster"`
Assegnare un utente a uno Studio quando si utilizza la modalità di autenticazione IAM. Per ulteriori informazioni, consulta Assegnare un utente o un gruppo a un EMR Studio.	No	No	No	`"elasticmapreduce:CreateStudioPresignedUrl"`

Azione	Base	Intermedia	Avanzata	Azioni associate

Descrivi oggetti di rete.	Sì	Sì	Sì	<pre>{
				    "Version": "2012-10-17",
				    "Statement": [
				        {
				            "Sid": "DescribeNetwork",
				            "Effect": "Allow",
				            "Action": [
				                "ec2:DescribeVpcs",
				                "ec2:DescribeSubnets",
				                "ec2:DescribeSecurityGroups"
				            ],
				            "Resource": "*"
				        }
				    ]
				}</pre>

Azione	Base	Intermedia	Avanzata	Azioni associate
Elenca i ruoli IAM.	Sì	Sì	Sì	`{ "Version": "2012-10-17", "Statement": [ { "Sid": "ListIAMRoles", "Effect": "Allow", "Action": [ "iam:ListRoles" ], "Resource": "*" } ] }`
Connettiti a EMR Studio da Amazon SageMaker Studio e usa l'interfaccia visiva di Data Wrangler.	No	No	Sì	`"sagemaker:CreatePresignedDomainUrl", "sagemaker:DescribeDomain", "sagemaker:ListDomains", "sagemaker:ListUserProfiles"`

Creazione di un EMR Studio

Puoi creare un EMR Studio per il tuo team utilizzando la console Amazon EMR o la AWS CLI. La creazione di un'istanza Studio fa parte della configurazione di Amazon EMR Studio.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

Prerequisiti

Prima di creare uno Studio, assicurati di aver completato i processi precedenti in Configurazione di un Amazon EMR Studio.

Per creare uno Studio utilizzando la AWS CLI sarebbe necessario installare la versione più recente. Per ulteriori informazioni, consulta Installare o aggiornare la versione più recente della AWS CLI.

> **Important**
>
> Disattivare gli strumenti di gestione proxy come FoxyProxy o SwitchyOmega nel browser prima di creare uno Studio. I proxy attivi possono generare un messaggio di errore Errore di rete quando scegli Crea Studio.

New console

Creazione di un EMR Studio con la nuova console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR Studio, nella barra di navigazione a sinistra, scegli Nozioni di base. È inoltre possibile creare un nuovo Studio dalla pagina Studio.
3. Seleziona Crea uno Studio per aprire la pagina Crea uno Studio.
4. Immetti un Nome Studio e una Descrizione facoltativa.
5. Se si utilizza l'autenticazione IAM per Studio, è possibile scegliere Aggiungi nuovo tag per aggiungere uno o più tag chiave-valore di tua scelta allo Studio. Utilizzate i tag per consentire agli utenti specificati l'accesso allo Studio. Per ulteriori informazioni, consulta Assegnare un utente o un gruppo a un EMR Studio.

   Puoi anche aggiungere tag per aiutarti a gestire, identificare, organizzare e filtrare Studio. Per ulteriori informazioni, consulta Assegnazione di tag alle risorse AWS.

6. In Networking (Reti), scegli un Amazon Virtual Private Cloud (VPC) per il Studio dall'elenco a discesa.
7. In Subnets (Sottoreti) seleziona un massimo di cinque sottoreti in VPC da associare allo Studio. È possibile aggiungere altre sottoreti dopo aver creato lo Studio.
8. Per Security groups (Gruppi di sicurezza), scegli i gruppi di sicurezza di default o i gruppi di sicurezza personalizzati. Per ulteriori informazioni, consulta Definizione di gruppi di sicurezza per controllare il traffico di rete EMR Studio.

Se scegli...	Esegui questa operazione…
I gruppi di sicurezza EMR Studio predefiniti	Per abilitare il collegamento del repository basato su Git per il Studio, scegli Enable clusters/endpoints and Git repository (Abilita cluster/endpoint e repository Git). In caso contrario, scegli Abilita cluster/endpoint.
Gruppi di sicurezza personalizzati per lo Studio	• In Cluster/endpoint security group (Gruppo di sicurezza cluster/endpoint), seleziona il gruppo di sicurezza del motore configurato dall'elenco a discesa. Lo Studio utilizza questo gruppo di sicurezza per consentire l'accesso in ingresso dai WorkSpace collegati. • In WorkSpace security group (Gruppo di sicurezza dell'istanza WorkSpace), seleziona il gruppo di sicurezza dell'istanza WorkSpace configurata dall'elenco a discesa. Il tuo Studio utilizza questo gruppo di sicurezza con WorkSpace per fornire l'accesso in uscita ai cluster Amazon EMR collegati e ai repository Git ospitati pubblicamente.

9. In Autenticazione, scegli una modalità di autenticazione per Studio e fornisci informazioni in base alla seguente tabella. Per ulteriori informazioni sull'autenticazione per EMR Studio, consulta Scelta di una modalità di autenticazione per Amazon EMR Studio.

Se utilizzi...	Esegui questa operazione…
Autenticazione o federazione IAM	Scegli un metodo di accesso per Studio. Se desideri che gli utenti federati accedano utilizzando l'URL di Studio e le credenzia li per il tuo provider di identità (IdP), seleziona il tuo IdP dall'elenco a discesa e inserisci il tuo URL di accesso del provider di identità (IdP) e il nome del parametro RelayState. Per un elenco degli URL di autenticazione IdP e dei nomi di RelayState, consulta Parametri RelayState del provider di identità e URL di autenticazione. Quindi, seleziona il EMR Studio Ruolo di servizio dall'elenco a discesa. Per ulteriori informazioni, consulta Creazione di un ruolo di servizio EMR Studio.
Autenticazione IAM Identity Center	Seleziona il Ruolo di servizio e il Ruolo utente di EMR Studio. Per ulteriori informazi oni, consulta Creazione di un ruolo di servizio EMR Studio e Creazione di un ruolo utente di EMR Studio per la modalità di autenticazione IAM Identity Center.

10. In WorkSpace storage (Archiviazione delle istanze WorkSpace), scegli Browse S3 (Sfoglia S3) per selezionare il bucket Amazon S3 designato per il backup delle istanze WorkSpace e i file notebook.

> **Note**
>
> Il ruolo di servizio EMR Studio deve disporre dell'accesso in lettura e scrittura al bucket selezionato.

11. Scegli Crea uno Studio per terminare e passare alla pagina Studio. Il nuovo Studio è visibile nell'elenco con dettagli quali Nome Studio, Data di creazione e URL di accesso allo Studio.

Dopo aver creato uno Studio, segui le istruzioni riportate in Assegnare un utente o un gruppo a un EMR Studio.

Old console

Creazione di un EMR Studio con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/home.
2. Seleziona EMR Studio dalla barra di navigazione a sinistra.
3. Seleziona Crea uno Studio per aprire la pagina Crea uno Studio.
4. Immetti un Nome Studio e una Descrizione facoltativa.
5. Se si utilizza l'autenticazione IAM per Studio, è possibile scegliere Aggiungi nuovo tag per aggiungere uno o più tag chiave-valore di tua scelta allo Studio. Utilizzate i tag per consentire agli utenti specificati l'accesso allo Studio. Per ulteriori informazioni, consulta Assegnare un utente o un gruppo a un EMR Studio.

   Puoi anche aggiungere tag per aiutarti a gestire, identificare, organizzare e filtrare Studio. Per ulteriori informazioni, consulta Assegnazione di tag alle risorse AWS.
6. In Networking (Reti), scegli un Amazon Virtual Private Cloud (VPC) per il Studio dall'elenco a discesa.
7. In Subnets (Sottoreti) seleziona un massimo di cinque sottoreti in VPC da associare allo Studio. È possibile aggiungere altre sottoreti dopo aver creato lo Studio.
8. Per Security groups (Gruppi di sicurezza), scegli i gruppi di sicurezza di default o i gruppi di sicurezza personalizzati. Per ulteriori informazioni, consulta Definizione di gruppi di sicurezza per controllare il traffico di rete EMR Studio.

Se scegli...	Esegui questa operazione…
I gruppi di sicurezza EMR Studio predefiniti	Per abilitare il collegamento del repository basato su Git per il Studio, scegli Enable clusters/endpoints and Git repository (Abilita cluster/endpoint e repository Git).

Se scegli...	Esegui questa operazione…
	In caso contrario, scegli Abilita cluster/endpoint.
Gruppi di sicurezza personalizzati per lo Studio	• In Cluster/endpoint security group (Gruppo di sicurezza cluster/endpoint), seleziona il gruppo di sicurezza del motore configurato dall'elenco a discesa. Lo Studio utilizza questo gruppo di sicurezza per consentire l'accesso in ingresso dai WorkSpace collegati. • In WorkSpace security group (Gruppo di sicurezza dell'istanza WorkSpace), seleziona il gruppo di sicurezza dell'istanza WorkSpace configurata dall'elenco a discesa. Il tuo Studio utilizza questo gruppo di sicurezza con WorkSpace per fornire l'accesso in uscita ai cluster Amazon EMR collegati e ai repository Git ospitati pubblicamente.

9. In Autenticazione, scegli una modalità di autenticazione per Studio e fornisci informazioni in base alla seguente tabella. Per ulteriori informazioni sull'autenticazione per EMR Studio, consulta Scelta di una modalità di autenticazione per Amazon EMR Studio.

Se utilizzi...	Esegui questa operazione…
Autenticazione o federazione IAM	Scegli un metodo di accesso per Studio. Se desideri che gli utenti federati accedano utilizzando l'URL di Studio e le credenziali per il tuo provider di identità (IdP), seleziona il tuo IdP dall'elenco a discesa e inserisci il tuo URL di accesso del provider di identità (IdP) e il nome del parametro RelayState.

Se utilizzi...	Esegui questa operazione…
	Per un elenco degli URL di autenticazione IdP e dei nomi di RelayState, consulta Parametri RelayState del provider di identità e URL di autenticazione. Quindi, seleziona il EMR Studio Ruolo di servizio dall'elenco a discesa. Per ulteriori informazioni, consulta Creazione di un ruolo di servizio EMR Studio.
Autenticazione IAM Identity Center	Seleziona il Ruolo di servizio e il Ruolo utente di EMR Studio. Per ulteriori informazioni, consulta Creazione di un ruolo di servizio EMR Studio e Creazione di un ruolo utente di EMR Studio per la modalità di autenticazione IAM Identity Center.

10. In WorkSpace storage (Archiviazione delle istanze WorkSpace), scegli Browse S3 (Sfoglia S3) per selezionare il bucket Amazon S3 designato per il backup delle istanze WorkSpace e i file notebook.

> **Note**
>
> Il ruolo di servizio EMR Studio deve disporre dell'accesso in lettura e scrittura al bucket selezionato.

11. Scegli Crea uno Studio per terminare e passare alla pagina Studio. Il nuovo Studio è visibile nell'elenco con dettagli quali Nome Studio, Data di creazione e URL di accesso allo Studio.

Dopo aver creato uno Studio, segui le istruzioni riportate in Assegnare un utente o un gruppo a un EMR Studio.

CLI

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

Example Comando CLI per creare un EMR Studio con modalità di autenticazione IAM

Il seguente comando di esempio della AWS CLI crea un EMR Studio che utilizza la modalità di autenticazione IAM. Quando utilizzi l'autenticazione o la federazione IAM per lo Studio, non specifichi un `--user-role`.

Per consentire agli utenti federati di accedere utilizzando l'URL di Studio e le credenziali per il tuo gestore dell'identità digitale (IdP), specifica `--idp-auth-url` e `--idp-relay-state-parameter-name`. Per avere un elenco degli URL di autenticazione IdP e dei nomi di RelayState, consulta Parametri RelayState del provider di identità e URL di autenticazione.

```
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 Il comando della CLI per creare un EMR Studio con modalità di autenticazione IAM Identity Center

Il seguente comando di esempio della AWS CLI crea un EMR Studio che utilizza la modalità di autenticazione IAM Identity Center. Quando utilizzi l'autenticazione IAM Identity Center, devi specificare un `--user-role`.

Per ulteriori informazioni sulla modalità di autenticazione IAM Identity Center, consulta la sezione Configurazione della modalità di autenticazione IAM Identity Center per 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 Output della CLI per **`aws emr create-studio`**

Di seguito è riportato un esempio dell'output visualizzato dopo la creazione di uno Studio.

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

Per ulteriori informazioni sul comando `create-studio`, consulta la Guida di riferimento ai comandi della AWS CLI.

Parametri RelayState del provider di identità e URL di autenticazione

Quando si utilizza la federazione IAM e si desidera che gli utenti accedano utilizzando l'URL di Studio e le credenziali per il proprio provider di identità (IdP), è possibile specificare l'URL di accesso del provider di identità (IdP) e nome del parametro RelayState quando si Creazione di un EMR Studio.

La tabella seguente mostra l'URL di autenticazione standard e il nome del parametro RelayState per alcuni provider di identità più diffusi.

Provider di identità	Parametro	Autenticazione URL
Auth0	`RelayState`	`https://<sub_domain> .auth0.com/samlp/<app_id>`
Account Google	`RelayState`	`https://accounts.google.com/o/saml2/initsso?i`

Provider di identità	Parametro	Autenticazione URL
		dpid= *<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>*

Assegnazione e gestione degli utenti di EMR Studio

Dopo aver creato un EMR Studio, è possibile assegnare utenti e gruppi ad esso. Il metodo utilizzato per assegnare, aggiornare e rimuovere utenti dipende dalla modalità di autenticazione di Studio.

- Quando utilizzi la modalità di autenticazione IAM, devi configurare l'assegnazione utente di EMR Studio e le autorizzazioni in IAM o con IAM e il provider di identità.
- Con la modalità di autenticazione IAM Identity Center, utilizzi la console di gestione Amazon EMR o la AWS CLI per gestire gli utenti.

Per ulteriori informazioni sull'autenticazione per Amazon EMR Studio, consulta Scelta di una modalità di autenticazione per Amazon EMR Studio.

Assegnare un utente o un gruppo a un EMR Studio

IAM

Quando si utilizza Configurare una modalità di autenticazione IAM per Amazon EMR Studio è necessario consentire l'operazione `CreateStudioPresignedUrl` nella policy di

autorizzazione IAM di un utente e limitare l'utente a un particolare Studio. È possibile includere `CreateStudioPresignedUrl` nell'Autorizzazioni utente per la modalità di autenticazione IAM o usare una policy separata.

Per limitare un utente a uno Studio (o a un set di Studio), è possibile utilizzare il controllo di accesso basato sugli attributi (ABAC) o specificare un nome della risorsa Amazon (ARN) di uno Studio nell'elemento `Resource` della policy di autorizzazione dell'utente.

Example Assegnazione di un utente a uno Studio con un ARN di Studio

La seguente policy di esempio fornisce a un utente l'accesso a un particolare EMR Studio consentendo l'operazione `CreateStudioPresignedUrl` e specificando il nome della risorsa Amazon (ARN) di Studio nel 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 Assegnazione di un utente a uno Studio con ABAC per l'autenticazione IAM

Esistono diversi modi per configurare il controllo di accesso basato sugli attributi (ABAC) per uno Studio. Ad esempio, è possibile allegare uno o più tag a un EMR Studio e quindi creare una policy IAM che limita l'operazione `CreateStudioPresignedUrl` per uno Studio particolare o un insieme di Studio con tali tag.

Puoi aggiungere tag durante o dopo la creazione di Studio. Per aggiungere tag a uno Studio esistente, puoi utilizzare il comando AWS CLI`emr add-tags`. L'esempio seguente aggiunge un tag con la coppia chiave-valore `Team = Data Analytics` a un EMR Studio.

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

La seguente policy di autorizzazione di esempio consente l'operazione `CreateStudioPresignedUrl` per EMR Studio con la coppia chiave-valore di tag `Team = DataAnalytics`. Per ulteriori informazioni sull'uso dei tag per controllare l'accesso, consulta Controllo dell'accesso a e per utenti e ruoli IAM mediante i tag o Controllo dell'accesso alle risorse AWS mediante i tag.

```
{
    "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 Assegnazione di un utente a uno Studio con la chiave di condizione globale aws:SourceIdentity

Quando usi la federazione IAM, puoi utilizzare la chiave di condizione globale `aws:SourceIdentity` in una policy di autorizzazione per consentire agli utenti di accedere a Studio quando assumono il ruolo IAM per la federazione.

È necessario innanzitutto configurare il proprio provider di identità (IdP) per restituire una stringa di identificazione, ad esempio un indirizzo e-mail o un nome utente, quando un utente autentica e assume il ruolo IAM per la federazione. IAM imposta la chiave di condizione globale `aws:SourceIdentity` alla stringa di identificazione restituita dal tuo IdP.

Per ulteriori informazioni, consulta Come mettere in relazione l'attività del ruolo IAM con l'identità aziendale post del blog nel Blog sulla sicurezza AWS e la voce aws:SourceIdentity nel Documentazione di riferimento delle chiavi di condizione globale.

La seguente policy di esempio consente l'operazione `CreateStudioPresignedUrl` e offre agli utenti un `aws:SourceIdentity` che corrisponde all'accesso *<example-source-identity>* all'EMR Studio specificato da*<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

Quando si assegna un utente o un gruppo a un EMR Studio, specificare una policy di sessione che definisca autorizzazioni granulari, ad esempio la possibilità di creare un nuovo cluster EMR, per tale utente o gruppo. Amazon EMR archivia le mappature delle policy di sessione. È possibile aggiornare le policy di sessione a un utente o a un gruppo dopo l'assegnazione.

> **Note**
>
> Le autorizzazioni finali per un utente o un gruppo sono un'intersezione tra le autorizzazioni definite nel ruolo utente di EMR Studio e le autorizzazioni definite nella policy di sessione per tale utente o gruppo. Se un utente appartiene a più gruppi assegnati allo Studio, EMR Studio utilizza un'unione di autorizzazioni per quel determinato utente.

Assegnazione di utenti o gruppi a un EMR Studio con la console Amazon EMR

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.

2. Scegli EMR Studio dalla barra di navigazione a sinistra.
3. Scegli il nome del tuo Studio dal menu Studios (Studio) oppure seleziona lo Studio e scegli View details (Visualizza dettagli) per aprire la pagina dei dettagli dello Studio.
4. Seleziona Add Users (Aggiungi utenti) per visualizzare la tabella di ricerca Users (Utenti) e Groups (Gruppi).
5. Seleziona la scheda Users (Utenti) oppure la scheda Groups (Gruppi) e inserisci un termine di ricerca nella barra di ricerca per trovare un utente o un gruppo.
6. Seleziona uno o più utenti o gruppi dall'elenco dei risultati di ricerca. Puoi passare dalla scheda Utenti alla scheda Gruppi e viceversa.
7. Dopo aver selezionato gli utenti e i gruppi da aggiungere allo Studio, scegli Add (Aggiungi). Dovresti visualizzare gli utenti e i gruppi nell'elenco Utenti dello Studio. Potrebbero essere necessari alcuni secondi prima che l'elenco venga aggiornato.
8. Segui le istruzioni in Aggiornamento delle autorizzazioni per un utente o un gruppo assegnato a uno Studio per perfezionare le autorizzazioni Studio per un utente o un gruppo.

Assegnazione di un utente o un gruppo a EMR Studio tramite la AWS CLI

Inserisci valori personalizzati per i seguenti argomenti `create-studio-session-mapping`. Per ulteriori informazioni sul comando `create-studio-session-mapping`, consulta la Guida di riferimento ai comandi della AWS CLI.

- **`--studio-id`**: l'ID dello Studio a cui si desidera assegnare l'utente o il gruppo. Per ricevere istruzioni su come recuperare l'ID dello Studio, consulta Visualizzazione dei dettagli dello Studio.
- **`--identity-name`**: il nome dell'utente o del gruppo dall'archivio identità. Per ulteriori informazioni, consulta UserName e DisplayName per i gruppi nella Guida di riferimento alle API dell'archivio identità .
- **`--identity-type`**: utilizza `USER` o `GROUP` per specificare il tipo di identità.
- **`--session-policy-arn`**: il nome della risorsa Amazon (ARN) per la policy di sessione che desideri associare all'utente o al gruppo. Ad esempio, **`arn:aws:iam::`*`<aws-account-id>`*`:policy/`*`EMRStudio_Advanced_User_Policy`**. Per ulteriori informazioni, consulta Creazione di policy di autorizzazione per gli utenti di 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**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

Utilizza il comando `get-studio-session-mapping` per verificare la nuova assegnazione. Sostituisci *<example-identity-name>* con il nome IAM Identity Center dell'utente o del gruppo aggiornato.

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

Aggiornamento delle autorizzazioni per un utente o un gruppo assegnato a uno Studio

IAM

Per aggiornare le autorizzazioni utente o di gruppo quando si utilizza la modalità di autenticazione IAM, utilizzare IAM per modificare le policy di autorizzazione IAM collegate alle identità IAM (utenti, gruppi o ruoli).

Per ulteriori informazioni, consulta Autorizzazioni utente per la modalità di autenticazione IAM.

IAM Identity Center

Aggiornamento delle autorizzazioni di EMR Studio per un utente o un gruppo mediante la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli EMR Studio dalla barra di navigazione a sinistra.

3. Scegli il nome del tuo Studio dal menu Studios (Studio) oppure seleziona lo Studio e scegli View details (Visualizza dettagli) per aprire la pagina dei dettagli dello Studio.
4. Nell'elenco Studio users (Utenti dello Studio) nella pagina dei dettagli dello Studio, cerca l'utente o il gruppo che desideri aggiornare. Puoi eseguire la ricerca per nome o tipo di identità.
5. Seleziona l'utente o il gruppo che desideri aggiornare e scegli Assign policy (Assegna policy) per aprire la finestra di dialogo Session policy (Policy di sessione).
6. Seleziona una policy da applicare all'utente o al gruppo scelto nella fase 5 e scegli Apply policy (Applica policy). L'elenco Utenti dello Studio dovrebbe mostrare il nome della policy nella colonna Policy di sessione per l'utente o il gruppo aggiornato.

Aggiornamento delle autorizzazioni di EMR Studio per un utente o un gruppo utilizzando AWS CLI

Inserisci valori personalizzati per i seguenti argomenti `update-studio-session-mappings`. Per ulteriori informazioni sul comando `update-studio-session-mappings`, consulta la Guida di riferimento ai comandi della 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> \
```

Utilizza il comando `get-studio-session-mapping` per verificare la nuova assegnazione della policy di sessione. Sostituisci *<example-identity-name>* con il nome IAM Identity Center dell'utente o del gruppo aggiornato.

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

Rimozione di un utente o un gruppo da uno Studio

IAM

Per rimuovere un utente o un gruppo da EMR Studio quando si utilizza la modalità di autenticazione IAM, è necessario revocare l'accesso dell'utente a Studio riconfigurando la policy di autorizzazione IAM dell'utente.

Nella seguente policy di esempio, supponiamo che tu disponga di un EMR Studio con la coppia chiave-valore di tag `Team = Quality Assurance`. Secondo la policy, l'utente può accedere a Studio taggati con la chiave `Team` il cui valore è uguale a `Data Analytics` o `Quality Assurance`. Per rimuovere l'utente dallo Studio taggato con `Team = Quality Assurance`, rimuovi `Quality Assurance` dall'elenco dei valori dei tag.

```
{
    "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

Rimozione di un utente o un gruppo da un EMR Studio mediante la console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli EMR Studio dalla barra di navigazione a sinistra.
3. Scegli il nome del tuo Studio dal menu Studios (Studio) oppure seleziona lo Studio e scegli View details (Visualizza dettagli) per aprire la pagina dei dettagli dello Studio.
4. Nell'elenco Studio users (Utenti dello Studio) nella pagina dei dettagli dello Studio, individua l'utente o il gruppo che desideri rimuovere dallo Studio. Puoi eseguire la ricerca per nome o tipo di identità.
5. Seleziona l'utente o il gruppo che desideri eliminare, quindi scegli Delete (Elimina) e conferma. L'utente o il gruppo eliminato scompare dall'elenco Utenti dello Studio.

Rimozione di un utente o di un gruppo da un EMR Studio utilizzando AWS CLI

Inserisci valori personalizzati per i seguenti argomenti `delete-studio-session-mapping`. Per ulteriori informazioni sul comando `delete-studio-session-mapping`, consulta la Guida di riferimento ai comandi della 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> \
```

Gestione di un Amazon EMR Studio

Questa sezione include istruzioni per monitorare, aggiornare o eliminare una risorsa EMR Studio. Per informazioni su come assegnare utenti o aggiornare le autorizzazioni utente, consulta Assegnazione e gestione degli utenti di EMR Studio.

Visualizzazione dei dettagli dello Studio

New console

Visualizzazione dei dettagli di un EMR Studio con la nuova console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR Studio, nella barra di navigazione a sinistra, scegli Studio.
3. Seleziona lo Studio dall'elenco Studio per aprire la pagina dei suoi dettagli. La pagina dei dettagli dello Studio include le informazioni di Configurazione dello Studio, quali Descrizione, VPC e Sottoreti dello Studio.

Old console

Visualizzazione dei dettagli di un EMR Studio con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/home.
2. Seleziona EMR Studio dalla barra di navigazione a sinistra.
3. Seleziona il Studio dall'elenco Studios (Studio) per aprire la pagina dei dettagli dello Studio. La pagina dei dettagli dello Studio include le informazioni di Configurazione dello Studio, quali Descrizione, VPC e Sottoreti dello Studio.

CLI

Recupero dei dettagli di un EMR Studio tramite l'ID dello Studio utilizzando AWS CLI

Utilizza il seguente comando `describe-studio` della AWS CLI per recuperare informazioni dettagliate su un EMR Studio specifico. Per ulteriori informazioni, consulta la Guida di riferimento ai comandi della AWS CLI.

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

Recupero di un elenco degli EMR Studio utilizzando AWS CLI

Utilizza il seguente comando `list-studios` della AWS CLI. Per ulteriori informazioni, consulta la Guida di riferimento ai comandi della AWS CLI.

```
aws emr list-studios
```

Di seguito è riportato un valore restituito di esempio per il comando `list-studios` in 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"
        }
    ]
}
```

Monitoraggio delle operazioni di Amazon EMR Studio

Visualizzazione dell'attività di EMR Studio e dell'API

EMR Studio è integrato con AWS CloudTrail, un servizio che offre un registro delle operazioni eseguite da un utente, da un ruolo IAM o da un altro servizio AWS in EMR Studio. CloudTrail acquisisce le chiamate API per EMR Studio come eventi. È possibile visualizzare gli eventi utilizzando la console CloudTrail all'indirizzo https://console.aws.amazon.com/cloudtrail/.

Gli eventi EMR Studio forniscono informazioni come quale utente Studio o IAM effettua una richiesta e il tipo di richiesta.

> **Note**
>
> Le operazioni on-cluster come l'esecuzione di processi notebook non emetto AWS CloudTrail.

È inoltre possibile creare un trail per la distribuzione continua di eventi CloudTrail EMR Studio in un bucket Amazon S3. Per ulteriori informazioni, consulta la Guida per l'utente di AWS CloudTrail.

Esempio di evento CloudTrail: un utente effettua una chiamata all'API DescribeStudio

Di seguito è riportato un esempio di evento AWS CloudTrail che viene creato quando un utente, `admin`, effettua una chiamata all'API DescribeStudio. CloudTrail registra il nome utente come `admin`.

> **Note**
>
> Per proteggere i dettagli di Studio, l'evento API EMR Studio per DescribeStudio esclude un valore per `responseElements`.


```
{
   "eventVersion":"1.08",
   "userIdentity":{
      "type":"IAMUser",
      "principalId":"AIDXXXXXXXXXXXXXXXXX",
      "arn":"arn:aws:iam::653XXXXXXXX:user/admin",
      "accountId":"653XXXXXXXX",
      "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-9O5XXXXXXXXXXXXXXXXXXX"
   },
   "responseElements":null,
   "requestID":"0fxxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "eventID":"b0xxxxxx-xxxx-xxxx-xxxx-xxxxxxxxxxxx",
   "readOnly":true,
   "eventType":"AwsApiCall",
   "managementEvent":true,
   "eventCategory":"Management",
   "recipientAccountId":"653XXXXXXXX"
}
```

Visualizzazione dell'attività degli utenti e dei processi Spark

Per visualizzare l'attività dei processi Spark da parte degli utenti di Amazon EMR Studio, è possibile configurare la rappresentazione utente in un cluster. Con la rappresentazione utente, ogni processo Spark inviato da un Workspace è associato all'utente Studio che ha eseguito il codice.

Quando la rappresentazione utente è abilitata, Amazon EMR crea una directory utente HDFS nel nodo primario del cluster per ogni utente che esegue il codice nel Workspace. Ad esempio, se l'utente `studio-user-1@example.com` esegue il codice, puoi collegarti al nodo primario per riscontrare che `hadoop fs -ls /user` ha una directory per `studio-user-1@example.com`.

Per impostare la rappresentazione utente di Spark, imposta le seguenti proprietà nelle classificazioni di configurazione:

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

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

Per visualizzare le pagine del server della cronologia, consulta Debug di applicazioni e processi con EMR Studio. È inoltre possibile connettersi al nodo primario del cluster utilizzando SSH per visualizzare le interfacce Web dell'applicazione. Per ulteriori informazioni, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Aggiornamento di un Amazon EMR Studio

Dopo aver creato un EMR Studio, puoi aggiornare i seguenti attributi utilizzando la AWS CLI:

- Nome
- Descrizione
- Percorso S3 predefinito
- Sottoreti

Aggiornamento di un EMR Studio mediante la AWS CLI

Utilizza il comando `update-studio` della AWS CLI per aggiornare un EMR Studio. Per ulteriori informazioni, consulta la sezione relativa alle informazioni di riferimento ai comandi della AWS CLI.

> **Note**
>
> Puoi associare uno Studio a un massimo di 5 sottoreti. Queste sottoreti devono appartenere allo stesso VPC dello Studio. L'elenco degli ID delle sottoreti inviati al comando `update-studio` può includere nuovi ID di sottorete, ma deve includere anche tutti gli ID di sottorete precedentemente associati allo Studio. Non è possibile rimuovere le sottoreti da uno Studio.

```
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> \
```

Per verificare le modifiche, utilizza il comando `describe-studio` della AWS CLI e specifica l'ID del tuo Studio. Per ulteriori informazioni, consulta la Guida di riferimento ai comandi della AWS CLI.

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

Eliminazione di un Amazon EMR Studio e di Workspace

Quando elimini uno Studio, EMR Studio elimina tutte le assegnazioni di utenti e gruppi IAM Identity Center che sono associate allo Studio.

> **Note**
>
> Quando elimini uno Studio, Amazon EMR non elimina i Workspace a esso associati. È necessario eliminare separatamente i Workspace nello Studio.

Eliminazione dei Workspace

Console

Poiché ogni Workspace EMR Studio è un'istanza di notebook EMR, è possibile utilizzare la console di gestione Amazon EMR per eliminare i Workspace. Puoi eliminare i Workspace utilizzando la console Amazon EMR prima o dopo l'eliminazione di Studio

Eliminazione di un Workspace mediante la console Amazon EMR

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Notebook.
3. Seleziona i Workspace che intendi eliminare.
4. Seleziona Elimina e quindi nuovamente Elimina per confermare.
5. Segui le istruzioni per l'Eliminazione di oggetti nella Guida per l'utente della console Amazon Simple Storage Service per rimuovere i file notebook associati al Workspace eliminato da Amazon S3.

EMR Studio UI

From the Workspace UI

Eliminazione di un Workspace e dei file di backup associati da EMR Studio

1. Accedi al tuo EMR Studio con l'URL di accesso allo Studio e seleziona Workspace dal riquadro di navigazione a sinistra.
2. Individua il Workspace nell'elenco, quindi seleziona la casella di spunta accanto al relativo nome. È possibile selezionare più Workspace da eliminare contemporaneamente.
3. Dall'elenco Workspace, seleziona Elimina in alto a destra per confermare che desideri eliminare i Workspace selezionati. Seleziona Elimina per confermare.

4. Se desideri rimuovere i file notebook associati al WorkSpace eliminato da Amazon S3, segui le istruzioni per l'Eliminazione di oggetti nella Guida per l'utente della console Amazon Simple Storage Service. Se non hai creato lo Studio, contatta l'amministratore dello Studio per determinare la posizione del backup di Amazon S3 per il Workspace eliminato.

From the Workspaces list

Eliminazione di un WorkSpace e dei file di backup associati dall'elenco dei WorkSpace

1. Vai all'elenco dei Workspace nella console.
2. Seleziona il WorkSpace che desideri eliminare dall'elenco, quindi scegli Azioni.
3. Scegli Elimina.
4. Se desideri rimuovere i file notebook associati al WorkSpace eliminato da Amazon S3, segui le istruzioni per l'Eliminazione di oggetti nella Guida per l'utente della console Amazon Simple Storage Service. Se non hai creato lo Studio, contatta l'amministratore dello Studio per determinare la posizione del backup di Amazon S3 per il Workspace eliminato.

Eliminazione di un EMR Studio

New console

Eliminazione di un EMR Studio con la nuova console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR Studio, nella barra di navigazione a sinistra, scegli Studio.
3. Seleziona lo Studio dall'elenco degli Studios (Studio) tramite l'interruttore a sinistra del nome dello Studio. Scegliere Delete (Elimina).

Old console

Eliminazione di un EMR Studio con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/home.
2. Seleziona EMR Studio dalla barra di navigazione a sinistra.

3. Seleziona lo Studio dall'elenco Studio e scegli Elimina.

CLI

Eliminazione di un EMR Studio con la AWS CLI

Utilizza il comando `delete-studio` della AWS CLI per eliminare un EMR Studio. Per ulteriori informazioni, consulta la sezione relativa alle informazioni di riferimento ai comandi della AWS CLI.

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

Definizione di gruppi di sicurezza per controllare il traffico di rete EMR Studio

Informazioni sui gruppi di sicurezza EMR Studio

Amazon EMR Studio utilizza due gruppi di sicurezza per controllare il traffico di rete tra Workspace in Studio e un cluster Amazon EMR collegato in esecuzione su Amazon EC2:

- Un gruppo di sicurezza del motore, che utilizza la porta 18888 per comunicare con un cluster Amazon EMR collegato in esecuzione su Amazon EC2.
- Un gruppo di sicurezza del Workspace associato ai Workspace in uno Studio. Questo gruppo di sicurezza include una regola HTTPS in uscita per consentire al Workspace di instradare il traffico a Internet e deve consentire il traffico in uscita verso Internet sulla porta 443 per abilitare il collegamento di repository Git a un Workspace.

EMR Studio utilizza questi gruppi di sicurezza oltre a tutti i gruppi di sicurezza associati a un cluster EMR collegato a un Workspace.

Devi creare questi gruppi di sicurezza quando utilizzi la AWS CLI per creare uno Studio.

> **Note**
>
> È possibile personalizzare i gruppi di sicurezza per EMR Studio con regole personalizzate per l'ambiente, ma è necessario includere le regole riportate in questa pagina. Il gruppo di

sicurezza del Workspace non può consentire traffico in entrata e il gruppo di sicurezza del motore deve consentire il traffico in entrata dal gruppo di sicurezza del Workspace.

Uso dei gruppi di sicurezza EMR Studio predefiniti

Quando utilizzi la console Amazon EMR, puoi scegliere i seguenti gruppi di sicurezza predefiniti. I gruppi di sicurezza predefiniti vengono creati da EMR Studio per tuo conto e includono le regole minime in entrata e in uscita richieste per i Workspace in un EMR Studio.

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

Prerequisiti

Per creare i gruppi di sicurezza per EMR Studio, è necessario un cloud privato virtuale (VPC) Amazon per lo Studio. Il VPC viene scelto quando crei i gruppi di sicurezza. Dovrebbe corrispondere al VPC specificato durante la creazione dello Studio. Se intendi utilizzare Amazon EMR su EKS con EMR Studio, scegli il VPC per i nodi worker del cluster Amazon EKS.

Istruzioni

Segui le istruzioni in Creazione di un gruppo di sicurezza nella Guida per l'utente di Amazon EC2 per le istanze Linux per creare un gruppo di sicurezza del motore e un gruppo di sicurezza del Workspace nel tuo VPC. I gruppi di sicurezza devono includere le regole riassunte nelle tabelle seguenti.

Quando crei gruppi di sicurezza per EMR Studio, prendi nota degli ID di entrambi. Quando crei uno Studio, puoi specificare ciascun gruppo di sicurezza in base all'ID.

Gruppo di sicurezza del motore

EMR Studio utilizza la porta 18888 per comunicare con un cluster collegato.

Regole in entrata

Tipo	Protocollo	Porta	Destinazione	Descrizione
TCP	TCP	18888	Il gruppo di sicurezza del	Consenti il traffico da qualsiasi risorsa nel

Tipo	Protocollo	Porta	Destinazione	Descrizione
			Workspace EMR Studio.	gruppo di sicurezza del Workspace per EMR Studio.

Gruppo di sicurezza del Workspace

Questo gruppo di sicurezza è associato ai Workspace in un EMR Studio.

Regole in uscita

Tipo	Protocollo	Porta	Destinazione	Descrizione
TCP	TCP	18888	Il gruppo di sicurezza del motore EMR Studio.	Consenti il traffico verso qualsiasi risorsa nel gruppo di sicurezza del motore per EMR Studio.
HTTPS	TCP	443	0.0.0.0/0	Consenti il traffico verso Internet per collegare repository Git in hosting pubblico ai Workspace.

Creazione di modelli AWS CloudFormation per Amazon EMR Studio

Informazioni sui modelli di cluster EMR Studio

È possibile creare modelli AWS CloudFormation per aiutare gli utenti di EMR Studio ad avviare nuovi cluster Amazon EMR in un Workspace. I modelli CloudFormation sono file di testo formattati in JSON o YAML. In un modello, descrivi una pila di risorse AWS e spiega a CloudFormation come effettuare il provisioning di tali risorse. Per EMR Studio, puoi creare uno o più modelli che descrivono un cluster Amazon EMR.

Organizzi i tuoi modelli in AWS Service Catalog.AWS Service Catalog consente di creare e gestire servizi IT comunemente distribuiti denominati prodotti su AWS. Raccogli i tuoi modelli come prodotti

in un portfolio che condividi con gli utenti di EMR Studio. Dopo aver creato modelli di cluster, gli utenti di Studio possono avviare un nuovo cluster per un Workspace con uno dei modelli. Gli utenti devono disporre delle autorizzazioni per creare nuovi cluster dai modelli. Puoi impostare le autorizzazioni utente nel tuo Policy di autorizzazione di EMR Studio.

Per ulteriori informazioni sui modelli CloudFormation, consulta Modelli nella Guida per l'utente di AWS CloudFormation. Per ulteriori informazioni su AWS Service Catalog, consulta Che cos'è AWS Service Catalog.

Il video seguente illustra come configurare modelli di cluster in AWS Service Catalog per EMR Studio. Per ulteriori informazioni, consulta il post del blog Build a self-service environment for each line of business using Amazon EMR and Service Catalog (Creazione di un ambiente self-service per ogni linea di business utilizzando Amazon EMR e Service Catalog).

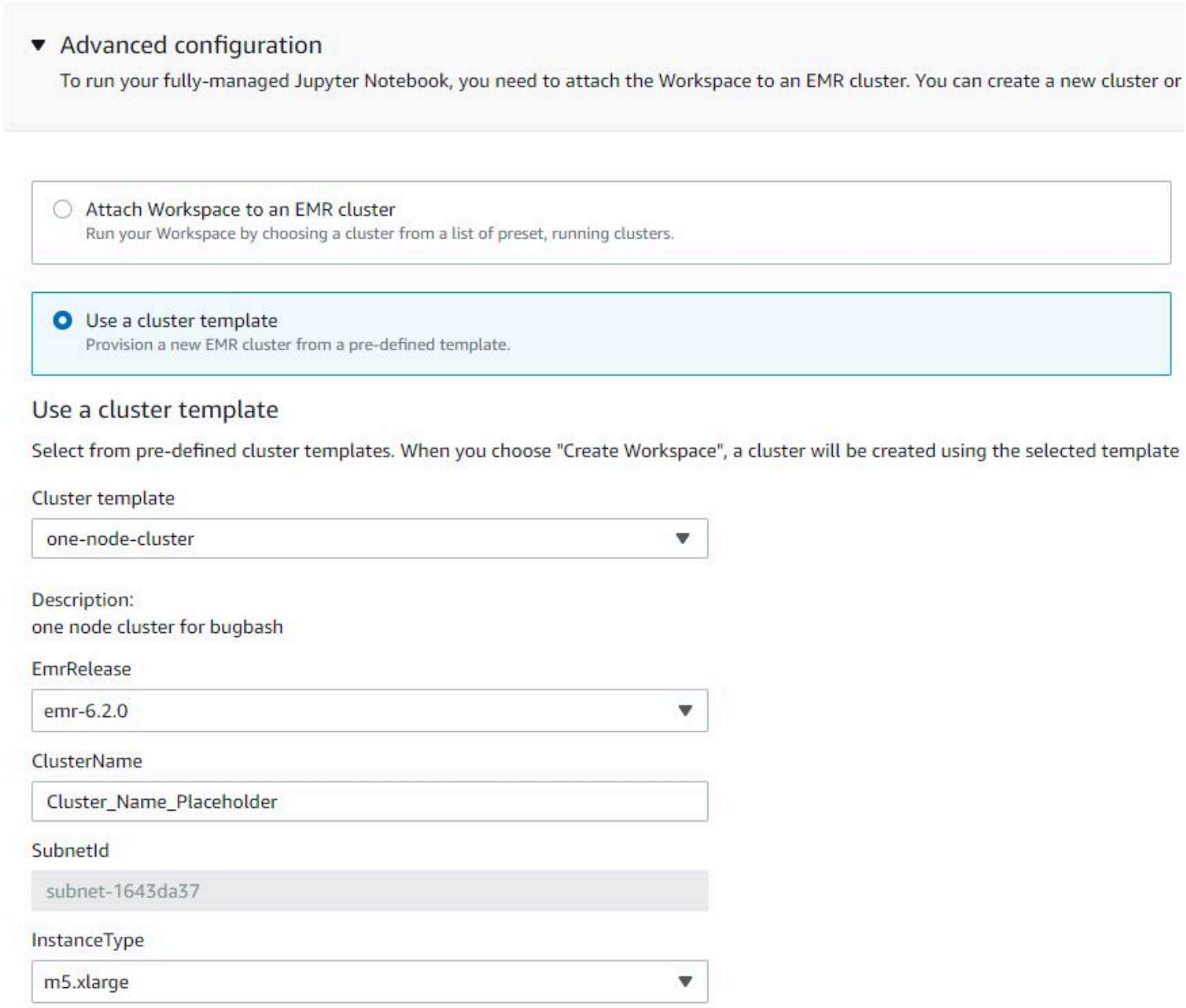
Parametri di modello opzionali

Puoi includere opzioni aggiuntive nella sezione `Parameters` del modello. Parametri consente agli utenti di Studio di immettere o selezionare valori personalizzati per un cluster. Ad esempio, prova a aggiungere un parametro che consente agli utenti di selezionare una release Amazon EMR specifica. Per ulteriori informazioni, consulta Parametri nella Guida per l'utente di AWS CloudFormation.

La sezione `Parameters` di esempio seguente definisce parametri di input aggiuntivi come il `ClusterName`, versione `EmrRelease` e `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"
```

Quando si aggiungono parametri, gli utenti di Studio visualizzano ulteriori opzioni di modulo dopo aver selezionato un modello di cluster. L'immagine seguente mostra opzioni di modulo aggiuntive per versione EmrRelease,ClusterName, e InstanceType.



Prerequisiti

Prima di creare un modello di cluster, assicurati di disporre delle autorizzazioni IAM per accedere alla vista console amministratore di Service Catalog. Sono inoltre necessarie le autorizzazioni IAM richieste per eseguire i processi amministrativi di Service Catalog. Per ulteriori informazioni, consulta Concessione di autorizzazioni ad amministratori Service Catalog.

Istruzioni

Creazione di modelli di cluster EMR con Service Catalog

1. Crea uno o più modelli CloudFormation. Il luogo in cui si archiviano i modelli dipende da te. Poiché i modelli sono file di testo formattati, puoi caricarli su Amazon S3 o conservarli nel file system locale. Per ulteriori informazioni sui modelli CloudFormation, consulta Modelli nella AWS CloudFormationGuida per l'utente di.

   Utilizza le regole seguenti per assegnare un nome ai modelli o controllare i nomi in base al pattern `[a-zA-Z0-9][a-zA-Z0-9._-]*`.

   - I nomi dei modelli devono iniziare con una lettera o un numero.
   - I nomi dei modelli possono includere solo di lettere, numeri, punti (.), trattini bassi (_) e trattini alti (-).

   Ogni modello del cluster che si crea deve includere le seguenti opzioni:

   Parametri di input

   - ClusterName: un nome che consente agli utenti di identificare il cluster dopo il provisioning.

   Output

   - `ClusterId`: l'ID del cluster EMR di cui è stato appena effettuato il provisioning.

   Di seguito è riportato un modello AWS CloudFormation di esempio in formato YAML per un cluster con due nodi. Il modello di esempio include le opzioni del modello richieste e definisce i parametri di input aggiuntivi per `EmrRelease` e `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. Crea un portfolio per i modelli di cluster nello stesso account AWS utilizzato per lo Studio.
   a. Apri la console AWS Service Catalog all'indirizzo https://console.aws.amazon.com/servicecatalog/.
   b. Dal menu di navigazione a sinistra, scegli Portfolio.
   c. Nella pagina Crea un portfolio, immetti le informazioni richieste.
   d. Scegli Crea. AWS Service Catalog crea il portfolio e ne visualizza i dettagli.
3. Utilizza la procedura seguente per aggiungere i modelli di cluster come prodotti AWS Service Catalog.
   a. Passare alla pagina Prodotti in Amministrazione nella console di gestione AWS Service Catalog
   b. Scegliere Upload new product (Carica un nuovo prodotto).
   c. Inserisci un Product name (Nome del prodotto) e Owner (Proprietario).
   d. Specificare il file modello in Version details (Dettagli della versione).
   e. Scegliere Review (Revisione) per rivedere le impostazioni del prodotto, quindi scegliere Create product (Crea prodotto).
4. Completa la procedura seguente per aggiungere i prodotti al portfolio.
   a. Passare alla pagina Products (Prodotti) della console di gestione AWS Service Catalog
   b. Scegli il tuo prodotto, scegli Actions (Operazioni), quindi scegli Add product to portfolio (Aggiungi prodotto al portfolio).
   c. Scegli un portfolio, quindi scegli Add product to portfolio (Aggiungi prodotto al portfolio).
5. Creare un vincolo di avvio per i prodotti. Un vincolo di avvio è un ruolo IAM che specifica le autorizzazioni utente per l'avvio di un prodotto. Puoi personalizzare i vincoli di avvio, ma devi consentire le autorizzazioni per l'uso di CloudFormation, Amazon EMR e AWS Service Catalog. Per ulteriori informazioni e istruzioni, consulta Vincoli di avvio di Service Catalog.
6. Applica il vincolo di avvio a ciascun prodotto del tuo portfolio. È necessario applicare il vincolo di avvio a ciascun prodotto singolarmente.
   a. Seleziona il tuo portfolio dal Portfolio della console di gestione AWS Service Catalog.
   b. Scegliere la scheda Vincoli, quindi Crea vincolo.
   c. Scegliere il tuo prodotto e scegliere Avvia in Tipo di vincolo. Scegli Continua.

d. Seleziona il ruolo di vincolo di avvio nella sezione Vincolo di avvio, quindi scegli Crea.

7. Concedere l'accesso al portfolio.

a. Seleziona il tuo portafoglio dal Portfolios (Portafogli) della console di gestione AWS Service Catalog

b. Espandi la scheda Gruppi, ruoli e utenti e scegli Aggiungi gruppi, ruoli, utenti.

c. Cerca il tuo ruolo EMR Studio IAM nella scheda Ruoli, seleziona il tuo ruolo e scegli Aggiungi accesso.

Se utilizzi...	Concedi l'accesso a...
Autenticazione IAM	I tuoi utenti nativi
Federazione IAM	Il tuo ruolo IAM per la federazione
Federazione IAM Identity Center	Il tuo ruolo utente di EMR Studio

Definizione di accesso e autorizzazioni per i repository basati su Git

EMR Studio supporta i seguenti servizi basati su Git:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

Per consentire agli utenti di EMR Studio di associare un repository Git a un Workspace, imposta i seguenti requisiti di accesso e autorizzazione. È inoltre possibile configurare i repository basati su Git ospitati in una rete privata seguendo le istruzioni riportate in Configurazione di un repository Git ospitato privatamente per EMR Studio.

Accesso a Internet del cluster

Entrambi i cluster Amazon EMR in esecuzione su cluster Amazon EC2 e Amazon EMR su EKS collegati a Workspace Studio devono trovarsi in una sottorete privata che utilizza un gateway

Network Address Translation (NAT) o devono essere in grado di accedere a Internet tramite un gateway virtuale privato. Per ulteriori informazioni, consulta Opzioni di Amazon VPC.

I gruppi di sicurezza utilizzati con EMR Studio devono includere anche una regola in uscita che consenta ai Workspace di instradare il traffico a Internet da un cluster EMR collegato. Per ulteriori informazioni, consulta Definizione di gruppi di sicurezza per controllare il traffico di rete EMR Studio.

> **Important**
>
> Se l'interfaccia di rete si trova in una sottorete pubblica, non sarà in grado di comunicare con Internet tramite un Gateway Internet (IGW).

Autorizzazioni per AWS Secrets Manager

Per consentire agli utenti di EMR Studio di accedere ai repository Git con segreti archiviati in AWS Secrets Manager, aggiungi una policy di autorizzazione al ruolo di servizio per EMR Studio che consente l'operazione `secretsmanager:GetSecretValue`.

Per ulteriori informazioni su come collegare repository basati su Git ai Workspace, consulta Collegamento di repository basati su Git a un WorkSpace EMR Studio.

Configurazione di un repository Git ospitato privatamente per EMR Studio

Utilizza le seguenti istruzioni per configurare repository ospitati privatamente per Amazon EMR Studio. Fornire un file di configurazione con informazioni sui server DNS e Git. EMR Studio utilizza queste informazioni per configurare Workspace in grado di instradare il traffico ai repository autogestiti.

> **Note**
>
> Se si configura `DnsServerIpV4`, EMR Studio utilizza il tuo server DNS per risolvere entrambi i `GitServerDnsName` e il tuo endpoint Amazon EMR, ad esempio `elasticmapreduce.us-east-1.amazonaws.com`. Per configurare un endpoint per Amazon EMR, connettiti al tuo endpoint tramite il VPC che stai utilizzando con Studio. Ciò garantisce che l'endpoint Amazon EMR viene risolto in un IP privato. Per ulteriori

informazioni, consulta Connessione ad Amazon EMR utilizzando un endpoint VPC di interfaccia.

Prerequisiti

Prima di configurare un repository Git ospitato in livello privato per EMR Studio, è necessario un archivio Amazon S3 in cui EMR Studio possa eseguire il backup di Workspace e file notebook nello Studio. Utilizza lo stesso bucket S3 specificato durante la creazione di uno Studio.

Configurazione di uno o più repository Git ospitati privatamente per EMR Studio

1. Crea un file di configurazione utilizzando il seguente modello. Includi i seguenti valori per ogni server Git che desideri specificare nella configurazione:

   - **`DnsServerIpV4`**: l'indirizzo IPv4 del server DNS. Se si forniscono valori per `DnsServerIpV4` e `GitServerIpV4List`, il valore per `DnsServerIpV4` ha la precedenza e EMR Studio utilizza `DnsServerIpV4` per risolvere il `GitServerDnsName`.

     > **Note**
     >
     > Per utilizzare repository Git ospitati privatamente, il server DNS deve consentire l'accesso in ingresso da EMR Studio. Si consiglia di proteggere il server DNS da altri accessi non autorizzati.

   - **`GitServerDnsName`**: il nome DNS del server Git. Ad esempio `"git.example.com"`.
   - **`GitServerIpV4List`**: un elenco di indirizzi IPv4 che appartengono ai server 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. Salva il file di configurazione come `configuration.json`.
3. Carica il file di configurazione nel percorso di archiviazione Amazon S3 in una cartella denominata `life-cycle-configuration`. Ad esempio, se la posizione S3 predefinita è `s3://`*`DOC-EXAMPLE-BUCKET`*`/studios`, il file di configurazione sarà in `s3://`*`DOC-EXAMPLE-BUCKET`*`/studios/`**`life-cycle-configuration/configuration.json`**.

   > **Important**
   >
   > Si consiglia di limitare l'accesso alla tua cartella `life-cycle-configuration` agli amministratori dello Studio e al ruolo di servizio EMR Studio, nonché di proteggere `configuration.json` contro l'accesso non autorizzato. Per ricevere istruzioni, consulta Controllo dell'accesso a un bucket con policy utente o Best practice di sicurezza per Amazon S3.

   Per istruzioni sul caricamento, consulta Creazione di una cartella e Caricamento degli oggetti nella Guida per l'utente di Amazon Simple Storage. Per applicare la configurazione a un Workspace esistente, chiudi e riavvia il Workspace dopo aver caricato il file di configurazione su Amazon S3.

Ottimizzazione dei processi Spark in EMR Studio

Quando si esegue un processo Spark utilizzando EMR Studio, è possibile eseguire alcune fase per garantire l'ottimizzazione delle risorse del cluster Amazon EMR.

Prolunga la tua sessione Livy

Se utilizzi Apache Livy insieme a Spark sul cluster Amazon EMR, ti consigliamo di aumentare il timeout della sessione Livy effettuando una delle seguenti operazioni:

- Quando crei un cluster Amazon EMR, imposta questa classificazione di configurazione nel campo Enter Configuration (Immettere la configurazione).

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

- Per un cluster EMR già in esecuzione, connettiti al cluster utilizzando `ssh` e imposta la classificazione di configurazione `livy-conf` in `/etc/livy/conf/livy.conf`.

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

  Potrebbe essere necessario riavviare Livy dopo aver modificato la configurazione.
- Se non vuoi che la tua sessione di Livy venga scaduta, imposta la proprietà `livy.server.session.timeout-check` a `false` in `/etc/livy/conf/livy.conf`.

Esecuzione di Spark in modalità cluster

In modalità cluster, il driver Spark viene eseguito su un nodo principale anziché sul nodo primario, il che migliora l'utilizzo delle risorse sul nodo primario.

Per eseguire l'applicazione Spark in modalità cluster anziché nella modalità client predefinita, scegli la modalità Cluster quando imposti Modalità di implementazione durante la configurazione della

fase Spark nel tuo nuovo cluster Amazon EMR. Per ulteriori informazioni, consulta Panoramica della modalità cluster nella documentazione di Apache Spark.

Aumento della memoria del driver Spark

Per aumentare la memoria del driver Spark, configura la sessione Spark utilizzando il comando magic `%%configure` nel notebook EMR, come nell'esempio seguente.

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

Utilizzare un Amazon EMR Studio

Questa sezione contiene argomenti che ti aiutano a configurare e interagire con Amazon EMR Studio.

Il video seguente illustra informazioni pratiche come creare un nuovo istanza WorkSpace e come avviare un nuovo cluster Amazon EMR utilizzando un modello di cluster. Il video mostra anche un notebook di esempio.

Questa sezione include i seguenti argomenti per informazioni su come lavorare in un EMR Studio:

- Informazioni sulle nozioni di base di WorkSpace
- Configurazione della collaborazione di WorkSpace
- Esecuzione di un WorkSpace EMR Studio con un ruolo di runtime
- Esecuzione di notebook Workspace a livello di programmazione
- Sfogliare i dati con SQL Explorer
- Collegamento di un calcolo a un WorkSpace EMR Studio
- Collegamento di repository basati su Git a un WorkSpace EMR Studio
- Debug di applicazioni e processi con EMR Studio
- Installazione di kernel e librerie in un'istanza WorkSpace di EMR Studio
- Migliora i kernel con comandi magic
- Usare notebook multilingue con i kernel Spark

Informazioni sulle nozioni di base di WorkSpace

Quando utilizzi un EMR Studio, puoi creare e configurare diversi Workspace per organizzare ed eseguire notebook. Questa sezione descrive la creazione e l'utilizzo delle istanze WorkSpace. Per una panoramica concettuale, consulta WorkSpace nella pagina Come funziona Amazon EMR Studio.

Questa sezione comprende i seguenti argomenti per informazioni su come utilizzare i WorkSpace EMR Studio:

- Creazione di un WorkSpace EMR Studio
- Avvio di un WorkSpace
- Informazioni sull'interfaccia utente WorkSpace
- Esplorazione di notebook di esempio
- Salvataggio del contenuto del WorkSpace
- Eliminazione dei file di un'istanza WorkSpace e di un notebook
- Informazioni sullo stato del WorkSpace
- Risolvi i problemi di connettività di WorkSpace

Creazione di un WorkSpace EMR Studio

È possibile creare WorkSpace EMR Studio per eseguire il codice del notebook utilizzando l'interfaccia EMR Studio.

Creazione di un WorkSpace in un EMR Studio

1. Accedi al tuo EMR Studio.
2. Scegli Crea un WorkSpace.
3. Immetti un WorkSpace name (Nome WorkSpace) e una Description (Descrizione). L'assegnazione di un nome a un'istanza WorkSpace consente di identificarlo facilmente nella pagina WorkSpaces (istanze WorkSpace).
4. Se desideri lavorare con altri utenti di Studio in questo WorkSpace in tempo reale, abilita la collaborazione di WorkSpace. Puoi configurare i collaboratori dopo aver avviato il WorkSpace.
5. Se desideri collegare un cluster a un WorkSpace, espandi la sezione Configurazione avanzata. Puoi collegare un cluster in un secondo momento, se preferisci. Per ulteriori informazioni, consulta Collegamento di un calcolo a un WorkSpace EMR Studio.

> **Note**
>
> Per eseguire il provisioning di un nuovo cluster, devi accedere alle autorizzazioni di accesso da parte dell'amministratore.

Scegli una delle opzioni del cluster per l'istanza WorkSpace e collega il cluster. Per ulteriori informazioni sul provisioning di un cluster dopo la creazione di un WorkSpace, consulta Creazione e collegamento di un nuovo cluster EMR a un Workspace EMR Studio.

6. Scegli Crea un WorkSpace in basso a destra nella pagina.

Dopo aver creato l'istanza WorkSpace, EMR Studio aprirà la pagina WorkSpaces (istanze WorkSpace). Verrà visualizzato un banner verde di esito positivo nella parte superiore della pagina e sarà possibile trovare l'istanza WorkSpace appena creato nell'elenco.

Per impostazione predefinita, un WorkSpace è condiviso e può essere visualizzato da tutti gli utenti di Studio. Tuttavia, solo un utente alla volta può aprire e lavorare in un WorkSpace. Lavorare contemporaneamente con altri utenti è possibile con Configurazione della collaborazione di WorkSpace

Avvio di un WorkSpace

Per iniziare a lavorare con i file notebook, avvia un WorkSpace per accedere all'editor del notebook. La pagina WorkSpace di uno Studio elenca tutti i WorkSpace ai quali hai accesso con dettagli quali Name (Nome), Status (Stato), Creation time (Ora di creazione) e Last Modified (Ultima modifica).

> **Note**
>
> Se nella vecchia console Amazon EMR utilizzavi notebook EMR, puoi trovarli nella nuova console come WorkSpace di EMR Studio. Gli utenti dei notebook EMR necessitano di autorizzazioni aggiuntive per i ruoli IAM per accedere o creare WorkSpace. Se di recente hai creato un notebook nella vecchia console, potrebbe essere necessario aggiornare l'elenco dei WorkSpace per visualizzarlo nella nuova console. Per ulteriori informazioni sulla transizione, consulta I notebook Amazon EMR sono disponibili come Workspace Amazon EMR Studio nella nuova console e Novità della console

Avvio di un'istanza WorkSpace per la modifica e l'esecuzione di notebook

1. Sulla pagina WorkSpaces (istanze WorkSpace) dello Studio, trova l'istanza WorkSpace. Puoi filtrare l'elenco in base alla parola chiave o al valore della colonna.
2. Scegli il nome dell'istanza WorkSpace per aprirla in una nuova scheda del browser. Potrebbero essere necessari alcuni minuti prima che il WorkSpace venga aperto se è Idle (Inattivo). In alternativa, seleziona la riga per il WorkSpace, quindi seleziona Avvia WorkSpace. Puoi scegliere tra le seguenti opzioni di avvio:

   - Avvio rapido: avvia rapidamente il tuo WorkSpace con le opzioni predefinite. Scegli Avvio rapido se desideri collegare i cluster al WorkSpace in JupyterLab.
   - Avvia con opzioni: avvia il tuo WorkSpace con opzioni personalizzate. Puoi scegliere di avviare in Jupyter o JupyterLab, collegare il tuo WorkSpace a un cluster EMR e selezionare i tuoi gruppi di sicurezza.

> **Note**
>
> Solo un utente alla volta può aprire e lavorare in un WorkSpace. Se si seleziona un'istanza WorkSpace già in uso, EMR Studio visualizza una notifica quando si tenta di aprirlo. La colonna User (Utente) sulla pagina WorkSpaces (istanze WorkSpace) mostra l'utente che sta attualmente utilizzando l'istanza WorkSpace.

Informazioni sull'interfaccia utente WorkSpace

L'interfaccia utente del WorkSpace EMR Studio si basa sull'interfaccia di JupyterLab, la quale si compone di schede contrassegnate da icone nella barra laterale sinistra. Spostando il mouse su un'icona, è possibile visualizzare un tooltip che mostra il nome della scheda. Seleziona le schede dalla barra laterale sinistra per accedere ai seguenti pannelli.

- Browser di file: visualizza i file e le directory nell'istanza WorkSpace, nonché i file e le directory dei repository Git collegati.
- Kernel e terminali in esecuzione: elenca tutti i kernel e i terminali in esecuzione nell'istanza WorkSpace. Per ulteriori informazioni, consulta Managing kernels and terminals (Gestione di kernel e terminali) nella documentazione ufficiale di JupyterLab.

- Git: fornisce un'interfaccia utente grafica per l'esecuzione di comandi nei repository Git collegati all'istanza WorkSpace. Questo pannello è un'estensione JupyterLab chiamata jupyterlab-git. Per ulteriori informazioni, consulta jupyterlab-git.
- Cluster EMR: consente di collegare o scollegare un cluster dal WorkSpace per eseguire il codice del notebook. Il pannello di configurazione del cluster EMR fornisce inoltre opzioni di configurazione avanzate che consentono di creare e collegare un nuovo cluster nell'istanza WorkSpace. Per ulteriori informazioni, consulta Creazione e collegamento di un nuovo cluster EMR a un Workspace EMR Studio.
- Repository Git Amazon EMR: consente di collegare il WorkSpace a un massimo di tre repository Git. Per dettagli e istruzioni, consulta Collegamento di repository basati su Git a un WorkSpace EMR Studio.
- Notebook di esempio: fornisce un elenco di notebook di esempio che è possibile salvare nell'istanza WorkSpace. Puoi accedere agli esempi anche scegliendo Notebook Examples (Notebook di esempio) sulla pagina Launcher (Utilità di avvio) dell'istanza WorkSpace.
- Comandi: offre un metodo basato su tastiera per cercare ed eseguire comandi JupyterLab. Per ulteriori informazioni, consulta la pagina Command palette (Palette dei comandi) nella documentazione di JupyterLab.
- Strumenti notebook: consente di selezionare e impostare opzioni quali il tipo di cella a scorrimento e i metadati. L'opzione Notebook Tools (Strumenti notebook) viene visualizzata nella barra laterale sinistra dopo aver aperto un file notebook.
- Open Tabs (Schede aperte): elenca i documenti e le attività aperti nell'area di lavoro principale in modo da poter passare a una scheda aperta. Per ulteriori informazioni, consulta la pagina Tabs and single-document mode (Schede e modalità documento singolo) nella documentazione di JupyterLab.
- Collaboration (Collaborazione): consente di abilitare o disabilitare la collaborazione di WorkSpace e di gestire i collaboratori. Per visualizzare il pannello Collaboration (Collaborazione), devi disporre delle autorizzazioni necessarie. Per ulteriori informazioni, consulta Imposta la proprietà per la collaborazione di WorkSpace.

Esplorazione di notebook di esempio

Ogni WorkSpace EMR Studio include una serie di notebook di esempio che è possibile utilizzare per esplorare le caratteristiche di EMR Studio. Per modificare o eseguire un notebook di esempio, è possibile salvarlo nell'istanza WorkSpace.

Salvataggio di un notebook di esempio nell'istanza WorkSpace

1. Nella barra laterale sinistra, scegli l'opzione Notebook Examples (Notebook di esempio) per aprire il riquadro Notebook Examples (Notebook di esempio). Puoi accedere agli esempi anche scegliendo Notebook Examples (Notebook di esempio) sulla pagina Launcher (Utilità di avvio) dell'istanza WorkSpace.
2. Scegli un notebook di esempio per visualizzarlo in anteprima nell'area di lavoro principale. L'esempio è di sola lettura.
3. Per salvare il notebook di esempio nell'istanza WorkSpace, seleziona Save to WorkSpace (Salva nell'istanza WorkSpace). EMR Studio salva l'esempio nella directory principale. Dopo aver salvato un notebook di esempio nell'istanza WorkSpace, è possibile rinominarlo, modificarlo ed eseguirlo.

Per ulteriori informazioni sugli esempi di notebook, consulta Repository GitHub degli esempi di EMR Studio Notebook.

Salvataggio del contenuto del WorkSpace

Quando lavori nell'editor del notebook di un'istanza WorkSpace, EMR Studio salva il contenuto delle celle del notebook e l'output nel percorso Amazon S3 associato al Studio. Questo processo di backup conserva il lavoro tra una sessione e l'altra.

Puoi anche salvare un notebook premendo CTRL+S nella scheda aperta del notebook o utilizzando una delle opzioni di salvataggio in File.

Un altro modo per eseguire il backup dei file notebook nell'istanza WorkSpace consiste nell'associare l'istanza WorkSpace a un repository basato su Git e sincronizzare le modifiche con il repository remoto. In questo modo, è inoltre possibile salvare e condividere i notebook con i membri del team che utilizzano un'istanza WorkSpace o Studio diverso. Per istruzioni, consultare Collegamento di repository basati su Git a un WorkSpace EMR Studio.

Eliminazione dei file di un'istanza WorkSpace e di un notebook

Quando elimini un file notebook da un'istanza WorkSpace di EMR Studio, elimini il file dal menu File browser (Browser di file) ed EMR Studio rimuove la sua copia di backup in Amazon S3. Non è necessario adottare ulteriori misure per evitare addebiti di archiviazione quando si elimina un file da un'istanza WorkSpace.

Quando elimini un intero WorkSpace, i file e le cartelle del notebook rimarranno nella posizione di archiviazione di Amazon S3. I file continuano ad aumentare i costi di archiviazione. Per evitare costi di archiviazione, rimuovi tutti i file e le cartelle di backup associati al WorkSpace eliminato da Amazon S3.

Eliminazione di un file notebook da un WorkSpace EMR Studio

1. Seleziona il riquadro File browser (Browser di file) dalla barra laterale sinistra nell'istanza WorkSpace.
2. Seleziona il file o la cartella da eliminare. Fai clic con il pulsante destro del mouse sulla selezione, quindi scegli Delete (Elimina). Il file scompare dall'elenco. EMR Studio rimuove il file o la cartella da Amazon S3.

From the Workspace UI

Eliminazione di un WorkSpace e dei file di backup associati da EMR Studio

1. Accedi al tuo EMR Studio con l'URL di accesso allo Studio e seleziona Workspace dal riquadro di navigazione a sinistra.
2. Individua il Workspace nell'elenco, quindi seleziona la casella di spunta accanto al relativo nome. È possibile selezionare più Workspace da eliminare contemporaneamente.
3. Dall'elenco Workspace, seleziona Elimina in alto a destra per confermare che desideri eliminare i Workspace selezionati. Seleziona Elimina per confermare.
4. Se desideri rimuovere i file notebook associati al WorkSpace eliminato da Amazon S3, segui le istruzioni per l'Eliminazione di oggetti nella Guida per l'utente della console Amazon Simple Storage Service. Se non hai creato lo Studio, contatta l'amministratore dello Studio per determinare la posizione del backup di Amazon S3 per il Workspace eliminato.

From the Workspaces list

Eliminazione di un WorkSpace e dei file di backup associati dall'elenco dei WorkSpace

1. Vai all'elenco dei Workspace nella console.
2. Seleziona il WorkSpace che desideri eliminare dall'elenco, quindi scegli Azioni.
3. Scegli Elimina.

4. Se desideri rimuovere i file notebook associati al WorkSpace eliminato da Amazon S3, segui le istruzioni per l'Eliminazione di oggetti nella Guida per l'utente della console Amazon Simple Storage Service. Se non hai creato lo Studio, contatta l'amministratore dello Studio per determinare la posizione del backup di Amazon S3 per il Workspace eliminato.

Informazioni sullo stato del WorkSpace

Una volta creato, l'istanza WorkSpace di EMR Studio viene visualizzata come una riga nell'elenco WorkSpaces (isteanze WorkSpace) nel tuo Studio con il nome, lo stato, l'ora di creazione e il timestamp dell'ultima modifica. La tabella seguente descrive i stati della istanza WorkSpace.

Stato	Descrizione
Avvio di	Il WorkSpace è in fase di preparazione, ma non è ancora pronto per l'uso. Non è possibile aprire un WorkSpace quando il relativo stato è Starting (Avvio in corso).
Pronto	È possibile aprire il WorkSpace per utilizzare l'editor del notebook, ma è necessario collegare il WorkSpace a un cluster EMR prima di poter eseguire il codice del notebook.
Collegamento	Il WorkSpace è collegato a un cluster.
Collegato	Il WorkSpace è collegato a un cluster EMR ed è pronto per la scrittura e l'esecuzione del codice del notebook. Se lo stato di un WorkSpace non è Attached (Collegato), è necessario collegarlo a un cluster prima di poter eseguire il codice del notebook.
Idle	Il WorkSpace viene arrestato. Per riattivare un WorkSpace inattivo, occorre selezionarlo dall'elenco WorkSpaces (WorkSpace). Quando si seleziona il WorkSpace, il relativo stato

Stato	Descrizione
	cambia da Idle (Inattivo) a Starting (Avvio in corso) a Ready (Pronto).
Stopping (In arresto)	Il WorkSpace è in fase di arresto e verrà impostato su Inattivo. Quando si arresta un WorkSpace, vengono terminati tutti i kernel dei notebook corrispondenti. EMR Studio arresta i notebook che sono rimasti inattivi per molto tempo.
Deleting (Eliminazione in corso)	Quando si elimina un WorkSpace, EMR Studio lo contrassegna per l'eliminazione e avvia il processo di eliminazione. Al termine del processo di eliminazione, il WorkSpace scompare dall'elenco. Quando elimini un WorkSpace, i file e le cartelle del relativo notebook rimarranno nella posizione di archiviazione di Amazon S3.

Risolvi i problemi di connettività di WorkSpace

Per risolvere i problemi di connettività di WorkSpace, è possibile arrestare e riavviare un WorkSpace. Quando si riavvia un workspace, EMR Studio avvia il WorkSpace in una zona di disponibilità diversa o in una subnet diversa associata a Studio.

Per arrestare e riavviare un EMR Studio WorkSpace

1. Chiudere WorkSpace nel browser.
2. Vai all'elenco dei WorkSpace nella console.
3. Selezionare WorkSpace dall'elenco e scegliere Operazioni.
4. Scegliere Stop e attendere che lo stato del WorkSpace passi da Stopping (In fase di arresto) a Idle (Inattivo).
5. Scegliere di nuovo Actions (Operazioni), quindi scegliere Start (Avvio) per riavviare il WorkSpace.

6. Attendere che lo stato del WorkSpace cambi da Starting (In avvio) a Ready (Pronto), quindi scegliere il nome del WorkSpace per riaprirlo in una nuova scheda del browser.

Configurazione della collaborazione di WorkSpace

La collaborazione di WorkSpace consente di scrivere ed eseguire il codice del notebook contemporaneamente con altri membri del team. Quando lavori nello stesso file del notebook, vedrai le modifiche man mano che i collaboratori le apportano. Puoi abilitare la collaborazione durante la creazione di un WorkSpace o attivare e disattivare la collaborazione in un WorkSpace esistente.

> **Note**
>
> La collaborazione con EMR Studio WorkSpace non è supportata con le applicazioni interattive EMR Serverless.

Prerequisiti

Prima di configurare la collaborazione per un WorkSpace, assicurati di completare le seguenti attività:

- Assicurati che l'amministratore di EMR Studio ti abbia fornito le autorizzazioni necessarie. Ad esempio, la seguente istruzione di esempio consente a un utente di configurare la collaborazione per tutti i WorkSpace con la chiave di tag `creatorUserId` il cui valore corrisponde all'ID dell'utente (indicato dalla variabile `aws:userId` della policy).

```
{
    "Sid": "UserRolePermissionsForCollaboration",
    "Action": [
        "elasticmapreduce:UpdateEditor",
        "elasticmapreduce:PutWorkspaceAccess",
        "elasticmapreduce:DeleteWorkspaceAccess",
        "elasticmapreduce:ListWorkspaceAccessIdentities"
    ],
    "Resource": "*",
    "Effect": "Allow",
    "Condition": {
        "StringEquals": {
            "elasticmapreduce:ResourceTag/creatorUserId": "${aws:userid}"
        }
    }
```

```
}
```

- Assicurati che il ruolo di servizio associato a EMR Studio disponga delle autorizzazioni necessarie per abilitare e configurare la collaborazione di WorkSpace, come nella seguente istruzione di esempio.

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

  Per ulteriori informazioni, consulta Creazione di un ruolo di servizio EMR Studio.

Abilitazione della collaborazione di WorkSpace e aggiunta di collaboratori

1. Nel tuo Workspace scegli l'icona Collaboration (Collaborazione) dalla schermata di avvio o nella parte inferiore del pannello a sinistra.

   > **Note**
   >
   > Non visualizzerai il pannello Collaboration (Collaborazione) a meno che l'amministratore di Studio non ti abbia concesso l'autorizzazione per configurare la collaborazione per il WorkSpace. Per ulteriori informazioni, consulta Imposta la proprietà per la collaborazione di WorkSpace.

2. Assicurati che l'interruttore Allow WorkSpace collaboration (Consenti collaborazione di WorkSpace) sia abilitato. Quando si abilita la collaborazione, solo tu e i collaboratori aggiunti potete vedere il WorkSpace nell'elenco nella pagina WorkSpace di Studio.
3. Immetti un Collaborator name (Nome collaboratore). Il WorkSpace può disporre di un massimo di cinque collaboratori incluso te stesso. Un collaboratore può essere qualsiasi utente con

accesso al tuo EMR Studio. Se non accedi a un collaboratore, Workspace è un workspace privato accessibile solo all'utente.

Nella seguente tabella vengono specificati i valori del collaboratore applicabili da inserire in base al tipo di identità del proprietario.

> **Note**
>
> Un proprietario può invitare solo collaboratori con lo stesso tipo di identità. Ad esempio, un utente può aggiungere solo altri utenti e un utente IAM Identity Center può aggiungere solo altri utenti IAM Identity Center.

Modalità di autenticazione	Valore da inserire per Collaborator name (Nome collaboratore)
Autenticazione IAM	un nome utente. Questo è il nome che un utente vede quando effettua l'accesso alla AWS Management Console.
Federazione IAM	Il nome di un ruolo IAM e un nome di sessione facoltativo. Per aggiungere tutti gli utenti federati che devono assumere lo stesso ruolo IAM, specifica il nome di un ruolo IAM per la federazione. Per aggiungere un singolo utente come collaboratore, specifica un ruolo e un nome di sessione. Ad esempio, `MyRoleNam e:MySessionName` .
SSO	Un nome utente IAM Identity Center come `user@example.com.`

4. Scegliere Add (Aggiungi). Ora il collaboratore può vedere il WorkSpace sulla propria pagina WorkSpace di EMR Studio e avviarlo per utilizzarlo in tempo reale con te.

> **Note**
>
> Se disabiliti la collaborazione di WorkSpace, il WorkSpace torna allo stato condiviso e può essere visualizzato da tutti gli utenti di Studio. Allo stato condiviso, solo un utente di Studio alla volta può aprire e lavorare in un WorkSpace.

Esecuzione di un WorkSpace EMR Studio con un ruolo di runtime

> **Note**
>
> La funzionalità del ruolo di runtime descritta in questa pagina si applica solo ad Amazon EMR in esecuzione su Amazon EC2 e non si riferisce alla funzionalità del ruolo di runtime nelle applicazioni interattive EMR Serverless. Per ulteriori informazioni su come utilizzare i ruoli di runtime in EMR Serverless, consulta Ruoli di runtime del processo nella Guida per l'utente di Amazon EMR serverless.

Un ruolo di runtime è un ruolo (IAM) AWS Identity and Access Management che puoi specificare quando invii un processo o una query a un cluster Amazon EMR. Il processo o la query inviata al cluster EMR utilizza il ruolo di runtime per accedere alle risorse AWS, ad esempio agli oggetti in Amazon S3.

Quando colleghi un WorkSpace EMR Studio a un cluster EMR che utilizza Amazon EMR 6.11 o versioni successive, puoi selezionare un ruolo di runtime per il processo o la query che invii da utilizzare quando accede alle risorse AWS. Tuttavia, se il cluster EMR non supporta i ruoli di runtime, il cluster EMR non assumerà il ruolo quando accede alle risorse AWS.

Per poter utilizzare un ruolo di runtime con un WorkSpace Amazon EMR Studio, un amministratore deve configurare le autorizzazioni utente in modo che l'utente Studio possa chiamare l'API `elasticmapreduce:GetClusterSessionCredentials` sul ruolo di runtime. Quindi, avvia un nuovo cluster con un ruolo di runtime che puoi utilizzare con il WorkSpace Amazon EMR Studio.

In questa pagina

- Considerazioni

Configurazione delle autorizzazioni utente per il ruolo di runtime

Configura le autorizzazioni utente in modo che l'utente Studio possa chiamare l'API `elasticmapreduce:GetClusterSessionCredentials` sul ruolo di runtime che l'utente desidera utilizzare. Devi inoltre configurare the section called “Autorizzazioni utente Studio (EC2, EKS)” prima che l'utente possa iniziare a utilizzare Studio.

> **Warning**
>
> Per concedere questa autorizzazione, crea una condizione basata sulla chiave di contesto `elasticmapreduce:ExecutionRoleArn` quando concedi l'accesso a un chiamante al fine di effettuare la chiamata alle API `GetClusterSessionCredentials`. L'esempio seguente mostra come fare.

```
{
      "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"
              ]
          }
      }
  }
```

L'esempio seguente mostra come consentire a un principale IAM di utilizzare un ruolo IAM denominato `test-emr-demo3` come ruolo di runtime. Inoltre, il titolare della policy potrà accedere ai cluster Amazon EMR solo con l'ID cluster `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"
            ]
        }
    }
}
```

L'esempio seguente consente a un principale IAM di utilizzare qualsiasi ruolo IAM con un nome che inizia con la stringa `test-emr-demo4` come ruolo di runtime. Inoltre, il titolare della policy potrà accedere solo ai cluster Amazon EMR contrassegnati con la coppia chiave-valore `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*"
            ]
        }
    }
}
```

Avvio di un nuovo cluster con un ruolo di runtime

Ora che disponi delle autorizzazioni necessarie, avvia un nuovo cluster con un ruolo di runtime da utilizzare con il tuo WorkSpace Amazon EMR Studio.

Se hai già avviato un nuovo cluster con un ruolo di runtime, puoi passare alla sezione the section called “Utilizzo del cluster con il WorkSpace”.

1. Innanzitutto, completa i prerequisiti nella sezione Ruoli di runtime per le fasi di Amazon EMR.
2. Quindi, avvia un cluster con le seguenti impostazioni per utilizzare i ruoli di runtime con i WorkSpace Amazon EMR Studio. Per istruzioni sull'avvio del cluster, consulta la sezione Impostazione di una configurazione di sicurezza per un cluster.
   - Scegli l'etichetta di release emr-6.11.0 o versioni successive.
   - Seleziona Spark, Livy e Jupyter Enterprise Gateway come applicazioni cluster.
   - Utilizza la configurazione di sicurezza creata nella fase precedente.
   - Facoltativamente, puoi abilitare Lake Formation per il cluster EMR. Per ulteriori informazioni, consulta Abilitazione di Amazon EMR con Lake Formation.

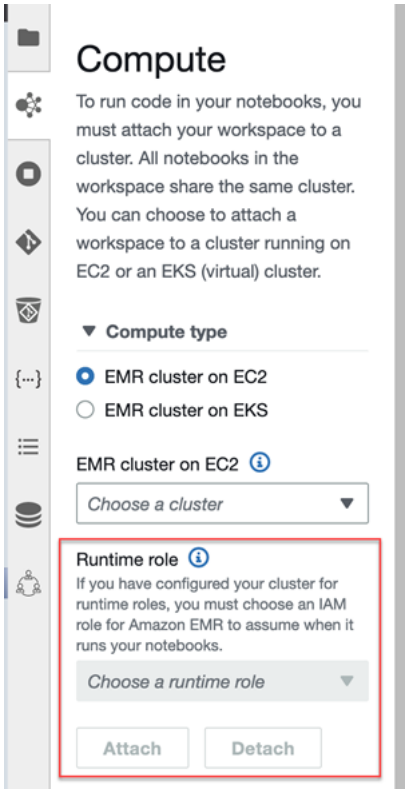
Dopo aver avviato il cluster, sei pronto per utilizzare il cluster con ruolo di runtime abilitato con un WorkSpace EMR Studio.

> **Note**
>
> Il valore ExecutionRoleArn non è attualmente supportato con l'operazione API StartNotebookExecution quando il valore `ExecutionEngineConfig.Type` è `EMR`.

Utilizzo del cluster EMR con un ruolo di runtime nei WorkSpace

Dopo aver configurato e avviato il cluster, puoi utilizzare il cluster abilitato al ruolo di runtime con il WorkSpace di EMR Studio.

1. Crea un nuovo workspace o avvia un workspace esistente. Per ulteriori informazioni, consulta Creazione di un WorkSpace EMR Studio.
2. Scegli la scheda Cluster EMR nella barra laterale sinistra del tuo WorkSpace aperto, espandi la sezione Tipo di calcolo e scegli il cluster dal menu Cluster EMR su EC2 e il ruolo di runtime dal menu Ruolo di runtime.



3. Scegli Collega per collegare il cluster con ruolo di runtime al tuo WorkSpace.

Considerazioni

Quando utilizzi un cluster abilitato al ruolo di runtime con il WorkSpace Amazon EMR Studio:

- Puoi selezionare un ruolo di runtime solo quando colleghi un WorkSpace EMR Studio a un cluster EMR che utilizza Amazon EMR rilascio 6.11 o successivi.
- La funzionalità del ruolo di runtime descritta in questa pagina è supportata solo con Amazon EMR in esecuzione su Amazon EC2 e non è supportata con le applicazioni interattive EMR Serverless. Per ulteriori informazioni sui ruoli di runtime per EMR Serverless, consulta Ruoli di runtime del processo nella Guida per l'utente di Amazon EMR serverless.
- Sebbene sia necessario configurare autorizzazioni aggiuntive per poter specificare un ruolo di runtime quando si invia un processo a un cluster, non sono necessarie autorizzazioni aggiuntive per accedere ai file generati da un WorkSpace EMR Studio. Le autorizzazioni per tali file sono le stesse dei file generati da cluster senza ruoli di runtime.
- Non è possibile utilizzare SQL Explorer in un WorkSpace EMR Studio con un cluster con un ruolo di runtime. Amazon EMR disabilita SQL Explorer nell'interfaccia utente quando un WorkSpace è collegato a un cluster EMR abilitato al ruolo di runtime.

- Non è possibile utilizzare la modalità di collaborazione in un WorkSpace EMR Studio con un cluster con un ruolo di runtime. Amazon EMR disabilita le funzionalità di collaborazione di un WorkSpace quando un WorkSpace è collegato a un cluster EMR abilitato al ruolo di runtime. Il WorkSpace rimarrà accessibile solo all'utente che lo ha collegato.
- Potresti ricevere un avviso "La pagina potrebbe non essere sicura!" dall'interfaccia utente di Spark per un cluster abilitato al ruolo di runtime. Se ciò accade, ignora l'avviso per continuare a visualizzare l'interfaccia utente di Spark.

Esecuzione di notebook Workspace a livello di programmazione

> **Note**
>
> L'esecuzione programmatica dei notebook non è supportata con le applicazioni interattive di Amazon EMR serverless.

Puoi eseguire i tuoi notebook Amazon EMR Studio Workspace a livello di programmazione con uno script o sul AWS CLI. Per informazioni su come eseguire il notebook in modo programmatico, consulta Comandi di esempio per eseguire Notebook EMR a livello di programmazione.

Sfogliare i dati con SQL Explorer

> **Note**
>
> SQL Explorer per EMR Studio non è supportato con le applicazioni interattive Amazon EMR Serverless.

In questo argomento vengono fornite informazioni utili per iniziare a utilizzare SQL Explorer in Amazon EMR Studio. SQL Explorer è uno strumento a pagina singola nel WorkSpace che aiuta a comprendere le origini dati nel catalogo dati del cluster EMR. Puoi utilizzare SQL Explorer per sfogliare i dati, eseguire query SQL per recuperare dati e scaricare i risultati delle query.

SQL Explorer supporta Presto. Prima di utilizzare SQL Explorer, assicurati di disporre di un cluster EMR che utilizzi Amazon EMR versione 5.34.0 o successive oppure versione 6.4.0 o successive con Presto installato. Amazon EMR Studio SQL Explorer non supporta i cluster Presto configurati con la crittografia in transito. Questo perché su questi cluster Presto viene eseguito in modalità TLS.

Sfogliare il catalogo dati del cluster

SQL Explorer fornisce un'interfaccia del browser di catalogo che è possibile utilizzare per esplorare e capire come sono organizzati i dati. Ad esempio, è possibile utilizzare il browser del catalogo dati per verificare i nomi di tabelle e colonne prima di scrivere una query SQL.

Sfogliare il catalogo dati

1. Apri SQL Explorer nel tuo WorkSpace.
2. Assicurati che il WorkSpace sia collegato a un cluster EMR in esecuzione su EC2 che utilizza Amazon EMR versione 6.4.0 o successive con Presto installato. Puoi selezionare un cluster esistente o crearne uno nuovo. Per ulteriori informazioni, consulta Collegamento di un calcolo a un WorkSpace EMR Studio.
3. Seleziona un Database dall'elenco a discesa per iniziare a sfogliarlo.
4. Espandi una tabella nel database per visualizzare i nomi delle colonne della tabella. Puoi anche immettere una parola chiave nella barra di ricerca per filtrare i risultati delle tabelle.

Esecuzione di una query SQL per recuperare i dati

Recupero di dati con una query SQL e download dei risultati

1. Apri SQL Explorer nel tuo WorkSpace.
2. Assicurati che il WorkSpace sia collegato a un cluster EMR in esecuzione su EC2 con Presto e Spark installati. Puoi selezionare un cluster esistente o crearne uno nuovo. Per ulteriori informazioni, consulta Collegamento di un calcolo a un WorkSpace EMR Studio.
3. Seleziona Open editor (Apri editor) per aprire una nuova scheda dell'editor nel WorkSpace.
4. Componi la query SQL nella scheda dell'editor.
5. Scegli Run (Esegui).
6. Visualizza i risultati della query in Result preview (Anteprima dei risultati). SQL Explorer visualizza i primi 100 risultati per impostazione predefinita. Puoi scegliere un numero diverso di risultati da visualizzare (fino a 1000) utilizzando il menu a discesa Preview first 100 query results (Anteprima dei primi 100 risultati della query).
7. Scegli Download results (Scarica risultati) per scaricare i risultati in formato CSV. È possibile scaricare fino a 1000 righe di risultati.

Collegamento di un calcolo a un WorkSpace EMR Studio

Amazon EMR Studio esegue comandi notebook utilizzando un kernel su un cluster EMR. Per poter selezionare un kernel, è necessario collegare il WorkSpace a un cluster che utilizza istanze Amazon EC2 o a un cluster Amazon EMR su EKS o a un'applicazione EMR Serverless. EMR Studio consente di collegare WorkSpace a cluster nuovi o esistenti e offre la flessibilità necessaria per modificare i cluster senza chiudere il WorkSpace.

Questa sezione comprende i seguenti argomenti per informazioni su come lavorare con i cluster ed effettuare il relativo provisioning per EMR Studio:

- Collegamento di un cluster Amazon EC2 a un WorkSpace EMR Studio
- Collegamento di un cluster Amazon EMR su EKS a un WorkSpace EMR Studio
- Collegamento di un'applicazione Amazon EMR serverless a un WorkSpace EMR Studio
- Creazione e collegamento di un nuovo cluster EMR a un Workspace EMR Studio
- Scollegamento di un calcolo da un WorkSpace EMR Studio

Collegamento di un cluster Amazon EC2 a un WorkSpace EMR Studio

Puoi collegare un cluster EMR in esecuzione su Amazon EC2 a un WorkSpace quando creai il WorkSpace oppure puoi collegare un cluster a un WorkSpace esistente. Se desideri creare e collegare un nuovo cluster, consulta Creazione e collegamento di un nuovo cluster EMR a un Workspace EMR Studio.

On create

Collegamento di un cluster di calcolo Amazon EMR quando si crea un WorkSpace

1. Nella finestra di dialogo Create a WorkSpace (Crea un WorkSpace), verifica di aver già selezionato una sottorete per la nuova istanza WorkSpace. Espandi la sezione Advanced configuration (Configurazione avanzata).
2. Nella finestra di dialogo Create a WorkSpace (Crea un WorkSpace), verifica di aver già selezionato una sottorete per la nuova istanza WorkSpace. Espandi la sezione Advanced configuration (Configurazione avanzata).
3. Scegli Attach WorkSpace to an EMR cluster (Collega WorkSpace a un cluster EMR).
4. Nell'elenco a discesa Cluster EMR, seleziona un cluster EMR esistente per collegarlo al WorkSpace.

Dopo aver collegato un cluster, completa il processo creando il WorkSpace. Quando apri il nuovo WorkSpace per la prima volta e scegli il riquadro Cluster EMR, dovresti visualizzare il cluster selezionato collegato.

On launch

Collegamento di un cluster di calcolo Amazon EMR quando si avvia il WorkSpace

1. Vai all'elenco dei WorkSpace e seleziona la riga relativa al WorkSpace che desideri avviare. Quindi, seleziona Avvia Workspace > Avvia con opzioni.
2. Scegli un cluster EMR da collegare al tuo WorkSpace.

Dopo aver collegato un cluster, completa il processo creando il WorkSpace. Quando apri il nuovo WorkSpace per la prima volta e scegli il riquadro Cluster EMR, dovresti visualizzare il cluster selezionato collegato.

In JupyterLab

Collega un WorkSpace a un cluster di calcolo Amazon EMR in JupyterLab

1. Seleziona il tuo WorkSpace, quindi seleziona Avvia Workspace > Avvio rapido.
2. All'interno di JupyterLab, apri la scheda Cluster nella barra laterale sinistra.
3. Seleziona il menu a discesa EMR sul cluster EC2 o seleziona un cluster Amazon EMR su EKS.
4. Seleziona Collega per collegare il cluster al tuo WorkSpace.

Dopo aver collegato un cluster, completa il processo creando il WorkSpace. Quando apri il nuovo WorkSpace per la prima volta e scegli il riquadro Cluster EMR, dovresti visualizzare il cluster selezionato collegato.

In the Workspace UI

Collegamento di un WorkSpace a un cluster di calcolo Amazon EMR dall'interfaccia utente del WorkSpace

1. Nel WorkSpace che desideri collegare a un cluster, scegli l'icona Cluster EMR dalla barra laterale sinistra per aprire il riquadro Cluster.
2. In Tipo di cluster, espandi il menu a discesa e seleziona Cluster EMR su EC2.

3. Scegli un cluster dall'elenco a discesa. Potrebbe essere necessario scollegare prima un cluster esistente per abilitare l'elenco a discesa di selezione del cluster.
4. Scegliere Attach (Collega). Quando il cluster è collegato, viene visualizzato un messaggio di esito positivo.

Collegamento di un cluster Amazon EMR su EKS a un WorkSpace EMR Studio

Oltre a utilizzare i cluster Amazon EMR in esecuzione su Amazon EC2, puoi collegare un WorkSpace a un cluster Amazon EMR su EKS per eseguire il codice del notebook. Per ulteriori informazioni su Amazon EMR su EKS, consulta Che cos'è Amazon EMR su EKS.

Per poter connettere un WorkSpace a un cluster Amazon EMR su EKS, l'amministratore di Studio deve concedere le autorizzazioni di accesso.

On create

Per collegare un cluster Amazon EMR su EKS quando si crea un WorkSpace

1. Nella finestra di dialogo Create a Workspace (Creazione di un WorkSpace), espandere la sezione Advanced configuration (Configurazione avanzata).
2. Scegli Collega il WorkSpace a un cluster Amazon EMR su EKS.
3. In Cluster Amazon EMR su EKS, scegli un cluster dall'elenco a discesa.
4. In Select an endpoint (Seleziona un endpoint), scegli un endpoint gestito da collegare all'istanza WorkSpace. Un endpoint gestito è un gateway che consente a EMR Studio di comunicare con il cluster scelto.
5. Scegli Crea un WorkSpace per completare il processo di creazione del WorkSpace e collegare il cluster selezionato.

Dopo aver collegato un cluster, è possibile completare il processo di creazione del WorkSpace. Quando apri il nuovo WorkSpace per la prima volta e selezioni il riquadro Cluster EMR, dovresti visualizzare il cluster selezionato collegato.

In the Workspace UI

Per collegare un cluster Amazon EMR su EKS dall'interfaccia utente del WorkSpace

1. Nel WorkSpace che desideri collegare a un cluster, scegli l'icona Cluster EMR dalla barra laterale sinistra per aprire il riquadro Cluster.

2. Espandi il menu a discesa Tipo di cluster e scegli Cluster EMR su EKS.
3. In Cluster EMR su EKS, scegli un cluster dall'elenco a discesa.
4. In Endpoint, scegli un endpoint gestito da collegare all'istanza WorkSpace. Un endpoint gestito è un gateway che consente a EMR Studio di comunicare con il cluster scelto.
5. Scegliere Attach (Collega). Quando il cluster è collegato, viene visualizzato un messaggio di esito positivo.

Collegamento di un'applicazione Amazon EMR serverless a un WorkSpace EMR Studio

È possibile collegare un WorkSpace a un'applicazione EMR Serverless per eseguire carichi di lavoro interattivi. Per ulteriori informazioni, consulta Utilizzo dei notebook per eseguire carichi di lavoro interattivi con EMR Serverless tramite EMR Studio.

Example Collegamento di un Workspace a un'applicazione Serverless in JupyterLab

Prima di poter connettere un'istanza WorkSpace a un'applicazione Serverless, l'amministratore dell'account deve concedere le autorizzazioni di accesso come descritto in Autorizzazioni richieste per i carichi di lavoro interattivi.

1. Vai su EMR Studio e seleziona il tuo WorkSpace, quindi seleziona Avvia Workspace > Avvio rapido.
2. All'interno di JupyterLab, apri la scheda Cluster nella barra laterale sinistra.
3. Seleziona EMR Serverless come opzione di calcolo, quindi seleziona un'applicazione EMR Serverless e un ruolo runtime.
4. Per collegare il cluster al tuo WorkSpace, scegli Collega.

Ora, quando apri questo WorkSpace, dovresti vedere l'applicazione selezionata collegata.

Creazione e collegamento di un nuovo cluster EMR a un Workspace EMR Studio

Gli utenti avanzati di EMR Studio possono eseguire il provisioning di nuovi cluster EMR in esecuzione su Amazon EC2 da utilizzare con un WorkSpace. Il nuovo cluster dispone di tutte le applicazioni Big Data necessarie per EMR Studio installate per impostazione predefinita.

Per creare cluster, l'amministratore dello Studio deve prima concedere l'autorizzazione a utilizzare una policy di sessione. Per ulteriori informazioni, consulta Creazione di policy di autorizzazione per gli utenti di EMR Studio.

È possibile creare un nuovo cluster nella finestra di dialogo Create a WorkSpace (Crea un WorkSpace) o nel riquadro Cluster dell'interfaccia utente di WorkSpace. In entrambi i casi, sono disponibili due opzioni di creazione del cluster:

1. Create an EMR cluster (Crea un cluster EMR): crea un cluster EMR scegliendo il tipo e il numero di istanze Amazon EC2.
2. Use a cluster template (Utilizza un modello di cluster): provisioning rapido di un cluster selezionando un modello di cluster predefinito. Questa opzione è disponibile se si dispone dell'autorizzazione per utilizzare i modelli di cluster.

Creazione di un cluster EMR fornendo una configurazione del cluster

1. Scegli un orario di inizio.

A…	Esegui questa operazione…
Crea il cluster durante la creazione di un WorkSpace tramite la finestra di dialogo Create a WorkSpace (Crea un WorkSpace).	Espandi la sezione Advanced configuration (Configurazione avanzata) nella finestra di dialogo Create a WorkSpace (Crea un WorkSpace) e seleziona Create an EMR cluster (Crea un cluster EMR).
Crea il cluster dal riquadro Cluster EMR nell'interfaccia utente del WorkSpace dopo aver creato un WorkSpace.	Scegli la scheda Cluster EMR nella barra laterale sinistra di un WorkSpace aperto, espandi la sezione Configurazione avanzata e scegli Crea cluster.

2. Immetti un Cluster name (Nome cluster). La denominazione del cluster consente di individuarlo successivamente nell'elenco Clusters (Cluster) di EMR Studio.
3. Per Rilascio di Amazon EMR, scegli una versione di rilascio di Amazon EMR per il cluster.
4. Per Instance (Istanza), seleziona il tipo e il numero di istanze Amazon EC2 per il cluster. Per ulteriori informazioni sulla selezione dei tipi di istanza, consulta Configurazione delle istanze Amazon EC2. Un'istanza sarà utilizzata come nodo primario.

5. Selezionare una Subnet (Sottorete) dove EMR Studio può lanciare il nuovo cluster. Ogni opzione di sottorete è pre-approvata dall'amministratore di Studio, pertanto WorkSpace dovrebbe essere in grado di connettersi a un cluster in qualsiasi sottorete elencata.
6. Scegli un S3 URI for log storage (URI S3 per l'archiviazione dei log).
7. Scegli Create EMR cluster (Crea cluster EMR) per eseguire il provisioning del cluster. Se utilizzi la finestra di dialogo Crea un WorkSpace, scegli Crea un WorkSpace per creare il WorkSpace ed eseguire il provisioning del cluster. Dopo che EMR Studio esegue il provisioning del nuovo cluster, il cluster viene collegato automaticamente all'istanza WorkSpace.

Creazione di un cluster tramite un modello di cluster

1. Scegli un orario di inizio.

A…	Esegui questa operazione…
Crea il cluster durante la creazione di un WorkSpace tramite la finestra di dialogo Create a WorkSpace (Crea un WorkSpace).	Espandi la sezione Advanced configuration (Configurazione avanzata) nella finestra di dialogo Create a WorkSpace (Crea un WorkSpace) e seleziona Use a cluster template (Utilizza un modello di cluster).
Crea il cluster dal riquadro Cluster EMR nell'interfaccia utente del WorkSpace.	Scegli la scheda Cluster EMR nella barra laterale sinistra di un WorkSpace aperto, espandi la sezione Configurazione avanzata e quindi scegli Modello di cluster.

2. Seleziona un modello di cluster dall'elenco a discesa. Ogni modello di cluster disponibile include una breve descrizione che ti aiuta a effettuare una selezione.
3. Il modello di cluster scelto potrebbe avere parametri aggiuntivi, ad esempio la versione di Amazon EMR o il nome del cluster. È possibile scegliere o inserire valori oppure utilizzare i valori predefiniti selezionati dall'amministratore.
4. Selezionare una Subnet (Sottorete) dove EMR Studio può lanciare il nuovo cluster. Ogni opzione di sottorete è pre-approvata dall'amministratore di Studio e WorkSpace dovrebbe essere in grado di connettersi a un cluster in qualsiasi sottorete.
5. Scegli Use cluster template (Utilizza modello cluster) per eseguire il provisioning del cluster e collegarlo all'istanza WorkSpace. La creazione del cluster da parte di EMR Studio richiederà

alcuni minuti. Se utilizzi la finestra di dialogo Crea un WorkSpace, scegli Crea un WorkSpace per creare il WorkSpace ed eseguire il provisioning del cluster. Dopo che EMR Studio esegue il provisioning del nuovo cluster, il cluster viene collegato automaticamente all'istanza WorkSpace.

Scollegamento di un calcolo da un WorkSpace EMR Studio

Per scambiare il cluster collegato a un'istanza WorkSpace, è possibile scollegare un cluster dall'interfaccia utente dell'istanza WorkSpace.

Scollegamento di un cluster da un'istanza WorkSpace

1. Nel WorkSpace che desideri scollegare da un cluster, scegli l'icona Cluster EMR dalla barra laterale sinistra per aprire il riquadro Cluster.
2. In Select cluster (Seleziona cluster), scegli Detach (Scollega) e attendi che EMR Studio scolleghi il cluster. Quando il cluster viene scollegato, visualizzerai un messaggio di esito positivo.

Scollegamento di un'applicazione EMR Serverless da un WorkSpace EMR Studio

Per scambiare il calcolo collegato a un'istanza WorkSpace, è possibile scollegare l'applicazione dall'interfaccia utente dell'istanza WorkSpace.

1. Nel WorkSpace che desideri scollegare da un cluster, scegli l'icona Calcolo Amazon EMR dalla barra laterale sinistra per aprire il riquadro Calcolo.
2. In Seleziona calcolo, scegli Scollega e attendi che EMR Studio scolleghi l'applicazione. Quando l'applicazione viene scollegata, visualizzerai un messaggio di esito positivo.

Collegamento di repository basati su Git a un WorkSpace EMR Studio

Informazioni sui repository Git per EMR Studio

È possibile associare un massimo di tre repository Git a un WorkSpace EMR Studio. Per impostazione predefinita, ogni istanza WorkSpace consente di scegliere da un elenco di repository Git associati allo stesso account AWS utilizzato dallo Studio. Puoi anche creare un nuovo repository Git come risorsa per un'istanza WorkSpace.

Puoi eseguire comandi Git come il seguente utilizzando un comando terminale mentre sei connesso al nodo primario di un cluster.

```
!git pull origin <branch-name>
```

In alternativa, puoi utilizzare l'estensione jupyterlab-git. Aprila dalla barra laterale sinistra scegliendo l'icona Git. Per informazioni sull'estensione jupyterlab-git per JupyterLab, consulta jupyterlab-git.

Prerequisiti

- Per associare un repository Git a un'istanza WorkSpace, il tuo Studio deve essere configurato per consentire il collegamento del repository Git. L'amministratore dello Studio dovrebbe adottare le misure necessarie per Definizione di accesso e autorizzazioni per i repository basati su Git.
- Se utilizzi un repository CodeCommit, usa le credenziali Git e HTTPS. Le chiavi SSH e HTTPS con l'assistente credenziali AWS Command Line Interface non sono supportate. Inoltre, CodeCommit non supporta i token di accesso personale (PAT). Per ulteriori informazioni, consulta Utilizzo di IAM con CodeCommit nella Guida per l'utente IAM e Configurazione degli utenti HTTPS con le credenziali Git nella Guida per l'utente IAMAWS CodeCommit.

Istruzioni

Collegamento di un repository Git associato a un'istanza WorkSpace

1. Apri l'istanza WorkSpace da collegare a un repository dall'elenco WorkSpaces (istanze WorkSpace) nel Studio.
2. Nella barra laterale sinistra, seleziona l'icona Repository Git Amazon EMR per aprire il pannello degli strumenti Repository Git.
3. In Git repositories (Repository Git), espandi l'elenco a discesa e seleziona un massimo di tre repository diversi da collegare all'istanza WorkSpace. EMR Studio registra la selezione e inizia a collegare ogni repository.

Il completamento del processo di collegamento potrebbe richiedere un po' di tempo. È possibile visualizzare lo stato di ogni repository selezionato nel pannello degli strumenti Git repository (Repository Git). Dopo che EMR Studio collega un repository all'istanza WorkSpace, i file che appartengono a tale repository dovrebbero essere visualizzati nel riquadro File browser (Browser di file).

Aggiunta di un nuovo repository Git all'istanza WorkSpace come risorsa

1. Apri il WorkSpace da collegare a un repository dall'elenco WorkSpaces (WorkSpace) nel tuo Studio.
2. Nella barra laterale sinistra, seleziona l'icona Repository Git Amazon EMR per aprire il pannello degli strumenti Repository Git.
3. Scegli Add new Git repository (Aggiungi nuovo repository Git).
4. Per Repository name (Nome repository), immetti un nome descrittivo da utilizzare per il repository in EMR Studio. I nomi possono contenere solo caratteri alfanumerici, trattini alti e trattini bassi.
5. Per URL repository Git), immetti l'URL del repository. Quando utilizzi un repository CodeCommit, questo è l'URL che viene copiato quando scegli Clone URL (Clona URL) e in seguito Clone HTTPS (Clona HTTPS). Ad esempio, `https://git-codecommit.us-west-2.amazonaws.com/v1/repos/[MyCodeCommitRepoName]`.
6. Per Branch (Ramo), immetti il nome di un ramo esistente da controllare.
7. Per Git credentials (Credenziali Git), scegli un'opzione in base alle seguenti linee guida. EMR Studio accede alle credenziali Git utilizzando i segreti archiviati in Secrets Manager.

> **Note**
>
> Se utilizzi un repository GitHub, ti consigliamo di utilizzare un token di accesso personale (PAT) per l'autenticazione. A partire dal 13 agosto 2021, GitHub richiede l'autenticazione basata su token e non accetta più password durante l'autenticazione delle operazioni Git. Per ulteriori informazioni, consulta il post Token authentication requirements for Git operations (Requisiti di autenticazione token per operazioni Git) nel Blog di GitHub.

Opzione	Descrizione
Crea un nuovo segreto	Scegli questa opzione per associare le credenziali Git esistenti a un nuovo segreto che verrà creato per te in AWS Secrets Manager. Esegui una delle seguenti operazioni in base alle credenziali Git utilizzate per il repository.

Opzione	Descrizione
	Se utilizzi un nome utente e una password Git per accedere al repository, seleziona Nome utente e password, immetti il Nome segreto da utilizzare in Secrets Manager e quindi il Nome utente e la Password da associare al segreto. –OPPURE– Se utilizzi un token di accesso personale per accedere al repository, seleziona Personal access token (PAT) (Token di accesso personale (PAT)), immetti il Secret name (Nome segreto) da utilizzare in Secrets Manager e, in seguito, immetti il tuo personal access token (token di accesso personale). Per ulteriori informazioni, consulta Creazione di un token di accesso personale per la riga di comando per GitHub e Token di accesso personale per Bitbucket. I repository CodeCommit non supportano questa opzione.
Utilizza un repository pubblico senza credenziali	Scegli questa opzione per accedere a un repository pubblico.

Opzione	Descrizione
Utilizzo di un segreto AWS esistente	Scegli questa opzione se le credenziali sono già state salvate come segreto in Secrets Manager, quindi seleziona il nome segreto dall'elenco. Se selezioni un segreto associato a un nome utente e una password Git, il segreto deve essere nel formato `{"gitUsername": "MyUserName ", "gitPassword": "MyPassword "}`.

8. Scegli Add repository (Aggiungi repository) per creare il nuovo repository. Dopo che EMR Studio ha creato il nuovo repository, visualizzerai un messaggio di esito positivo. Il nuovo repository viene visualizzato nell'elenco a discesa in Git repositories (Repository Git).
9. Per collegare il nuovo repository alla tua istanza WorkSpace, selezionalo dall'elenco a discesa in Git repositories (Repository Git).

Il completamento del processo di collegamento potrebbe richiedere un po' di tempo. Dopo che EMR Studio collega il nuovo repository all'istanza WorkSpace, verrà visualizzata una nuova cartella con lo stesso nome del repository nel riquadro File Browser (Browser di file).

Per aprire un repository collegato differente, passare alla relativa cartella nel File browser (Browser di file).

Debug di applicazioni e processi con EMR Studio

Con Amazon EMR Studio, è possibile avviare interfacce di applicazioni dati per analizzare le applicazioni e le esecuzioni di processo nel browser.

Inoltre, è possibile avviare le interfacce utente persistenti fuori cluster per Amazon EMR in esecuzione su cluster EC2 dalla console di Amazon EMR. Per ulteriori informazioni, consulta Visualizzazione di interfacce utente delle applicazioni persistenti.

> **Note**
>
> A seconda delle impostazioni del browser, potrebbe essere necessario abilitare i popup per l'apertura dell'interfaccia utente di un'applicazione.

Per informazioni sulla configurazione e sull'utilizzo delle interfacce delle applicazioni, consulta Il Timeline Server di YARN, Monitoraggio e strumentazione o Panoramica dell'interfaccia utente Tez.

Esecuzione del debug di Amazon EMR su processi Amazon EC2

Workspace UI

Avvio di un'interfaccia utente su cluster da un file notebook

Quando si utilizza Amazon EMR versione 5.33.0 e successive, è possibile avviare l'interfaccia utente Web Spark (l'interfaccia utente Spark o Spark History Server) da un notebook nel WorkSpace.

Le interfacce utente su cluster funzionano con i kernel PySpark, Spark o SparkR. La dimensione massima del file visualizzabile per i log eventi o i log del container di Spark è di 10 MB. Se i file di log superano i 10 MB, si consiglia di utilizzare lo Spark History Server persistente anziché l'interfaccia utente Spark su cluster per eseguire il debug dei processi.

> **Important**
>
> Affinché EMR Studio possa avviare le interfacce utente delle applicazioni su cluster da un Workspace, il cluster deve essere in grado di comunicare con il Gateway Amazon API. È necessario configurare il cluster EMR per consentire il traffico di rete in uscita verso Amazon API Gateway e assicurarsi che Amazon API Gateway sia raggiungibile dal cluster.
> L'interfaccia utente Spark accede ai log del container risolvendo i nomi host. Se si utilizza un nome di dominio personalizzato, è necessario assicurarsi che i nomi host dei nodi del cluster possano essere risolti da Amazon DNS o dal server DNS specificato. A tale scopo, imposta le opzioni DHCP (Dynamic Host Configuration Protocol) per l'Amazon Virtual Private Cloud (VPC) associato al cluster. Per ulteriori informazioni sulle opzioni DHCP, consulta Set di opzioni DHCP nella Guida per l'utente di Amazon Virtual Private Cloud.

1. In EMR Studio, apri il WorkSpace che desideri utilizzare e assicurati che sia collegato a un cluster Amazon EMR in esecuzione su EC2. Per istruzioni, consultare Collegamento di un calcolo a un WorkSpace EMR Studio.
2. Apri un file notebook e utilizza il kernel PySpark, Spark o SparkR. Per selezionare un kernel, scegli il nome del kernel in alto a destra della barra degli strumenti del notebook per aprire la finestra di dialogo Select Kernel (Seleziona kernel). Il nome viene visualizzato come No Kernel! (Nessun kernel!) se non è stato selezionato alcun kernel.
3. Esegui il codice del notebook. Quando avvii il contesto Spark, viene visualizzato come output nel notebook. Potrebbero essere necessari alcuni secondi prima di visualizzarlo. Se hai avviato il contesto Spark, è possibile eseguire il comando `%%info` per accedere a un collegamento all'interfaccia utente Spark in qualsiasi momento.

   > **Note**
   >
   > Se i collegamenti dell'interfaccia utente Spark non funzionano o non vengono visualizzati dopo alcuni secondi, crea una nuova cella del notebook ed esegui il comando `%%info` per rigenerare i collegamenti.

4. Per avviare l'interfaccia utente Spark, seleziona Link (Collegamento) in Spark UI (Interfaccia utente Spark). Se l'applicazione Spark è in esecuzione, l'interfaccia utente Spark si apre in una nuova scheda. Se l'applicazione è stata completata, si apre Spark History Server.

   Dopo aver avviato l'interfaccia utente Spark, è possibile modificare l'URL nel browser per aprire il ResourceManager o il Timeline Server di YARN. Aggiungi uno dei percorsi seguenti dopo `amazonaws.com`.

Interfaccia utente Web	Path	Esempio di URL modificato
YARN ResourceManager	/rm	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*rm*
Timeline Server di Yarn	/yts	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*yts*
Spark History Server	/shs	https://*j-examplebby5ij* .emrappui-prod.*eu-west-1* .amazonaws.com/*shs*

Studio UI

Avvio dell'interfaccia utente persistente del Timeline Server di YARN, di Spark History Server o di Tez dall'interfacia utente di EMR Studio

1. In EMR Studio, seleziona Amazon EMR su EC2 a sinistra della pagina per aprire l'elenco di cluster Amazon EMR su EC2.
2. Filtra l'elenco dei cluster per name (nome), state (stato) oppure ID immettendo valori nella casella di ricerca. Puoi anche effettuare una ricerca per time range (intervallo temporale) di creazione.
3. Seleziona un cluster e scegli Launch application UIs (Avvia interfacce utente applicazioni) per selezionare l'interfaccia utente di un'applicazione. L'interfaccia utente dell'applicazione si apre in una nuova scheda del browser e potrebbe richiedere del tempo per il caricamento.

Esegui il debug di EMR Studio in esecuzione su EMR Serverless

Analogamente ad Amazon EMR in esecuzione su Amazon EC2, puoi utilizzare l'interfaccia utente di WorkSpace per analizzare le tue applicazioni EMR Serverless. Dall'interfaccia utente WorkSpaces, quando utilizzi le versioni 6.14.0 e successive di Amazon EMR, è possibile avviare l'interfaccia utente Web Spark (l'interfaccia utente Spark o Spark History Server) da un notebook nel WorkSpace. Per comodità, forniamo anche un collegamento al log dei driver per accedere rapidamente ai log dei driver Spark.

Esecuzione del debug delle esecuzioni di processo Amazon EMR su EKS con Spark History Server

Quando invii un processo eseguito a un cluster Amazon EMR su EKS, puoi accedere ai registri per quel processo eseguito utilizzando Spark History Server. Spark History Server fornisce strumenti per il monitoraggio delle applicazioni Spark, come un elenco di fasi e processi di pianificazione, un riepilogo delle dimensioni RDD e dell'utilizzo della memoria e informazioni ambientali. È possibile avviare Spark History Server per Amazon EMR sulle esecuzioni del processo EKS nei seguenti modi:

- Quando invii l'esecuzione di un processo utilizzando EMR Studio con un endpoint gestito da Amazon EMR su EKS, puoi avviare Spark History Server da un file notebook nel tuo Workspace.
- Quando si invia un processo eseguito utilizzando la AWS CLI o AWS SDK per Amazon EMR su EKS, è possibile avviare Spark History Server dall'interfaccia utente di EMR Studio.

Per informazioni su come utilizzare Spark History Server, consulta Monitoraggio e strumentazionenella documentazione di Apache Spark. Per ulteriori informazioni sulle esecuzioni di processo, consulta Concetti e componenti nella Guida allo sviluppo di Amazon EMR su EKS.

Per l'avvio dello Spark History Server persistente da un file notebook nella tua istanza WorkSpace di EMR Studio

1. Aprire un'istanza WorkSpace connessa a un cluster Amazon EMR su EKS.
2. Seleziona e apri il file del notebook nell'istanza WorkSpace.
3. Scegli Spark UI (Interfaccia utente Spark) nella parte superiore di un file notebook per aprire lo Spark History Server persistente in una nuova scheda.

Per avviare Spark History Server dall'interfaccia utente di EMR Studio

> **Note**
>
> L'elenco Jobs (Processi) nell'interfaccia utente di EMR Studio visualizza solo le esecuzioni di processi inviati utilizzando la AWS CLI o AWS SDK per Amazon EMR su EKS.

1. In EMR Studio, seleziona Amazon EMR su EKS a sinistra della pagina.
2. Cerca il cluster virtuale Amazon EMR su EKS utilizzato per inviare l'esecuzione del processo. Puoi filtrare l'elenco dei cluster per status (stato) o ID immettendo valori nella casella di ricerca.

3. Seleziona il cluster per aprire la relativa pagina dei dettagli. Nella pagina dei dettagli vengono visualizzate informazioni sul cluster, ad esempio ID, spazio dei nomi e stato. La pagina mostra anche un elenco di tutti i processi eseguiti inviati a quel cluster.
4. Dalla pagina dei dettagli del cluster, seleziona un processo da sottoporre a debug.
5. In alto a destra dell'elenco Jobs (Processi), seleziona Launch Spark History Server (Avvia Spark History Server) per aprire l'interfaccia dell'applicazione in una nuova scheda del browser.

Installazione di kernel e librerie in un'istanza WorkSpace di EMR Studio

Ogni istanza WorkSpace di Amazon EMR Studio viene fornito con un set di librerie e kernel preinstallati.

Kernel e librerie sui cluster che eseguono su Amazon EC2

È inoltre possibile personalizzare l'ambiente per EMR Studio nei seguenti modi quando si utilizzano cluster EMR in esecuzione su Amazon EC2:

- Installazione dei kernel Jupyter Notebook e delle librerie Python su un nodo primario del cluster: quando si installano le librerie utilizzando questa opzione, tutti i WorkSpace collegati allo stesso cluster condividono quelle librerie. È possibile installare kernel o librerie all'interno di una cella del notebook o mentre si è connessi tramite SSH al nodo primario di un cluster.
- Utilizzare librerie con ambito notebook: quando gli utenti delle istanze WorkSpace installano e utilizzano le librerie all'interno di una cella del notebook, tali librerie sono disponibili solo per quel notebook. Questa opzione consente a diversi notebook utilizzando lo stesso cluster di lavorare senza preoccuparsi di versioni di libreria in conflitto.

I WorkSpace EMR Studio hanno la stessa architettura sottostante dei notebook EMR. È possibile installare e utilizzare i kernel Jupyter Notebook e le librerie Python con EMR Studio allo stesso modo dei notebook EMR. Per istruzioni, consultare Installazione e uso di kernel e librerie.

Kernel e librerie sui cluster Amazon EMR su EKS

I cluster Amazon EMR su EKS includono i kernel PySpark e Python 3.7 con un set di librerie preinstallate. Amazon EMR su EKS non supporta l'installazione di librerie o cluster aggiuntivi.

Ogni cluster Amazon EMR su EKS viene fornito con le seguenti librerie Python e PySpark installate:

- 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

Kernel e librerie sulle applicazioni EMR Serverless

Ogni applicazione EMR Serverless viene fornito con le seguenti librerie Python e PySpark installate:

- Python – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn
- PySpark – ggplot, matplotlib,numpy, pandas, plotly, bokeh, scikit-learn, scipy, seaborn

Migliora i kernel con comandi magic

Panoramica

EMR Studio e EMR Notebooks supportano i comandi magic, Magic o magic, sono miglioramenti che il kernel IPython fornisce per facilitare l'esecuzione e l'analisi dei dati. IPython è un ambiente shell (interprete di comandi) interattivo creato con Python.

Amazon EMR supporta anche Sparkmagic, un pacchetto che fornisce kernel relativi a Spark (kernel PySpark, SparkR e Scala) con comandi magic specifici e che utilizza Livy sul cluster per inviare processi Spark.

È possibile utilizzare i comandi magic purché si disponga di un kernel Python nel notebook EMR. Analogamente, qualsiasi kernel relativo a Spark supporta i comandi Sparkmagic.

Comandi Magic, chiamati anche magic, sono disponibili in due varietà:

- Line magic: questi comandi magic sono denotati da un singolo prefisso `%` e funzionano su una singola riga di codice
- Cell magic: questi comandi magic sono indicati con un doppio prefisso `%%` e funzionano su più righe di codice

Per tutti i magic disponibili, consulta Elenco magic e comandi Sparkmagic.

Considerazioni e limitazioni

- EMR Serverless non supporta il `%%sh` per eseguire `spark-submit`. Non supporta i magic di EMR Notebooks.
- I cluster Amazon EMR su EKS non supportano i comandi Sparkmagic per EMR Studio. Questo perché i kernel Spark utilizzati con gli endpoint gestiti sono integrati in Kubernetes e non sono supportati da Sparkmagic e Livy. Puoi impostare la configurazione Spark direttamente nell'oggetto SparkContext come soluzione alternativa, come dimostra l'esempio seguente.

  ```
  spark.conf.set("spark.driver.maxResultSize", '6g')
  ```
- I seguenti comandi e azioni magic sono proibiti da: AWS
  - `%alias`
  - `%alias_magic`
  - `%automagic`
  - `%macro`
  - Modifica `proxy_user` con `%configure`
  - Modifica `KERNEL_USERNAME` con `%env` o `%set_env`

Elenco magic e comandi Sparkmagic

Utilizza i comandi seguenti per visualizzare un elenco di tutti i comandi magic disponibili:

- `%lsmagic` elenca tutte le funzioni magic attualmente disponibili.
- `%%help` elenca le funzioni magic attualmente disponibili su Spark fornite dal pacchetto Sparkmagic.

Utilizza **`%%configure`** per configurare Spark

Uno dei comandi Sparkmagic più utili è il comando `%%configure`, che configura i parametri di creazione della sessione. Utilizzando le impostazioni `conf`, è possibile configurare qualsiasi configurazione Spark menzionata nella documentazione di configurazione di Apache Spark.

Example Aggiungi un file JAR esterno ai EMR Notebooks dal repository Maven o da Amazon S3

È possibile utilizzare il seguente approccio per aggiungere una dipendenza da file JAR esterno a qualsiasi kernel relativo a Spark supportato da 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 : Configura Hudi

È quindi possibile utilizzare l'editor del notebook per configurare EMR Notebooks per l'utilizzo di 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"
     }
}
```

Utilizza `%%sh` per eseguire `spark-submit`

Il `%%sh` magic esegue comandi shell (interprete di comandi) in un sottoprocesso su un'istanza del cluster collegato. In genere, si utilizza uno dei kernel relativi a Spark per eseguire applicazioni Spark sul cluster collegato. Tuttavia, se si desidera utilizzare un kernel Python per inviare un'applicazione Spark, è possibile utilizzare la seguente magic, sostituendo il nome del bucket con il nome del bucket in minuscolo.

```
%%sh
spark-submit --master yarn --deploy-mode cluster s3://DOC-EXAMPLE-BUCKET/test.py
```

In questo esempio, il cluster deve accedere alla posizione di `s3://DOC-EXAMPLE-BUCKET/test.py` o il comando avrà esito negativo.

È possibile utilizzare qualsiasi comando Linux con `%%sh` magic. Per eseguire qualsiasi comando Spark o YARN, utilizza una delle seguenti opzioni per creare un Utente Hadoop `emr-notebook` e concedi all'utente le autorizzazioni per eseguire i comandi:

- Puoi creare esplicitamente un nuovo utente eseguendo i seguenti comandi.

```
hadoop fs -mkdir /user/emr-notebook
hadoop fs -chown emr-notebook /user/emr-notebook
```

- Puoi attivare la rappresentazione dell'utente in Livy, che crea automaticamente l'utente. Per ulteriori informazioni, consulta Abilitazione della rappresentazione utente per monitorare l'attività dell'utente e dei processi Spark.

Utilizza `%%display` per visualizzare i dataframe Spark

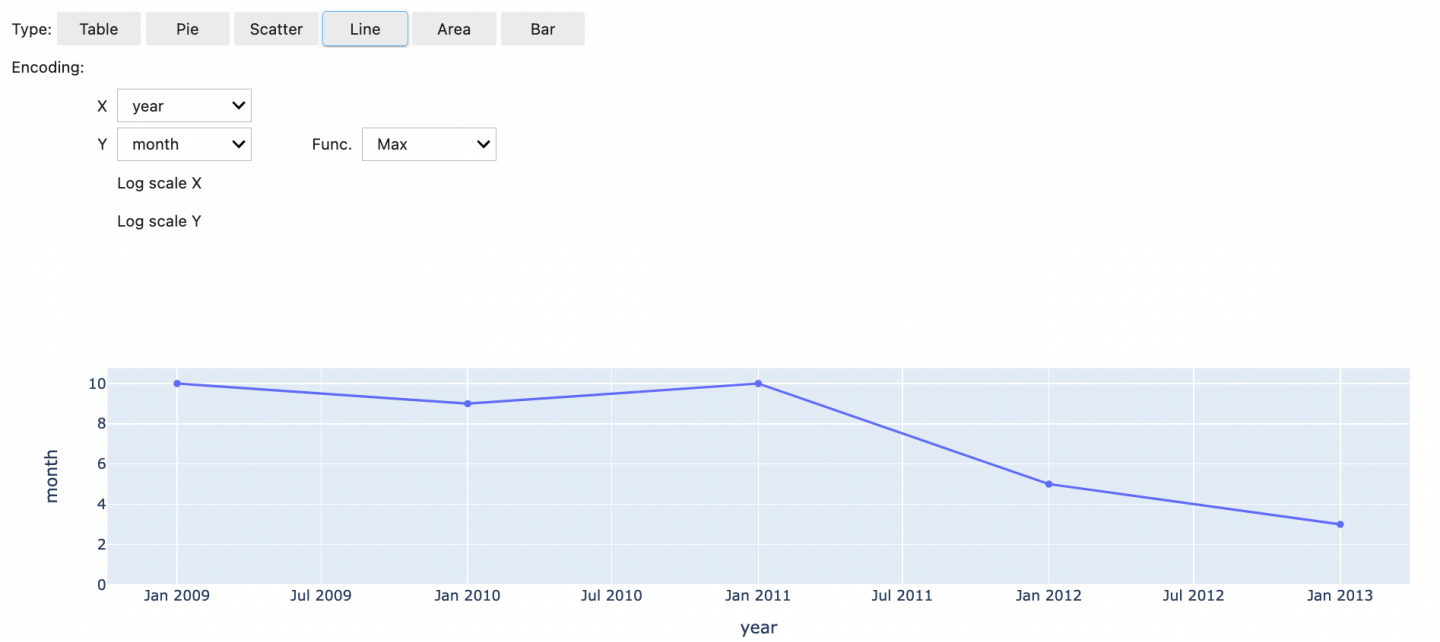
Puoi usare `%%display` magic per visualizzare un dataframe Spark. Per usare questo magic, eseguire il seguente comando.

```
%%display df
```

Scegli di visualizzare i risultati in formato tabella, come illustrato nell'immagine seguente.

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

È inoltre possibile scegliere di visualizzare i dati con cinque tipi di grafici. Le opzioni includono grafici a torta, a dispersione, a linee, ad area e a barre.



Utilizza i magic di EMR Notebooks

Amazon EMR fornisce i seguenti magic di EMR Notebooks che possono essere utilizzati con i kernel basati su Python3 e Spark:

- `%mount_workspace_dir` - Monta la directory WorkSpace sul cluster in modo da poter importare ed eseguire codice da altri file nel WorkSpace

> **Note**
>
> con `%mount_workspace_dir`, solo il kernel Python 3 può accedere ai file system locali. Gli executor Spark non avranno accesso alla directory montata con questo kernel.

- `%umount_workspace_dir` - Smonta la directory WorkSpace dal cluster
- `%generate_s3_download_url` - Genera un link per il download temporaneo nell'output del notebook per un oggetto Amazon S3

Prerequisiti

Prima di installare i magic di EMR Notebooks, completare i seguenti processi:

- Assicurarsi che il proprio Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2) disponga dell'accesso in lettura per Amazon S3. Il `EMR_EC2_DefaultRole` con la policy gestita

`AmazonElasticMapReduceforEC2Role` soddisfa questo requisito. Se si utilizza un ruolo o una policy personalizzati, assicurarsi che dispongano delle autorizzazioni S3 necessarie.

> **Note**
>
> I magic di EMR Notebooks vengono eseguiti su un cluster come utente del notebook e utilizzano il profilo dell'istanza EC2 per interagire con Amazon S3. Quando si monta una directory WorkSpace su un cluster EMR, tutti i notebook WorkSpace e EMR con il permesso di collegarsi a tale cluster possono accedere alla directory montata.
> Le directory sono montate in sola lettura per impostazione predefinita. Mentre `s3fs-fuse` e `goofys` permettono il supporto in lettura-scrittura, consigliamo vivamente di non modificare i parametri di montaggio per montare le directory in modalità di lettura-scrittura. Se si permette l'accesso in scrittura, le modifiche apportate alla directory vengono scritte nel bucket S3. Per evitare l'eliminazione o la sovrascrittura accidentale, è possibile abilitare il controllo delle versioni per il bucket S3. Per ulteriori informazioni, consulta Utilizzo della funzione controllo delle versioni nei bucket S3.

- Eseguire uno dei seguenti script sul cluster per installare le dipendenze per i magic di EMR Notebooks. Per eseguire uno script, è possibile eseguire Utilizzo di operazioni di bootstrap personalizzate oppure segui le istruzioni in Run commands and scripts on an Amazon EMR cluster (Esegui comandi e script su un cluster Amazon EMR) quando si dispone già di un cluster in esecuzione.

  È possibile scegliere quale dipendenza installare. Entrambi s3fs-fuse e goofys sono strumenti FUSE (Filesystem in Userspace) che consentono di montare un bucket Amazon S3 come file system locale su un cluster. Lo strumento `s3fs` offre un'esperienza simile a POSIX. Lo strumento `goofys` è una buona scelta quando si preferiscono le prestazioni rispetto a un file system conforme a POSIX.

  ```
  #!/bin/sh

  # Install the s3fs dependency for EMR Notebooks magics sudo amazon-linux-extras
   install epel -y
  sudo yum install s3fs-fuse -y
  ```

  OPPURE

  ```
  #!/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
```

Installare i magic di EMR Notebooks

> **Note**
>
> Con i rilasci di Amazon EMR da 6.0 a 6.9.0 e da 5.0 a 5.36.0, solo il pacchetto `emr-notebooks-magics` versione 0.2.0 e successive supporta `%mount_workspace_dir` magic.

Per installare i magic di EMR Notebooks, completare i seguenti passaggi.

1. Nel notebook, esegui i comandi seguenti per installare il pacchetto `emr-notebooks-magics`.

   ```
   %pip install boto3 --upgrade
   %pip install botocore --upgrade
   %pip install emr-notebooks-magics --upgrade
   ```

2. Riavviare il kernel per caricare i magic di EMR Notebooks.
3. Verificare l'installazione con il seguente comando, che dovrebbe visualizzare il testo della guida di output per `%mount_workspace_dir`.

   ```
   %mount_workspace_dir?
   ```

Montare una directory WorkSpace con **`%mount_workspace_dir`**

Il `%mount_workspace_dir` magic consente di montare la directory WorkSpace sul cluster EMR in modo da poter importare ed eseguire altri file, moduli o pacchetti archiviati nella directory.

L'esempio seguente monta l'intera directory di WorkSpace su un cluster e specifica l'argomento opzionale *<--fuse-type>* affinché vengano utilizzati goofys per il montaggio della directory.

```
%mount_workspace_dir . <--fuse-type goofys>
```

Per verificare che la directory WorkSpace sia montata, utilizzare l'esempio seguente per visualizzare l'attuale directory di lavoro con il comando `ls`. L'output dovrebbe visualizzare tutti i file nel WorkSpace.

```
%%sh
ls
```

Una volta che sono state apportate le modifiche nel WorkSpace, puoi smontare la directory del WorkSpace con il seguente comando:

> **Note**
>
> La directory WorkSpace rimane montata sul cluster anche quando il WorkSpace viene arrestato o distaccato. È necessario smontare esplicitamente la directory di WorkSpace.

```
%umount_workspace_dir
```

Scaricare un oggetto Amazon S3 con **`%generate_s3_download_url`**

Il comando `generate_s3_download_url` crea un URL prefirmato per un oggetto archiviato in Amazon S3. È possibile utilizzare l'URL preimpostato per scaricare l'oggetto sul computer locale. Ad esempio si potrebbe eseguire `generate_s3_download_url` per scaricare il risultato di una query SQL che il codice scrive su Amazon S3.

L'URL prefirmato è valido per 60 minuti per impostazione predefinita. È possibile modificare il tempo di scadenza specificando un numero di secondi per il flag `--expires-in`. Ad esempio, `--expires-in 1800` crea un URL valido per 30 minuti.

L'esempio seguente genera un collegamento per il download di un oggetto specificando il percorso completo di Amazon S3: *`s3://EXAMPLE-DOC-BUCKET/path/to/my/object`*.

```
%generate_s3_download_url s3://EXAMPLE-DOC-BUCKET/path/to/my/object
```

Per ulteriori informazioni sull'uso di `generate_s3_download_url`, eseguire il comando di seguito per visualizzare il testo della guida.

```
%generate_s3_download_url?
```

Gestione di un notebook in modalità headless con `%execute_notebook`

Con `%execute_notebook` magic, puoi eseguire un altro notebook in modalità headless e visualizzare l'output di ogni cella che hai utilizzato. Questo magic richiede autorizzazioni aggiuntive per il ruolo dell'istanza condiviso da Amazon EMR e Amazon EC2. Per ulteriori dettagli su come concedere autorizzazioni aggiuntive, esegui il comando `%execute_notebook?`.

Durante un processo di lunga durata, il sistema potrebbe andare in modalità di sospensione a causa dell'inattività o perdere temporaneamente la connettività Internet. Ciò potrebbe interrompere la connessione tra il browser e il server Jupyter. In questo caso, potresti perdere l'output delle celle che hai eseguito e inviato dal server Jupyter.

Se utilizzi il notebook in modalità headless con `%execute_notebook` magic, EMR Notebooks acquisisce l'output dalle celle che hai eseguito, anche in caso di interruzione della rete locale. EMR Notebooks salva l'output in modo incrementale in un nuovo notebook con lo stesso nome del notebook che hai utilizzato. EMR Notebooks inserisce quindi il notebook in una nuova cartella all'interno del workspace. Le esecuzioni headless avvengono sullo stesso cluster e utilizzano il ruolo di servizio `EMR_Notebook_DefaultRole`, ma argomenti aggiuntivi possono modificare i valori predefiniti.

Per eseguire un notebook in modalità headless, usa il seguente comando:

```
%execute_notebook <relative-file-path>
```

Per specificare un ID cluster e il ruolo di servizio per un'esecuzione headless, usa il seguente comando:

```
%execute_notebook <notebook_name>.ipynb --cluster-id <emr-cluster-id> --service-role
 <emr-notebook-service-role>
```

Quando Amazon EMR e Amazon EC2 condividono un ruolo di istanza, il ruolo richiede le seguenti autorizzazioni aggiuntive:

```
{
    "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**
>
> Per utilizzare `%execute_notebook` magic, installa il pacchetto `emr-notebooks-magics`, versione 0.2.3 o successive.

Usare notebook multilingue con i kernel Spark

Ogni kernel notebook di Jupyter ha un linguaggio di default. Ad esempio, il linguaggio di default del kernel Spark è Scala e il linguaggio di default del kernel PySpark è Python. Con Amazon EMR 6.4.0 e versioni successive, EMR Studio supporta notebook multilingue. Ciò significa che ogni kernel di EMR Studio può supportare le seguenti linguaggi oltre al linguaggio di default: Python, Spark, R e Spark SQL.

Per attivare questa funzionalità, specificare uno dei seguenti comandi magic all'inizio di qualsiasi cella.

Linguaggio	Comando
Python	`%%pyspark`
Scala	`%%scalaspark`
R	`%%rspark`

Linguaggio	Comando
	Non è supportato per i carichi di lavoro interattivi con EMR Serverless.
Spark SQL	`%%sql`

Quando vengono invocati, questi comandi eseguono l'intera cella all'interno della stessa sessione Spark utilizzando l'interprete del linguaggio corrispondente.

La cella `%%pyspark` magic consente agli utenti di scrivere codice PySpark in tutti i kernel Spark.

```
%%pyspark
a = 1
```

La cella `%%sql` magic consente agli utenti di eseguire codice Spark-SQL in tutti i kernel Spark.

```
%%sql
SHOW TABLES
```

La cella `%%rspark` magic consente agli utenti di eseguire codice SparkR in tutti i kernel Spark.

```
%%rspark
a <- 1
```

La cella `%%scalaspark` magic consente agli utenti di eseguire il codice Spark Scala in tutti i kernel Spark.

```
%%scalaspark
val a = 1
```

Condividere i dati tra interpreti del linguaggio utilizzando tabelle temporanee

È inoltre possibile condividere dati tra interpreti del linguaggio utilizzando tabelle temporanee. Nell'esempio seguente viene utilizzato `%%pyspark` in una cella per creare una tabella temporanea in Python e utilizza `%%scalaspark` nella cella seguente per leggere i dati di quella tabella in 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)
```

Panoramica di Notebook Amazon EMR

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

È possibile utilizzare Notebook Amazon EMR insieme ai cluster Amazon EMR che eseguono Apache Spark per creare e aprire interfacce di notebook Jupyter e JupyterLab all'interno della console di Amazon EMR. Un notebook EMR è un notebook "serverless" che puoi utilizzare per eseguire query e codice. A differenza di un notebook tradizionale, i contenuti di un notebook EMR, ossia le equazioni, le query, i modelli, il codice e il testo narrativo all'interno delle celle del notebook, vengono eseguiti in un client. I comandi vengono eseguiti utilizzando un kernel sul cluster EMR. I contenuti del notebook vengono salvati in Amazon S3 separatamente dai dati del cluster per una maggiore durata e un riutilizzo flessibile.

È possibile avviare un cluster, collegare un notebook EMR per l'analisi e infine terminare il cluster. Inoltre, puoi chiudere un notebook collegato a un cluster in esecuzione e passare a un altro. Più utenti possono collegare notebook allo stesso cluster contemporaneamente e condividere i file dei notebook tra loro in Amazon S3. Queste funzionalità consentono di eseguire cluster on demand per risparmiare sui costi e ridurre il tempo richiesto per riconfigurare i notebook per cluster e set di dati diversi.

Inoltre, puoi eseguire un notebook EMR a livello di programmazione utilizzando l'API Amazon EMR, senza la necessità di interagire con la console Amazon EMR ("esecuzione headless"). Nel notebook EMR, dovrai includere una cella contenente un tag di parametri. Questa cella consente a uno script di passare nuovi valori di input al notebook. I notebook parametrizzati possono essere riutilizzati con diversi set di valori di input. Non è necessario creare copie dello stesso notebook per modificarlo ed eseguirlo con nuovi valori di input. Amazon EMR crea e salva il notebook di output su S3 per ogni esecuzione del notebook parametrizzato. Per visualizzare alcuni esempi di codice API per notebook EMR, consulta Comandi di esempio per eseguire Notebook EMR a livello di programmazione.

> **Important**
>
> La funzionalità Notebook EMR supporta cluster che utilizzano i rilasci 5.18.0 e successivi di Amazon EMR. Consigliamo di utilizzare Notebook EMR con cluster che utilizzano la versione più recente di Amazon EMR o almeno 5.30.0, 5.32.0 o 6.2.0. Con questi rilasci, i kernel Jupyter vengono eseguiti sul cluster collegato anziché su un'istanza Jupyter. Ciò migliora le prestazioni e aumenta la possibilità di personalizzare kernel e librerie. Per ulteriori informazioni, consulta Differenze nelle funzionalità in base alla versione del cluster.

Si applicano i costi previsti per l'archiviazione Amazon S3 e per i cluster Amazon EMR.

I notebook Amazon EMR sono disponibili come Workspace Amazon EMR Studio nella nuova console

Transizione da Notebook EMR ai Workspace

Nella nuova console Amazon EMR, abbiamo unito Notebook EMR ai Workspace Amazon EMR Studio in un'unica esperienza. Quando utilizzi un EMR Studio, puoi creare e configurare diversi Workspace per organizzare ed eseguire notebook. Se nella vecchia console avevi notebook Amazon EMR, questi sono disponibili come Workspace EMR Studio nella nuova console.

Amazon EMR ha creato questi nuovi Workspace EMR Studio per tuo conto. Il numero di istanze Studio create corrisponde al numero di VPC diversi che utilizzi da Notebook EMR. Ad esempio, se ti connetti ai cluster EMR in due diversi VPC da Notebook EMR, significa che abbiamo creato due nuovi EMR Studio. I tuoi notebook vengono distribuiti tra i nuovi Studio.

> **Important**
>
> Abbiamo disattivato la possibilità di creare nuovi notebook nella vecchia console Amazon EMR. Utilizza invece il pulsante Crea Workspace nella nuova console Amazon EMR.

Per ulteriori informazioni sui Workspace Amazon EMR Studio, consulta Informazioni sulle nozioni di base di WorkSpace. Per una panoramica concettuale di EMR Studio, consulta WorkSpace nella pagina Come funziona Amazon EMR Studio.

Che cosa occorre fare?

Sebbene sia ancora possibile utilizzare i notebook esistenti nella vecchia console, ti consigliamo di utilizzare i Workspace Amazon EMR Studio nella nuova console. Devi configurare le autorizzazioni di ruolo aggiuntive per attivare le funzionalità di EMR Studio che non sono disponibili in Notebook EMR.

> **Note**
>
> Come minimo, per visualizzare i notebook EMR esistenti come Workspace EMR Studio e per crearne di nuovi, gli utenti devono disporre delle autorizzazioni `elasticmapreduce:ListStudios` e `elasticmapreduce:CreateStudioPresignedUrl` relative ai propri ruoli. Per accedere a tutte le funzionalità di EMR Studio, consulta Abilitazione delle funzionalità di EMR Studio per gli utenti di Notebook EMR per l'elenco completo delle autorizzazioni aggiuntive necessarie agli utenti di Notebook EMR.

Funzionalità avanzate in EMR Studio oltre Notebook EMR

Con Amazon EMR Studio, puoi configurare e utilizzare le seguenti funzionalità che non sono disponibili con Notebook EMR:

- Sfogliare e collegare i cluster EMR dall'interno di Jupyterlab
- Sfogliare e collegare i cluster virtuali di Notebook EMR dall'interno di Jupyterlab
- Connettersi ai repository Git dall'interno di Jupyterlab
- Collaborare con altri membri del team per scrivere ed eseguire il codice del notebook
- Sfogliare i dati con SQL Explorer
- Eseguire il provisioning di cluster EMR con Service Catalog

Per un elenco completo delle funzionalità di Amazon EMR Studio, consulta Funzionalità principali di EMR Studio.

Abilitazione delle funzionalità di EMR Studio per gli utenti di Notebook EMR

Le nuove istanze EMR Studio che creeremo nell'ambito di questa unione utilizzano il ruolo IAM `EMR_Notebooks_DefaultRole` esistente come ruolo di servizio EMR Studio.

Gli utenti che effettuano il passaggio da Notebook EMR a EMR Studio e desiderano utilizzare le funzionalità aggiuntive di EMR Studio devono richiedere nuove autorizzazioni di ruolo. Aggiungi le seguenti autorizzazioni ai ruoli degli utenti di Notebook EMR che intendono utilizzare EMR Studio.

> **Note**
>
> Come minimo, per visualizzare i notebook EMR esistenti come Workspace EMR Studio e per crearne di nuovi, gli utenti devono disporre delle autorizzazioni `elasticmapreduce:ListStudios` e `elasticmapreduce:CreateStudioPresignedUrl` relative ai propri ruoli. Per utilizzare tutte le funzionalità di EMR Studio, aggiungi tutte le autorizzazioni elencate di seguito. Gli utenti amministratori devono inoltre disporre dell'autorizzazione per creare e gestire un EMR Studio. Per ulteriori informazioni, consulta Autorizzazioni di amministratore per creare e gestire 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"
```

Le seguenti autorizzazioni sono necessarie anche per utilizzare le funzionalità di collaborazione in EMR Studio, ma non erano richieste con i notebook EMR.

```
"sso-directory:SearchUsers",
"iam:GetUser",
"iam:GetRole",
"iam:ListUsers",
"iam:ListRoles",
"sso:GetManagedApplicationInstance"
```

Considerazioni su quando utilizzare Notebook EMR

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Quando crei cluster e sviluppi soluzioni utilizzando notebook EMR, tieni a mente i seguenti requisiti.

Requisiti del cluster

- Attivazione del blocco dell'accesso pubblico Amazon EMR: l'accesso in ingresso a un cluster consente agli utenti del cluster di eseguire i kernel dei notebook. Assicurati che solo gli utenti autorizzati possano accedere al cluster. Consigliamo vivamente di lasciare abilitato il blocco dell'accesso pubblico e di limitare il traffico SSH in ingresso solo a origini affidabili. Per ulteriori informazioni, consulta Utilizzo del blocco dell'accesso pubblico di Amazon EMR e Controllo del traffico di rete con gruppi di sicurezza.
- Utilizzo di un cluster compatibile: un cluster collegato a un notebook deve soddisfare i seguenti requisiti:
  - Sono supportati solo i cluster creati utilizzando Amazon EMR. È possibile creare un cluster in modo indipendente all'interno di Amazon EMR e, successivamente, collegare un notebook EMR, oppure è possibile creare un cluster compatibile durante la creazione di un notebook EMR.

- Solo i cluster creati utilizzando Amazon EMR versione 5.18.0 e successive sono supportati. Per informazioni, consultare the section called “Differenze nelle funzionalità in base alla versione del cluster”.
- I cluster creati utilizzando istanze Amazon EC2 con processori AMD EPYC, ad esempio i tipi di istanza m5a.* e r5a.*, non sono supportati.
- Notebook EMR funziona solo con cluster creati con `VisibleToAllUsers` impostato su `true`. `VisibleToAllUsers` è `true` per impostazione predefinita.
- Il cluster deve essere avviato all'interno di un EC2-VPC. Sono supportate sottoreti pubbliche e private. La piattaforma EC2-Classic non è supportata.
- I cluster devono essere avviati con Hadoop, Spark e Livy installati. Possono essere installate altre applicazioni, ma attualmente Notebook EMR supporta solo i cluster Spark.

  > **Important**
  >
  > Per le versioni di Amazon EMR 5.32.0 e successive, o 6.2.0 e successive, il cluster deve eseguire anche l'applicazione Jupyter Enterprise Gateway per poter lavorare con Notebook EMR.

- I cluster che utilizzano l'autenticazione Kerberos non sono supportati.
- I cluster integrati con AWS Lake Formation supportano solo l'installazione di librerie con ambito notebook. L'installazione di kernel e librerie nel cluster non è supportata.
- I cluster con più nodi primari non sono supportati.
- I cluster che utilizzano istanze Amazon EC2 basate su AWS Graviton2 non sono supportati.

Differenze nelle funzionalità in base alla versione del cluster

Consigliamo di utilizzare Notebook EMR con cluster creati utilizzando Amazon EMR versione 5.30.0, 5.32.0 o successive oppure 6.2.0 o successive. Con queste versioni, Notebook EMR esegue i kernel sul cluster Amazon EMR collegato. I kernel e le librerie possono essere installati direttamente sul nodo primario del cluster. L'uso di EMR Notebooks con queste versioni del cluster presenta i seguenti vantaggi:

- Prestazioni migliorate: i kernel dei notebook vengono eseguiti su cluster con i tipi di istanza EC2 selezionati. Le versioni precedenti eseguono i kernel su un'istanza specializzata che non è ridimensionabile, accessibile o personalizzabile.

- Possibilità di aggiungere e personalizzare i kernel: è possibile connettersi al cluster per installare i pacchetti kernel utilizzando `conda` e `pip`. Inoltre, l'installazione `pip` è supportata utilizzando i comandi del terminale all'interno delle celle di notebook. Nelle versioni precedenti, erano disponibili solo i kernel preinstallati (Python, PySpark, Spark e SparkR). Per ulteriori informazioni, consulta Installazione di kernel e librerie Python su un nodo primario del cluster.
- Possibilità di installare librerie Python: è possibile installare librerie Python sul nodo primario del cluster utilizzando `conda` e `pip`. Consigliamo l'uso di `conda`. Con le versioni precedenti, sono supportate solo le librerie con ambito notebook per PySpark.

Funzionalità EMR Notebooks supportate dalla versione del cluster

Versione di rilascio del cluster	Librerie con ambito notebook per PySpark	Installazione del kernel sul cluster	Installazione della libreria Python sul nodo primario
Precedente a 5.18.0	Notebook EMR non supportato		
5.18.0-5.25.0	No	No	No
5.26.0–5.29.0	Sì	No	No
5.30.0	Sì	Sì	Sì
6.0.0	No	No	No
5.32.0 e versioni successive e 6.2.0 e versioni successive	Sì	Sì	Sì

Limiti di notebook EMR collegati contemporaneamente

Quando crei un cluster che supporta i notebook, considera il tipo di istanza EC2 per il nodo primario del cluster. I limiti di memoria di questa istanza EC2 determinano il numero di notebook che possono essere pronti simultaneamente per l'esecuzione di codice e di query nel cluster.

Tipo di istanza EC2 del nodo primario	Numero di notebook EMR
*.medium	2
*.large	4
*.xlarge	8
*.2xlarge	16
*.4xlarge	24
*.8xlarge	24
*.16xlarge	24

Versioni Jupyter Notebook e Python

EMR Notebooks esegue Jupyter Notebook versione 6.0.2 e Python 3.6.5 a prescindere dalla versione Amazon EMR del cluster collegato.

Considerazioni relative alla sicurezza

Utilizzo di posizioni S3 crittografate

Se si specifica un percorso crittografato in Amazon S3 per archiviare i file del notebook, è necessario impostare Ruolo di servizio per EMR Notebooks come un utente chiave. Il ruolo di servizio predefinito è `EMR_Notebooks_DefaultRole`. Se utilizzi una chiave AWS KMS per la crittografia, consulta Utilizzo delle policy delle chiavi in AWS KMS nella Guida per gli sviluppatori di AWS Key Management Service e nell'Articolo di supporto per l'aggiunta di utenti della chiave.

Creazione di un notebook

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per

accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Per creare un notebook EMR, puoi utilizzare la vecchia console Amazon EMR. La creazione di notebook con AWS CLI o l'API di Amazon EMR non è supportata.

Creazione di un notebook EMR

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/.
2. Scegli Notebook, Crea un notebook.
3. Immetti un Nome notebook e una Descrizione notebook facoltativa.
4. Se disponi di un cluster attivo a cui desideri collegare il notebook, lascia selezionata l'impostazione Scegli un cluster esistente predefinita, fai clic su Scegli, seleziona un cluster dall'elenco, quindi fai clic su Scegli cluster. Per informazioni sui requisiti del cluster per Notebook EMR, consulta Considerazioni su quando utilizzare Notebook EMR.

   —oppure—

   Scegli Crea un cluster, immetti un Nome cluster e scegli le opzioni in base alle linee guida riportate di seguito. Il cluster viene creato nel VPC predefinito per l'account utilizzando istanze on demand.

Impostazione	Descrizione
Nome cluster	Nome descrittivo utilizzato per identificare il cluster.
Versione	Non si può modificare. L'impostazione predefinita è la versione di rilascio Amazon EMR più recente (5.36.1).
Applicazioni	Non si può modificare. Elenca le applicazioni installate nel cluster.
Istanza	Immetti il numero di istanze e seleziona il tipo di istanza EC2. Per il nodo primario

Impostazione	Descrizione
	viene utilizzata un'istanza. Le altre vengono utilizzate per i nodi principali. Il tipo di istanza determina il numero di notebook che possono si possono collegare simultane amente al cluster. Per ulteriori informazioni, consulta Limiti di notebook EMR collegati contemporaneamente.
Ruolo EMR	Lascia l'impostazione predefinita o scegli il collegamento per specificare un ruolo di servizio personalizzato per Amazon EMR. Per ulteriori informazioni, consulta Ruolo di servizio per Amazon EMR (ruolo EMR).
Profilo dell'istanza EC2	Lascia l'impostazione predefinita o scegli il collegamento per specificare un ruolo di servizio personalizzato per le istanze EC2. Per ulteriori informazioni, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2).
Coppia di chiavi EC2	Scegli una coppia di chiavi EC2 per poterti connettere alle istanze del cluster. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH.
Terminazione automatica	La terminazione automatica è supportata per Amazon EMR versione 5.30.0, 6.1.0 e successive. Seleziona la casella di spunta per abilitare la terminazione automatica, quindi specifica il tempo di inattività dopo il quale il cluster dovrebbe spegnersi in automatico. Per ulteriori informazioni, consulta Utilizzo di una policy di terminazione automatica.

5. Per Gruppo di sicurezza, scegli Usa gruppi di sicurezza predefiniti. In alternativa, fai clic su Scegli gruppi di sicurezza e seleziona i gruppi di sicurezza personalizzati disponibili nel VPC del cluster. Selezionane uno per l'istanza primaria e un altro per l'istanza client del notebook. Per ulteriori informazioni, consulta the section called “Gruppi di sicurezza per EMR Notebooks”.
6. Per Ruolo di servizio AWS, lascia l'impostazione predefinita o scegli un ruolo personalizzato dall'elenco. L'istanza client per il notebook utilizza questo ruolo. Per ulteriori informazioni, consulta Ruolo di servizio per EMR Notebooks.
7. Per Notebook location (Percorso notebook) scegli il percorso Amazon S3 in cui è salvato il file del notebook, oppure specifica un percorso. Se il bucket e la cartella non esistono, Amazon EMR li crea.

   Amazon EMR crea una cartella con Notebook ID (ID notebook) come nome della cartella e salva il notebook in un file denominato *`NotebookName`*`.ipynb`. Ad esempio, se si specifica il percorso Amazon S3 `s3://MyBucket/MyNotebooks` per un notebook denominato `MyFirstEMRManagedNotebook`, il file del notebook viene salvato in `s3://MyBucket/MyNotebooks/`*`NotebookID`*`/MyFirstEMRManagedNotebook.ipynb`.

   Se si specifica un percorso crittografato in Amazon S3, è necessario impostare Ruolo di servizio per EMR Notebooks come utente chiave. Il ruolo di servizio predefinito è `EMR_Notebooks_DefaultRole`. Se utilizzi una chiave AWS KMS per la crittografia, consulta Utilizzo delle policy delle chiavi in AWS KMS nella Guida per gli sviluppatori di AWS Key Management Service e nell'Articolo di supporto per l'aggiunta di utenti della chiave.
8. Facoltativamente, se hai aggiunto ad Amazon EMR un repository basato su Git che desideri associare a questo notebook, scegli Repository Git, fai clic su Scegli repository e seleziona un repository dall'elenco. Per ulteriori informazioni, consulta Associazione di repository basati su Git con Notebook EMR.
9. Facoltativamente, puoi scegliere Tag e quindi aggiungere ulteriori tag chiave-valore per il notebook.

> **Important**
>
> Ai fini dell'accesso vengono applicati un tag predefinito con la stringa Chiave impostata su `creatorUserID` e il valore impostato per l'ID dell'utente IAM. Consigliamo di non modificare o rimuovere questo tag perché può essere utilizzato per controllare l'accesso. Per ulteriori informazioni, consulta Utilizzo di tag di cluster e notebook con policy IAM per il controllo degli accessi.

10. Scegli Crea un notebook.

Lavorare con EMR Notebooks

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Dopo aver creato un notebook EMR, questo si avvia poco dopo. Lo Stato nell'elenco Notebook mostra Avvio in corso. È possibile aprire un notebook quando il suo stato è Pronto. Potrebbe essere necessario più tempo a un notebook per raggiungere lo stato Pronto se hai creato un cluster insieme a esso.

> **Tip**
>
> Aggiorna il browser oppure scegli l'icona di aggiornamento sopra l'elenco dei notebook per aggiornare lo stato.

Comprensione dello stato dei notebook

Un notebook EMR può avere il seguente Status (Stato) nell'elenco Notebooks (Notebook).

Stato	Significato
Pronto	Puoi aprire il notebook utilizzando l'editor di notebook. Quando un notebook è in stato Pronto, è possibile arrestarlo o eliminarlo. Per modificare i cluster, devi prima arrestare il notebook. Un notebook in stato Pronto che

Stato	Significato
	rimane inattivo a lungo viene arrestato in automatico.
Avvio in corso	Il notebook viene creato e collegato al cluster. Mentre un notebook è in fase di avvio, non è possibile aprire l'editor di notebook, arrestarlo, eliminarlo o modificare i cluster.
In attesa	Il notebook è stato creato ed è in attesa dell'integrazione con il cluster per il completamento. È possibile che il cluster stia ancora eseguendo il provisioning delle risorse o rispondendo ad altre richieste. Puoi aprire l'editor di notebook con il notebook in modalità locale. Il codice che si affida a processi di cluster non viene eseguito e dà esito negativo.
In arresto	Il notebook è in fase di arresto o il cluster a cui è collegato il notebook è in fase di terminazione. Mentre un notebook è in fase di arresto, non è possibile aprire l'editor di notebook, arrestarlo, eliminarlo o modificare i cluster.
Arrestato	Il notebook è stato arrestato. È possibile avviare il notebook sullo stesso cluster, purché quest'ultimo sia ancora in esecuzione. Puoi modificare i cluster ed eliminare il cluster.
Eliminazione in corso	Il cluster è in fase di rimozione dall'elenco dei cluster disponibili. Il file del notebook, *NotebookName* `.ipynb` , resta in Amazon S3 e continua ad accumulare addebiti di archiviazione applicabili.

Utilizzo dell'editor di notebook

Uno dei vantaggi offerti dall'utilizzo di un notebook EMR è che è possibile avviare il notebook in Jupyter o JupyterLab direttamente dalla console.

Con EMR Notebooks, l'editor di notebook a cui si accede dalla console di Amazon EMR è il noto editor di notebook Jupyter open source o JupyterLab. Poiché l'editor di notebook viene avviato dalla console di Amazon EMR, la configurazione dell'accesso è più efficiente di quanto non sia con un notebook ospitato in un cluster Amazon EMR. Non è necessario configurare un client dell'utente per avere l'accesso Web attraverso SSH, le regole per i gruppi di sicurezza e le configurazioni del proxy. Se un utente dispone di autorizzazioni sufficienti, basta semplicemente aprire l'editor di notebook nella console di Amazon EMR.

EMR Notebooks può essere aperto da un solo utente alla volta da Amazon EMR. Se un altro utente cerca di aprire un notebook EMR già aperto, si verifica un errore.

> **Important**
>
> Amazon EMR crea un URL prefirmato univoco per ogni sessione di editor notebook, valido solo per un breve periodo di tempo. Consigliamo di non condividere l'URL dell'editor di notebook. Ciò comporta un rischio per la sicurezza, perché i destinatari dell'URL adottano le autorizzazioni per modificare il notebook ed eseguire il codice del notebook per tutta la durata dell'URL. Se altri utenti necessitano di accedere a un notebook, fornisci loro le autorizzazioni attraverso le policy di autorizzazione e assicurati che il ruolo di servizio di Notebook EMR disponga dell'accesso al percorso di Amazon S3. Per ulteriori informazioni, consulta the section called “Sicurezza” e Ruolo di servizio per EMR Notebooks.

Apertura dell'editor di notebook per un notebook EMR

1. Seleziona un notebook con lo Stato su Pronto o In attesa dall'elenco Notebook.
2. Scegli Apri in JupyterLab o Apri in Jupyter.

   Si apre una nuova scheda del browser con l'editor di notebook JupyterLab o Jupyter.
3. Dal menu Kernel, scegli Cambia kernel, quindi seleziona il kernel per il tuo linguaggio di programmazione.

   Ora è tutto pronto per scrivere ed eseguire il codice dall'interno dell'editor di notebook.

Salvataggio dei contenuti di un notebook

Quando utilizzi l'editor di notebook, i contenuti delle celle di notebook e l'output vengono salvati in automatico nel file del notebook in Amazon S3 con cadenza periodica. Un notebook che non ha avuto modifiche dall'ultima volta che è una cella stata modificata mostra la dicitura (salvato in automatico) accanto al nome del notebook nell'editor. Se le modifiche non sono state ancora salvate, viene visualizzato modifiche non salvate.

È possibile salvare un notebook manualmente. Dal menu File, scegli Salva ed esegui il checkpoint o premi CTRL+S. In questo modo viene creato un file denominato *NotebookName*`.ipynb` in una cartella checkpoints all'interno della cartella del notebook in Amazon S3. Ad esempio, `s3://`*MyBucket*`/`*MyNotebookFolder*`/`*NotebookID*`/checkpoints/`*NotebookName*`.ipynb`. Solo i file di checkpoint più recenti vengono salvati in questa posizione.

Modifica dei cluster

È possibile modificare il cluster a cui è collegato un notebook EMR senza modificare i contenuti del notebook stesso. È possibile modificare i cluster solo per i notebook che hanno lo stato Arrestato.

Modifica del cluster di un notebook EMR

1. Se il notebook che desideri modificare è in esecuzione, selezionalo dall'elenco Notebook e scegli Arresta.
2. Quando lo stato del notebook è Arrestato, seleziona il notebook dall'elenco Notebook, quindi scegli Visualizza dettagli.
3. Seleziona Modifica cluster.
4. Se disponi di un cluster attivo che esegue Hadoop, Spark e Livy a cui desideri collegare il notebook, lascia l'impostazione predefinita e seleziona un cluster dall'elenco. Sono elencati solo i cluster che soddisfano i requisiti.

   —oppure—

   Seleziona Crea un cluster e quindi scegli le opzioni del cluster. Per ulteriori informazioni, consulta Requisiti del cluster.
5. Scegli un'opzione per i Gruppi di sicurezza, quindi scegli Modifica il cluster e avvia il notebook.

Eliminazione dei notebook e dei relativi file

Quando si elimina un notebook EMR mediante la console di Amazon EMR, il notebook viene eliminato dall'elenco dei notebook disponibili. Tuttavia, i file del notebook restano in Amazon S3 e continuano ad accumulare costi di archiviazione.

Eliminazione di un notebook e rimozione dei file associati

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/.
2. Scegli Notebook, seleziona il notebook dall'elenco e quindi scegli Visualizza dettagli.
3. Scegli l'icona della cartella accanto a Posizione del notebook e copia l'URL, che si trova nel pattern `s3://MyNotebookLocationPath/NotebookID/`.
4. Scegliere Delete (Elimina).

   Il notebook viene rimosso dall'elenco e i dettagli del notebook non possono più essere visualizzati.
5. Per istruzioni, consulta Come eliminare cartelle da un bucket S3 nella Guida per l'utente di Amazon Simple Storage Service. Passa al bucket e alla cartella della fase 3.

   —oppure—

   Se hai installato la AWS CLI, apri un prompt dei comandi e digita il comando al termine di questo paragrafo. Sostituisci il percorso Amazon S3 con quello copiato in precedenza. Verifica che la AWS CLI sia configurata con le chiavi di accesso di un utente con le autorizzazioni necessarie per eliminare il percorso Amazon S3. Per ulteriori informazioni, consulta Configurazione della AWS CLI nella Guida per l'utente di AWS Command Line Interface.

   ```
   aws s3 rm s3://MyNotebookLocationPath/NotebookID
   ```

Condivisione di file del notebook

Ogni notebook EMR viene salvato su Amazon S3 come file denominato *NotebookName*`.ipynb`. Finché un file del notebook è compatibile con la stessa versione del notebook Jupyter su cui è basato EMR Notebooks, è possibile aprire il notebook come un notebook EMR.

Il modo più semplice per aprire un file notebook da un altro utente è salvare il file *.ipynb da un altro utente nel file system locale, quindi utilizzare la funzionalità di caricamento negli editor Jupyter e JupyterLab.

È possibile utilizzare questo processo per l'uso di EMR Notebooks condivisi da altri utenti, notebook condivisi nella community Jupyter o per ripristinare un notebook che è stato eliminato dalla console quando si dispone ancora del file del notebook.

Uso di un altro file del notebook come base per un notebook EMR

1. Prima di procedere, chiudi l'editor di notebook per tutti notebook da utilizzare, quindi arresta il notebook se è un notebook EMR.
2. Crea un notebook EMR e assegnagli un nome. Il nome assegnato al notebook sarà il nome del file da sostituire. Il nuovo nome del file deve corrispondere esattamente a questo.
3. Annota il percorso scelto per il notebook in Amazon S3. Il file sostituito si trova in una cartella con un percorso e un nome di file simile al pattern seguente: `s3://`*`MyNotebookLocation`*`/`*`NotebookID`*`/`*`MyNotebookName`*`.ipynb`.
4. Arresta il notebook.
5. Sostituisci il vecchio file del notebook nel percorso Amazon S3 con il nuovo file, utilizzando esattamente lo stesso nome.

   Il seguente comando della AWS CLI per Amazon S3 sostituisce un file salvato su un computer locale denominato `SharedNotebook.ipynb` con un notebook EMR con il nome MyNotebook, un ID `e-12A3BCDEFJHIJKLMNO45PQRST` e creato con `MyBucket/MyNotebooksFolder` specificato in Amazon S3. Per informazioni sull'uso della console di Amazon S3 per copiare e sostituire i file, consulta Caricamento, download e gestione di oggetti nella Guida per l'utente di Amazon Simple Storage Service.

   ```
   aws s3 cp SharedNotebook.ipynb s3://MyBucket/
   MyNotebooksFolder/-12A3BCDEFJHIJKLMNO45PQRST/MyNotebook.ipynb
   ```

Comandi di esempio per eseguire Notebook EMR a livello di programmazione

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Panoramica

È possibile eseguire notebook EMR con API di esecuzione da uno script o dalla riga di comando. Quando avvii, arresti, elenchi e descrivi le esecuzioni di notebook EMR all'esterno della console AWS, puoi controllare a livello di programmazione un notebook EMR. Puoi trasmettere diversi valori di parametro a un notebook con una cella di notebook parametrizzata. Questa opzione elimina la necessità di creare una copia del notebook per ogni nuovo set di valori di parametro. Per ulteriori informazioni, consulta le operazioni dell'API di Amazon EMR.

Puoi pianificare o eseguire in batch le esecuzioni di notebook EMR con Eventi Amazon CloudWatch e AWS Lambda. Per ulteriori informazioni, consulta Uso di AWS Lambda con Eventi Amazon CloudWatch.

Autorizzazioni di ruolo per l'esecuzione a livello di programmazione

Per utilizzare l'esecuzione a livello di programmazione con Notebook EMR, è necessario configurare le autorizzazioni utente con le policy seguenti:

```
{
    "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"
        }
    ]
}
```

Quando esegui Notebook EMR a livello di programmazione su un cluster Notebook EMR, devi aggiungere queste autorizzazioni supplementari:

```
{
    "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"
            ]
        }
    ]
}
```

Limitazioni relative all'esecuzione a livello di programmazione

- Sono supportate un massimo di 100 esecuzioni simultanee per Regione AWS per ogni account.
- Un'esecuzione viene terminata se dura più di 30 giorni.
- L'esecuzione programmatica dei notebook non è supportata con le applicazioni interattive di Amazon EMR serverless.

Esempi di esecuzione di notebook EMR a livello di programmazione

In questa sezione vengono forniti diversi esempi di esecuzione di notebook EMR a livello di programmazione con la AWS CLI, l'SDK Boto3 (Python) e Ruby:

- Esempi di comandi della CLI per l'esecuzione di notebook
- Esempi Python di esecuzione di notebook
- Esempi Ruby di esecuzione di notebook

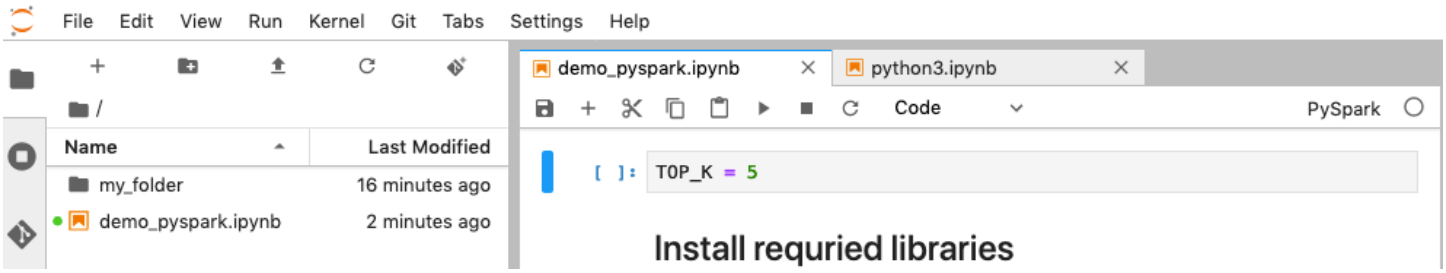
Puoi anche eseguire notebook parametrizzati nell'ambito dei flussi di lavoro pianificati con uno strumento di orchestrazione come Apache Airflow o Amazon Managed Workflows per Apache Airflow (MWAA). Per ulteriori informazioni, consulta Orchestrazione dei processi di analisi su Notebook EMR con MWAA nel blog AWS Big Data.

Esempi di comandi della CLI per l'esecuzione di notebook

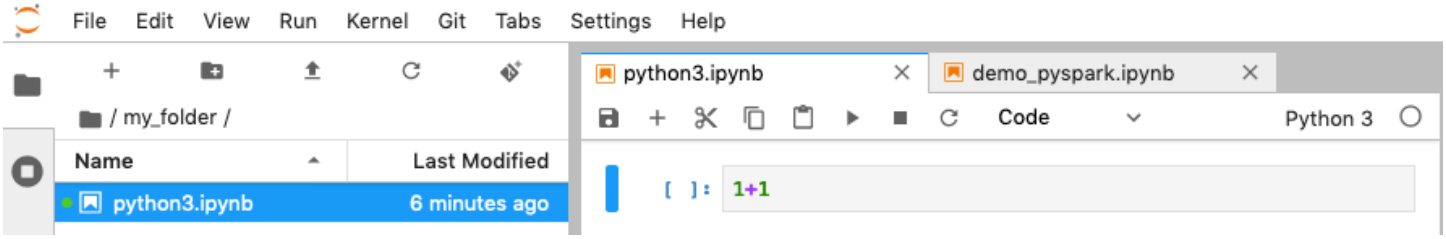
> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

L'esempio seguente si basa sul notebook dimostrativo della console Notebook EMR. Per individuare il notebook, utilizza il percorso di file relativo alla home directory. In questo esempio, ci sono due file di notebook che puoi eseguire: `demo_pyspark.ipynb` e `my_folder/python3.ipynb`.

Il percorso relativo del file `demo_pyspark.ipynb` è `demo_pyspark.ipynb`, mostrato di seguito.



Il percorso relativo per `python3.ipynb` è `my_folder/python3.ipynb`, mostrato di seguito.



Per ulteriori informazioni sulle operazioni `NotebookExecution` dell'API di Amazon EMR, consulta Operazioni dell'API di Amazon EMR.

Esecuzione di un notebook

Puoi utilizzare la AWS CLI per eseguire il notebook con l'operazione `start-notebook-execution`, come dimostra l'esempio seguente.

Example - Esecuzione di un notebook EMR in un Workspace EMR Studio con un cluster Amazon EMR (in esecuzione su 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 - Esecuzione di un notebook EMR in un Workspace EMR Studio con un cluster Notebook 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
```

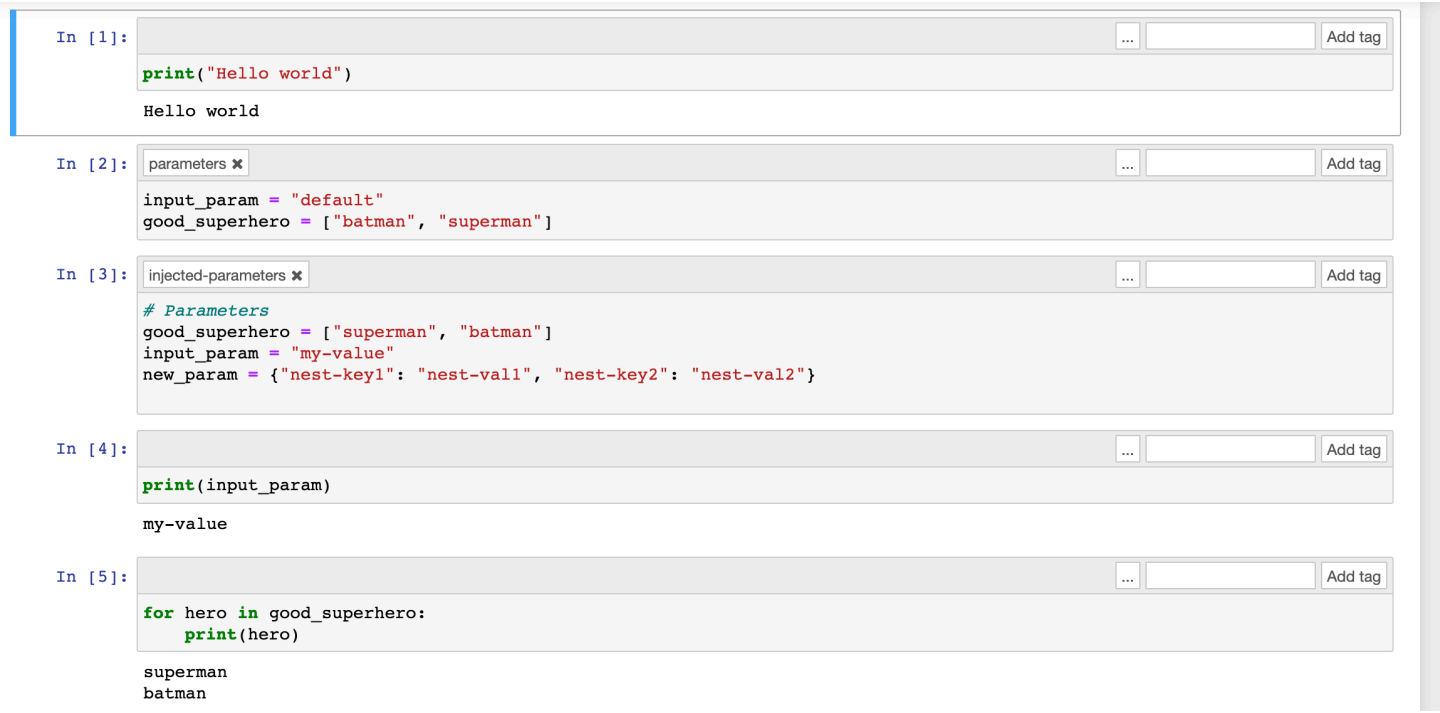
Example - Esecuzione di un notebook EMR specificando la sua posizione Amazon S3

```
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"}'
```

Output del notebook

Ecco l'output di un notebook campione. La cella 3 mostra i valori di parametro appena inseriti.



Descrizione di un notebook

Puoi utilizzare l'operazione `describe-notebook-execution` per accedere alle informazioni sull'esecuzione di un notebook specifico.

```
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": []
    }
}
```

Arrestare un notebook

Se il notebook sta eseguendo un'esecuzione che desideri arrestare, puoi farlo con il 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": []
    }
}
```

Elencazione delle esecuzioni di un notebook per ora di inizio

Puoi trasmettere un parametro `--from` a `list-notebook-executions` per elencare le esecuzioni del notebook in base all'ora di inizio.

```
# 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
        }
    ]
}
```

Elencazione delle esecuzioni di un notebook per ora di inizio e stato

Il comando `list-notebook-executions` può anche filtrare i risultati in base a un parametro `--status`.

```
# 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
        }
    ]
}
```

Esempi Python di esecuzione di notebook

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

L'esempio di codice seguente è un file SDK per Python (Boto3) denominato `demo.py` che mostra le API di esecuzione di notebook.

Per ulteriori informazioni sulle operazioni `NotebookExecution` dell'API di Amazon EMR, consulta Operazioni dell'API di 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")
```

Quello mostrato è l'output dall'esecuzione di `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}}
```

Esempi Ruby di esecuzione di notebook

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Di seguito sono riportati esempi di codice Ruby che dimostrano l'uso dell'API di esecuzione di notebook.

```
# prepare an Amazon EMR client

emr = Aws::EMR::Client.new(
  region: 'us-east-1',
  access_key_id: 'AKIA...JKPKA',
  secret_access_key: 'rLMeu...vU0OLrAC1',
)
```

Avvio dell'esecuzione di notebook e ottenimento dell'ID di esecuzione

In questo esempio, l'editor Amazon S3 e il notebook EMR sono `s3://mybucket/notebooks/e-EA8VGAA429FEQTC8HC9ZHWISK/test.ipynb`.

Per ulteriori informazioni sulle operazioni `NotebookExecution` dell'API di Amazon EMR, consulta Operazioni dell'API di 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
```

Descrizione dell'esecuzione di notebook e stampa dei dettagli

```
describe_resp = emr.describe_notebook_execution({
    notebook_execution_id: notebook_execution_id
})
puts describe_resp.notebook_execution
```

L'output dei comandi precedenti sarà il seguente.

```
{
: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=>[]
}
```

Filtri del notebook

```
"EditorId": "e-XXXX",           [Optional]
"From" : "1593400000.000",    [Optional]
"To" :
```

Arresto dell'esecuzione di notebook

```
stop_resp = emr.stop_notebook_execution({
    notebook_execution_id: notebook_execution_id
```

```
})
```

Abilitazione della rappresentazione utente per monitorare l'attività dell'utente e dei processi Spark

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Notebook EMR ti consente di configurare la rappresentazione utente in un cluster Spark. Questa funzionalità consente di monitorare le attività dei processi avviati dall'interno dell'editor di notebook. Inoltre, Notebook EMR dispone di un widget di notebook Jupyter integrato per visualizzare i dettagli del processo Spark insieme all'output delle query nell'editor di notebook. Il widget è disponibile per impostazione predefinita e non richiede alcuna configurazione speciale. Tuttavia, per visualizzare i server della cronologia, il client deve essere configurato per visualizzare le interfacce Web Amazon EMR ospitate nel nodo primario.

Impostazione della rappresentazione utente Spark

Per impostazione predefinita, i processi Spark inviati dagli utenti mediante l'editor di notebook risultano provenienti da un'identità utente `livy` indistinta. È possibile configurare la rappresentazione utente per il cluster in modo che tali processi siano associati piuttosto all'identità utente che ha eseguito il codice. Le directory utente HDFS nel nodo primario vengono create per ciascuna identità utente che esegue codice nel notebook. Ad esempio, se l'utente `NbUser1` esegue il codice dall'editor di notebook, puoi collegarti al nodo primario e vedere che `hadoop fs -ls /user` mostra la directory `/user/user_NbUser1`.

Puoi abilitare questa funzionalità impostando le proprietà nelle classificazioni di configurazione `core-site` e `livy-conf`. Questa funzionalità non è disponibile per impostazione predefinita quando Amazon EMR crea un cluster insieme a un notebook. Per ulteriori informazioni su come utilizzare

le classificazioni di configurazione per personalizzare le applicazioni, consulta Configurazione delle applicazioni nella Guida ai rilasci di Amazon EMR.

Utilizza i seguenti valori e classificazioni di configurazione per abilitare la rappresentazione degli utenti per EMR Notebooks:

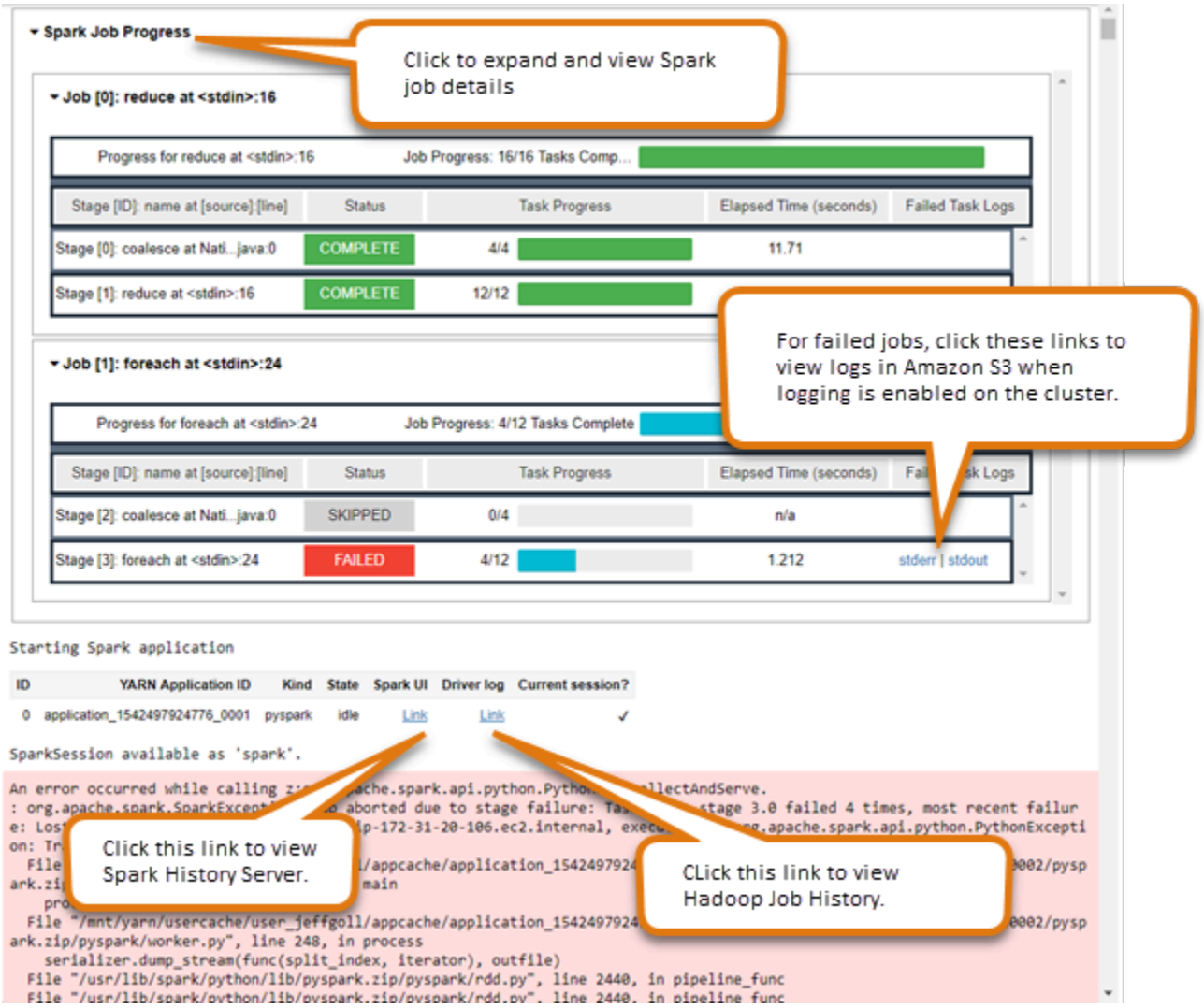
```
[
    {
        "Classification": "core-site",
        "Properties": {
          "hadoop.proxyuser.livy.groups": "*",
          "hadoop.proxyuser.livy.hosts": "*"
        }
    },
    {
        "Classification": "livy-conf",
        "Properties": {
          "livy.impersonation.enabled": "true"
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

Utilizzo del widget di monitoraggio dei processi Spark

Quando esegui codice nell'editor di notebook che esegue i processi Spark nel cluster EMR, l'output include un widget di notebook Jupyter per il monitoraggio dei processi Spark. Il widget fornisce i dettagli del processo e collegamenti utili alla pagina dei server della cronologia di Spark e alla pagina della cronologia di Hadoop, insieme a collegamenti ai log dei processi in Amazon S3 per tutti i processi non riuscito.

Per visualizzare le pagine dei server della cronologia nel nodo primario del cluster, devi configurare un client SSH e un proxy adeguati. Per ulteriori informazioni, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR. Per visualizzare i log nel cluster Amazon S3 la registrazione deve essere abilitata, il che corrisponde all'impostazione predefinita per i nuovi cluster. Per ulteriori informazioni, consulta Visualizzazione dei file di log archiviati in Amazon S3.

Di seguito è riportato un esempio del monitoraggio dei processi Spark.



Sicurezza e controllo dell'accesso a EMR Notebooks

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Sono disponibili alcune funzionalità che aiutano a personalizzare l'assetto di sicurezza di Notebook EMR. Ciò contribuisce a garantire che solo gli utenti autorizzati possano accedere a un notebook EMR, lavorare con i notebook e utilizzare l'editor di notebook per l'esecuzione del codice nel cluster. Queste funzionalità si accompagnano alle funzionalità di sicurezza disponibili per Amazon EMR e per i cluster Amazon EMR. Per ulteriori informazioni, consulta Sicurezza in Amazon EMR.

- È possibile utilizzare le istruzioni delle policy di AWS Identity and Access Management con i tag dei notebook per limitare l'accesso. Per ulteriori informazioni, consulta Funzionamento di Amazon EMR con IAM e Dichiarazioni di policy basate su identità di esempio per EMR Notebooks.
- I gruppi di sicurezza Amazon EC2 fungono da firewall virtuali che controllano il traffico di rete tra l'istanza primaria del cluster e l'editor di notebook. Puoi utilizzare le impostazioni predefinite o personalizzare questi gruppi di sicurezza. Per ulteriori informazioni, consulta Specifica dei gruppi di sicurezza EC2 per EMR Notebooks.
- Specifica un ruolo di servizio AWS che determina le autorizzazioni di cui dispone un notebook EMR durante l'interazione con altri servizi AWS. Per ulteriori informazioni, consulta Ruolo di servizio per EMR Notebooks.

Installazione e uso di kernel e librerie

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Ogni notebook EMR viene fornito con un set di librerie e kernel preinstallati. Puoi installare librerie e kernel aggiuntivi in un cluster EMR se il cluster dispone dell'accesso al repository in cui si trovano i kernel e le librerie. Ad esempio, per i cluster in sottoreti private, potrebbe essere necessario configurare Network Address Translation (NAT) e fornire un percorso per il cluster per accedere al repository PyPI pubblico per installare una libreria. Per ulteriori informazioni sulla configurazione dell'accesso esterno per diverse configurazioni di rete, consulta Scenari ed esempi nella Guida per l'utente di Amazon VPC.

Le applicazioni EMR Serverless vengono fornite con le seguenti librerie preinstallate per Python e PySpark:

- Librerie Python – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, scipy
- Librerie PySpark – ggplot, matplotlib, numpy, pandas, plotly, bokeh, scikit-learn, scipy, scipy

Installazione di kernel e librerie Python su un nodo primario del cluster

Con Amazon EMR versione 5.30.0 e successive, esclusa la versione 6.0.0, puoi installare librerie e kernel Python aggiuntivi sul nodo primario del cluster. Dopo l'installazione, questi kernel e librerie sono disponibili a ogni utente che esegue un notebook EMR collegato al cluster. Le librerie Python installate in questo modo sono disponibili solo per i processi in esecuzione sul nodo primario. Le librerie non sono installate nei nodi principali o attività e non sono disponibili per gli executor in esecuzione su tali nodi.

> **Note**
>
> Per le versioni Amazon EMR 5.30.1, 5.31.0 e 6.1.0, devi eseguire ulteriori passaggi per installare kernel e librerie nel nodo primario di un cluster.
> Per abilitare questa funzione, procedi come segue:
>
> 1. Assicurati che le policy di autorizzazione associate al ruolo di servizio per EMR Notebooks consentano l'operazione seguente:
>
>    ```
>    elasticmapreduce:ListSteps
>    ```
>
>   Per ulteriori informazioni, consulta il Ruolo di servizio per EMR Notebooks.
> 2. Utilizza la AWS CLI per eseguire un passaggio sul cluster che imposti EMR Notebooks come mostrato nell'esempio seguente. È necessario utilizzare il nome della fase `EMRNotebooksSetup`. Sostituisci *`us-east-1`* con la Regione in cui risiede il cluster. Per ulteriori informazioni, consulta Aggiunta di fasi a un cluster con 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"]
>    ```

Puoi installare kernel e librerie eseguendo `pip` o `conda` nella directory `/emr/notebook-env/bin` sul nodo primario.

Example - Installazione di librerie Python

Dal kernel Python3, esegui il `%pip` magic come un comando dall'interno di una cella del notebook per installare librerie Python.

```
%pip install pmdarima
```

Potrebbe essere necessario riavviare il kernel per utilizzare i pacchetti aggiornati. Puoi anche utilizzare `%%sh` magic Spark per invocare `pip`.

```
%%sh
/emr/notebook-env/bin/pip install -U matplotlib
/emr/notebook-env/bin/pip install -U pmdarima
```

Quando utilizzi un kernel PySpark, puoi installare librerie sul cluster utilizzando comandi `pip` o utilizzare librerie con ambito notebook dall'interno di un notebook PySpark.

Per eseguire comandi `pip` sul cluster dal terminale, prima esegui la connessione al nodo primario utilizzando SSH, come illustrato nei seguenti comandi.

```
sudo pip3 install -U matplotlib
sudo pip3 install -U pmdarima
```

In alternativa, puoi utilizzare librerie con ambito notebook. Con le librerie con ambito notebook, l'installazione della libreria è limitata all'ambito della sessione e si verifica su tutti gli executor Spark. Per ulteriori informazioni, consulta Utilizzo di librerie con ambito notebook.

Se vuoi creare un pacchetto di più librerie Python all'interno di un kernel PySpark, puoi anche creare un ambiente virtuale Python isolato. Per esempi di utilizzo, consulta Utilizzo di Virtualenv.

Per creare un ambiente virtuale Python in una sessione, utilizza la proprietà Spark `spark.yarn.dist.archives` dal comando `%%configure` magic nella prima cella di un notebook, come illustrato nell'esempio seguente.

```
%%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"
   }
}
```

Puoi creare allo stesso modo un ambiente executor 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"
   }
}
```

Puoi utilizzare anche `conda` per installare librerie Python. Per utilizzare `conda` non è necessario l'accesso sudo. Devi connetterti al nodo primario con SSH e quindi eseguire `conda` dal terminale. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH.

Example - Installazione di kernel

L'esempio seguente illustra l'installazione del kernel Kotlin utilizzando un comando del terminale mentre è connesso al nodo primario di un cluster:

```
sudo /emr/notebook-env/bin/conda install kotlin-jupyter-kernel -c jetbrains
```

> **Note**
>
> Queste istruzioni non prevedono l'installazione delle dipendenze del kernel. Se il kernel ha dipendenze di terze parti, potrebbe essere necessario eseguire ulteriori passaggi di configurazione prima di poterlo utilizzare con il notebook.

Considerazioni e limitazioni relative alle librerie con ambito notebook

Quando utilizzi le librerie con ambito notebook, tieni presente quanto segue:

- Le librerie con ambito notebook sono disponibili con i cluster creati utilizzando le versioni 5.26.0 e successive di Amazon EMR.
- Le librerie con ambito notebook sono pensate per l'uso solo con il kernel PySpark.
- Qualsiasi utente può installare ulteriori librerie con ambito notebook dall'interno di una cella di notebook. Queste librerie sono disponibili solo per tale utente del notebook durante una singola sessione di notebook. Se altri utenti richiedono le stesse librerie o lo stesso utente richiede le stesse librerie in una sessione diversa, è necessario reinstallare la libreria.
- È possibile disinstallare solo le librerie che sono state installate con l'API `install_pypi_package`. Non è possibile disinstallare le librerie preinstallate nel cluster.
- Se nel cluster sono installate versioni diverse delle stesse librerie come librerie con ambito notebook, la versione della libreria con ambito notebook sostituisce la versione della libreria del cluster.

Lavorare con le librerie con ambito notebook

Per installare le librerie, il cluster Amazon EMR deve avere accesso al repository PyPI in cui queste si trovano.

Gli esempi riportati di seguito illustrano semplici comandi per elencare, installare e disinstallare le librerie da una cella di notebook utilizzando il kernel e le API di PySpark. Per ulteriori esempi, consulta il post Install Python libraries on a running cluster with EMR Notebooks (Installazione di librerie Python su un cluster in esecuzione con EMR Notebooks) del blog sui Big Data AWS.

Example - Elenco delle librerie correnti

Il comando riportato di seguito elenca i pacchetti Python disponibili per la sessione Spark corrente del notebook. Questo comando elenca le librerie installate nel cluster e le librerie con ambito notebook.

```
sc.list_packages()
```

Example - Installazione della libreria Celery

Il comando seguente installa la libreria Celery come libreria con ambito notebook.

```
sc.install_pypi_package("celery")
```

Dopo l'installazione della libreria, il comando seguente conferma che la libreria è disponibile sul driver e sugli executor Spark.

```
import celery
sc.range(1,10000,1,100).map(lambda x: celery.__version__).collect()
```

Example - Installazione della libreria Arrow, specifica della versione e del repository

Il comando seguente installa la libreria Arrow come libreria con ambito notebook, con una specifica dell'URL del repository e della versione della libreria.

```
sc.install_pypi_package("arrow==0.14.0", "https://pypi.org/simple")
```

Example - Disinstallazione di una libreria

Il comando seguente disinstalla la libreria Arrow rimuovendola come libreria con ambito notebook dalla sessione corrente.

```
sc.uninstall_package("arrow")
```

Associazione di repository basati su Git con Notebook EMR

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Puoi associare repository basati su Git ai notebook Amazon EMR per salvare i notebook in un ambiente con controllo della versione. È possibile associare fino a tre repository a un notebook. Sono supportati i seguenti servizi basati su Git:

- AWS CodeCommit
- GitHub
- Bitbucket
- GitLab

L'associazione di repository basati su GIT al notebook presenta i seguenti vantaggi.

- Controllo della versione: è possibile registrare le modifiche al codice in un sistema di controllo della versione in modo da poter rivedere la cronologia delle modifiche e annullarle selettivamente.
- Collaborazione: i colleghi che lavorano in notebook diversi possono condividere il codice tramite i repository basati su Git. I notebook possono clonare o unire il codice proveniente da repository remoti e inviare nuovamente le modifiche agli stessi.
- Riutilizzo del codice: molti notebook Jupyter che presentano tecniche di analisi dei dati o di machine learning sono disponibili in repository ospitati pubblicamente, ad esempio su GitHub. Puoi associare i notebook a un repository per riutilizzare i notebook Jupyter contenuti in un repository.

Per utilizzare i repository basati su Git con EMR Notebooks, aggiungi i repository come risorse nella console di Amazon EMR, associa le credenziali per i repository che richiedono l'autenticazione e collegali ai notebook. È possibile visualizzare un elenco di repository che sono archiviati nell'account e informazioni dettagliate su ogni repository nella console di Amazon EMR. Puoi associare un repository basato su Git esistente a un notebook quando lo crei.

Argomenti

- Prerequisiti e considerazioni
- Aggiunta di un repository basato su Git ad Amazon EMR
- Aggiornamento o eliminazione di un repository basato su Git
- Collegamento o scollegamento di un repository basato su Git
- Creazione di un nuovo Notebook con un repository Git associato
- Uso di repository Git in un Notebook

Prerequisiti e considerazioni

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Quando pianifichi di integrare un repository basato su Git con Notebook EMR, considera quanto segue.

AWS CodeCommit

Se utilizzi un repository CodeCommit, usa le credenziali Git e HTTPS con CodeCommit. Le chiavi SSH e HTTPS con l'assistente credenziali AWS CLI non sono supportate. Inoltre, CodeCommit non supporta i token di accesso personale (PAT). Per ulteriori informazioni, consulta Utilizzo di IAM con CodeCommit: credenziali Git, chiavi SSH e chiavi di accesso AWS nella Guida per l'utente IAM e Configurazione degli utenti HTTPS con le credenziali Git nella Guida per l'utente di AWS CodeCommit.

Considerazioni su accesso e autorizzazione

Prima di associare un repository al notebook, assicurati che il cluster, il ruolo IAM per EMR Notebooks e i gruppi di sicurezza dispongano delle impostazioni e delle autorizzazioni corrette. È inoltre possibile configurare i repository basati su Git ospitati in una rete privata seguendo le istruzioni riportate in Configurazione di un repository Git ospitato privatamente per EMR Notebooks.

- Accesso a Internet del cluster: l'interfaccia di rete avviata dispone di solo un indirizzo IP privato. Ciò significa che il cluster a cui il notebook si connette deve trovarsi in una sottorete privata con un gateway Network Address Translation (NAT) o deve essere in grado di accedere a Internet attraverso un gateway privato virtuale. Per ulteriori informazioni, consulta la sezione relativa alle Opzioni di Amazon VPC.

I gruppi di sicurezza del notebook devono includere una regola in uscita che consenta al notebook di instradare il traffico a Internet dal cluster. È consigliabile creare gruppi di sicurezza personali. Per ulteriori informazioni, consulta Specifica dei gruppi di sicurezza EC2 per EMR Notebooks.

> **Important**
>
> Se l'interfaccia di rete viene avviata in una sottorete pubblica, non sarà in grado di comunicare con Internet tramite un gateway Internet (IGW).

- Autorizzazioni per AWS Secrets Manager: se si utilizza Secrets Manager per archiviare i segreti usati per accedere a un repository, al the section called “Ruolo EMR Notebooks” deve essere collegata una policy di autorizzazione che consenta l'operazione `secretsmanager:GetSecretValue`.

Configurazione di un repository Git ospitato privatamente per EMR Notebooks

Utilizza le istruzioni seguenti per configurare repository ospitati privatamente per EMR Notebooks. È necessario fornire un file di configurazione con informazioni sui server DNS e Git. Amazon EMR utilizza queste informazioni per configurare EMR Notebooks in grado di instradare il traffico ai repository ospitati privatamente.

Prerequisiti

Prima di configurare un repository Git ospitato privatamente per EMR Notebooks, devi disporre di quanto segue:

- Un percorso Amazon S3 Control in cui verranno salvati i file del notebook EMR.

Configurazione di uno o più repository Git ospitati privatamente per EMR Notebooks

1. Crea un file di configurazione utilizzando il modello fornito. Includi i seguenti valori per ogni server Git che desideri specificare nella configurazione:

   - **`DnsServerIpV4`**: l'indirizzo IPv4 del server DNS. Se si forniscono valori per `DnsServerIpV4` e `GitServerIpV4List`, il valore per `DnsServerIpV4` ha la precedenza e verrà utilizzato per risolvere il `GitServerDnsName`.

> **Note**
>
> Per utilizzare repository Git ospitati privatamente, il server DNS deve consentire l'accesso in ingresso da EMR Notebooks. Si consiglia di proteggere il server DNS da altri accessi non autorizzati.

- **`GitServerDnsName`**: il nome DNS del server Git. Ad esempio `"git.example.com"`.
- **`GitServerIpV4List`**: un elenco di indirizzi IPv4 che appartengono ai server 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. Salva il file di configurazione come `configuration.json`.
3. Carica il file di configurazione nel percorso di archiviazione Amazon S3 designato in una cartella denominata `life-cycle-configuration`. Ad esempio, se il percorso S3 predefinito è `s3://DOC-EXAMPLE-BUCKET/notebooks`, il file di configurazione dovrebbe trovarsi in `s3://DOC-EXAMPLE-BUCKET/notebooks/`**`life-cycle-configuration/configuration.json`**.

> **Important**
>
> Si consiglia fortemente di limitare l'accesso alla cartella `life-cycle-configuration` solo agli amministratori di EMR Notebooks e al ruolo di servizio per EMR Notebooks. Dovresti inoltre proteggere `configuration.json` contro l'accesso non autorizzato. Per istruzioni, consulta Controllo dell'accesso a un bucket con policy utente o Best practice di sicurezza per Amazon S3.

Per istruzioni sul caricamento, consulta Creazione di una cartella e Caricamento degli oggetti nella Guida per l'utente di Amazon Simple Storage Service.

Aggiunta di un repository basato su Git ad Amazon EMR

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Fai riferimento alle sezioni seguenti per avere istruzioni sull'aggiunta di un repository basato su Git a un notebook EMR nella vecchia console o a un Workspace EMR Studio nella nuova console.

New console

Poiché i notebook EMR sono Workspace EMR Studio nella nuova console, puoi seguire le istruzioni riportate in Collegamento di repository basati su Git a un WorkSpace EMR Studio per associare fino a tre repository Git al tuo Workspace.

In alternativa, puoi utilizzare l'estensione JupyterLab Git. Scegli l'icona Git dalla barra laterale sinistra del notebook Jupyterlab per accedere all'estensione. Per avere informazioni sull'estensione, consulta il repository GitHub jupyterlab-git.

Per associare un repository Git a un Workspace, l'amministratore dello Studio deve adottare misure per configurare lo Studio in modo da consentire il collegamento del repository Git. Per ulteriori informazioni, consulta Definizione di accesso e autorizzazioni per i repository basati su Git.

Old console

Aggiunta di un repository basato su Git come risorsa nell'account Amazon EMR con la vecchia console

1. Apri la vecchia console Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce.
2. Scegli Repository Git e quindi Aggiungi repository.
3. Per Repository name (Nome repository), immetti un nome da utilizzare per il repository in Amazon EMR.

   I nomi possono contenere solo caratteri alfanumerici, trattini (-) o trattini bassi (_).
4. Per URL repository Git), immetti l'URL del repository. Quando si utilizza un repository CodeCommit, questo è l'URL che viene copiato quando si sceglie Clone URL (Clona URL) e in seguito Clone HTTPS (Clona HTTPS), ad esempio `https://git-codecommit.us-west-2.amazonaws.com/v1/repos/MyCodeCommitRepoName`.
5. In Ramo, immetti un nome di ramo.
6. In Credenziali Git, scegli le opzioni in base alle seguenti linee guida. È possibile utilizzare un nome utente e una password Git o un token di accesso personale (PAT) per autenticarsi nel repository. EMR Notebooks accede alle credenziali Git utilizzando i segreti archiviati in Secrets Manager.

   > **Note**
   >
   > Se utilizzi un repository GitHub, ti consigliamo di utilizzare un token di accesso personale (PAT) per l'autenticazione. A partire dal 13 agosto 2021, GitHub non accetta più password durante l'autenticazione delle operazioni Git. Per ulteriori informazioni, consulta il post Token authentication requirements for Git operations (Requisiti di autenticazione token per operazioni Git) nel Blog di GitHub.

Opzione	Descrizione
Usa un segreto AWS esistente	Scegli questa opzione se le credenziali sono già state salvate come segreto in Secrets Manager, quindi seleziona il nome segreto dall'elenco. Se selezioni un segreto associato a un nome utente e una password Git, il segreto deve essere nel formato `{"gitUsername": " MyUserName ", "gitPassword": " MyPassword "}`.

Opzione	Descrizione
Crea un nuovo segreto	Scegli questa opzione per associare le credenziali Git esistenti a un nuovo segreto creato in Secrets Manager. Esegui una delle seguenti operazioni in base alle credenziali Git utilizzate per il repository. Se utilizzi un nome utente e una password Git per accedere al repository, seleziona Nome utente e password, immetti il Nome segreto da utilizzare in Secrets Manager e quindi il Nome utente e la Password da associare al segreto. –OPPURE– Se utilizzi un token di accesso personale per accedere al repository, seleziona Personal access token (PAT) (Token di accesso personale (PAT)), immetti il Secret name (Nome segreto) da utilizzare in Secrets Manager e, in seguito, immetti il tuo personal access token (token di accesso personale). Per ulteriori informazioni, consulta Creazione di un token di accesso personale per la riga di comando per GitHub e Token di accesso personale per Bitbucket. I repository CodeCommit non supportano questa opzione.
Utilizza un repository pubblico senza credenziali	Scegli questa opzione per accedere a un repository pubblico.

7. Scegli Aggiungi repository.

Aggiornamento o eliminazione di un repository basato su Git

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Fai riferimento alle sezioni seguenti per avere istruzioni sull'eliminazione di un repository basato su Git da un notebook EMR nella vecchia console o da un Workspace EMR Studio nella nuova console.

New console

Poiché i Notebook EMR sono Workspace EMR Studio nella nuova console, puoi fare riferimento a Collegamento di repository basati su Git a un WorkSpace EMR Studio per avere ulteriori informazioni sull'uso dei repository Git nel tuo Workspace. Al momento, tuttavia, non puoi eliminare i repository Git dai Workspace.

Old console

Aggiornamento di un repository basato su Git nella vecchia console

1. Nella pagina Repository Git scegli il repository che desideri aggiornare.
2. Nella pagina del repository, scegli Modifica repository.
3. Aggiorna Credenziali Git nella pagina del repository.

Eliminazione di un repository Git nella vecchia console

1. Nella pagina Repository Git, scegli il repository che desideri eliminare.
2. Nella pagina del repository, scegli tutti i notebook attualmente collegati al repository. Scegli Scollega notebook.
3. Nella pagina del repository, scegli Elimina.

> **Note**
>
> Per eliminare il repository Git locale da Amazon EMR, è necessario prima scollegare tutti i notebook dal repository. Per ulteriori informazioni, consulta Collegamento o scollegamento di un repository basato su Git. L'eliminazione di un repository Git non eliminerà alcun segreto creato per il repository. Puoi eliminare il segreto in AWS Secrets Manager.

Collegamento o scollegamento di un repository basato su Git

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Utilizza le seguenti fasi per collegare o scollegare un repository basato su Git a/da un notebook EMR nella vecchia console o a/da un Workspace EMR Studio nella nuova console.

New console

Poiché i Notebook EMR sono Workspace EMR Studio nella nuova console, puoi fare riferimento a Collegamento di repository basati su Git a un WorkSpace EMR Studio per avere ulteriori informazioni sull'uso dei repository Git nel tuo Workspace. Al momento, tuttavia, non puoi eliminare i repository Git dai Workspace.

Old console

Collegamento di un repository basato su Git a un notebook EMR

Il repository può essere collegato a un notebook una volta che il notebook è Pronto.

1. Dall'elenco Notebook, scegli il notebook che desideri aggiornare.
2. Nella sezione Repository Git nella pagina Notebook, scegli Collega nuovo repository.

3. Nell'elenco dei repository della finestra Collega il repository Git al notebook, seleziona uno o più repository che desideri collegare al notebook, quindi scegli Collega repository.

Oppure

1. Nella pagina Repository Git, scegli il repository che desideri collegare al notebook.
2. Nell'elenco Notebook EMR, scegli Collega nuovo notebook per collegare questo repository a un notebook esistente.

Scollegamento di un repository Git da un notebook EMR

1. Dall'elenco Notebook, scegli il notebook che desideri aggiornare.
2. Nell'elenco Repository Git, seleziona il repository che desideri scollegare dal notebook, quindi scegli Scollega repository.

Oppure

1. Nella pagina Repository Git, scegli il repository che desideri aggiornare.
2. Nell'elenco Notebook EMR, seleziona il notebook che desideri scollegare dal repository, quindi scegli Scollega notebook.

> **Note**
>
> Il collegamento di un repository Git a un notebook esegue la clonazione del repository remoto sul notebook Jupyter locale. Lo scollegamento del repository Git da un notebook disconnette solo il notebook dal repository remoto ma non elimina il repository Git locale.

Informazioni sullo stato del repository

Un repository Git può presentare uno dei seguenti stati nell'elenco dei repository. Per ulteriori informazioni sul collegamento dei EMR Notebooks con i repository Git, consulta Collegamento o scollegamento di un repository basato su Git.

Stato	Significato
Collegamento	Il repository Git è in fase di collegamento al notebook. Mentre il repository è nello stato Collegamento, non è possibile arrestarlo.
Collegato	Il repository Git è collegato al notebook. Quando è nello stato Collegato, il repository è collegato al repository remoto.
Collegamento non riuscito	Il repository Git non è riuscito a collegarsi al notebook. È possibile riprovare a collegarlo.
Scollegamento	Il repository Git è in fase di scollegamento dal notebook. Mentre il repository è nello stato Scollegamento, non è possibile arrestarlo. Lo scollegamento di un repository Git da un notebook lo disconnette solo dal repository remoto; non elimina alcun codice dal notebook.
Scollegamento non riuscito	Il repository Git non è riuscito a scollegarsi dal notebook. È possibile riprovare a scollegarlo.

Creazione di un nuovo Notebook con un repository Git associato

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Creazione di un notebook e relativa associazione ai repository Git nella vecchia console Amazon EMR

1. Segui le istruzioni riportate in Creazione di un notebook.
2. Per Gruppo di sicurezza, scegli Utilizza gruppo di sicurezza personale.

> **Note**
>
> I gruppi di sicurezza del notebook devono includere una regola in uscita per consentire al notebook di instradare il traffico a Internet attraverso il cluster. È consigliabile creare gruppi di sicurezza personali. Per ulteriori informazioni, consulta Specifica dei gruppi di sicurezza EC2 per Notebook EMR.

3. Per Repository Git, seleziona Scegli repository per selezionare il repository da associare al notebook.
    1. Scegli un repository archiviato come risorsa nel tuo account e quindi Salva.
    2. Per aggiungere un nuovo repository come risorsa nell'account, scegli Aggiungi un nuovo repository. Completa il flusso di lavoro Aggiungi repository in una nuova finestra.

Uso di repository Git in un Notebook

> **Note**
>
> I Notebook Amazon EMR sono disponibili come Workspace EMR Studio nella nuova console. Nella vecchia console puoi ancora utilizzare i notebook esistenti, ma non crearne di nuovi. Il pulsante Crea workspace nella nuova console sostituisce questa funzionalità. Per accedere ai Workspace o crearne di nuovi, gli utenti di Notebook EMR necessitano di ulteriori autorizzazioni per i ruoli IAM. Per ulteriori informazioni, consulta I Notebook Amazon EMR sono Workspace Amazon EMR Studio nella nuova console e Novità della console.

Quando apri un notebook, puoi scegliere Apri in JupyterLab o Apri in Jupyter.

Se scegli di aprire il notebook in Jupyter, viene visualizzato un elenco di file e cartelle espandibili all'interno del notebook. È possibile eseguire manualmente comandi Git come il seguente in una cella di notebook.

```
!git pull origin primary
```

Per aprire uno dei repository aggiuntivi, passa ad altre cartelle.

Se apri il notebook con un'interfaccia JupyterLab, puoi utilizzare l'estensione Git di JupyterLab preinstallata. Per avere informazioni sull'estensione, consulta jupyterlab-git.

Pianificazione e configurazione di cluster

Questa sezione illustra le opzioni di configurazione e le istruzioni per la pianificazione, la configurazione e l'avvio di cluster tramite Amazon EMR. Prima di avviare un cluster, puoi definire le scelte riguardanti il sistema in base ai dati che stai elaborando e alle tue esigenze in termini di velocità, capacità, disponibilità, sicurezza e gestibilità. Le opzioni disponibili includono:

- La regione in cui eseguire un cluster, dove e come archiviare i dati e quale output preferire per i risultati. Per informazioni, consultare Configurazione del percorso del cluster e dell'archiviazione dati.
- Sia che si eseguano i cluster Amazon EMR su Outposts o Local Zones. Vedi Cluster EMR in AWS Outposts o Cluster EMR sulle AWS Local Zones.
- Se un cluster è a lunga durata o transitoria e quale software vi viene eseguito. Consultare Configurazione di un cluster per continuare o terminare dopo l'esecuzione della fase e Configurazione del software cluster.
- Se un cluster dispone di un singolo nodo primario o tre nodi primari. Per informazioni, consultare Pianificazione e configurazione dei nodi primari.
- Le opzioni hardware e di rete che ottimizzano costi, prestazioni e disponibilità per la tua applicazione. Per informazioni, consultare Configurazione di hardware e reti cluster.
- Come configurare i cluster in modo da poter gestire più facilmente e monitorare le attività, le prestazioni e lo stato. Consultare Configurazione della registrazione e del debug di cluster e Aggiunta di tag ai cluster.
- Come autenticare e autorizzare l'accesso alle risorse del cluster e come crittografare i dati. Per informazioni, consultare Sicurezza in Amazon EMR.
- Come eseguire l'integrazione con altri software e servizi. Per informazioni, consultare Driver e integrazione delle applicazioni di terze parti.

Avvio rapido di un cluster

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Avvio rapido di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr/clusters.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona Create cluster (Crea cluster).
3. Nella pagina Create Cluster (Crea cluster), inserisci o seleziona i valori per i campi forniti. Il pannello di riepilogo persistente consente di visualizzare in tempo reale le opzioni del cluster attualmente selezionate. Seleziona un'intestazione nel pannello di riepilogo per accedere alla sezione corrispondente e apportare modifiche. È necessario completare tutte le configurazioni richieste prima di poter scegliere Create cluster (Crea cluster).
4. Scegli Create cluster (Crea cluster) per accettare la configurazione come mostrato.
5. Si apre la pagina dei dettagli del cluster. Cerca l'indicazione Stato accanto al nome del cluster. Lo stato dovrebbe cambiare da Starting (Avvio in corso) a Running (In esecuzione) a Waiting (In attesa) durante il processo di creazione del cluster. Potrebbe essere necessario scegliere l'icona di aggiornamento in alto a destra o aggiornare il browser per ricevere gli aggiornamenti.

   Quando lo stato cambia in Waiting (In attesa), il cluster è attivo, in esecuzione e pronto per accettare fasi e connessioni SSH.

Old console

Vai alla pagina Create Cluster - Quick Options (Crea cluster - Opzioni rapide) nella console Amazon EMR per creare rapidamente un cluster per svolgere attività semplici o per scopi di valutazione o test. Quick Options (Opzioni rapide) utilizza i valori predefiniti per le opzioni di configurazione come il software cluster, la rete e la sicurezza.

Avvio di un cluster con Quick Options (Opzioni rapide) con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Clusters (Cluster), quindi seleziona Create cluster (Crea cluster) per aprire la pagina Quick Options (Opzioni rapide).

3. Sulla pagina Create Cluster - Quick Options (Crea cluster - Opzioni rapide, immetti o seleziona i valori per i campi forniti.
4. Scegli Create cluster (Crea cluster) per avviare il cluster e aprire la pagina di stato del cluster.
5. Nella pagina di stato del cluster, cerca l'indicazione Status (Stato) accanto al nome del cluster. Lo stato dovrebbe cambiare da Starting (Avvio in corso) a Running (In esecuzione) a Waiting (In attesa) durante il processo di creazione del cluster. Potrebbe essere necessario scegliere l'icona di aggiornamento a destra o aggiornare il browser per ricevere gli aggiornamenti.

   Quando lo stato cambia in Waiting (In attesa), il cluster è attivo, in esecuzione e pronto per accettare fasi e connessioni SSH.

Configurazione del percorso del cluster e dell'archiviazione dati

Questa sezione descrive come configurare la regione per un cluster, i diversi file system disponibili per l'utilizzo con Amazon EMR e come utilizzarli. Descrive inoltre come preparare o caricare dati su Amazon EMR, se necessario, nonché come preparare un percorso di output per i file di log e qualsiasi file di dati di output configurato.

Argomenti

Scelta di una Regione AWS

Amazon Web Services viene eseguito su server in data center di tutto il mondo. I data center sono organizzati per Regione geografica. Quando si avvia un cluster Amazon EMR, occorre specificare una Regione. Puoi scegliere una Regione per ridurre la latenza, minimizzare i costi o rispondere ai requisiti normativi. Per un elenco di tutte le regioni e di tutti gli endpoint supportati da Amazon EMR, consulta Regioni ed endpoint in Riferimenti generali di Amazon Web Services.

Per ottenere il massimo delle prestazioni, devi avviare il cluster nella stessa Regione in cui si trovano i dati. Ad esempio, se il bucket Amazon S3 di archiviazione dei dati di input si trova nella Regione Stati Uniti occidentali (Oregon), il cluster deve essere avviato nella Regione Stati Uniti occidentali

(Oregon) per evitare i costi del trasferimento di dati tra Regioni. Se utilizzi un bucket Amazon S3 per ricevere l'output del cluster, è opportuno crearlo nella Regione Stati Uniti occidentali (Oregon).

Se prevedi di associare una coppia di chiavi Amazon EC2 con il cluster (obbligatorio per utilizzare SSH per accedere al nodo core), la coppia di chiavi deve essere creata nella stessa Regione del cluster. Analogamente, i gruppi di sicurezza creati da Amazon EMR per gestire il cluster si trovano nella stessa Regione del cluster.

Se ti sei registrato per un Account AWS a partire dal 17 maggio 2017 incluso, la Regione predefinita quando si accede a una risorsa dalla AWS Management Console è Stati Uniti orientali (Ohio) (us-east-2); per account meno recenti, l'area predefinita è Stati Uniti occidentali (Oregon) (us-west-2) o Stati Uniti orientali (Virginia settentrionale) (us-east-1). Per ulteriori informazioni, consulta Regioni ed endpoint di .

Alcune caratteristiche di AWS sono disponibili solo in alcune Regioni. Ad esempio, le istanze Cluster Compute sono disponibili solo nella Regione Stati Uniti orientali (Virginia settentrionale), mentre la Regione Asia Pacifico (Sydney) supporta solo Hadoop rilascio 1.0.3 e successivi. Quando scegli una Regione, controlla che supporti le caratteristiche che desideri utilizzare.

Per ottenere il massimo delle prestazioni, utilizza la stessa Regione per tutte le risorse AWS che verranno impiegate con il cluster. La tabella seguente mappa i nomi delle Regioni tra i servizi. Per un elenco delle regioni di Amazon EMR, consulta la sezione Regioni AWS ed endpoint in Riferimenti generali di Amazon Web Services.

Scelta di una Regione con la console

La tua Regione predefinita viene visualizzata a sinistra delle informazioni del tuo account nella barra di navigazione. Per cambiare Regione, sia nella nuova console sia in quella vecchia, scegli il menu a discesa Region (Regione) e seleziona una nuova opzione.

Scelta di una Regione con la AWS CLI

Puoi specificare una Regione predefinita in AWS CLI utilizzando il comando aws configure o la variabile di ambiente `AWS_DEFAULT_REGION`. Per ulteriori informazioni, consulta la sezione relativa alla Configurazione della Regione AWS nella Guida per l'utente della AWS Command Line Interface.

Scelta di una Regione con un SDK o l'API

Per scegliere una Regione utilizzando un SDK, configura l'applicazione per utilizzare l'endpoint di tale Regione. Se stai creando un'applicazione client utilizzando un SDK AWS, puoi modificare l'endpoint client chiamando `setEndpoint`, come mostrato nell'esempio seguente:

```
client.setEndpoint("elasticmapreduce.us-west-2.amazonaws.com");
```

Dopo che nell'applicazione è stata specificata una Regione impostando l'endpoint, puoi impostare la zona di disponibilità per le istanze EC2 del cluster. Le zone di disponibilità sono posizioni geografiche distinte sviluppate per essere isolate dai guasti in altre zone di disponibilità e offrono una connettività di rete economica a bassa latenza ad altre zone di disponibilità nella stessa Regione. Una Regione contiene una o più zone di disponibilità. Per ottimizzare le prestazioni e ridurre la latenza, tutte le risorse devono trovarsi nella stessa zona di disponibilità del cluster che le utilizza.

Utilizzo di archiviazione e file system

Amazon EMR e Hadoop offrono un'ampia gamma di file system che puoi utilizzare durante l'elaborazione delle fasi del cluster. Puoi specificare il file system da utilizzare tramite il prefisso dell'URI di accesso ai dati. Ad esempio, `s3://DOC-EXAMPLE-BUCKET1/path` fa riferimento a un bucket Amazon S3 che utilizza EMRFS. La tabella seguente elenca i file system disponibili, con suggerimenti riguardo all'utilizzo più appropriato per ciascuno.

Amazon EMR e Hadoop in genere usano due o più dei seguenti file system durante l'elaborazione di un cluster. HDFS ed EMRFS sono i due principali file system utilizzati con Amazon EMR.

> **Important**
>
> A partire da Amazon EMR rilascio 5.22.0, Amazon EMR utilizza AWS Signature Version 4 esclusivamente per autenticare le richieste inviate ad Amazon S3. I rilasci precedenti di Amazon EMR impiegano AWS Signature Version 2 in alcuni casi, a meno che le note di rilascio indichino l'utilizzo esclusivo di Signature Version 4. Per ulteriori informazioni, consulta Autenticazione delle richieste (AWS Signature Version 4) e Autenticazione delle richieste (AWS Signature Version 2) nella Guida per gli sviluppatori di Amazon Simple Storage.

File system	Prefix	Descrizione
HDFS	`hdfs://` (o nessun prefisso)	HDFS è un file system distribuito, scalabile e portatile per Hadoop. HDFS ha il vantaggio di garantire la consapevolezza dei dati tra i nodi del cluster Hadoop che gestiscono i cluster e i nodi del cluster Hadoop che gestiscono le singole fasi. Per ulteriori informazioni, consulta la documentazione di Hadoop.

File system	Prefix	Descrizione
		HDFS viene utilizzato per i nodi master e principal i. Uno dei vantaggi è la sua rapidità, mentre ha lo svantaggio di essere uno storage temporaneo che viene recuperato quando il cluster termina. Trova il miglior utilizzo nel caching dei risultati prodotti dalle fasi intermedie del flusso di elaborazione.
EMRFS	`s3://`	EMRFS è un'implementazione del file system Hadoop utilizzato per leggere e scrivere file standard da Amazon EMR direttamente su Amazon S3. EMRFS offre il vantaggio di archiviare dati persistenti in Amazon S3 per l'utilizzo con Hadoop fornendo nel contempo caratteristiche quali crittografia lato server Amazon S3, coerenza lettura dopo scrittura e coerenza degli elenchi. **Note** In precedenza, Amazon EMR utilizzava i file system `s3n` e `s3a`. Benché entrambe le opzioni siano funzionanti, è preferibile utilizzare lo schema URI `s3` per ottenere prestazioni migliori, sicurezza e affidabilità.

File system	Prefix	Descrizione
File system locale		Il file system locale fa riferimento a un disco con connessione locale. Al momento della creazione di un cluster Hadoop, ogni nodo viene creato da un'istanza EC2 fornita con un blocco preconfigurato di storage su disco precollegato, chiamato instance store. I dati in qualsiasi volume instance store sono persistenti solo durante il ciclo di vita della relativa istanza EC2. I volumi instance store sono ideali per lo storage di dati temporanei in continua evoluzione, come buffer, cache, dati Scratch e altri contenuti temporanei. Per ulteriori informazioni, consulta Archiviazione dell'istanza Amazon EC2. Il file system locale viene utilizzato da HDFS, ma Python viene eseguito anche dal file system locale ed è possibile scegliere di archiviare file di applicazione aggiuntivi sui volumi archivio istanza.
(Legacy) File system a blocchi Amazon S3	`s3bfs://`	Il file system a blocchi Amazon S3 è un file storage system legacy. e ne sconsigliamo caldamente l'utilizzo. **Important** Consigliamo di non utilizzare questo file system perché può attivare una race condition che può causare errori nel cluster. Tuttavia, può essere richiesto da applicazioni legacy.

Accesso ai file system

Puoi specificare il file system da utilizzare tramite il prefisso dell'URI (Uniform Resource Identifier) di accesso ai dati. Le seguenti procedure illustrano come fare riferimento a diversi tipi di file system.

Per accedere all'HDFS locale

- Specifica il prefisso `hdfs:///` nell'URI. Amazon EMR risolve i percorsi che non specificano un prefisso nell'URI dell'HDFS locale. Ad esempio, entrambi i seguenti URI verrebbero risolti alla stessa posizione in HDFS.

```
hdfs:///path-to-data

/path-to-data
```

Per accedere all'HDFS remoto

- Includi l'indirizzo IP del nodo master nell'URI, come illustrato negli esempi seguenti.

```
hdfs://master-ip-address/path-to-data

master-ip-address/path-to-data
```

Accesso ad Amazon S3

- Utilizza il prefisso `s3://`.

```
s3://bucket-name/path-to-file-in-bucket
```

Accesso al file system a blocchi Amazon S3

- Utilizzalo solo per le applicazioni legacy che richiedono il file system a blocchi Amazon S3. Per accedere o archiviare i dati con questo file system, utilizza il prefisso `s3bfs://` nell'URI.

  Il file system a blocchi Amazon S3 è un file system legacy che veniva utilizzato per supportare caricamenti con una dimensione maggiore di 5 GB su Amazon S3. Con la funzionalità di caricamento in più parti offerta da Amazon EMR tramite l'SDK AWS per Java, puoi caricare file

fino a 5 TB per il file system nativo Amazon S3. Pertanto, il file system a blocchi Amazon S3 è ormai obsoleto.

> **Warning**
>
> Poiché questo file system legacy può creare race condition che possono danneggiare il file system, è opportuno evitarlo e utilizzare invece EMRFS.

```
s3bfs://bucket-name/path-to-file-in-bucket
```

Preparazione dei dati di input

La maggior parte dei cluster carica i dati di input e li elabora. Per caricare i dati, deve essere in una posizione alla quale il cluster possa accedere e in un formato che il cluster possa elaborare. Lo scenario più comune è quello di caricare i dati di input in Amazon S3. Amazon EMR fornisce strumenti per il cluster per importare o leggere dati da Amazon S3.

Il formato di input predefinito in Hadoop è quello dei file di testo, anche se è possibile personalizzare Hadoop e utilizzare strumenti per importare i dati memorizzati in altri formati.

Argomenti

- Tipi di input che Amazon EMR può accettare
- Come ottenere i dati in Amazon EMR

Tipi di input che Amazon EMR può accettare

Il formato di input predefinito per un cluster è un file di testo con ogni riga separata da un carattere di newline (\n), che è il formato di input più comunemente usato.

Se i dati di input sono in un formato diverso da quello predefinito dei file di testo, è possibile utilizzare l'interfaccia di Hadoop `InputFormat` per specificare altri tipi di input. Puoi persino creare una sottoclasse della classe `FileInputFormat` per gestire i tipi di dati personalizzati. Per ulteriori informazioni, consulta http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/InputFormat.html.

Se si utilizza Hive, è possibile utilizzare un serializzatore/deserializzatore (SerDe) per leggere i dati in un determinato formato in HDFS. Per ulteriori informazioni, consulta https://cwiki.apache.org/confluence/display/Hive/SerDe.

Come ottenere i dati in Amazon EMR

Amazon EMR fornisce diversi modi per ottenere dati su un cluster. Il modo più comune è quello di caricare i dati in Amazon S3 e utilizzare le funzionalità integrate di Amazon EMR per caricare i dati sul cluster. È inoltre possibile utilizzare la funzione di cache distribuita di Hadoop per trasferire i file da un file system distribuito al file system locale. L'implementazione di Hive fornita da Amazon EMR (Hive versione 0.7.1.1 e successive) include funzionalità che è possibile utilizzare per importare ed esportare i dati tra DynamoDB e un cluster Amazon EMR. Se si dispone di grandi quantità di dati da elaborare, il servizio AWS Direct Connect può essere utile.

Argomenti

- Caricamento dei dati su Amazon S3
- Carica i dati con AWS DataSync
- Importazione di file con la cache distribuita
- Come elaborare file compressi
- Importazione di dati DynamoDB in Hive
- Connessione ai dati con AWS Direct Connect
- Caricamento di grandi quantità di dati con AWS Snowball

Caricamento dei dati su Amazon S3

Per istruzioni su come caricare oggetti su Amazon S3, consulta Aggiunta di un oggetto a un bucket nella Guida per l'utente di Amazon Simple Storage Service. Per ulteriori informazioni su come utilizzare Amazon S3 con Hadoop, consulta http://wiki.apache.org/hadoop/AmazonS3.

Argomenti

- Creazione e configurazione di un bucket Amazon S3
- Configurazione del caricamento in più parti per Amazon S3
- Best practice

Creazione e configurazione di un bucket Amazon S3

Amazon EMR utilizza AWS SDK for Java con Amazon S3 per archiviare dati di input, file di log e dati di output. Amazon S3 fa riferimento a questi percorsi di archiviazione come bucket. I bucket presentano determinate restrizioni e limitazioni in conformità con i requisiti di Amazon S3 e DNS. Per ulteriori informazioni, consulta Restrizioni e limitazioni dei bucket nella Guida per l'utente di Amazon Simple Storage Service.

In questa sezione viene descritto come utilizzare la AWS Management Console Amazon S3 per creare e impostare le autorizzazioni per un bucket Amazon S3. È anche possibile creare e impostare autorizzazioni per un bucket Amazon S3 utilizzando l'API Amazon S3 o la AWS CLI. È anche possibile utilizzare Curl insieme a una modifica per passare i parametri di autenticazione appropriati per Amazon S3.

Consulta le seguenti risorse:

- Per creare un bucket utilizzando la console, consultare Creare un bucket nella Guida per l'utente di Amazon S3.
- Per creare e lavorare con i bucket utilizzando la AWS CLI, consulta Utilizzo dei comandi S3 di alto livello con la AWS Command Line Interface nella Guida per l'utente di Amazon S3.
- Per creare un bucket tramite un SDK, consulta Esempi di creazione di un bucket nella Guida per l'utente di Amazon Simple Storage Service.
- Per utilizzare i bucket mediante Curl, consulta Strumento di autenticazione Amazon S3 per Curl.
- Per ulteriori informazioni sulla definizione di bucket specifici per Regione, consulta Accesso a un bucket nella Guida per l'utente di Amazon Simple Storage Service.
- Per lavorare con bucket utilizzando Amazon S3 Access Points, consulta Utilizzo di un alias in stile bucket per il punto di accesso nella Guida per l'utente di Amazon S3. Puoi utilizzare facilmente Amazon S3 Access Points con l'alias di Access Point di Amazon S3 anziché il nome del bucket Amazon S3. Puoi utilizzare l'alias di Access Point di Amazon S3 per applicazioni esistenti e nuove, tra cui Spark, Hive, Presto e altre.

> **Note**
>
> Se si abilita la registrazione per un bucket, saranno abilitati solo i log di accesso al bucket e non i log del cluster Amazon EMR.

Durante o dopo la creazione del bucket, è possibile impostare le autorizzazioni appropriate per accedere al bucket a seconda dell'applicazione. In genere, consenti a te stesso (il proprietario) l'accesso in lettura e scrittura e il solo accesso in lettura agli utenti autenticati.

Per poter creare un cluster, sono necessari i bucket Amazon S3 richiesti. È necessario caricare in Amazon S3 tutti gli script e i dati a cui viene fatto riferimento nel cluster. Nella seguente tabella vengono descritti dati, script e ubicazioni di file di log esempio.

Configurazione del caricamento in più parti per Amazon S3

Amazon EMR supporta il caricamento in più parti di Amazon S3 attraverso SDK AWS per Java. Il caricamento in più parti consente di caricare un singolo oggetto come un insieme di parti. È possibile caricare queste parti dell'oggetto in modo indipendente e in qualsiasi ordine. Se la trasmissione di una parte non riesce, è possibile ritrasmettere tale parte senza influire sulle altre. Una volta caricate tutte le parti dell'oggetto, Amazon S3 le assembla e crea l'oggetto.

Per ulteriori informazioni, consulta Panoramica del caricamento in più parti nella Guida per l'utente di Amazon Simple Storage Service.

Inoltre, Amazon EMR include proprietà che consentono di controllare con maggiore precisione il cleanup delle parti di un caricamento in più parti.

Nella tabella seguente vengono descritte le proprietà di configurazione di Amazon EMR per il caricamento in più parti. Per la configurazione, utilizza la classificazione di configurazione `core-site`. Per ulteriori informazioni, consulta Configurazione delle applicazioni nella Guida ai rilasci di Amazon EMR.

Nome del parametro di configurazione	Valore predefinito	Descrizione
`fs.s3n.multipart.uploads.enabled`	`true`	Un tipo booleano che indica se abilitare il caricamento in più parti. Quando la visualizzazione coerente EMRFS è abilitata, i caricamenti in più parti sono abilitati per impostazione predefinita e l'impostazione di questo valore su `false` viene ignorata.
`fs.s3n.multipart.uploads.split.size`	`134217728`	Specifica la dimensione massima di una parte, in byte, prima che EMRFS avvii il

Nome del parametro di configurazione	Valore predefinito	Descrizione
		caricamento di una nuova parte quando è abilitato il caricamento in più parti. Il valore minimo è `5242880` (5 MB). Se viene specificato un valore più basso, viene utilizzato `5242880`. Il valore massimo è `5368709120` (5 GB). Se viene specificato un valore più alto, viene utilizzato `5368709120` . Se la crittografia lato client di EMRFS è disattivata e il committer ottimizzato Amazon S3 è disabilitato, questo valore controlla anche la dimensione massima di un file di dati che può aumentare finché EMRFS utilizza i caricamenti in più parti, anziché una richiesta `PutObject` per caricare il file. Per ulteriori informazioni, consulta
`fs.s3n.ssl.enabled`	`true`	Un tipo booleano che indica se utilizzare http o https.
`fs.s3.buckets.create.enabled`	`false`	Un tipo booleano che indica se è necessario creare un bucket in caso non esista. L'impostazione su `false` causa un'eccezione nelle operazioni `CreateBucket` .
`fs.s3.multipart.clean.enabled`	`false`	Un tipo booleano che indica se abilitare il cleanup periodico in background dei caricamenti in più parti incompleti.

Nome del parametro di configurazione	Valore predefinito	Descrizione
`fs.s3.multipart.clean.age.threshold`	`604800`	Un tipo long che specifica la durata minima, in secondi, di un caricamento in più parti, prima che venga considerato idonea per il cleanup. Il valore predefinito è una settimana.
`fs.s3.multipart.clean.jitter.max`	`10000`	Un numero intero che specifica il ritardo massimo, in secondi, nel jitter casuale aggiunto al ritardo fisso di 15 minuti prima di pianificare la fase successiva di cleanup.

Disabilitazione dei caricamenti in più parti

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Disabilitazione dei caricamenti in più parti con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Software Settings (Modifica delle impostazioni), inserisci la seguente configurazione: `classification=core-site,properties=[fs.s3n.multipart.uploads.enabled=false]`.
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Disabilitazione dei caricamenti in più parti con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. In Edit Software Settings (Modifica impostazioni software), inserisci la seguente configurazione: `classification=core-site,properties=[fs.s3n.multipart.uploads.enabled=false]`
4. Procedere con la creazione del cluster.

CLI

Per disabilitare il caricamento in più parti utilizzando la AWS CLI

Questa procedura illustra come disabilitare il caricamento in più parti utilizzando la AWS CLI. Per disabilitare il caricamento in più parti, digitare il comando `create-cluster` con il parametri `--bootstrap-actions`.

1. Crea un file, `myConfig.json`, con il seguente contenuto e salvalo nella stessa directory in cui esegui il comando:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "fs.s3n.multipart.uploads.enabled": "false"
    }
  }
]
```

2. Digita il comando seguente e sostituisci *myKey* con il nome della coppia di chiavi EC2.

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

```
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

Disabilitazione dei caricamento in più parti con l'API

- Per informazioni sull'utilizzo dei caricamenti in più parti di Amazon S3 a livello di programmazione, consulta Utilizzo dell'SDK AWS per Java per il caricamento in più parti nella Guida per l'utente di Amazon Simple Storage Service.

  Per ulteriori informazioni su come utilizzare l'SDK AWS per Java, consulta SDK AWS per Java.

Best practice

Di seguito sono riportate le raccomandazioni per l'utilizzo di bucket Amazon S3 con cluster EMR.

Abilita il controllo delle versioni

La funzione Versioni multiple è una configurazione consigliata per il bucket Amazon S3. Abilitando la funzione Versioni multiple, si garantisce che anche se i dati vengono eliminati o sovrascritti involontariamente, possono essere ripristinati. Per ulteriori informazioni, consulta Utilizzo del controllo delle versioni nella Guida per l'utente di Amazon Simple Storage Service.

Eliminazione dei caricamenti in più parti non riusciti

I componenti del cluster EMR utilizzano caricamenti in più parti mediante l'SDK AWS per Java con le API Amazon S3 per scrivere i file di log e i dati di output su Amazon S3 per impostazione predefinita.

Per ulteriori informazioni sulla modifica delle proprietà relative a questa configurazione utilizzando Amazon EMR, consulta Configurazione del caricamento in più parti per Amazon S3. A volte il caricamento di un file di grandi dimensioni può risultare in un caricamento in più parti incompleto Amazon S3. Quando un caricamento in più parti non riesce a essere completato con successo, il caricamento in più parti in corso continua a occupare il bucket e comporta costi di storage. Per evitare lo storage eccessivo dei dati, consigliamo le seguenti opzioni:

- Per i bucket utilizzati con Amazon EMR, utilizza una regola di configurazione del ciclo di vita in Amazon S3 per rimuovere i caricamenti in più parti incompleti tre giorni dopo la data di inizio del caricamento. Le regole di configurazione del ciclo di vita consentono di controllare la classe di storage e la durata degli oggetti. Per ulteriori informazioni, consulta Gestione del ciclo di vita degli oggetti e Interruzione dei caricamenti in più parti incompleti utilizzando una policy per il ciclo di vita del bucket.
- Abilita la funzione di eliminazione in più parti di Amazon EMR impostando `fs.s3.multipart.clean.enabled` su `TRUE` e l'ottimizzazione di altri parametri di eliminazione. Questa funzione è utile in caso di volume elevato, su larga scala e con cluster con tempi di attività limitati. In questo caso, il parametro `DaysAfterIntitiation` di una regola di configurazione del ciclo di vita potrebbe essere troppo lungo, anche se impostato sul valore minimo, causando picchi di archiviazione in Amazon S3. L'eliminazione multiparte di Amazon EMR consente un controllo più preciso. Per ulteriori informazioni, consulta Configurazione del caricamento in più parti per Amazon S3.

Gestione dei contrassegni di versione

Consigliamo di abilitare una regola di configurazione del ciclo di vita in Amazon S3 per rimuovere i contrassegni di eliminazione dell'oggetto scaduto per i bucket con versione utilizzati con Amazon EMR. Quando si elimina un oggetto in un bucket con versione, viene creato un contrassegno di eliminazione. Se tutte le versioni precedenti dell'oggetto scadono successivamente, nel bucket viene lasciato un contrassegno di eliminazione dell'oggetto scaduto. Sebbene non sia previsto alcun costo per questi contrassegni di eliminazione, la loro rimozione può migliorare le prestazioni delle richieste di LIST. Per ulteriori informazioni, consulta Configurazione del ciclo di vita per un bucket con controllo delle versioni nella Guida per l'utente di Amazon Simple Storage Service.

Best practice sulle prestazioni

A seconda dei carichi di lavoro, tipi specifici di utilizzo dei cluster EMR e delle applicazioni su di essi possono portare a un elevato numero di richieste contro una bucket. Per maggiori informazioni,

consulta Considerazioni sulla percentuale di richieste e sulle prestazioni nella Guida per l'utente di Amazon Simple Storage Service.

Carica i dati con AWS DataSync

AWS DataSync è un servizio di trasferimento dati online che semplifica, automatizza e accelera il processo di spostamento dei dati tra l'archiviazione on-premise e i servizi di archiviazione AWS o tra servizi di archiviazione AWS. DataSync supporta una varietà di sistemi di archiviazione on-premise come Hadoop Distributed File System (HDFS), file server NAS e archiviazione di oggetti autogestita.

Il modo più comune per inserire dati in un cluster è quello di caricare i dati in Amazon S3 e utilizzare le funzionalità integrate di Amazon EMR per caricare i dati sul cluster.

DataSync può aiutarti a eseguire i seguenti processi:

- Replica HDFS sul tuo cluster Hadoop in Amazon S3 per la continuità aziendale
- Copia HDFS in Amazon S3 per popolare i data lake
- Trasferisci i dati tra HDFS del cluster Hadoop e Amazon S3 per l'analisi e l'elaborazione

Per caricare i dati nel bucket S3, è necessario innanzitutto distribuire uno o più agenti DataSync nella stessa rete dell'archiviazione on-premise. Un agent (agente) è una macchina virtuale (VM) utilizzata per leggere o scrivere dati in una posizione autogestita. Quindi attivi i tuoi agenti nell'Account AWS e nella Regione AWS dove si trova il bucket S3.

Dopo aver attivato l'agente, crei una posizione di origine per l'archiviazione on-premise, una posizione di destinazione per il bucket S3 e un processo. Un'attivitàè costituita da un set di due percorsi (origine e destinazione) e un set di opzioni predefinite che permettono di controllarne il comportamento.

Infine, esegui il processo DataSync per trasferire i dati dall'origine alla destinazione.

Per ulteriori informazioni, consulta la pagina Getting started with AWS DataSync.

Importazione di file con la cache distribuita

Argomenti

- Tipi di file supportati
- Percorso dei file memorizzati nella cache

- Accesso ai file della cache dalle applicazioni in streaming
- Accesso ai file della cache dalle applicazioni in streaming

DistributedCache è una funzione Hadoop in grado di aumentare l'efficienza quando una mappa o un'attività di riduzione richiede l'accesso a dati comuni. Se il cluster dipende da applicazioni esistenti o da file binari non installati al momento della creazione, è possibile utilizzare DistributedCache per importare questi file. Questa funzione consente a un nodo del cluster di leggere i file importati dal file system locale, invece di recuperare i file da altri nodi del cluster.

Per ulteriori informazioni, vai alla pagina http://hadoop.apache.org/docs/stable/api/org/apache/hadoop/filecache/DistributedCache.html.

È possibile chiamare DistributedCache quando si crea il cluster. I file vengono messi in cache poco prima dell'avvio del lavoro Hadoop e rimangono in cache per tutta la durata del lavoro. È possibile mettere in cache i file archiviati su qualsiasi file system compatibile con Hadoop, ad esempio HDFS o Amazon S3. La dimensione predefinita della cache dei file è 10 GB. Per modificare la dimensione della cache, riconfigurare il parametro Hadoop, `local.cache.size` usando l'operazione di bootstrap. Per ulteriori informazioni, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo.

Tipi di file supportati

DistributedCache consente di archiviare sia singoli file che archivi. I singoli file vengono memorizzati in cache come file di sola lettura. Gli eseguibili e i file binari hanno le autorizzazioni di esecuzione impostate.

Gli archivi sono uno o più file pacchettizzati utilizzando un'utilità, ad esempio `gzip`. DistributedCache trasferisce i file compressi in ogni nodo principale e decomprime l'archivio come parte dell'inserimento nella cache. DistributedCache supporta i seguenti formati di compressione:

- zip
- tgz
- tar.gz
- tar
- jar

Percorso dei file memorizzati nella cache

DistributedCache copia i file solo nei nodi principali. Se non sono presenti nodi core nel cluster, DistributedCache copia i file nel nodo primario.

DistributedCache associa i file della cache alla directory di lavoro corrente del mappatore e del riduttore utilizzando collegamenti simbolici. Un collegamento simbolico è un alias a una posizione del file, non la posizione effettiva del file. Il valore del parametro, `yarn.nodemanager.local-dirs` in `yarn-site.xml`, specifica il percorso dei file temporanei. Amazon EMR imposta il parametro su `/mnt/mapred` o alcune variazioni in base al tipo di istanza e alla versione EMR. Ad esempio, un'impostazione può avere `/mnt/mapred` e `/mnt1/mapred` perché il tipo di istanza ha due volumi effimeri. I file della cache si trovano in una sottodirectory della posizione del file temporaneo in `/mnt/mapred/taskTracker/archive`.

Se si memorizza nella cache un singolo file, DistributedCache inserisce il file nella directory `archive`. Se si mette in cache un archivio, DistributedCache decomprime il file, crea una sottodirectory in `/archive` con lo stesso nome del file di archivio. I singoli file si trovano nella nuova sottodirectory.

È possibile utilizzare la cache distribuita solo quando si utilizza lo streaming.

Accesso ai file della cache dalle applicazioni in streaming

Per accedere ai file memorizzati nella cache dalle applicazioni mappatore o riduttore, accertati di aver aggiunto la directory di lavoro corrente (. /) nel percorso dell'applicazione e di aver fatto riferimento ai file memorizzati nella cache come se fossero presenti nella directory di lavoro corrente.

Accesso ai file della cache dalle applicazioni in streaming

È possibile utilizzare la AWS Management Console e la AWS CLI per creare cluster che utilizzano la cache distribuita.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Specifica dei file della cache distribuita con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Steps (Fasi), scegli Add step (Aggiungi fase). Si apre la finestra di dialogo Add step (Aggiungi fase). Nel campo Arguments (Argomenti), includi i file e gli archivi da salvare nella cache. La dimensione del file (o la dimensione totale dei file in un file di archivio) deve essere inferiore alla dimensione della cache assegnata.

   Se desideri aggiungere un singolo file alla cache distribuita, specifica `-cacheFile` seguito dal nome e dalla posizione del file, dal segno cancelletto (#) e quindi dal nome che desideri assegnare al file quando viene collocato nella cache locale. L'esempio seguente mostra come aggiungere un singolo file alla cache distribuita.

   ```
   -cacheFile \
   s3://DOC-EXAMPLE-BUCKET/file-name#cache-file-name
   ```

   Se desideri aggiungere un file di archivio alla cache distribuita, inserisci `-cacheArchive` seguito dal percorso dei file in Amazon S3, dal segno cancelletto (#) e quindi dal nome che desideri assegnare alla raccolta di file nella cache locale. L'esempio seguente mostra come aggiungere un file di archivio alla cache distribuita.

   ```
   -cacheArchive \
   s3://DOC-EXAMPLE-BUCKET/archive-name#cache-archive-name
   ```

   Inserisci i valori appropriati negli altri campi di dialogo. Le opzioni variano a seconda del tipo di fase. Per aggiungere la fase e uscire dalla finestra di dialogo, scegli Add step (Aggiungi fase).
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica dei file della cache distribuita con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Step execution (Esecuzione fase) come modalità di avvio.
4. Nella sezione Steps (Fasi), nel campo Add Step (Aggiungi fase), scegli Streaming Program (Programma Streaming) dall'elenco e quindi fai clic su Configure and add (Configura e aggiungi).
5. Nel campo Arguments (Argomenti), includi i file e gli archivi da salvare nella cache e fai clic su Add (Aggiungi). La dimensione del file (o la dimensione totale dei file in un file di archivio) deve essere inferiore alla dimensione della cache assegnata.

   Se desideri aggiungere un singolo file alla cache distribuita, specifica `-cacheFile` seguito dal nome e dalla posizione del file, dal segno cancelletto (#) e quindi dal nome che desideri assegnare al file quando viene collocato nella cache locale. L'esempio seguente mostra come aggiungere un singolo file alla cache distribuita.

   ```
   -cacheFile \
   s3://DOC-EXAMPLE-BUCKET/file_name#cache_file_name
   ```

   Se desideri aggiungere un file di archivio alla cache distribuita, inserisci `-cacheArchive` seguito dal percorso dei file in Amazon S3, dal segno cancelletto (#) e quindi dal nome che desideri assegnare alla raccolta di file nella cache locale. L'esempio seguente mostra come aggiungere un file di archivio alla cache distribuita.

   ```
   -cacheArchive \
   s3://DOC-EXAMPLE-BUCKET/archive_name#cache_archive_name
   ```
6. Procedi con la configurazione e l'avvio del cluster. Il cluster copia i file nella posizione di cache prima di elaborare qualsiasi fase del cluster.

CLI

Specifica dei file della cache distribuita con la AWS CLI

- Per inviare una fase di Streaming quando un cluster è stato creato, digita il comando `create-cluster` con il parametro `--steps`. Per specificare i file della cache distribuita utilizzando la AWS CLI, specifica gli argomenti appropriati quando invii una fase di streaming.

  Se desideri aggiungere un singolo file alla cache distribuita, specifica `-cacheFile` seguito dal nome e dalla posizione del file, dal segno cancelletto (#) e quindi dal nome che desideri assegnare al file quando viene collocato nella cache locale.

  Se desideri aggiungere un file di archivio alla cache distribuita, inserisci `-cacheArchive` seguito dal percorso dei file in Amazon S3, dal segno cancelletto (#) e quindi dal nome che desideri assegnare alla raccolta di file nella cache locale. L'esempio seguente mostra come aggiungere un file di archivio alla cache distribuita.

  Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Example 1

Digita il comando seguente per avviare un cluster e invia una fase di Streaming che utilizza `-cacheFile` per aggiungere un file, `sample_dataset_cached.dat`, alla cache.

```
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"]
```

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi core. Tutti i nodi utilizzeranno il tipo di istanza specificato nel comando.

Se in precedenza non sono stati creati il ruolo di servizio EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Example 2

Il comando seguente mostra la creazione di un cluster di streaming e si avvale di `-cacheArchive` per aggiungere un archivio di file nella cache.

```
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"]
```

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi core. Tutti i nodi utilizzeranno il tipo di istanza specificato nel comando.

Se in precedenza non sono stati creati il ruolo di servizio EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Come elaborare file compressi

Hadoop controlla l'estensione del file per rilevare i file compressi. I tipi di compressione supportati da Hadoop sono: gzip, bzip2 e LZO. Non è necessario intraprendere alcuna azione aggiuntiva per estrarre i file utilizzando questo tipo di compressione; Hadoop lo gestisce per voi.

Per indicizzare i file LZO, è possibile utilizzare la libreria hadoop-lzo che può essere scaricata da https://github.com/kevinweil/hadoop-lzo. Trattandosi di una libreria di terze parti, Amazon EMR non offre supporto agli sviluppatori su come utilizzare questo strumento. Per informazioni sull'utilizzo, consulta il file readme di hadoop-lzo.

Importazione di dati DynamoDB in Hive

L'implementazione di Hive fornita da Amazon EMR include funzionalità che è possibile utilizzare per importare ed esportare i dati tra DynamoDB e un cluster Amazon EMR. Questa funzione è utile se i dati di input vengono memorizzati in DynamoDB. Per ulteriori informazioni, consulta Esportazione, importazione, esecuzione di query e unione di tabelle in DynamoDB con Amazon EMR.

Connessione ai dati con AWS Direct Connect

AWS Direct Connect è un servizio che può essere utilizzato per stabilire una connessione di rete dedicata privata ad Amazon Web Services dal data center, dall'ufficio o dall'ambiente di co-locazione. Se si dispone di grandi quantità di dati di input, l'utilizzo di AWS Direct Connect può consentire di ridurre i costi di rete, di aumentare il throughput della larghezza di banda e di offrire un'esperienza di rete più coerente rispetto alle connessioni basate su Internet. Per ulteriori informazioni, consulta la Guida per l'utente AWS Direct Connect.

Caricamento di grandi quantità di dati con AWS Snowball

AWS Snowball è un servizio che puoi utilizzare per trasferire grandi quantità di dati tra Amazon Simple Storage Service (Amazon S3) e il percorso di archiviazione dei dati in loco a velocità superiori a Internet. Snowball supporta due tipi di processo: i processi di importazione e i processi di esportazione. I processi di importazione comportano un trasferimento di dati da un'origine on-premise a un bucket Amazon S3. I processi di esportazione comportano un trasferimento di dati da bucket Amazon S3 a un'origine on-premise. Per entrambi i tipi di processo, i dispositivi Snowball proteggono i tuoi dati mentre i corrieri regionali li trasportano tra Amazon S3 e il percorso di archiviazione dei dati in loco. I dispositivi Snowball sono fisicamente robusti e protetti dal AWS Key Management Service (AWS KMS). Per ulteriori informazioni, consulta la Guida per gli sviluppatori di AWS Snowball Edge.

Configurazione di un percorso di output

Il formato di output più comune di un cluster Amazon EMR sono i file di testo, compressi o decompressi. Di solito, tali file vengono scritti su un bucket Amazon S3. Occorre creare questo bucket prima dell'avvio del cluster. Specifica il bucket S3 come percorso di output al momento dell'avvio del cluster.

Per ulteriori informazioni, consulta i seguenti argomenti:

Argomenti

- Creazione e configurazione di un bucket Amazon S3
- Quali formati Amazon EMR è in grado di restituire?
- Come scrivere dati in un bucket Amazon S3 di cui non si è proprietari
- Compressione dell'output del cluster

Creazione e configurazione di un bucket Amazon S3

Amazon EMR usa Amazon S3 per archiviare dati di input, file di log e dati di output. Amazon S3 fa riferimento a questi percorsi di archiviazione come bucket. I bucket presentano determinate restrizioni e limitazioni in conformità con i requisiti di Amazon S3 e DNS. Per ulteriori informazioni, consulta Restrizioni e limitazioni dei bucket nella Guida per gli sviluppatori di Amazon Simple Storage Service.

Per creare un bucket Amazon S3, segui le istruzioni nella pagina Creazione di un bucket della Guida per gli sviluppatori di Amazon Simple Storage Service.

> **Note**
>
> Se abiliti la registrazione nella procedura guidata Create a Bucket (Crea un bucket), sono abilitati solo i log di accesso al bucket e non i log del cluster.

> **Note**
>
> Per ulteriori informazioni sulla definizione di bucket specifici di una regione, consulta la sezione relativa a Bucket e Regioni nella Guida per gli sviluppatori di Amazon Simple Storage Service e la pagina relativa agli Endpoint di Regione disponibili per gli SDK AWS.

Dopo aver creato il bucket è possibile impostare le autorizzazioni appropriate su di esso. In genere, consenti a te stesso (il proprietario) l'accesso in lettura e scrittura. Si consiglia di seguire Best practice di sicurezza per Amazon S3 durante la configurazione del bucket.

Per poter creare un cluster, sono necessari i bucket Amazon S3 richiesti. È necessario caricare in Amazon S3 tutti gli script e i dati a cui viene fatto riferimento nel cluster. Nella seguente tabella vengono descritti dati, script e ubicazioni di file di log esempio.

Informazioni	Esempio di percorso su Amazon S3
script o programma	`s3://DOC-EXAMPLE-BUCKET1 /script/MapperScript.py`
file di log	`s3://DOC-EXAMPLE-BUCKET1 /logs`
dati di input	`s3://DOC-EXAMPLE-BUCKET1 /input`

Informazioni	Esempio di percorso su Amazon S3
dati di output	`s3://DOC-EXAMPLE-BUCKET1/output`

Quali formati Amazon EMR è in grado di restituire?

Il formato di output predefinito per un cluster è il testo con coppie di chiavi e valori scritte su singole righe nei file di testo. Questo è il formato di output utilizzato più comunemente.

Se devi scrivere i dati di output in un formato diverso da quello predefinito dei file di testo, puoi utilizzare l'interfaccia di Hadoop `OutputFormat` per specificare altri tipi di output. Puoi persino creare una sottoclasse della classe `FileOutputFormat` per gestire i tipi di dati personalizzati. Per ulteriori informazioni, consulta http://hadoop.apache.org/docs/current/api/org/apache/hadoop/mapred/OutputFormat.html.

Se avvii un cluster Hive, puoi utilizzare un serializzatore/deserializzatore (SerDe) per l'output dei dati da HDFS in un determinato formato. Per ulteriori informazioni, consulta https://cwiki.apache.org/confluence/display/Hive/SerDe.

Come scrivere dati in un bucket Amazon S3 di cui non si è proprietari

Quando scrivi un file in un bucket Amazon Simple Storage Service (Amazon S3), per impostazione predefinita sei l'unico in grado di leggere tale file. Il presupposto è che scriverai file nel tuo bucket e questa impostazione predefinita protegge la privacy dei file.

Tuttavia, se stai eseguendo un cluster e desideri che l'output venga scritto nel bucket Amazon S3 di un altro utente AWS e che l'altro utente AWS sia in grado di leggere tale output, devi eseguire due operazioni:

- Chiedi all'altro utente AWS di concederti le autorizzazioni in scrittura per il suo bucket Amazon S3. Il cluster che avvii viene eseguito con le tue credenziali AWS, perciò qualunque sia il cluster avviato, sarà in grado di scrivere nel bucket dell'altro utente AWS.
- Imposta le autorizzazioni di lettura per l'altro utente AWS sui file scritti da te o dal cluster nel bucket Amazon S3. Il modo più semplice per impostare queste autorizzazioni di lettura è di utilizzare le liste di controllo accessi (ACL) predefinite, un insieme di policy d'accesso predefinite definite da Amazon S3.

Per ulteriori informazioni su come l'altro utente AWS può concederti autorizzazioni per scrivere file nel bucket Amazon S3 dell'altro utente, consulta Modifica delle autorizzazioni per il bucket nella Guida per l'utente di Amazon Simple Storage Service.

Affinché il tuo cluster utilizzi liste di controllo accessi (ACL) predefinite quando scrive file in Amazon S3, imposta l'opzione di configurazione del cluster `fs.s3.canned.acl` sulla lista di controllo accessi (ACL) predefinita da utilizzare. Nella tabella seguente vengono elencate le liste di controllo degli accessi predefinite attualmente definite.

ACL predefinita	Descrizione
`AuthenticatedRead`	Specifica che al proprietario viene concesso l'accesso `Permission.FullControl` e al gruppo `GroupGrantee.AuthenticatedUsers` assegnatario viene concesso l'accesso `Permission.Read`.
`BucketOwnerFullControl`	Specifica che al proprietario del bucket viene concesso l'accesso `Permission.FullControl`. Il proprietario del bucket non coincide necessariamente con il proprietario dell'oggetto.
`BucketOwnerRead`	Specifica che al proprietario del bucket viene concesso l'accesso `Permission.Read`. Il proprietario del bucket non coincide necessariamente con il proprietario dell'oggetto.
`LogDeliveryWrite`	Specifica che al proprietario viene concesso l'accesso `Permission.FullControl` e al gruppo `GroupGrantee.LogDelivery` assegnatario viene concesso l'accesso `Permission.Write`, in modo da poter recapitare i log di accesso.
`Private`	Specifica che al proprietario viene concesso l'accesso `Permission.FullControl`.
`PublicRead`	Specifica che al proprietario viene concesso l'accesso `Permission.FullControl` e al gruppo `GroupGran`

ACL predefinita	Descrizione
	`tee.AllUsers` assegnatario viene concesso l'accesso `Permission.Read` .
`PublicReadWrite`	Specifica che al proprietario viene concesso l'accesso `Permission.FullControl` e al gruppo `GroupGran tee.AllUsers` assegnatario viene concesso l'accesso `Permission.Read` e `Permission.Write` .

A seconda del tipo di cluster in esecuzione, esistono molti modi per impostare le opzioni di configurazione del cluster. Nelle procedure seguenti viene mostrato come impostare l'opzione per i casi comuni.

Per scrivere file utilizzando liste di controllo degli accessi predefinite in Hive

- Dal prompt dei comandi di Hive, imposta l'opzione di configurazione `fs.s3.canned.acl` per la lista di controllo accessi (ACL) predefinita che deve essere impostata dal cluster sui file scritti in Amazon S3. Per accedere al prompt dei comandi di Hive, esegui la connessione al nodo master tramite SSH e digita Hive al prompt dei comandi di Hadoop. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH.

  Nell'esempio seguente, l'opzione di configurazione `fs.s3.canned.acl` viene impostata su `BucketOwnerFullControl`. In questo modo, al proprietario del bucket Amazon S3 viene assegnato il controllo completo sul file. Si noti che il comando set rileva la distinzione tra maiuscole e minuscole e non contiene né virgolette né spazi.

  ```
  hive> set fs.s3.canned.acl=BucketOwnerFullControl;
  create table acl (n int) location 's3://acltestbucket/acl/';
  insert overwrite table acl select count(*) from acl;
  ```

  Nelle ultime due righe dell'esempio viene creata una tabella che è archiviata in Amazon S3 e vengono scritti dati nella tabella.

Per scrivere file utilizzando liste di controllo accessi (ACL) predefinite in Pig

- Dal prompt dei comandi di Pig, imposta l'opzione di configurazione `fs.s3.canned.acl` per la lista di controllo accessi (ACL) predefinita che deve essere impostata dal cluster sui file scritti in Amazon S3. Per accedere al prompt dei comandi di Pig, esegui la connessione al nodo master utilizzando SSH e digita Pig al prompt dei comandi di Hadoop. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH.

  Nell'esempio seguente, l'opzione di configurazione `fs.s3.canned.acl` viene impostata su BucketOwnerFullControl. In questo modo, al proprietario del bucket Amazon S3 viene assegnato il controllo completo sul file. Si noti che il comando set include uno spazio prima del nome della lista di controllo accessi (ACL) predefinita e non contiene virgolette.

```
pig> set fs.s3.canned.acl BucketOwnerFullControl;
store some data into 's3://acltestbucket/pig/acl';
```

Per scrivere file utilizzando liste di controllo accessi (ACL) predefinite in un JAR personalizzato

- Imposta l'opzione di configurazione `fs.s3.canned.acl` utilizzando Hadoop con il flag -D. Questo viene mostrato nell'esempio sottostante.

```
hadoop jar hadoop-examples.jar wordcount
-Dfs.s3.canned.acl=BucketOwnerFullControl s3://mybucket/input s3://mybucket/output
```

Compressione dell'output del cluster

Argomenti

- Compressione dei dati di output
- Compressione dei dati intermedi
- Utilizzo della libreria Snappy con Amazon EMR

Compressione dei dati di output

In questo modo si comprime l'output del processo Hadoop. Se si utilizza TextOutputFormat il risultato è un file di testo con compressione gzip. Se si scrive in SequenceFile, il risultato è un SequenceFile che viene compresso internamente. Questo può essere abilitato impostando la configurazione mapred.output.compress su "true".

Se si sta eseguendo un processo in streaming, è possibile abilitarlo trasmettendo il processo in streaming a questi argomenti.

```
-jobconf mapred.output.compress=true
```

È anche possibile utilizzare un'operazione di bootstrap per comprimere automaticamente tutti gli output del lavoro. Ecco come procedere con il client Ruby.

```
--bootstrap-actions s3://elasticmapreduce/bootstrap-actions/configure-hadoop \
--args "-s,mapred.output.compress=true"
```

Infine, se si sta scrivendo un Jar personalizzato, è possibile attivare la compressione dell'output con la seguente riga quando si crea un lavoro.

```
FileOutputFormat.setCompressOutput(conf, true);
```

Compressione dei dati intermedi

Se il processo mischia una quantità significativa di dati dai mappatori ai riduttori, è possibile vedere un miglioramento delle prestazioni grazie all'abilitazione della compressione intermedia. Comprime l'output della mappa e lo decomprime quando arriva sul nodo principale. L'impostazione di configurazione è mapred.compress.map.output. È possibile abilitarlo in modo simile alla compressione di output.

Quando si scrive un Jar personalizzato, utilizzare il seguente comando:

```
conf.setCompressMapOutput(true);
```

Utilizzo della libreria Snappy con Amazon EMR

Snappy è una libreria di compressione e decompressione ottimizzata per la velocità. È disponibile sugli AMI Amazon EMR versione 2.0 e successive ed è utilizzato come elemento predefinito per la compressione intermedia. Per ulteriori informazioni su Snappy, vai alla pagina http://code.google.com/p/snappy/.

Pianificazione e configurazione dei nodi primari

Quando avvii un cluster Amazon EMR, puoi scegliere di avere uno o tre nodi primari nel cluster. L'avvio di un cluster con tre nodi primari è supportato solo da Amazon EMR versione 5.23.0 e successive. Amazon EMR può trarre vantaggio dai gruppi di collocamento EC2 per garantire che i nodi primari siano posizionati su hardware sottostante distinto al fine di migliorare ulteriormente la disponibilità del cluster. Per ulteriori informazioni, consulta Integrazione di Amazon EMR con gruppi di collocamento EC2.

Un cluster Amazon EMR con più nodi primari fornisce i seguenti vantaggi chiave:

- Il nodo primario non rappresenta più un singolo punto di errore. Se uno dei nodi primari riscontra errori, il cluster utilizza gli altri due nodi primari e viene eseguito senza interruzioni. Nel frattempo, Amazon EMR sostituisce automaticamente i nodi primari con errori con un nuovo nodo dotato delle stesse operazioni di configurazione e bootstrap.
- Amazon EMR abilita le caratteristiche ad alta disponibilità Hadoop di HDFS NameNode e YARN ResourceManager e supporta l'alta disponibilità per altre applicazioni open source.

  Per ulteriori informazioni su come un cluster EMR con più nodi primari supporta le applicazioni open source e altre caratteristiche Amazon EMR, consulta la sezione Applicazioni e caratteristiche supportate.

> **Note**
>
> I cluster possono trovarsi in una sola zona di disponibilità o sottorete.

Questa sezione fornisce informazioni sulle applicazioni e sulle caratteristiche supportate di un cluster Amazon EMR con più nodi primari nonché i dettagli di configurazione, le best practice e le considerazioni in merito all'avvio del cluster.

Argomenti

Applicazioni e caratteristiche supportate

Questo argomento fornisce informazioni sulle caratteristiche di alta disponibilità Hadoop di HDFS NameNode e YARN ResourceManager in un cluster Amazon EMR e su come le caratteristiche di alta disponibilità funzionano con le applicazioni open source e altre funzionalità Amazon EMR.

HDFS a disponibilità elevata

Un cluster Amazon EMR con più nodi primari abilita la caratteristica di disponibilità elevata HDFS NameNode in Hadoop. Per ulteriori informazioni, consulta l'argomento relativo alla Disponibilità elevata HDFS.

In un cluster Amazon EMR, due o più nodi separati sono configurati come NameNodes. Un NameNode è nello stato `active` e gli altri sono nello stato `standby`. Se il nodo con NameNode `active` ha esito negativo, Amazon EMR avvia automaticamente un processo di failover HDFS. Un nodo con NameNode `standby` diventa `active` e assume tutte le operazioni client nel cluster. Amazon EMR sostituisce il nodo con esito negativo con uno nuovo che quindi si ricollega come `standby`.

> **Note**
>
> In Amazon EMR versioni 5.23.0 fino alla versione 5.30.1 inclusa, solo due dei tre nodi primari eseguono HDFS NameNode.

Per sapere quale NameNode è `active`, puoi utilizzare SSH per connetterti a qualsiasi nodo primario nel cluster ed eseguire il seguente comando:

```
hdfs haadmin -getAllServiceState
```

L'output elenca i nodi in cui NameNode è installato e il relativo stato. Ad esempio,

```
ip-##-#-#-##1.ec2.internal:8020 active
ip-##-#-#-##2.ec2.internal:8020 standby
ip-##-#-#-##3.ec2.internal:8020 standby
```

ResourceManager YARN a disponibilità elevata

Un cluster Amazon EMR con più nodi primari abilita la caratteristica di disponibilità elevata YARN ResourceManager in Hadoop. Per ulteriori informazioni, consulta Disponibilità elevata di ResourceManager.

In un cluster Amazon EMR con più nodi primari, YARN ResourceManager viene eseguito su tutti e tre i nodi primari. Un ResourceManager è nello stato `active` e gli altri due sono nello stato `standby`. Se il nodo primario con ResourceManager `active` riscontra un errore, Amazon EMR avvia un processo di failover automatico. Un nodo primario con un ResourceManager `standby` assume tutte le operazioni. Amazon EMR sostituisce il nodo primario con errori con uno nuovo, che quindi si ricollega al quorum ResourceManager come `standby`.

Puoi connetterti a "http://*master-public-dns-name*:8088/cluster" per qualsiasi nodo primario, che indirizza automaticamente al gestore risorse `active`. Per sapere quale gestore è `active`, utilizza SSH per connetterti a un nodo primario del cluster. Quindi esegui il comando seguente per ottenere un elenco dei tre nodi primari e il relativo stato:

```
yarn rmadmin -getAllServiceState
```

Applicazioni supportate in un cluster Amazon EMR con più nodi primari

È possibile installare ed eseguire le seguenti applicazioni in un cluster Amazon EMR con più nodi primari. Per ogni applicazione, il processo di failover del nodo primario varia.

Applicazione	Disponibilità durante il failover del nodo primario	Note
Flink	Disponibilità non interessata dal failover del nodo primario	I processi Flink su Amazon EMR vengono eseguiti come applicazioni YARN. I JobManage

Applicazione	Disponibilità durante il failover del nodo primario	Note
		rs di Flink vengono eseguiti come Applicati onMasters di YARN sui nodi principali. Il JobManager non è interessato dal processo di failover del nodo primario. Se si utilizza la versione 5.27.0 di Amazon EMR o precedente, JobManager è un singolo punto di errore. Quando il JobManager ha esito negativo, perde tutti gli stati del processo e non riprende i processi in esecuzione. Puoi abilitare l'elevata disponibilità di JobManager configura ndo il numero di tentativi delle applicazioni, il checkpoint e l'abilitazione di ZooKeeper come storage di stato per Flink. Per ulteriori informazi oni, consulta la sezione Configuring Flink on an Amazon EMR Cluster with multiple primary nodes (Configurazione di Flink su un cluster Amazon EMR con più nodi primari). A partire dalla versione 5.28.0 di Amazon EMR, non è necessaria alcuna configurazione manuale per abilitare la disponibilità elevata di JobManager.
Ganglia	Disponibilità non interessata dal failover del nodo primario	Ganglia è disponibile su tutti i nodi primari, quindi può continuare a funzionare durante il processo di failover del nodo primario.
Hadoop	Elevata disponibilità	HDFS NameNode e YARN ResourceManager eseguono automaticamente il failover sul nodo di standby quando il nodo primario attivo riscontra un errore.

Applicazione	Disponibilità durante il failover del nodo primario	Note
HBase	Elevata disponibilità	HBase esegue automaticamente il failover sul nodo di standby quando il nodo primario attivo riscontra un errore. Se ti connetti a HBase tramite un server REST o Thrift, devi passare a un nodo primario diverso quando il nodo primario attivo riscontra un errore.
HCatalog	Disponibilità non interessata dal failover del nodo primario	HCatalog si basa sul metastore Hive che esiste al di fuori del cluster. HCatalog rimane disponibi le durante il processo di failover del nodo primario.
JupyterHub	Elevata disponibilità	JupyterHub è installato su tutte e tre le istanze primarie. Si consiglia vivamente di configura re la persistenza del notebook per evitare la perdita del notebook in caso di errore del nodo primario. Per ulteriori informazioni, consulta la sezione Configurazione della persistenza per i notebook in Amazon S3.
Livy	Elevata disponibilità	Livy è installato su tutti e tre i nodi primari. Quando il nodo primario attivo riscontra un errore, perdi l'accesso alla sessione Livy corrente e devi creare una nuova sessione Livy su un altro nodo primario o sul nuovo nodo sostitutivo.
Mahout	Disponibilità non interessata dal failover del nodo primario	Poiché Mahout non prevede il daemon, non viene interessato dal processo di failover del nodo primario.

Applicazione	Disponibilità durante il failover del nodo primario	Note
MXNet	Disponibilità non interessata dal failover del nodo primario	Poiché MXNet non prevede il daemon, non viene interessato dal processo di failover del nodo primario.
Phoenix	Elevata disponibilità	QueryServer di Phoenix viene eseguito solo su uno dei tre nodi primari. Phoenix su tutti e tre i master è configurato per collegare il Phoenix QueryServer. È possibile trovare l'IP privato del server Query di Phoenix utilizzando il file `/etc/phoenix/conf/phoenix-env.sh`
Pig	Disponibilità non interessata dal failover del nodo primario	Poiché Pig non prevede il daemon, non viene interessato dal processo di failover del nodo primario.
Spark	Elevata disponibilità	Tutte le applicazioni Spark vengono eseguite in container YARN e possono reagire al failover del nodo primario allo stesso modo delle caratteristiche YARN a disponibilità elevata.
Sqoop	Elevata disponibilità	Per impostazione predefinita, sqoop-job e sqoop-metastore memorizzano i dati (descrizi oni del lavoro) sul disco locale del master che esegue il comando; se si desidera salvare i dati del metastore su database esterno, fare riferimento alla documentazione di apache Sqoop
Tez	Elevata disponibilità	Poiché i container Tez funzionano su YARN, Tez si comporta allo stesso modo di YARN durante il processo di failover del nodo primario.

Applicazione	Disponibilità durante il failover del nodo primario	Note
TensorFlow	Disponibilità non interessata dal failover del nodo primario	Poiché TensorFlow non prevede il daemon, non viene interessato dal processo di failover del nodo primario.
Zeppelin	Elevata disponibilità	Zeppelin è installato su tutti e tre i nodi primari. Zeppelin memorizza note e configurazioni di interpreti in HDFS per impostazione predefini ta per evitare la perdita di dati. Le sessioni di interprete sono completamente isolate in tutte e tre le istanze primarie. I dati della sessione andranno persi in caso di errore principale. Si consiglia di non modificare la stessa nota contemporaneamente su istanze primarie diverse.
ZooKeeper	Elevata disponibilità	ZooKeeper è la base della caratteristica di failover automatico HDFS. ZooKeeper fornisce un servizio altamente disponibile per il mantenimento dei dati di coordinamento, la notifica ai client delle modifiche apportate ai dati e il monitoraggio dei client in caso di errori. Per ulteriori informazioni, consulta la sezione relativa al Failover automatico di HDFS.

Per eseguire le seguenti applicazioni in un cluster Amazon EMR con più nodi primari, è necessario configurare un database esterno. Il database esterno esiste al di fuori del cluster e rende i dati persistenti durante il processo di failover del nodo primario. Per le applicazioni seguenti, i componenti del servizio verranno ripristinati automaticamente durante il processo di failover del nodo primario, ma i processi attivi potrebbero non riuscire e dovranno essere riattivati.

Applicazione	Disponibilità durante il failover del nodo primario	Note
Hive	Disponibilità elevata solo per i componenti di servizio	È necessario un metastore esterno per Hive. Deve trattarsi di un metastore esterno MySQL, poiché PostgreSQL non è supportato per cluster multi-master. Per ulteriori informazioni, consulta Configurazione di un metastore esterno per Hive.
Hue	Disponibilità elevata solo per i componenti di servizio	È necessario un database esterno per Hue. Per ulteriori informazioni, consulta Utilizzo di Hue con un database remoto in Amazon RDS.
Oozie	Disponibilità elevata solo per i componenti di servizio	È necessario un database esterno per Oozie. Per ulteriori informazioni, consulta Utilizzo di Oozie con un database remoto in Amazon RDS. Oozie-server e oozie-client sono installati su tutti e tre i nodi primari. I client oozie-sono configurati per connettersi al server oozie-cor retto per impostazione predefinita.
PrestoDB o PrestoSQL/ Trino	Disponibilità elevata solo per i componenti di servizio	È necessario un metastore Hive esterno per PrestoDB (PrestoSQL su Amazon EMR 6.1.0-6.3.0 o Trino su Amazon EMR 6.4.0 e versioni successive). È possibile utilizzare Presto con AWS Glue Data Catalog o utilizzare un database MySQL esterno per Hive. La CLI di Presto è installata su tutti e tre i nodi primari, quindi è possibile utilizzarla per accedere a Presto Coordinator da uno qualsiasi dei nodi primari. Presto Coordinat or è installato su un solo nodo primario. Puoi

Applicazione	Disponibilità durante il failover del nodo primario	Note
		trovare il nome DNS del nodo primario in cui è installato Presto Coordinator chiamando l'API `describe-cluster` di Amazon EMR e la lettura del valore restituito del campo `MasterPublicDnsName` nella risposta.

> **Note**
>
> Quando un nodo primario riscontra un errore, Java Database Connectivity (JDBC) o Open Database Connectivity (ODBC) termina la connessione al nodo primario. Puoi connetterti a uno dei nodi primari rimanenti per continuare il lavoro dal momento che il daemon metastore Hive viene eseguito su tutti i nodi primari. In alternativa, puoi attendere che il nodo primario con errori venga sostituito.

Funzionamento delle caratteristiche di Amazon EMR in un cluster con più nodi primari

Connessione ai nodi primari tramite SSH

È possibile connettersi a uno dei tre nodi primari in un cluster Amazon EMR utilizzando SSH nello stesso modo in cui ci si connette a un singolo nodo primario. Per ulteriori informazioni, consulta la sezione Connect to the primary node using SSH (Connessione al nodo primario tramite SSH).

Se un nodo primario riscontra un errore, la connessione SSH per quel nodo primario viene terminata. Per continuare il lavoro, puoi connetterti a uno degli altri due nodi primari. In alternativa, puoi accedere al nuovo nodo primario dopo che Amazon EMR sostituisce quello con errori con un nuovo nodo.

> **Note**
>
> L'indirizzo IP privato per il nodo primario sostitutivo rimane uguale a quello precedente. L'indirizzo IP pubblico per il nodo primario sostitutivo potrebbe cambiare. Puoi recuperare i nuovi indirizzi IP nella console o utilizzando il comando `describe-cluster` nella CLI AWS.

NameNode viene eseguito solo su due dei nodi primari. Tuttavia, è possibile eseguire i comandi della CLI `hdfs` e gestire i processi per accedere a HDFS su tutti e tre i nodi primari.

Utilizzo delle fasi in un cluster Amazon EMR con più nodi primari

Puoi inviare le fasi a un cluster Amazon EMR con più nodi primari nello stesso modo in cui utilizzi le fasi in un cluster con un singolo nodo primario. Per ulteriori informazioni, consulta Invio di lavoro a un cluster.

Di seguito sono riportate le considerazioni sull'utilizzo di fasi in un cluster Amazon EMR con più nodi primari:

- Se un nodo primario riscontra un errore, le fasi in esecuzione sul nodo primario vengono contrassegnate come FAILED. I dati scritti in locale andranno perduti. Tuttavia, lo stato FAILED potrebbe non riflettere lo stato reale delle fasi.
- Se una fase in esecuzione ha avviato un'applicazione YARN quando il nodo primario riscontra un errore, la fase può continuare e avere esito positivo a causa del failover automatico del nodo primario.
- È consigliabile controllare lo stato delle fasi consultando l'output dei processi. Ad esempio, i processi MapReduce utilizzano un file `_SUCCESS` per determinare se il processo viene completato correttamente.
- È consigliabile impostare il parametro ActionOnFailure su CONTINUE o CANCEL_AND_WAIT, invece di TERMINATE_JOB_FLOW o TERMINATE_CLUSTER.

Protezione da cessazione automatica

Amazon EMR abilita automaticamente la protezione da terminazione per tutti i cluster con più nodi primari e sostituisce tutte le impostazioni di esecuzione di fasi fornite durante la creazione del cluster. È possibile disattivare la protezione da cessazione dopo l'avvio del cluster. Per informazioni, consultare Configurazione della protezione da cessazione per i cluster in esecuzione. Per chiudere un cluster con più nodi primari, è necessario modificare gli attributi del cluster per disabilitare la protezione da terminazione. Per istruzioni, consultare Terminazione di un cluster Amazon EMR con più nodi primari.

Per ulteriori informazioni sulla protezione da cessazione, consulta Utilizzo della protezione da cessazione.

Caratteristiche non supportate in un cluster Amazon EMR con più nodi primari

Le seguenti caratteristiche Amazon EMR non sono attualmente disponibili in un cluster EMR con più nodi primari:

- Notebook EMR
- Parchi istanze
- Accesso con un clic al server della cronologia Spark persistente
- Interfacce utente delle applicazioni persistenti
- L'accesso con un clic alle interfacce utente delle applicazioni persistenti non è attualmente disponibile per i cluster Amazon EMR con più nodi primari o per i cluster Amazon EMR integrati con AWS Lake Formation.

> **Note**
>
> Per utilizzare l'autenticazione Kerberos nel cluster, è necessario configurare un KDC esterno. A partire dalla versione 5.27.0 di Amazon EMR, puoi configurare la crittografia trasparente HDFS su un cluster EMR con più nodi primari. Per ulteriori informazioni, consulta l'argomento relativo alla Crittografia trasparente in HDFS su Amazon EMR.

Avvio di un cluster Amazon EMR con più nodi primari

Questo argomento contiene i dettagli di configurazione e gli esempi per avviare un cluster Amazon EMR con più nodi primari.

> **Note**
>
> Amazon EMR abilita automaticamente la protezione da terminazione per tutti i cluster con più nodi primari e sostituisce tutte le impostazioni di terminazione automatica fornite durante la creazione del cluster. Per chiudere un cluster con più nodi primari, è necessario modificare gli attributi del cluster per disabilitare la protezione da terminazione. Per istruzioni, consultare Terminazione di un cluster Amazon EMR con più nodi primari.

Prerequisiti

- Puoi avviare un cluster Amazon EMR con più nodi primari in sottoreti VPC pubbliche e private. EC2-Classic non è supportato. Per avviare un cluster Amazon EMR con più nodi primari in una sottorete pubblica, è necessario abilitare le istanze in questa sottorete per ricevere un indirizzo IP pubblico selezionando Auto-assign IPv4 (Assegna automaticamente IPv4) nella console o eseguendo il seguente comando. Sostituisci *22XXXX01* con l'ID della sottorete.

```
aws ec2 modify-subnet-attribute --subnet-id subnet-22XXXX01 --map-public-ip-on-launch
```

- Per eseguire Hive oppure Oozie in un cluster Amazon EMR con più nodi primari, è necessario creare un metastore esterno per Hive. Per ulteriori informazioni, consulta la sezione Configurazione di un metastore esterno per Hive, Utilizzo di Hue con un database remoto in Amazon RDS o Apache Oozie.
- Per utilizzare l'autenticazione Kerberos nel cluster, è necessario configurare un KDC esterno. Per ulteriori informazioni, consulta Configurazione di Kerberos su Amazon Amazon EMR.

Avvio di un cluster Amazon EMR con più nodi primari

È necessario specificare un valore della conta di istanze pari a tre per il gruppo di istanze del nodo primario all'avvio di un cluster Amazon EMR con più nodi primari. I seguenti esempi illustrano come avviare il cluster utilizzando l'AMI predefinita o un'AMI personalizzata.

> **Note**
>
> È necessario specificare l'ID della sottorete quando si avvia un cluster Amazon EMR con più nodi primari mediante la AWS CLI. Sostituisci *22XXXX01* con l'ID della sottorete negli esempi seguenti.

Example - Avvio di un cluster Amazon EMR con più nodi primari utilizzando un'AMI predefinita

```
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 - Avvio di un cluster Amazon EMR con più nodi primari utilizzando un'AMI personalizzata

```
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 - Avvio di un cluster Amazon EMR con più nodi primari con un metastore Hive esterno

Per eseguire Hive su un cluster Amazon EMR con più nodi primari, è necessario specificare un metastore esterno per Hive, come dimostrato dal seguente esempio.

1. Creare un file hiveConfiguration.json temporaneo che contenga le credenziali per il metastore 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. Avviare il cluster con il metastore 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
```

Terminazione di un cluster Amazon EMR con più nodi primari

Per terminare un cluster Amazon EMR con più nodi primari, è necessario prima disabilitare la protezione dalla terminazione, come dimostrato nel seguente esempio. Sostituisci *j-3KVTXXXXXX7UG* con l'ID del cluster.

```
aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-
protected
aws emr terminate-clusters --cluster-id j-3KVTXXXXXX7UG
```

Integrazione di Amazon EMR con gruppi di collocamento EC2

Quando avvii un cluster con più nodi primari Amazon EMR su Amazon EC2, hai la possibilità di utilizzare le strategie dei gruppi di collocamento per specificare come vuoi che le istanze del nodo primario vengano distribuite per assicurare la protezione dai guasti hardware.

Le strategie dei gruppi di collocamento sono supportate a partire da Amazon EMR versione 5.23.0 come opzione per i cluster con più nodi primari. Attualmente, solo i tipi di nodi primari sono supportati dalla strategia del gruppo di collocamento e la strategia `SPREAD` viene applicata a tali nodi primari. La strategia `SPREAD` colloca un piccolo gruppo di istanze su hardware sottostante separato per evitare la perdita di più nodi primari in caso di guasto hardware. Una richiesta di avvio di un'istanza potrebbe non riuscire se l'hardware univoco è insufficiente per l'esecuzione della richiesta. Per ulteriori informazioni sulle strategie di collocamento EC2 e sulle limitazioni, consulta Gruppi di collocamento nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Esiste un limite iniziale da parte di Amazon EC2 di 500 cluster abilitati alla strategia per gruppi di collocamento che possono essere avviati per Regione AWS. Contatta il supporto AWS per richiedere un aumento del numero di gruppi di collocamento consentiti. Puoi identificare i gruppi di collocamento EC2 creati da Amazon EMR monitorando la coppia chiave-valore che Amazon EMR associa alla strategia del gruppo di collocamento Amazon EMR. Per ulteriori informazioni sui tag delle istanze per i cluster EC2, consulta Visualizzazione di istanze cluster in Amazon EC2.

Collegamento della policy gestita del gruppo di collocamento al ruolo Amazon EMR

La strategia del gruppo di collocamento richiede una policy gestita denominata `AmazonElasticMapReducePlacementGroupPolicy` per consentire ad Amazon EMR di creare, eliminare e descrivere gruppi di collocamento su Amazon EC2. È necessario allegare `AmazonElasticMapReducePlacementGroupPolicy` per il ruolo di servizio per Amazon EMR prima di avviare un cluster con più nodi master Amazon EMR

La policy gestita da Amazon EMR, `AmazonEMRServicePolicy_v2`, può essere associata al ruolo di Amazon EMR anziché alla policy gestita dal gruppo di collocamento. `AmazonEMRServicePolicy_v2` consente lo stesso accesso ai gruppi di collocamento su Amazon EC2 di `AmazonElasticMapReducePlacementGroupPolicy`. Per ulteriori informazioni, consulta Ruolo di servizio per Amazon EMR (ruolo EMR).

La policy gestita da `AmazonElasticMapReducePlacementGroupPolicy` è il seguente testo JSON che viene creato e gestito da Amazon EMR.

> **Note**
>
> Poiché la policy `AmazonElasticMapReducePlacementGroupPolicy` gestita viene aggiornata automaticamente, tale contenuto potrebbe essere obsoleto. Utilizza la Console di gestione AWS per visualizzare la policy corrente.

```
{
   "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"
         ]
      }
   ]
}
```

Avvio di un cluster con più nodi master Amazon EMR con strategia per gruppi di collocamento

Per avviare un cluster con più nodi principali Amazon EMR con una strategia per gruppi di collocamento, allega la policy gestita dal gruppo di collocamento `AmazonElasticMapReducePlacementGroupPolicy` al ruolo Amazon EMR. Per ulteriori informazioni, consulta Collegamento della policy gestita del gruppo di collocamento al ruolo Amazon EMR.

Ogni volta che utilizzi questo ruolo per avviare un cluster con più nodi primari Amazon EMR, Amazon EMR tenta di avviare un cluster con la strategia `SPREAD` applicata ai suoi nodi primari. Se utilizzi un ruolo che non dispone della policy gestita dal gruppo di collocamento `AmazonElasticMapReducePlacementGroupPolicy` ad esso collegato, Amazon EMR tenta di avviare un cluster con più nodi principali Amazon EMR senza una strategia per gruppi di collocamento.

Se avvii un cluster con più nodi primari Amazon EMR con il parametro `placement-group-configs` utilizzando l'API o la CLI di Amazon EMR, Amazon EMR avvia il cluster solo se al ruolo Amazon EMR è collegata la policy gestita del gruppo di collocamento `AmazonElasticMapReducePlacementGroupPolicy`. Se al ruolo Amazon EMR non è associata la policy, l'avvio del cluster con più nodi primari di Amazon EMR riscontra un errore.

Example - Avvio di un cluster con più nodi primari Amazon EMR con strategia del gruppo di collocamento utilizzando l'API di Amazon EMR.

Quando utilizzi l'operazione RunJobFlow per creare un cluster con più nodi master di Amazon EMR, devi impostare la proprietà `PlacementGroupConfigs` come segue. Attualmente, il ruolo dell'istanza `MASTER` utilizza automaticamente `SPREAD` come strategia del gruppo di collocamento.

```
{
   "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"
}
```

- Sostituisci *ha-cluster* con il nome del cluster ad alta disponibilità.
- Sostituisci *subnet-22XXXX01* con il tuo ID sottorete.
- Sostituisci *ec2_key_pair_name* con il nome della coppia di chiavi EC2 per questo cluster. La coppia di chiavi EC2 è facoltativa e obbligatoria solo se si desidera utilizzare SSH per accedere al cluster.

Example - Avvio di un cluster con più nodi primari con strategia del gruppo di collocamento utilizzando la CLI 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
```

- Sostituisci *`ha-cluster`* con il nome del cluster ad alta disponibilità.
- Sostituisci *`subnet-22XXXX01`* con il tuo ID sottorete.
- Sostituisci *`ec2_key_pair_name`* con il nome della coppia di chiavi EC2 per questo cluster. La coppia di chiavi EC2 è facoltativa e obbligatoria solo se si desidera utilizzare SSH per accedere al cluster.

Avvio di un cluster con più nodi primari Amazon EMR senza strategia del gruppo di collocamento

Affinché un cluster con più nodi primari possa avviare nodi primari senza la strategia del gruppo di collocamento, dovrai effettuare una delle seguenti operazioni:

- Rimuovi la policy gestita del gruppo di collocamento `AmazonElasticMapReducePlacementGroupPolicy` dal ruolo Amazon EMR, oppure
- Avvia un cluster con più nodi primari Amazon EMR con il parametro `placement-group-configs` utilizzando l'API o la CLI di Amazon EMR oppure scegliendo `NONE` come strategia del gruppo di collocamento.

Example - Avvio di un cluster con più nodi primari senza strategia del gruppo di collocamento utilizzando l'API di Amazon EMR.

Quando utilizzi l'operazione RunJobFlow per creare un cluster con più nodi primari, devi impostare la proprietà `PlacementGroupConfigs` come segue.

```
{
   "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"
}
```

- Sostituisci *ha-cluster* con il nome del cluster ad alta disponibilità.
- Sostituisci *subnet-22XXXX01* con il tuo ID sottorete.
- Sostituisci *ec2_key_pair_name* con il nome della coppia di chiavi EC2 per questo cluster. La coppia di chiavi EC2 è facoltativa e obbligatoria solo se si desidera utilizzare SSH per accedere al cluster.

Example - Avvio di un cluster con più nodi primari senza strategia del gruppo di collocamento utilizzando la CLI 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
```

- Sostituisci *ha-cluster* con il nome del cluster ad alta disponibilità.
- Sostituisci *subnet-22XXXX01* con il tuo ID sottorete.
- Sostituisci *ec2_key_pair_name* con il nome della coppia di chiavi EC2 per questo cluster. La coppia di chiavi EC2 è facoltativa e obbligatoria solo se si desidera utilizzare SSH per accedere al cluster.

Controllo della configurazione della strategia del gruppo di collocamento associata al cluster con più nodi primari di Amazon EMR

Puoi utilizzare l'API describe cluster di Amazon EMR per visualizzare la configurazione della strategia del gruppo di collocamento associata al cluster con più nodi primari di Amazon EMR.

Example

```
aws emr describe-cluster --cluster-id "j-xxxxx"
{
   "Cluster":{
      "Id":"j-xxxxx",
      ...
      ...
      "PlacementGroups":[
         {
            "InstanceRole":"MASTER",
            "PlacementStrategy":"SPREAD"
         }
      ]
   }
}
```

Considerazioni e best practice

Limitazioni di un cluster Amazon EMR con più nodi primari:

- Non puoi utilizzare un cluster Amazon EMR con più nodi primari con parchi istanze. Per ulteriori informazioni sulle funzionalità di Amazon EMR con più nodi primari, consulta Applicazioni e caratteristiche supportate.
- Se due nodi primari riscontrano errori contemporaneamente, Amazon EMR non può ripristinare il cluster.
- I cluster Amazon EMR con più nodi primari non possono tollerare gli errori della zona di disponibilità. Nel caso di un'interruzione nella zona di disponibilità, perdi l'accesso ai cluster EMR.
- Amazon EMR non garantisce le funzionalità di disponibilità elevata di applicazioni open-source diverse da quelle specificate in Applicazioni supportate in un cluster Amazon EMR con più nodi primari.
- In Amazon EMR dal rilascio 5.23.0 fino al rilascio 5.30.1, solo due dei tre nodi primari eseguono HDFS NameNode.

Considerazioni per la configurazione della sottorete:

- Un cluster Amazon EMR con più nodi primari può trovarsi in una sola zona di disponibilità o sottorete. Amazon EMR non è in grado di sostituire un nodo primario con errori se la sottorete è completamente utilizzata o sovrascritta in caso di failover. Per evitare questo scenario, è opportuno dedicare un'intera sottorete a un cluster Amazon EMR. Inoltre, assicurati che nella sottorete siano disponibili sufficienti indirizzi IP privati.

Considerazioni per la configurazione dei nodi core:

- Per garantire che anche il gruppo di istanze del nodo principale sia altamente disponibile, ti consigliamo di avviare almeno quattro nodi principali. Se decidi di avviare un cluster più piccolo con tre nodi principali (o un numero minore), imposta `dfs.replication parameter` almeno su 2 in modo che HDFS abbia una replica DFS sufficiente. Per ulteriori informazioni, consulta la sezione dedicata alla Configurazione di HDFS.

> **Warning**
>
> 1. L'impostazione di `dfs.replication` su 1 per i cluster con meno di quattro nodi può causare la perdita di dati HDFS in caso di disattivazione anche di un singolo nodo. Ti consigliamo di utilizzare un cluster con almeno quattro nodi principali per i carichi di lavoro di produzione.

2. Amazon EMR non consente ai cluster di dimensionare i nodi principali al di sotto di `dfs.replication`. Ad esempio, se `dfs.replication = 2`, il numero minimo di nodi principali è 2.
3. Quando utilizzi il dimensionamento gestito, il dimensionamento automatico o scegli di dimensionare manualmente il cluster, ti consigliamo di impostare `dfs.replication` su 2 o su un valore superiore.

Considerazioni per l'impostazione di allarmi sui parametri:

- Amazon EMR non fornisce parametri specifici dell'applicazione su HDFS o YARN. Ti consigliamo di configurare gli allarmi per monitorare il conteggio delle istanze dei nodi primari. Configura gli allarmi usando i seguenti parametri di Amazon CloudWatch: `MultiMasterInstanceGroupNodesRunning`, `MultiMasterInstanceGroupNodesRunningPercentage` o `MultiMasterInstanceGroupNodesRequested`. CloudWatch ti avviserà in caso di errore e sostituzione del nodo primario.
  - Se il `MultiMasterInstanceGroupNodesRunningPercentage` è inferiore a 1.0 e superiore a 0.5, il cluster può avere perso un nodo primario. In questo caso, Amazon EMR tenta di sostituire un nodo primario.
  - Se il `MultiMasterInstanceGroupNodesRunningPercentage` è inferiore a 0.5, due nodi primari potrebbero avere riscontrato errori. In questo caso, il quorum viene perso e il cluster non può essere recuperato. È necessario eseguire manualmente la migrazione dei dati al di fuori del cluster.

  Per ulteriori informazioni, consulta Impostazione di allarmi per i parametri.

Cluster EMR in AWS Outposts

A partire da Amazon EMR versione 5.28.0, è possibile creare ed eseguire cluster EMR in AWS Outposts. AWS Outposts abilita i servizi nativi di AWS, l'infrastruttura e i modelli operativi nelle strutture in locale. Negli ambienti AWS Outposts è possibile utilizzare le stesse API, strumenti e infrastruttura AWS utilizzati in AWS Cloud. Amazon EMR su AWS Outposts è ideale per carichi di lavoro a bassa latenza che devono essere eseguiti in prossimità di dati e applicazioni on-premise. Per ulteriori informazioni su AWS Outposts, consultare la Guida per l'utente di AWS Outposts.

Prerequisiti

Di seguito sono indicati i prerequisiti per l'utilizzo di Amazon EMR in AWS Outposts:

- Devi avere installato e configurato AWS Outposts nel data center in locale.
- Devi disporre di una connessione di rete affidabile tra l'ambiente Outpost in uso e una Regione AWS.
- Devi disporre di una capacità sufficiente per i tipi di istanza supportati da EMR disponibili nel tuo Outpost.

Limitazioni

Di seguito sono riportate le limitazioni dell'utilizzo di Amazon EMR su AWS Outposts:

- Le istanze on demand sono l'unica opzione supportata per le istanze Amazon EC2. Le istanze Spot non sono disponibili per Amazon EMR in AWS Outposts.
- Se sono necessari volumi di archiviazione Amazon EBS aggiuntivi, è supportato solo General Purpose SSD (GP2).
- I bucket S3 che memorizzano oggetti in una Regione AWS specificata sono l'unica opzione S3 supportata per Amazon EMR su Outposts. S3 su Outposts non è supportato per Amazon EMR su AWS Outposts.
- Solo i seguenti tipi di istanza sono supportati da Amazon EMR in AWS Outposts:

Classe istanza	Tipi di istanza
Uso generale	`m5.xlarge |m5.2xlarge | m5.4xlarge |m5.12xlarge | m5.24xlarge |m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge |m5d.24xlarge`
Ottimizzato per il calcolo	`c5.xlarge |c5.2xlarge | c5.4xlarge |c5.18xlarge | c5d.xlarge |c5d.2xlarge | c5d.4xlarge | c5d.18xlarge`
Ottimizzato per la memoria	

Classe istanza	Tipi di istanza
	`r5.xlarge | r5.2xlarge | r5.4xlarge | r5.12xlarge | r5d.xlarge | r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge`
Ottimizzato per lo storage	`i3en.xlarge | i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge | i3en.24xlarge`

Considerazioni sulla connettività di rete

- Se la connettività di rete tra l'Outpost e la Regione AWS viene persa, i cluster continueranno a funzionare. Tuttavia, non potrai creare nuovi cluster o intraprendere nuove azioni nei cluster esistenti fino a quando la connettività non viene ripristinata. In caso di errori di istanza, l'istanza non verrà sostituita automaticamente. Inoltre, le azioni come l'aggiunta di fasi a un cluster in esecuzione, il controllo dello stato di esecuzione della fase e l'invio di parametri ed eventi CloudWatch verranno ritardate.
- Raccomandiamo di assicurarsi una connettività di rete affidabile e altamente disponibile tra il tuo Outpost e la Regione AWS. Se la connettività di rete tra Outpost e la Regione AWS viene persa per più di alcune ore, i cluster con una protezione dalle interruzioni abilitata continueranno a funzionare e i cluster con una protezione dalle interruzioni disabilitata saranno terminati.
- Se la connettività di rete subirà danni a causa della manutenzione di routine, suggeriamo di abilitare proattivamente una protezione dalle interruzioni. Più in generale, l'interruzione della connettività significa che eventuali dipendenze esterne che non sono locali per Outpost o la rete clienti non saranno accessibili. Sono inclusi Amazon S3, DynamoDB utilizzato con la visualizzazione coerente EMRFS e Amazon RDS se un'istanza interna alla Regione è utilizzata per un cluster Amazon EMR con più nodi primari.

Creazione di un cluster Amazon EMR su AWS Outposts

La creazione di un cluster Amazon EMR su AWS Outposts è simile alla creazione di un cluster Amazon EMR nel cloud AWS. Quando crei un cluster Amazon EMR in AWS Outposts, devi specificare una sottorete Amazon EC2 associata all'Outpost.

Un Amazon VPC può estendersi in tutte le zone di disponibilità in una Regione AWS. Gli AWS Outposts sono estensioni delle zone di disponibilità ed è possibile estendere un Amazon VPC in un

account per l'estensione su più zone di disponibilità e percorsi Outpost associati. Quando si configura l'Outpost, si associa a esso un gruppo di sottoreti per estendere l'ambiente regionale del VPC alla struttura locale. Le istanze Outpost e i servizi correlati fanno parte del VPC regionale, in modo simile a una zona di disponibilità con le sottoreti associate. Per ulteriori informazioni, consulta la Guida per l'utente di AWS Outposts.

Console

Per creare un nuovo cluster Amazon EMR in AWS Outposts con la AWS Management Console, specifica una sottorete Amazon EC2 associata al tuo Outpost.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Creazione di un cluster su AWS Outposts con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Cluster configuration (Configurazione del cluster), seleziona Instance groups (Gruppi di istanze) oppure Instance fleets (Parchi istanze). Quindi, scegli un tipo di istanza dal menu a discesa Choose EC2 instance type (Scegli il tipo di istanza EC2) o seleziona Actions (Operazioni) e scegli Add EBS volumes (Aggiungi volumi EBS). Amazon EMR su AWS Outposts supporta tipi di istanze e volumi limitati di Amazon EBS.
4. In Networking (Reti), seleziona una sottorete EC2 con un ID dell'Outpost in questo formato: op-123456789.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Creazione di un cluster su AWS Outposts con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate).
4. In Software Configuration (Configurazione software), per Release (Versione), scegliere 5.28.0 o versioni successive.
5. In Hardware Configuration (Configurazione hardware), per EC2 Subnet (Sottorete EC2), seleziona una sottorete EC2 con un ID dell'Outpost in questo formato: op-123456789.
6. Scegli un tipo di istanza o aggiungi volumi di archiviazione Amazon EBS per i gruppi di istanze uniformi o i parchi istanze. I tipi di istanze e volumi limitati di Amazon EBS sono supportati per Amazon EMR su AWS Outposts.

CLI

Creazione di un cluster su AWS Outposts con la AWS CLI

- Per creare un nuovo cluster Amazon EMR in AWS Outposts con la AWS CLI, specifica una sottorete EC2 associata al tuo Outpost come nell'esempio seguente. Sostituisci *subnet-22XXXX01* con l'ID della tua sottorete 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
```

Cluster EMR sulle AWS Local Zones

A partire dalla versione Amazon EMR 5.28.0, è possibile creare ed eseguire i cluster Amazon EMR in una sottorete di AWS Local Zones come estensione logica di una Regione AWS che supporta le Local Zones. Una Local Zone consente alle caratteristiche Amazon EMR e a un sottoinsieme di

servizi AWS, come i servizi di calcolo e archiviazione, di essere posizionati più vicini agli utenti per fornire accesso a bassissima latenza alle applicazioni in esecuzione localmente. Per un elenco delle zone locali disponibili, consulta Local Zones AWS. Per informazioni sull'accesso disponibile per le AWS Local Zones, consulta Regioni, zone di disponibilità e Local Zones.

Tipi di istanze supportati

I seguenti tipi di istanza sono disponibili per i cluster Amazon EMR sulle Local Zones. La disponibilità del tipo di istanza può variare in base all'area geografica.

Classe istanza	Tipi di istanza
Uso generale	m5.xlarge | m5.2xlarge | m5.4xlarge | m5.12xlarge | m5.24xlarge | m5d.xlarge | m5d.2xlarge | m5d.4xlarge | m5d.12xlarge | m5d.24xlarge
Ottimizzato per il calcolo	c5.xlarge | c5.2xlarge | c5.4xlarge | c5.9xlarge | c5.18xlarge | c5d.xlarge | c5d.2xlarge | c5d.4xlarge| c5d.9xlarge | c5d.18xlarge
Ottimizzato per la memoria	r5.xlarge | r5.2xlarge | r5.4xlarge | r5.12xlarge | r5d.xlarge | r5d.2xlarge | r5d.4xlarge | r5d.12xlarge | r5d.24xlarge
Ottimizzato per lo storage	i3en.xlarge | i3en.2xlarge | i3en.3xlarge | i3en.6xlarge | i3en.12xlarge | i3en.24xlarge

Creazione di un cluster Amazon EMR sulle Local Zones

Crea un cluster Amazon EMR sulle AWS Local Zones avviando il cluster Amazon EMR in una sottorete Amazon VPC associata a una Local Zone. È possibile accedere al cluster utilizzando il nome della Local Zone, ad esempio us-west-2-lax-1a nella console US West (Oregon) (Stati Uniti occidentali (Oregon)).

Attualmente, le Local Zones non supportano Amazon EMR Notebooks o connessioni effettuate direttamente ad Amazon EMR utilizzando l'interfaccia endpoint VPC (AWS PrivateLink).

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Creazione di un cluster su una zona locale con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Networking (Reti), seleziona una sottorete EC2 con un ID della zona locale in questo formato: subnet 123abc | us-west-2-lax-1a.
4. Scegli un tipo di istanza o aggiungi volumi di archiviazione Amazon EBS per i gruppi di istanze uniformi o i parchi istanze.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Creazione di un cluster su una zona locale con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate).
4. In Software Configuration (Configurazione software), per Release (Versione), scegliere 5.28.0 o versioni successive.
5. In Hardware Configuration (Configurazione hardware), per EC2 Subnet (Sottorete EC2), seleziona una sottorete EC2 con un ID della Local Zone in questo formato: subnet 123abc | us-west-2-lax-1a.

6. Aggiungi volumi di archiviazione Amazon EBS per i gruppi di istanze uniformi o i parchi istanze e scegli un tipo di istanza.

CLI

Creazione di un cluster su una zona locale con la AWS CLI

- Utilizza il comando create-cluster, insieme al SubnetId per la Local Zone come illustrato nell'esempio seguente. Sostituisci subnet-22XXXX1234567 con il SubnetId della Local Zone e sostituisci altre opzioni se necessario. Per ulteriori informazioni, consulta 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
```

Configura Docker

Amazon EMR 6.x supporta Hadoop 3, che consente a YARN NodeManager di avviare container direttamente sul cluster Amazon EMR o all'interno di un container Docker. I contenitori Docker forniscono ambienti di esecuzione personalizzati in cui viene eseguito il codice dell'applicazione. L'ambiente di esecuzione personalizzato è isolato dall'ambiente di esecuzione di YARN NodeManager e altre applicazioni.

I contenitori Docker possono includere librerie speciali utilizzate dall'applicazione e possono fornire diverse versioni di strumenti e librerie native, come R e Python. È possibile utilizzare gli strumenti Docker familiari per definire librerie e dipendenze di runtime per le applicazioni.

I cluster Amazon EMR 6.x sono configurati per impostazione predefinita per consentire alle applicazioni YARN, ad esempio Spark, di essere eseguite utilizzando container Docker. Per personalizzare la configurazione del container, modificare le opzioni di supporto Docker definite nei file `yarn-site.xml` e `container-executor.cfg` disponibili nella directory `/etc/hadoop/conf`. Per informazioni dettagliate su ciascuna opzione di configurazione e su come viene utilizzata, consulta Avvio di applicazioni utilizzando contenitori Docker.

È possibile scegliere di utilizzare Docker quando si invia un processo. Utilizzare le variabili seguenti per specificare il runtime Docker e l'immagine Docker.

- `YARN_CONTAINER_RUNTIME_TYPE=docker`
- `YARN_CONTAINER_RUNTIME_DOCKER_IMAGE={DOCKER_IMAGE_NAME}`

Quando si utilizzano contenitori Docker per eseguire le applicazioni YARN, YARN scarica l'immagine Docker specificata quando si invia il lavoro. Per fare in modo che YARN risolva questa immagine Docker, è necessario configurarla con un Registro di sistema Docker. Le opzioni di configurazione per un Registro di sistema Docker dipendono dal fatto che si distribuisca il cluster utilizzando una sottorete pubblica o privata.

Registri Docker

Un registro Docker è un sistema di archiviazione e distribuzione per le immagini Docker. Per Amazon EMR ti consigliamo di usare Amazon ECR, che è un registro del container Docker completamente gestito, che consente di creare le proprie immagini personalizzate e ospitarle in un'architettura altamente disponibile e scalabile.

Considerazioni sulla distribuzione

I registri Docker richiedono l'accesso alla rete da ciascun host nel cluster. Questo perché ogni host scarica immagini dal Registro di sistema Docker quando l'applicazione YARN è in esecuzione nel cluster. Questi requisiti di connettività di rete possono limitare la scelta del Registro di sistema Docker, a seconda che si distribuisca il cluster Amazon EMR in una sottorete pubblica o privata.

Public subnet (Sottorete pubblica)

Quando i cluster EMR vengono distribuiti in una sottorete pubblica, i nodi che eseguono YARN NodeManager possono accedere direttamente a qualsiasi Registro di sistema disponibile su Internet.

Sottorete privata

Quando i cluster EMR vengono distribuiti in una sottorete privata, i nodi che eseguono YARN NodeManager non hanno accesso diretto a Internet. Le immagini Docker possono essere ospitate in Amazon ECR e accessibili tramite AWS PrivateLink.

Per ulteriori informazioni su come utilizzare AWS PrivateLink per consentire l'accesso ad Amazon ECR in uno scenario di sottorete privata, consulta Impostazione di AWS PrivateLink per Amazon ECS e Amazon ECR.

Configurazione dei registri Docker

Per utilizzare i registri Docker con Amazon EMR, è necessario configurare Docker per considerare attendibile il Registro di sistema specifico che si desidera utilizzare per risolvere le immagini Docker. I registri di trust predefiniti sono locali (privati) e centos. Per utilizzare altri repository pubblici oppure Amazon ECR, è possibile ignorare le impostazioni `docker.trusted.registries` in `/etc/hadoop/conf/container-executor.cfg` utilizzando l'API di classificazione EMR con la chiave di classificazione `container-executor`.

Nell'esempio seguente viene illustrato come configurare il cluster per considerare attendibile sia un repository pubblico, denominato `your-public-repo`, sia un endpoint del Registro di sistema ECR, denominato `123456789123.dkr.ecr.us-east-1.amazonaws.com`. Se si utilizza ECR, sostituire questo endpoint con l'endpoint ECR specifico.

```
[
  {
    "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"
            }
        }
    ]
  }
]
```

Per avviare un cluster Amazon EMR 6.0.0 con questa configurazione utilizzando la AWS Command Line Interface (AWS CLI), crea un file denominato `container-executor.json` con il contenuto della configurazione JSON container-executor precedente. Quindi, utilizzare i seguenti comandi per avviare il cluster.

```
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
```

Configurazione di YARN per accedere ad Amazon ECR su EMR 6.0.0 e versioni precedenti

Se sei nuovo su Amazon ECR, segui le istruzioni riportate in Guida introduttiva di Amazon ECR e verifica di avere accesso ad Amazon ECR da ogni istanza del cluster Amazon EMR.

Su EMR 6.0.0 e versioni successive, per accedere ad Amazon ECR utilizzando il comando Docker è necessario innanzitutto generare le credenziali. Per verificare se YARN può accedere alle immagini da Amazon ECR, utilizza la variabile di ambiente container `YARN_CONTAINER_RUNTIME_DOCKER_CLIENT_CONFIG` per passare un riferimento alle credenziali generate.

Eseguire il seguente comando su uno dei nodi principali per ottenere la riga di accesso per l'account ECR.

```
aws ecr get-login --region us-east-1 --no-include-email
```

Il comando `get-login` genera il comando CLI Docker corretto da eseguire per creare le credenziali. Copiare ed eseguire l'output da `get-login`.

```
sudo docker login -u AWS -p <password> https://<account-id>.dkr.ecr.us-
east-1.amazonaws.com
```

Questo comando genera un file `config.json` nella cartella `/root/.docker`. Copia questo file in HDFS in modo che i processi inviati al cluster possano utilizzarlo per autenticarsi in Amazon ECR.

Eseguire i comandi riportati di seguito per copiare il file `config.json` nella directory home.

```
mkdir -p ~/.docker
sudo cp /root/.docker/config.json ~/.docker/config.json
sudo chmod 644 ~/.docker/config.json
```

Eseguire i comandi riportati di seguito per inserire config.json in HDFS in modo che possa essere utilizzato dai processi in esecuzione nel cluster.

```
hadoop fs -put ~/.docker/config.json /user/hadoop/
```

YARN può accedere a ECR come registro immagini Docker ed estrarre contenitori durante l'esecuzione del processo.

Dopo aver configurato i registri Docker e YARN, è possibile eseguire applicazioni YARN utilizzando contenitori Docker. Per ulteriori informazioni, consulta Esecuzione di applicazioni Spark con Docker utilizzando Amazon EMR 6.0.0.

In EMR 6.1.0 e versioni successive, non è necessario configurare manualmente l'autenticazione su Amazon ECR. Se viene rilevato un registro Amazon ECR nella chiave di classificazione `container-executor`, viene attivata la funzione di autenticazione automatica Amazon ECR e YARN gestisce il processo di autenticazione quando si invia un processo Spark con un'immagine ECR. È possibile verificare se l'autenticazione automatica è abilitata controllando `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` in yarn-site. L'autenticazione automatica è abilitata e l'impostazione di autenticazione YARN è impostata su `true` se `docker.trusted.registries` contiene un URL del registro ECR.

Prerequisiti per l'utilizzo dell'autenticazione automatica per Amazon ECR

- EMR versione 6.1.0 o versioni successive
- Il registro ECR incluso nella configurazione si trova nella stessa Regione con il cluster
- Ruolo IAM con autorizzazioni per ottenere token di autorizzazione ed estrarre qualsiasi immagine

Per ulteriori informazioni, consulta Configurazione di Amazon ECR.

Come abilitare l'autenticazione automatica

Segui Configurazione dei registri Docker per impostare un registro Amazon ECR come registro attendibile e assicurarti che il repository Amazon ECR e il cluster si trovino nella stessa Regione.

Per attivare questa caratteristica anche quando il registro ECR non è impostato nel registro attendibile, utilizza la classificazione di configurazione per impostare `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` su `true`.

Come disabilitare l'autenticazione automatica

Per impostazione predefinita, l'autenticazione automatica viene disattivata se non viene rilevato alcun registro Amazon ECR nel registro attendibile.

Per disattivare l'autenticazione automatica anche quando il registro Amazon ECR è impostato nel registro attendibile, utilizza la classificazione di configurazione per impostare `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled` su `false`.

Come verificare se l'autenticazione automatica è abilitata in un cluster

Sul nodo master, utilizza un editor di testo come `vi` per visualizzare il contenuto del file: `vi /etc/hadoop/conf.empty/yarn-site.xml`. Verifica il valore di `yarn.nodemanager.runtime.linux.docker.ecr-auto-authentication.enabled`.

Controllo della terminazione di un cluster

Questa sezione descrive le opzioni per la arresta dei cluster Amazon EMR. Copre la protezione di terminazione automatica e terminazione e il modo in cui interagiscono con altre caratteristiche di Amazon EMR.

Puoi arrestare un cluster Amazon EMR nei seguenti modi:

- Terminazione dopo l'esecuzione dell'ultima fase: crea un cluster transitorio che si arresta dopo il completamento di tutti i fasi.
- Terminazione automatica (dopo inattività): crea un cluster con una policy di terminazione automatica che si arresta dopo un periodo di inattività specificato. Per ulteriori informazioni, consulta Utilizzo di una policy di terminazione automatica.
- Terminazione manuale: crea un cluster di lunga durata che continua a essere eseguito finché non viene terminato deliberatamente. Per informazioni su come terminare un cluster manualmente, consulta Terminazione di un cluster.

È inoltre possibile impostare la protezione di terminazione su un cluster per evitare di arrestare le istanze EC2 per sbaglio o per errore.

Quando Amazon EMR arresta il cluster, tutte le istanze Amazon EC2 nel cluster vengono arrestate. I dati dell'istanza e i volumi EBS non sono più disponibili né recuperabili. La comprensione e la gestione della terminazione dei cluster è fondamentale per sviluppare una strategia di gestione e conservazione dei dati per le operazioni di scrittura su Amazon S3 e il bilanciamento dei costi.

Argomenti

Configurazione di un cluster per continuare o terminare dopo l'esecuzione della fase

Questo argomento spiega le differenze tra l'utilizzo di un cluster di lunga durata e la creazione di un cluster transitorio che si arresta dopo l'esecuzione dell'ultimo passaggio. Viene inoltre descritto come configurare l'esecuzione delle fasi per un cluster.

Crea un cluster di lunga durata

Per impostazione predefinita, i cluster creati con la console o la AWS CLI sono di lunga durata. I cluster di lunga durata continuano a funzionare, accettano il lavoro e accumulano addebiti fino a quando non si interviene per arrestarli.

Un cluster di lunga durata è efficace nelle situazioni seguenti:

- Quando è necessario eseguire una query sull dati in modo interattivo o automatico.
- Quando è necessario interagire regolarmente con applicazioni Big Data ospitate sul cluster.
- Quando se si esegue un'elaborazione periodica su un set di dati di dimensioni talmente grandi che risulterebbe inefficace avviare nuovi cluster e caricare ogni volta i dati.

È inoltre possibile impostare la protezione di terminazione su un cluster di lunga durata per evitare di arrestare le istanze EC2 accidentalmente o per errore. Per ulteriori informazioni, consulta Utilizzo della protezione da cessazione.

> **Note**
>
> Amazon EMR abilita automaticamente la protezione da terminazione per tutti i cluster con più nodi primari e sostituisce tutte le impostazioni di esecuzione di fasi fornite durante la creazione del cluster. È possibile disattivare la protezione da cessazione dopo l'avvio del cluster. Per informazioni, consultare Configurazione della protezione da cessazione per i cluster in esecuzione. Per chiudere un cluster con più nodi primari, è necessario modificare gli attributi del cluster per disabilitare la protezione da terminazione. Per istruzioni, consultare Terminazione di un cluster Amazon EMR con più nodi primari.

Configurare un cluster da terminare dopo l'esecuzione della fase

Quando si configura la terminazione dopo l'esecuzione della fase, il cluster viene avviato, esegue le azioni bootstrap e quindi esegue le fasi specificate. Non appena viene completato l'ultima fase, Amazon EMR termina le istanze Amazon EC2 del cluster. Per i cluster avviati utilizzando l'API di Amazon EMR, l'esecuzione delle fasi è abilitata per impostazione predefinita.

La terminazione dopo l'esecuzione della fase è efficace per i cluster che esegue un'operazione di elaborazione periodica, ad esempio un'esecuzione giornaliera dell'elaborazione dei dati. L'esecuzione delle fasi consente inoltre di garantire che viene fatturato solo il tempo necessario per elaborare i dati. Per ulteriori informazioni sulle fasi, consulta Invio di lavoro a un cluster.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Attivazione dell'esecuzione delle fasi con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.

3. In Steps (Fasi), scegli Add step (Aggiungi fase). Inserisci i valori appropriati nei campi della finestra di dialogo Add step (Aggiungi fase). Le opzioni variano a seconda del tipo di fase. Per aggiungere la fase e uscire dalla finestra di dialogo, scegli Add step (Aggiungi fase).
4. In Cluster termination (Terminazione del cluster), seleziona la casella di controllo Terminate cluster after last step completes (Termina il cluster dopo il completamento dell'ultima fase).
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Attivazione dell'esecuzione delle fasi con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Step execution (Esecuzione fase).
4. Scegli le altre impostazioni appropriate per l'applicazione in uso, quindi scegli Create Cluster (Crea cluster).

AWS CLI

Attivazione dell'esecuzione delle fasi con la AWS CLI

- Specifica il parametro `--auto-terminate` quando utilizzi il comando `create-cluster` per creare un cluster transitorio.

  L'esempio seguente illustra come utilizzare il parametro `--auto-terminate`. È possibile digitare il comando seguente e sostituire *myKey* con il nome della coppia di chiavi EC2.

  > **Note**
  >
  > I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

```
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

Disattivazione dell'esecuzione delle fasi con l'API di Amazon EMR

- Quando si utilizza l'operazione RunJobFlow per creare un cluster, è necessario impostare la proprietà KeepJobFlowAliveWhenNoSteps su `true`.

Utilizzo di una policy di terminazione automatica

Una policy di terminazione automatica consente di orchestrare la pulizia del cluster senza la necessità di monitorare e terminare manualmente i cluster inutilizzati. Quando si aggiunge una policy di terminazione automatica a un cluster, si specifica la quantità di tempo di inattività dopo il quale il cluster deve arrestarsi automaticamente.

A seconda della versione di rilascio, Amazon EMR utilizza criteri diverse per contrassegnare un cluster come inattivo. Nella tabella seguente viene illustrato come Amazon EMR determina l'inattività del cluster.

Quando utilizzi...	Un cluster è considerato inattivo quando...
Amazon EMR versione 5.34.0 e successive, e versione 6.4.0 e successive	• Non esistono applicazioni YARN attive • L'utilizzo dell'HDFS è inferiore al 10% • Non ci sono connessioni attive per EMR Notebooks o EMR Studio •

Quando utilizzi...	Un cluster è considerato inattivo quando...
	Non sono in uso interfacce utente dell'appl icazione on-cluster
Amazon EMR versioni 5.30.0 - 5.33.0 e 6.1.0 - 6.3.0	• Non esistono applicazioni YARN attive • Il cluster non ha processi Spark attivi **Note** Amazon EMR contrassegna un cluster come inattivo e potrebbe terminare automaticamente il cluster anche se si dispone di un kernel Python3 attivo. Questo perché l'esecuzione di un kernel Python3 non invia un processo Spark sul cluster. Per utilizzar e la terminazione automatica con un kernel Python3, consigliamo di utilizzare Amazon EMR versione 6.4.0 o successive.

Note

Amazon EMR versione 6.4.0 e successive supportano un file su cluster per il rilevamento dell'attività sul nodo primario: `/emr/metricscollector/isbusy`. Quando si utilizza un cluster per eseguire script di shell o applicazioni non YARN, è possibile toccare o aggiornare periodicamente `isbusy` per indicare ad Amazon EMR che il cluster non è inattivo.

È possibile allegare una policy di terminazione automatica quando si crea un cluster o si aggiunge una policy a un cluster esistente. Per modificare o disabilitare la terminazione automatica, è possibile aggiornare o rimuovere la policy.

Considerazioni

Prima di utilizzare una policy di terminazione automatica, considera le seguenti caratteristiche e limitazioni:

- In Asia Pacific (Giacarta), la terminazione automatica di Amazon EMR è disponibile con Amazon EMR 6.14.0 e versioni successive.
- Nelle seguenti Regioni AWS, la terminazione automatica di Amazon EMR è disponibile con Amazon EMR versione 5.30.0, 6.1.0 e successive:

  Stati Uniti orientali (Virginia settentrionale e Ohio), Stati Uniti occidentali (Oregon e California settentrionale), Sud America (San Paolo), Europa (Francoforte, Irlanda, Londra, Milano, Parigi e Stoccolma), Canada (Centrale), Asia Pacifico (Hong Kong), Mumbai, Seoul, Singapore, Sydney e Tokyo), Medio Oriente (Bahrein), Africa (Città del Capo), AWS GovCloud (Stati Uniti orientali), AWS GovCloud (Stati Uniti occidentali),Cina (Pechino) gestito da Sinnet, e Cina (Ningxia) gestito da NWCD.
- Il timeout inattivo è predefinito di 60 minuti (un'ora) quando non si specifica un importo. È possibile specificare un timeout minimo di inattività di un minuto e un timeout massimo di 7 giorni.
- Con Amazon EMR versioni 6.4.0 e successive, la terminazione automatica è abilitata per impostazione predefinita quando si crea un nuovo cluster tramite la console Amazon EMR.
- Amazon EMR pubblica metriche Amazon CloudWatch ad alta risoluzione quando si abilita la terminazione automatica per un cluster. Puoi utilizzare queste metriche per monitorare l'attività e l'inattività del cluster. Per ulteriori informazioni, consulta Parametri della capacità del cluster.
- La terminazione automatica non è supportata quando si utilizzano applicazioni non basate su YARN come Presto, Trino o HBase.
- Per utilizzare la terminazione automatica in API Gateway, il processo di raccolta dei parametri deve essere in grado di connettersi all'endpoint API pubblico. Se utilizzi un nome DNS privato con Amazon Virtual Private Cloud, la terminazione automatica non funzionerà correttamente. Per garantire che la terminazione automatica funzioni, è consigliabile eseguire una delle seguenti operazioni:
  - Rimuovi l'endpoint VPC dell'interfaccia API Gateway dal tuo Amazon VPC.
  - Segui le istruzioni in Why do I get an HTTP 403 Forbidden error when connecting to my API Gateway APIs from a VPC? (Perché visualizzo un errore HTTP 403 Forbidden durante la connessione alle API di API Gateway da un VPC?) per disabilitare l'impostazione del nome DNS privato.

  - In alternativa, avvia il cluster in una sottorete privata. Per ulteriori informazioni, consulta l'argomento in Sottoreti private.
- (EMR rilascio 5.30.0 e successivi) Se si rimuove la regola predefinita Allow All (Consenti tutto) in uscita su 0.0.0.0/ nel gruppo di sicurezza primario, è necessario aggiungere una regola per consentire la connettività TCP in uscita al gruppo di sicurezza di accesso al servizio sulla porta 9443. Inoltre, il gruppo di sicurezza di accesso al servizio deve consentire il traffico TCP in ingresso sulla porta 9443 dal gruppo di sicurezza primario. Per ulteriori informazioni sulla configurazione dei gruppi di sicurezza, consulta la sezione Amazon EMR-managed security group for the primary instance (private subnets) (Gruppo di sicurezza gestito da Amazon EMR per l'istanza primaria [sottoreti private]).

Autorizzazioni per l'utilizzo della terminazione automatica

Per poter applicare e gestire le policy di terminazione automatica per Amazon EMR, devi collegare le autorizzazioni elencate nell'esempio seguente sulle policy di autorizzazioni IAM alle risorse IAM che gestiscono il cluster EMR.

```
{
  "Version": "2012-10-17",
  "Statement": {
      "Sid": "AllowAutoTerminationPolicyActions",
      "Effect": "Allow",
      "Action": [
        "elasticmapreduce:PutAutoTerminationPolicy",
        "elasticmapreduce:GetAutoTerminationPolicy",
        "elasticmapreduce:RemoveAutoTerminationPolicy"
      ],
      "Resource": "<your-resources>"
    }
}
```

Allega, aggiorna o rimuovi una policy di terminazione automatica

Questa sezione include istruzioni per allegare, aggiornare o rimuovere una policy di terminazione automatica da un cluster Amazon EMR. Prima di lavorare con le policy di terminazione automatica, assicurati di disporre delle autorizzazioni IAM necessarie. Per informazioni, consultare Autorizzazioni per l'utilizzo della terminazione automatica.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Collegamento di una policy di terminazione automatica durante la creazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Cluster termination (Terminazione del cluster), seleziona Terminate cluster after idle time (Termina il cluster dopo il tempo di inattività).
4. Specifica il numero di ore e minuti di inattività dopo i quali il cluster deve terminare automaticamente. Il tempo di inattività predefinito è 1 ora.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Collegamento, aggiornamento o rimozione di una policy di terminazione automatica su un cluster in esecuzione con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Nella scheda Properties (Proprietà) della pagina dei dettagli del cluster, cerca Cluster termination (Terminazione del cluster) e seleziona Edit (Modifica).
4. Seleziona o deseleziona Enable auto-termination (Abilita la terminazione automatica) per attivare o disattivare la caratteristica. Se attivi la terminazione automatica, specifica il numero di ore e minuti di inattività dopo il quale il cluster termina automaticamente. Quindi seleziona Save changes (Salva modifiche) per confermare.

Old console

Collegamento di una policy di terminazione automatica durante la creazione di un cluster con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. In Configurazione hardware, seleziona Terminazione automatica.
4. Specificare il numero di ore e minuti di inattività dopo i quali il cluster deve terminare automaticamente. Il tempo di inattività di default è un'ora.
5. Scegli le altre impostazioni appropriate per l'applicazione in uso, quindi scegli Create Cluster (Crea cluster).

Collegamento, aggiornamento o rimozione di una policy di terminazione automatica su un cluster in esecuzione con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Cluster e scegli il cluster da aggiornare.
3. Nella scheda Hardware nella pagina cluster detail.
4. Seleziona o deseleziona Enable auto-termination (Abilita la terminazione automatica) per attivare o disattivare la caratteristica. Se attivi la terminazione automatica, specifica il numero di ore e minuti di inattività dopo il quale il cluster deve terminare automaticamente.

AWS CLI

Prima di iniziare

Prima di utilizzare le policy di terminazione automatica, si consiglia di eseguire l'aggiornamento alla versione più recente della AWS CLI. Per le istruzioni, consulta Installazione, aggiornamento e disinstallazione della AWS CLI.

Per allegare o aggiornare una policy di terminazione automatica utilizzando la AWS CLI

- Puoi utilizzare il comando `aws emr put-auto-termination-policy` per allegare o aggiornare una policy di terminazione automatica su un cluster.

  L'esempio seguente specifica 3600 secondi per *`IdleTimeout`*. Se non indichi *`IdleTimeout`*, il valore di default è di un'ora.

  ```
  aws emr put-auto-termination-policy \
  --cluster-id <your-cluster-id> \
  --auto-termination-policy IdleTimeout=3600
  ```

  > **Note**
  >
  > I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

  È anche possibile specificare un valore per `--auto-termination-policy` quando si utilizza il comando `aws emr create-cluster`. Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta la Guida di riferimento ai comandi della AWS CLI.

Rimozione di una policy di terminazione automatica con la AWS CLI

- Utilizzo il comando `aws emr remove-auto-termination-policy` per rimuovere una policy di terminazione automatica da un cluster. Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta la Guida di riferimento ai comandi della AWS CLI.

  ```
  aws emr remove-auto-termination-policy --cluster-id <your-cluster-id>
  ```

Utilizzo della protezione da cessazione

Quando la protezione da cessazione è abilitata su un cluster di lunga durata, è necessario rimuovere esplicitamente la protezione da cessazione dal cluster prima di poterlo terminare. Ciò aiuta a

garantire che le istanze EC2 non verranno arrestate in modo accidentale o per errore. Questo tipo di protezione è particolarmente utile nel caso in cui il cluster includa dati archiviati sui dichi locali che devono essere recuperati prima che le istanze vengano terminate. È possibile abilitare la protezione da cessazione al momento della creazione di un cluster ed è possibile modificare l'impostazione su un cluster in esecuzione.

Se la protezione da cessazione è abilitata, l'operazione `TerminateJobFlows` nell'API di Amazon EMR non funziona. Gli utenti non sono in grado di terminare il cluster utilizzando questa API né il comando `terminate-clusters` di AWS CLI. L'API restituisce un errore e l'interfaccia a riga di comando (CLI) si chiude con un codice restituito diverso da zero. Se utilizzi la console Amazon EMR per terminare un cluster, ti viene richiesto di eseguire una fase supplementare per disattivare la protezione da terminazione.

> **Warning**
>
> La protezione da cessazione non garantisce che i dati vengano conservati in caso di errore umano o di soluzione alternativa, ad esempio se un comando di riavvio viene emesso dalla riga di comando mentre è connesso all'istanza tramite SSH, se un'applicazione o uno script in esecuzione sull'istanza emette un comando di riavvio o se l'API Amazon EC2 o Amazon EMR viene utilizzata per disattivare la protezione da cessazione. Anche se la protezione da cessazione è abilitata, i dati salvati nell'archiviazione dell'istanza, inclusi i dati HDFS, possono andare persi. Scrivi l'output dei dati nei percorsi Amazon S3 e crea strategie di backup appropriate per i tuoi requisiti di business continuity.

Questo tipo di protezione non pregiudica la possibilità di dimensionare le risorse del cluster mediante una delle seguenti operazioni:

- Ridimensionamento manuale di un cluster mediante la AWS Management Console o AWS CLI. Per ulteriori informazioni, consulta Ridimensionamento manuale di un cluster in esecuzione.
- Rimozione delle istanze da un gruppo di istanze principali o di task utilizzando una policy di riduzione con scalabilità automatica. Per ulteriori informazioni, consulta Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze .
- Rimozione delle istanze da un parco istanze mediante la riduzione della capacità di destinazione. Per ulteriori informazioni, consulta Opzioni del parco istanze.

Protezione da cessazione e Amazon EC2

Un cluster Amazon EMR con protezione da cessazione abilitata ha l'attributo `disableAPITermination` impostato per tutte le istanze Amazon EC2 nel cluster. Se una richiesta di cessazione proviene da Amazon EMR e le impostazioni Amazon EMR e Amazon EC2 per un'istanza sono in conflitto, l'impostazione di Amazon EMR sostituisce l'impostazione Amazon EC2. Ad esempio, se si utilizza la console Amazon EC2 per abilitare la protezione da terminazione su un'istanza Amazon EC2 in un cluster per il quale la protezione da terminazione è disabilitata, quando si utilizza la console Amazon EMR, i comandi della AWS CLI per Amazon EMR o l'API Amazon EMR per terminare il cluster, Amazon EMR imposta `DisableApiTermination` su `false` e termina l'istanza insieme a tutte le altre.

> **Important**
>
> Se un'istanza viene creata come parte di un cluster Amazon EMR con protezione da cessazione e l'API Amazon EC2 o i comandi della AWS CLI vengono utilizzati per modificare l'istanza in modo che `DisableApiTermination` sia impostata su `false`, l'API Amazon EC2 o i comandi della AWS CLI eseguono l'operazione `TerminateInstances` e l'istanza Amazon EC2 viene terminata.

Protezione da cessazione e nodi YARN con stato Unhealthy (Non integro)

Amazon EMR controlla periodicamente lo stato dei nodi Apache Hadoop YARN in esecuzione sulle istanze Amazon EC2 principali o attività in un cluster. Lo stato di integrità viene segnalato dal servizio Health Checker Service di NodeManager. Se un nodo viene segnalato come `UNHEALTHY`, il controller dell'istanza Amazon EMR elenca il nodo come negato e non assegna i container YARN a tale nodo finché non assume di nuovo lo stato "healthy (integro)". La mancata integrità dei nodi è dovuta in genere a un utilizzo del disco al di sopra del 90%. Per ulteriori informazioni su come identificare e recuperare i nodi con stato Unhealthy, consulta Errori nelle risorse.

Se il nodo rimane nello stato `UNHEALTHY` per più di 45 minuti, Amazon EMR effettua la seguente operazione in base allo stato di protezione da cessazione.

Termination protection (Protezione da cessazione)	Risultato
Abilitata (consigliata)	

Termination protection (Protezione da cessazione)	Risultato
	Le istanze principali di Amazon EC2 rimangono nello stato elencato come negato e continuano a essere conteggiate per il raggiungimento della capacità del cluster. È possibile connettersi all'istanza core Amazon EC2 per eseguire la configurazione e il recupero dei dati e ridimensionare il cluster per aggiungere capacità. Per ulteriori informazioni, consulta Errori nelle risorse. I nodi attività non integri sono esenti dalla protezione da cessazione e verranno terminati.

Termination protection (Protezione da cessazione)	Risultato
Disabilitato	L'istanza Amazon EC2 è terminata. Amazon EMR esegue il provisioning di una nuova istanza in base al numero di istanze specificato nel gruppo di istanze o alla capacità target per il parco istanze. Se tutti i nodi principali presentano lo stato `UNHEALTHY` per più di 45 minuti, il cluster viene terminato e viene restituito lo stato `NO_SLAVES_LEFT` . **Important** Si potrebbe verificare la perdita dei dati HDFS se un'istanza principale termina a causa di uno stato Unhealthy. Se il nodo ha archiviato blocchi non replicati su altri nodi, tali blocchi vengono persi e ciò potrebbe determinare la perdita dei dati. È consigliabile utilizzare la protezione da cessazione in modo da connettersi alle istanze e ripristinare i dati, quando necessario.

Protezione della terminazione ed esecuzione delle fasi

Quando si abilita l'esecuzione delle fasi e si abilita anche la protezione della terminazione, Amazon EMR ignora la protezione della terminazione.

Quando si inviano fasi a un cluster, è possibile impostare la proprietà `ActionOnFailure` per stabilire cosa accede se la fase non può completare l'esecuzione a causa di un errore. I valori possibile per questa impostazione sono `TERMINATE_CLUSTER` (`TERMINATE_JOB_FLOW` con le versioni precedenti), `CANCEL_AND_WAIT` e `CONTINUE`. Per ulteriori informazioni, consulta Invio di lavoro a un cluster.

Se una fase, configurata con `ActionOnFailure` impostato su `CANCEL_AND_WAIT`, non riesce e se l'esecuzione della fase è attivata, il cluster viene terminato senza eseguire le fasi successive.

Se una fase, configurata con `ActionOnFailure` impostato su `TERMINATE_CLUSTER`, non riesce, è possibile utilizzare la tabella delle impostazioni riportata di seguito per determinare il risultato.

ActionOnFailure	Esecuzione della fase	Termination protection (Protezione da cessazione)	Risultato
`TERMINATE_CLUSTER`	Abilitato	Disabilitato	Il cluster viene terminato
	Abilitato	Abilitato	Il cluster viene terminato
	Disabilitato	Abilitato	Il cluster non viene terminato
	Disabilitato	Disabilitato	Il cluster viene terminato

Protezione da cessazione e istanze Spot

L'opzione di protezione da cessazione di Amazon EMR non impedisce l'arresto di un'istanza Spot di Amazon EC2 quando il prezzo Spot supera il prezzo Spot massimo.

Configurazione della protezione da cessazione all'avvio di un cluster

È possibile attivare o disattivare la protezione da cessazione quando si avvia un cluster utilizzando la console, AWS CLI o l'API.

L'impostazione di protezione da cessazione predefinita dipende dalla modalità di avvio del cluster:

- Console (nuova versione) di Amazon EMR: la protezione da terminazione è abilitata per impostazione predefinita.

- Opzioni rapide della console (vecchia versione) di Amazon EMR: la protezione da terminazione è disabilitata per impostazione predefinita.
- Opzioni avanzate della console (vecchia versione) di Amazon EMR: la protezione da terminazione è abilitata per impostazione predefinita.
- AWS CLI `aws emr create-cluster` - La protezione da cessazione è disabilitata a meno che non sia specificato `--termination-protected`.
- Comando RunJobFlow dell'API di Amazon EMR: la protezione da cessazione è disabilitata a meno che il valore booleano `TerminationProtected` non sia impostato su `true`.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Attivazione o disattivazione della protezione da terminazione durante la creazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. Per EMR release version (Versione del rilascio EMR), scegli emr-6.6.0 o un rilascio successivo.
4. In Cluster termination (Terminazione del cluster), assicurati che l'opzione Use termination protection (Utilizza la protezione da terminazione) sia preselezionata oppure annulla la selezione per disattivarla.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Attivazione o disattivazione della protezione da terminazione durante la creazione di un cluster con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate).
4. Per Step 3: General Cluster Settings (Fase 3: impostazioni generali del cluster), in General Options (Opzioni generali) accertati che l'opzione Termination protection (Protezione da cessazione) sia selezionata per abilitarla oppure deselezionala per disattivarla.
5. Scegli le altre impostazioni appropriate per l'applicazione, scegli Next (Avanti), quindi completa la configurazione del cluster.

AWS CLI

Attivazione o disattivazione della protezione da terminazione durante la creazione di un cluster con la AWS CLI

- Con AWS CLI, è possibile avviare un cluster con protezione da terminazione abilitata digitando il comando `create-cluster` con il parametro `--termination-protected`. Per impostazione predefinita, la protezione da cessazione è disabilitata.

  Nell'esempio seguente viene creato un cluster con protezione da cessazione abilitata:

  > **Note**
  >
  > I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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
```

Per ulteriori informazioni sull'utilizzo dei comandi Amazon EMR nella AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Configurazione della protezione da cessazione per i cluster in esecuzione

È possibile configurare la protezione da terminazione per un cluster in esecuzione con la console o la AWS CLI.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Attivazione o disattivazione della protezione da terminazione per un cluster in esecuzione con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Nella scheda Properties (Proprietà) della pagina dei dettagli del cluster, cerca Cluster termination (Terminazione del cluster) e seleziona Edit (Modifica).
4. Seleziona o deseleziona la casella di controllo Use termination protection (Utilizza la protezione da terminazione) per attivare o disattivare la caratteristica. Quindi seleziona Save changes (Salva modifiche) per confermare.

Old console

Abilitazione o disabilitazione della protezione da terminazione per un cluster in esecuzione con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Nella pagina Clusters (Cluster), scegli il Name (Nome) del cluster.
3. Nella scheda Summary (Riepilogo), per Termination protection (Protezione da cessazione), scegli Change (Modifica).
4. Per abilitare la protezione da cessazione, scegli On (Attiva). Per disabilitare la protezione da cessazione, scegli Off (Disattivata). Seleziona quindi il segno di spunta verde per confermare.

AWS CLI

Abilitazione o disabilitazione della protezione da terminazione per un cluster in esecuzione con la AWS CLI

- Per abilitare la protezione da terminazione su un cluster in esecuzione con la AWS CLI, utilizza il comando `modify-cluster-attributes` con il parametro `--termination-protected`. Per disabilitarla, utilizza il parametro `--no-termination-protected`.

  Nell'esempio seguente viene abilitata la protezione da cessazione sul cluster con ID *j-3KVTXXXXXX7UG*:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --termination-protected
  ```

  Nell'esempio seguente viene disabilitata la protezione da cessazione sullo stesso cluster:

  ```
  aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
  ```

Lavorare con le AMI Amazon Linux in Amazon EMR

Amazon Machine Image (AMI) di Amazon Linux

Amazon EMR utilizza un'Amazon Machine Image (AMI) Amazon Linux per inizializzare le istanze Amazon EC2 al momento della creazione e dell'avvio di un cluster. L'AMI contiene il sistema operativo Amazon Linux, altro software e le configurazioni richieste per l'hosting su ogni istanza delle applicazioni del cluster.

Per impostazione predefinita, al momento della creazione di un cluster, Amazon EMR utilizza un'AMI Amazon Linux predefinita creata appositamente per la versione di Amazon EMR che utilizzi. Per ulteriori informazioni sull'AMI Amazon Linux di default, consulta Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR. Se utilizzi Amazon EMR 5.7.0 o versioni successive, puoi scegliere di specificare un'AMI Amazon Linux personalizzata anziché l'AMI Amazon Linux predefinita per Amazon EMR. Un'AMI personalizzata consente di crittografare il volume dispositivo root e di personalizzare le applicazioni e le configurazioni come alternativa all'utilizzo delle operazioni di bootstrap. È possibile specificare un'AMI personalizzata per ogni tipo di istanza nella configurazione del gruppo di istanze o del parco istanze di un cluster Amazon EMR. Il supporto AMI personalizzato multiplo offre la flessibilità di utilizzare più tipi di architettura in un cluster. Per informazioni, consultare Utilizzo di un'AMI personalizzata.

Amazon EMR allega automaticamente un volume General Purpose SSD di Amazon EBS come dispositivo di root per tutte le AMI. L'utilizzo di un'AMI supportata da EBS consente di migliorare le prestazioni. I costi Amazon EBS vengono calcolati all'ora in base alle tariffe Amazon EBS mensili per volumi gp2 nella Regione in cui viene eseguito il cluster. Ad esempio, il costo all'ora per il volume root su ogni istanza del cluster in una Regione che addebita $0,10/GB/mese è di circa $0,00139 all'ora ($0,10/GB/mese diviso per 30 giorni diviso per 24 ore moltiplicato per 10 GB). Sia che venga utilizzata un'AMI Amazon Linux predefinita o personalizzata, è possibile specificare per il volume dispositivo di root Amazon EBS una dimensione compresa tra 10 e 100 GiB.

Per ulteriori informazioni sull'AMI Amazon Linux, consulta Amazon Machine Images (AMI).
Per ulteriori informazioni sull'archiviazione dell'istanza di Amazon EMR, consulta Archiviazione dell'istanza.

Utilizzo dell'AMI Amazon Linux predefinita per Amazon EMR

Ogni versione di Amazon EMR utilizza un'AMI Amazon Linux predefinita per Amazon EMR a meno che non venga specificata un'AMI personalizzata. A partire dai rilasci di Amazon EMR 5.36 e Amazon

EMR 6.6, il comportamento predefinito per l'aggiornamento di Amazon Linux 2 (AL2) in un'AMI Amazon EMR predefinita consiste nell'applicare automaticamente l'ultimo rilascio AL2 all'AMI Amazon EMR predefinita.

Aggiornamenti automatici di Amazon Linux 2 per i rilasci di Amazon EMR

Quando avvii un cluster con l'ultima versione di patch di Amazon EMR 5.36 o successiva, o 6.6 o successiva, Amazon EMR utilizza l'ultima versione di Amazon Linux 2 per l'AMI Amazon EMR predefinita. Ad esempio:

- Se è presente un rilascio `x.x.0` e un rilascio `x.x.1`, il rilascio `x.x.0` non riceve più gli aggiornamenti per l'AMI all'avvio di `x.x.1`.
- Allo stesso modo, `x.x.1` non riceve più gli aggiornamenti per l'AMI all'avvio di `x.x.2`.
- Successivamente, con i rilasci `x.y.0`, `x.x.[latest]` continua a ricevere gli aggiornamenti per l'AMI insieme a `x.y.[latest]`.

Per verificare se stai utilizzando l'ultimo rilascio della patch indicato dal numero dopo il secondo punto decimale (`6.8.1`) per un rilascio di Amazon EMR, consulta i rilasci disponibili nella Guida ai rilasci di Amazon EMR, controlla il menu a discesa dei rilasci di Amazon EMR quando crei un cluster nella console o utilizza l'API `ListReleaseLabels` o l'azione della CLI `list-release-labels`. Per ricevere aggiornamenti quando avvii un nuovo rilascio di Amazon EMR, iscriviti al feed RSS nella pagina Novità della Guida ai rilasci.

Se lo desideri, puoi scegliere di avviare il cluster con la versione di Amazon Linux fornita per la prima volta con il rilascio di Amazon EMR. Per informazioni su come specificare il rilascio di Amazon Linux per il cluster, consulta Modifica del rilascio di Amazon Linux durante la creazione di un cluster.

Versioni Amazon Linux predefinite

La seguente tabella riporta informazioni su Amazon Linux per la l'ultima versione della patch di Amazon EMR 6.x rilasci 6.6 e successivi.

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 727.0	4.14.320	14 agosto 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+ e 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+ e 5.36.1)
2.0.2023 719.0	4.14.320	2 agosto 2023	us-east-1 , us-east-2 , us-west-1 , us-west-2 , eu-north-1 , eu-west-1 ,

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			`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+ e 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+ e 5.36.1)

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 628.0	4.14.318	12 luglio 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

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 612.0	4.14.314	23 giugno 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

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 504.1	4.14.313	16 maggio 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

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 418.0	4.14.311	3 maggio 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

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.20230404.1	4.14.311	18 aprile 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 aprile 2023	us-east-1 , eu-west-3

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 320.0	4.14.309	30 marzo 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

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 307.0	4.14.305	15 marzo 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-centra l-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-centra l-1 , me-south-1 , ca-central-1

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 207.0	4.14.304	3 marzo 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

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 119.1	4.14.301	9 febbraio 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 (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 210.1	4.14.301	12 gennaio 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 (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 103.3	4.14.296	5 dicembre 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 (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022004.0	4.14.294	2 novembre 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 (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	7 ottobre 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 805.0	4.14.287	30 agosto 2022	us-west-1

OsRelea Label (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 719.0	4.14.287	10 agosto 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 (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 426.0	4.14.281	10 giugno 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 (Versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 406.1	4.14.275	2 maggio 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

La seguente tabella riporta informazioni su Amazon Linux per l'ultima versione della patch di Amazon EMR 5.x rilasci 5.36 e successivi.

OsRelea Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 504.1	4.14.313	16 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapor e), Asia Pacifico (Sydney), Asia Pacific (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)
2.0.2023 418.0	4.14.311	3 maggio 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti

OsRelea Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
			occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapor e), Asia Pacifico (Sydney), Asia Pacific (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale), Medio Oriente (Emirati Arabi Uniti)

OsRelea Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 404.1	4.14.311	18 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapor e), Asia Pacifico (Sydney), Asia Pacific (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)
2.0.2023 404.0	4.14.311	10 aprile 2023	Stati Uniti orientali (Virginia settentrionale), Europa (Parigi)

OsRelea Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 320.0	4.14.309	30 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapor e), Asia Pacifico (Sydney), Asia Pacific (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsRelea Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 307.0	4.14.305	15 marzo 2023	Stati Uniti orientali (Virginia settentrionale), Stati Uniti orientali (Ohio), Stati Uniti occidentali (California settentrionale), Stati Uniti occidentali (Oregon), Europa (Stoccolma), Europa (Milano), Europa (Francoforte), Europa (Irlanda), Europa (Londra), Europa (Parigi), Asia Pacifico (Hong-Kong), Asia Pacifico (Mumbai), Asia Pacifico (Tokyo), Asia Pacifico (Seoul), Asia Pacifico (Osaka-Locale), Asia Pacifico (Singapor e), Asia Pacifico (Sydney), Asia Pacific (Giacarta), Africa (Città del Capo), Sud America (San Paolo), Medio Oriente (Bahrein), Canada (Centrale)

OsRelea Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2023 207.0	4.14.304	3 marzo 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 (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 210.1	4.14.301	12 gennaio 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 (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 103.3	4.14.296	5 dicembre 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 (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 004.0	4.14.294	2 novembre 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-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 (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 912.1	4.14.291	7 ottobre 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-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
2.0.2022 719.0	4.14.287	10 agosto 2022	us-west-1 , eu-west-3 , eu-north-1 , eu-central-1 , ap-south-1 , me-south-1

OsRelea Label (versione di Amazon Linux)	Versione del kernel di Amazon Linux	Data di disponibilità	Regioni supportate
2.0.2022 426.0	4.14.281	14 giugno 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-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

Considerazioni sull'aggiornamento del software

Di seguito sono riportati i comportamenti di aggiornamento software predefiniti di cui tenere conto:

Amazon EMR 5.35.0 e versioni precedenti, e 6.5.0 e versioni precedenti: AMI di Amazon Linux "bloccata" alla versione di rilascio di Amazon EMR

Per Amazon EMR 5.35.0 e versioni precedenti, e 6.5.0 e versioni precedenti, l'AMI predefinita si basa sull'AMI Amazon Linux più aggiornata disponibile al momento del rilascio di Amazon EMR. L'AMI è testata per la compatibilità con le applicazioni Big Data e le caratteristiche di Amazon EMR incluse in tale versione.

Per garantire la compatibilità, ogni versione di rilascio Amazon EMR 5.35.0 e precedenti e Amazon EMR 6.5.0 e precedenti è "bloccata" alla versione dell'AMI Amazon Linux rispettiva assegnata. Ciò significa che le versioni precedenti dell'AMI Amazon Linux vengono utilizzate per le versioni di rilascio precedenti di Amazon EMR anche quando sono disponibili nuove AMI Amazon Linux. Per questo motivo, consigliamo di utilizzare la versione di rilascio di Amazon EMR più recente, a meno che non sia necessaria una versione precedente per la compatibilità e non sia possibile eseguire la migrazione. Se è necessario utilizzare una versione di rilascio precedente di Amazon EMR per la compatibilità, consigliamo di utilizzare il rilascio più recente di una serie. Ad esempio, se è necessario utilizzare la serie 5.12, utilizzare la serie 5.12.2 anziché la 5.12.0 o 5.12.1. Se è disponibile una nuova versione in una serie, si consiglia di eseguire la migrazione delle applicazioni alla nuova release.

Per ulteriori informazioni sul comportamento di aggiornamento automatico introdotto con Amazon EMR 5.36.0 e versioni successive e 6.6.0 e versioni successive, consulta Aggiornamenti automatici di Amazon Linux 2 per i rilasci di Amazon EMR.

Il comportamento di avvio predefinito esclude gli aggiornamenti del kernel

Quando un'istanza Amazon EC2 in un cluster basato sull'AMI Amazon Linux predefinita per Amazon EMR viene avviata per la prima volta, verifica se nei repository di pacchetti per Amazon Linux e Amazon EMR sono presenti aggiornamenti software applicabili alla versione dell'AMI. Come con altre istanze Amazon EC2, gli aggiornamenti di sicurezza critici e importanti provenienti da questi repository vengono installati automaticamente. Tuttavia, se utilizzi una versione precedente di Amazon Linux AMI, l'aggiornamento della sicurezza più recente potrebbe non essere installato automaticamente. Questo perché i repository a cui fa riferimento EMR sono corretti per ogni versione di Amazon Linux AMI. Si noti anche che la tua configurazione di rete deve consentire l'uscita HTTP e HTTPS ai repository Amazon Linux in Amazon S3, altrimenti gli aggiornamenti della sicurezza non avranno esito positivo. Per ulteriori informazioni, consulta Repository di pacchetti nella Guida per l'utente di Amazon EC2 per le istanze Linux. Per impostazione predefinita, altri pacchetti software e aggiornamenti del kernel che richiedono un riavvio, inclusi NVIDIA e CUDA, sono esclusi dal download automatico al primo avvio.

> **Important**
>
> I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon

Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi personalizzare l'AMI di Amazon EMR per utilizzare l'AMI di Amazon Linux più recente.

Il cluster viene avviato con o senza aggiornamenti

Tieni presente che se gli aggiornamenti software non possono essere installati perché i repository dei pacchetti non sono raggiungibili al primo avvio del cluster, l'istanza del cluster continua a completare il suo avvio. Ad esempio, i repository potrebbero non essere raggiungibili perché S3 non è temporaneamente disponibile oppure è possibile che siano configurate regole VPC o firewall per bloccare l'accesso.

Non eseguire **`sudo yum update`**

Al momento della connessione a un'istanza cluster tramite SSH, nelle prime righe dell'output dello schermo vengono forniti un link alle note di rilascio per l'AMI Amazon Linux utilizzata dall'istanza, un avviso per indicare la versione AMI Amazon Linux più recente, un avviso indicante il numero di pacchetti disponibili per l'aggiornamento dai repository abilitati e una direttiva per l'esecuzione di `sudo yum update`.

> **Important**
>
> Consigliamo vivamente di non eseguire `sudo yum update` su istanze cluster, durante la connessione con SSH o mediante un'operazione di bootstrap. Questo potrebbe causare incompatibilità perché tutti i pacchetti vengono installati indistintamente.

Best practice per l'aggiornamento del software

Best practice per la gestione degli aggiornamenti software

- Se utilizzi una versione di rilascio precedente di Amazon EMR, prendi in considerazione e verifica la possibilità di effettuare una migrazione al rilascio più recente prima di aggiornare i pacchetti software.

- Se esegui la migrazione a una versione di rilascio più recente oppure aggiorni pacchetti software, verifica prima l'implementazione in un ambiente non di produzione. A tale scopo, è utile l'opzione di clonazione dei cluster con la console Amazon EMR.
- Valuta gli aggiornamenti software per le applicazioni in uso e per la versione dell'AMI Amazon Linux su base individuale. Verifica e installa i pacchetti solo negli ambienti di produzione dove sia assolutamente necessario per la sicurezza, la funzionalità o le prestazioni dell'applicazione.
- Verifica se sono disponibili aggiornamenti nel Centro di sicurezza Amazon Linux.
- Evita di installare pacchetti eseguendo il collegamento a singole istanze cluster mediante SSH. Utilizza invece un'operazione di bootstrap per installare e aggiornare i pacchetti su tutte le istanze cluster in base alle necessità. Per questa operazione è necessario chiudere e riavviare un cluster. Per ulteriori informazioni, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo.

Utilizzo di un'AMI personalizzata

Se utilizzi Amazon EMR 5.7.0 o versioni successive, puoi scegliere di specificare un'AMI Amazon Linux personalizzata anziché l'AMI Amazon Linux predefinita per Amazon EMR. Un'AMI personalizzata è utile se desideri effettuare le seguenti operazioni:

- Preinstallazione delle applicazioni ed esecuzione di altre personalizzazioni invece di utilizzare operazioni di bootstrap. In questo modo è possibile migliorare il tempo di avvio del cluster e semplificare il flusso di lavoro all'avvio. Per ulteriori informazioni e un esempio, consulta Creazione di un'AMI Amazon Linux personalizzata da un'istanza preconfigurata.
- Implementazione di configurazioni per cluster e nodo più sofisticate di quelle consentite dalle operazioni di bootstrap.
- Esegui la crittografia dei volumi del dispositivo root EBS (volumi di avvio) di istanze EC2 nel cluster se utilizzi una versione di Amazon EMR precedente a 5.24.0. Come per l'AMI predefinita, la dimensione minima del volume di root per un'AMI personalizzata è 10 GiB. Per ulteriori informazioni, consulta Creazione di un'AMI personalizzata con un volume del dispositivo di root Amazon EBS crittografato.

> **Note**
>
> A partire dalla versione 5.24.0 di Amazon EMR, puoi utilizzare un'opzione di configurazione di sicurezza per crittografare il dispositivo di root EBS e i volumi di archiviazione quando

specifichi AWS KMS come provider di chiavi Per ulteriori informazioni, consulta Crittografia del disco locale.

Un'AMI personalizzata deve trovarsi nella stessa Regione AWS in cui viene creato il cluster. Dovrebbe anche corrispondere all'architettura dell'istanza EC2. Ad esempio, un'istanza m5.xlarge ha un'architettura x86_64. Pertanto, per effettuare il provisioning di un m5.xlarge utilizzando un'AMI personalizzata, anche l'AMI personalizzata dovrebbe avere un'architettura x86_64. Allo stesso modo, per eseguire il provisioning di un'istanza m6g.xlarge, con architettura arm64, l'AMI personalizzata dovrebbe avere l'architettura arm64. Per ulteriori informazioni sull'identificazione di un'AMI Linux per il tipo di istanza, consulta Trovare un'AMI Linux nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Important

I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi personalizzare l'AMI di Amazon EMR per utilizzare l'AMI di Amazon Linux più recente.

Creazione di un'AMI Amazon Linux personalizzata da un'istanza preconfigurata

Di seguito è riportata la procedura di base per la preinstallazione di software e l'esecuzione di altre configurazioni per creare un'AMI Amazon Linux personalizzata per Amazon EMR:

- Avvia un'istanza dall'AMI Amazon Linux di base.
- Connessione all'istanza per l'installazione di software e l'esecuzione di altre personalizzazioni.
- Creazione di una nuova immagine (snapshot AMI) dell'istanza configurata.

Una volta creata l'immagine in base alla tua istanza personalizzata, puoi copiare tale immagine in una destinazione crittografata, come descritto in Creazione di un'AMI personalizzata con un volume del dispositivo di root Amazon EBS crittografato.

Tutorial: creazione di un'AMI da un'istanza con software personalizzato installato

Per avviare un'istanza EC2 in base alla più recente AMI Amazon Linux

1. Utilizzare AWS CLI per eseguire il comando seguente, che crea un'istanza da un'AMI esistente. Sostituisci *MyKeyName* con la coppia di chiavi utilizzata per la connessione all'istanza e *MyAmiId* con l'ID di un'AMI Amazon Linux appropriata. Per gli ID AMI più recenti, consulta AMI Amazon Linux.

   > **Note**
   >
   > I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

   ```
   aws ec2 run-instances --image-id MyAmiID \
   --count 1 --instance-type m5.xlarge \
   --key-name MyKeyName --region us-west-2
   ```

   Il valore di output `InstanceId` viene usato come *MyInstanceId* nella fase successiva.
2. Esegui il comando seguente:

   ```
   aws ec2 describe-instances --instance-ids MyInstanceId
   ```

   Il valore di output `PublicDnsName` viene usato per connettersi all'istanza nella fase successiva.

Per connettersi all'istanza e installare il software

1. Utilizza una connessione SSH che consente di eseguire comandi shell sull'istanza di Linux. Per ulteriori informazioni, consulta Connessione all'istanza Linux tramite SSH nella Guida per l'utente di Amazon EC2 per le istanze Linux.
2. Esegui eventuali personalizzazioni necessarie. Ad esempio:

```
sudo yum install MySoftwarePackage
sudo pip install MySoftwarePackage
```

Per creare una snapshot dall'immagine personalizzata

- Dopo aver personalizzato l'istanza, utilizzare il comando `create-image` per creare un'AMI dall'istanza.

  ```
  aws ec2 create-image --no-dry-run --instance-id MyInstanceId --name MyEmrCustomAmi
  ```

  Il valore di output `imageID` viene utilizzato all'avvio del cluster o alla creazione di una snapshot crittografata. Per ulteriori informazioni, consulta Utilizzare un'AMI personalizzata singola in un cluster EMR e Creazione di un'AMI personalizzata con un volume del dispositivo di root Amazon EBS crittografato.

Come utilizzare un'AMI personalizzata in un cluster Amazon EMR

Puoi utilizzare un'AMI personalizzata per eseguire il provisioning di un cluster Amazon EMR in due modi:

- Usa un'AMI personalizzata per tutte le istanze EC2 nel cluster.
- Utilizzare AMI personalizzate diverse per i diversi tipi di istanza EC2 utilizzati nel cluster.

È possibile utilizzare solo una delle due opzioni durante il provisioning di un cluster EMR e non è possibile modificarlo una volta avviato il cluster.

Considerazioni sull'utilizzo di AMI personalizzate singole o multiple in un cluster Amazon EMR

Considerazione	AMI personalizzata singola	AMI personalizzate multiple
Utilizzare processori x86 e Graviton2 con AMI personalizzate nello stesso cluster	✗ Non supportato	✓ Supportato

Considerazione	AMI personalizzata singola	AMI personalizzate multiple
La personalizzazione di AMI varia a seconda dei tipi di istanza	✗ Non supportato	✓ Supportato
Modificare le AMI personalizzate quando si aggiungono nuovi gruppi/parchi istanze di processo a un cluster in esecuzione. Nota: non è possibile modificare l'AMI personalizzata dei gruppi/parchi istanze esistenti.	✗ Non supportato	✓ Supportato
Utilizza la console AWS per avviare un cluster	✓ Supportato	✗ Non supportato
Utilizza AWS CloudFormation per avviare un cluster	✓ Supportato	✓ Supportato

Utilizzare un'AMI personalizzata singola in un cluster EMR

Per specificare un ID AMI personalizzato al momento della creazione di un cluster, utilizzare una delle seguenti opzioni:

- AWS Management Console
- AWS CLI
- Amazon EMR SDK
- Amazon EMR API RunJobFlow
- AWS CloudFormation (consulta la proprietà `CustomAmiID` in Cluster InstanceGroupConfig, Cluster InstanceTypeConfig, Resource InstanceGroupConfig o Resource InstanceFleetConfig-InstanceTypeConfig)

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Specifica di una singola AMI personalizzata con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Name and applications (Nome e applicazioni), cerca Operating system options (Opzioni del sistema operativo). Scegli Custom AMI (AMI personalizzata) e inserisci l'ID della tua AMI nel campo Custom AMI (AMI personalizzata).
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica di una singola AMI personalizzata con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. In Configurazione software, per Rilascio, scegli emr-5.7.0 o successivi e quindi scegli le altre opzioni appropriate per l'applicazione. Seleziona Successivo.
4. In Hardware Configuration (Configurazione hardware), seleziona i valori appropriati per l'applicazione e scegli Next (Avanti).
5. In Additional Options (Opzioni aggiuntive), per Custom AMI ID (ID AMI personalizzata), inserisci un valore e assicurati che l'opzione Aggiorna tutti i pacchetti installati al riavvio è

selezionata. Per ulteriori informazioni sulla modifica dell'opzione di aggiornamento, consulta Gestione degli aggiornamenti per i repository di pacchetti AMI.

6. Per avviare il cluster, scegli Next (Avanti) e completa le altre opzioni di configurazione.

AWS CLI

Specifica di una singola AMI personalizzata con la AWS CLI

- Utilizza il parametro `--custom-ami-id` per specificare l'ID dell'AMI durante l'esecuzione del comando `aws emr create-cluster`.

  L'esempio seguente specifica un cluster che utilizza un'AMI personalizzata singola con 20 GiB di volume di avvio. Per ulteriori informazioni, consulta Specifica della dimensione del volume del dispositivo di root Amazon EBS.

  > **Note**
  >
  > I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

```
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
```

Usa AMI personalizzate multiple in un cluster Amazon EMR

Per creare un cluster utilizzando più AMI personalizzate, utilizzare una delle seguenti:

- AWS CLI versione 1.20.21 o successiva
- AWS SDK
- Amazon EMR RunJobFlow nella Guida di riferimento alle API Amazon EMR

- AWS CloudFormation (consulta la proprietà `CustomAmiID` in Cluster InstanceGroupConfig, Cluster InstanceTypeConfig, Resource InstanceGroupConfig o Resource InstanceFleetConfig-InstanceTypeConfig)

La Console di gestione AWS non supporta la creazione di un cluster utilizzando AMI personalizzate multiple.

Example - Usa la CLI AWS per creare un cluster di gruppi di istanze utilizzando AMI personalizzate multiple

Utilizzando l'AWS CLI versione 1.20.21 o successive, è possibile assegnare una singola AMI personalizzata all'intero cluster oppure è possibile assegnare più AMI personalizzate a ogni nodo di istanza del cluster.

L'esempio seguente mostra un cluster di gruppi di istanze uniforme creato con due tipi di istanza (m5.xlarge) utilizzati tra i tipi di nodi (master, core, task). Ogni nodo ha AMI personalizzate multiple. L'esempio illustra diverse caratteristiche della configurazione AMI personalizzata multipla:

- Non è stata assegnata alcuna AMI personalizzata a livello di cluster. Questo per evitare conflitti tra le AMI personalizzate multiple e una AMI personalizzata singola, il che causerebbe il fallimento del avvio del cluster.
- Il cluster può avere AMI personalizzate multiple tra nodi principali, core e nodi attività singoli. Ciò consente personalizzazioni AMI individuali, come applicazioni preinstallate, configurazioni cluster sofisticate e volumi di dispositivi root Amazon EBS crittografati.
- Il nodo principale del gruppo di istanze può avere un solo tipo di istanza e una corrispondente AMI personalizzata. Analogamente, il nodo principale può avere un solo tipo di istanza e una corrispondente AMI personalizzata.
- Il cluster può avere più nodi attività.

```
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 - Usa l'AWS CLI versione 1.20.21 o successive per aggiungere un nodo attività a un cluster di gruppi di istanze in esecuzione con più tipi di istanza e più AMI personalizzate

Utilizzando l'AWS CLI versione 1.20.21 o successive, è possibile aggiungere più AMI personalizzate a un gruppo di istanze aggiunto a un cluster in esecuzione. L'argomento `CustomAmiId` può essere utilizzato con il comando `add-instance-groups` come mostrato nell'esempio seguente. Si noti che lo stesso ID AMI personalizzate multiple (ami-123456) viene utilizzato in più di 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 - Usa l'AWS CLI versione 1.20.21 o successive per creare un cluster di parchi istanze, AMI personalizzate multiple, tipi di istanza multipli, nodi master On-Demand, nodi core On-Demand, nodi core multipli e nodi attività multipli

```
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 - Usa l'AWS CLI versione 1.20.21 o successive per aggiungere nodi attività a un cluster in esecuzione con più tipi di istanza e più AMI personalizzate

```
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}']
```

Gestione degli aggiornamenti per i repository di pacchetti AMI

Per impostazione predefinita, al primo avvio le AMI Amazon Linux si connettono ai repository di pacchetti per installare gli aggiornamenti di sicurezza prima dell'avvio di altri servizi. A seconda dei requisiti, puoi scegliere di disabilitare gli aggiornamenti quando specifichi un'AMI personalizzata per Amazon EMR. L'opzione di disattivazione di questa funzione è disponibile solo con le AMI personalizzate. Per impostazione predefinita, il kernel di Amazon Linux e altri pacchetti software che richiedono un riavvio non vengono aggiornati. Tieni presente che la tua configurazione di rete deve consentire l'uscita HTTP e HTTPS ai repository Amazon Linux in Amazon S3, altrimenti gli aggiornamenti della sicurezza non avranno esito positivo.

> **Warning**
>
> Specificando un'AMI personalizzata, consigliamo vivamente di scegliere di aggiornare tutti i pacchetti installati al riavvio. La scelta di non aggiornare pacchetti crea maggiori rischi per la sicurezza.

Con la AWS Management Console, puoi selezionare l'opzione per disabilitare gli aggiornamenti al momento della scelta di una Custom AMI (AMI personalizzata).

Con la AWS CLI, puoi specificare `--repo-upgrade-on-boot NONE` insieme a `--custom-ami-id` quando utilizzi il comando create-cluster.

Con l'API di Amazon EMR, puoi specificare `NONE` per il parametro RepoUpgradeOnBoot.

Creazione di un'AMI personalizzata con un volume del dispositivo di root Amazon EBS crittografato

Per crittografare il volume del dispositivo di root Amazon EBS di un'AMI Amazon Linux per Amazon EMR, copia un'immagine snapshot da un'AMI non crittografata a una destinazione crittografata. Per ulteriori informazioni sulla creazione di volumi EBS crittografati, consulta Crittografia Amazon EBS nella Guida per l'utente di Amazon EC2 per le istanze Linux. L'AMI di origine per lo snapshot può essere l'AMI Amazon Linux di base; in alternativa, poi copiare uno snapshot da un'AMI ottenuta da un'AMI Amazon Linux di base personalizzata.

> **Note**
>
> A partire dalla versione 5.24.0 di Amazon EMR, puoi utilizzare un'opzione di configurazione di sicurezza per crittografare il dispositivo di root EBS e i volumi di archiviazione quando specifichi AWS KMS come provider di chiavi Per ulteriori informazioni, consulta Crittografia del disco locale.

Puoi utilizzare un fornitore di chiavi esterno o una chiave KMS AWS per crittografare il volume root EBS. Affinché Amazon EMR possa creare un cluster utilizzando l'AMI, il ruolo di servizio utilizzato da Amazon EMR (in genere il ruolo `EMR_DefaultRole` predefinito) deve essere autorizzato a crittografare e decrittare il volume. Se utilizzi AWS KMS come fornitore di chiavi, ciò significa che devono essere consentite le seguenti azioni:

- `kms:encrypt`
- `kms:decrypt`
- `kms:ReEncrypt*`
- `kms:CreateGrant`
- `kms:GenerateDataKeyWithoutPlaintext"`
- `kms:DescribeKey"`

Il modo più semplice per farlo consiste nell'aggiungere il ruolo come utente chiave, come descritto nel seguente tutorial. L'esempio seguente di istruzione per policy viene fornito qualora fosse necessario personalizzare le policy di ruolo.

```
{
```

```
"Version": "2012-10-17",
"Statement": [
   {
   "Sid": "EmrDiskEncryptionPolicy",
   "Effect": "Allow",
   "Action": [
      "kms:Encrypt",
      "kms:Decrypt",
      "kms:ReEncrypt*",
      "kms:CreateGrant",
      "kms:GenerateDataKeyWithoutPlaintext",
      "kms:DescribeKey"
      ],
   "Resource": [
      "*"
      ]
   }
]
}
```

Tutorial: creazione di un'AMI personalizzata con un volume dispositivo di root crittografato tramite una chiave KMS

La prima cosa da fare in questo esempio è trovare l'ARN di una chiave KMS o crearne una nuova. Per ulteriori informazioni sulla creazione di chiavi, consulta Creazione di chiavi nella Guida per gli sviluppatori di AWS Key Management Service. La procedura riportata di seguito illustra come aggiungere il ruolo di servizio `EMR_DefaultRole` come utente chiave alla policy chiavi. Annota il valore ARN per la chiave quando la crei o la modifichi. L'ARN servirà in seguito, durante la creazione dell'AMI.

Aggiunta del ruolo di servizio per Amazon EC2 all'elenco degli utenti chiave di crittografia con la console

1. Accedi alla AWS Management Console e apri la console AWS Key Management Service (AWS KMS) all'indirizzo https://console.aws.amazon.com/kms.
2. Per modificare la Regione AWS, utilizza il Selettore di regione nell'angolo in alto a destra della pagina.
3. Scegliere l'alias della chiave KMS da utilizzare.
4. Nella pagina dei dettagli della chiave, in Key Users (Utenti di chiavi), scegli Add (Aggiungi).

5. Nella finestra di dialogo Attach (Collega), scegli il ruolo di servizio Amazon EMR. Il nome del ruolo di default è `EMR_DefaultRole`.
6. Scegliere Attach (Collega).

Creazione di un'AMI crittografata con la AWS CLI

- Utilizza il comando `aws ec2 copy-image` di AWS CLI per creare un'AMI con un volume dispositivo root EBS crittografato e la chiave modificata. Sostituisci il valore `--kms-key-id` specificato con l'ARN completo della chiave creata o modificata in precedenza.

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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
```

L'output del comando fornisce l'ID dell'AMI creata, che è possibile specificare quando si crea un cluster. Per ulteriori informazioni, consulta Utilizzare un'AMI personalizzata singola in un cluster EMR. È anche possibile scegliere di personalizzare questa AMI installando software ed eseguendo altre configurazioni. Per ulteriori informazioni, consulta Creazione di un'AMI Amazon Linux personalizzata da un'istanza preconfigurata.

Best practice e considerazioni

Quando crei un'AMI personalizzata per Amazon EMR, tieni in considerazione quanto segue:

- Amazon EMR 5.30.0 e versioni successive e le serie Amazon EMR 6.x sono basate su Amazon Linux 2. Per queste versioni Amazon EMR, è necessario utilizzare immagini basate su Amazon Linux 2 per le AMI personalizzate. Per trovare un'AMI personalizzata di base, consulta Ricerca di un'AMI Linux.

- Per le versioni di Amazon EMR precedenti alla 5.30.0 e alla 6.x, le AMI di Amazon Linux 2 non sono supportate.
- È necessario utilizzare un'AMI Amazon Linux a 64 bit. Non è supportata un'AMI a 32 bit.
- Le AMI Amazon Linux non sono supportate con più volumi Amazon EBS.
- La personalizzazione deve basarsi sulla versione dell'AMI Amazon Linux più recente supportata da EBS. Per un elenco delle AMI Amazon Linux e dei corrispondenti ID, consulta AMI Amazon Linux.
- Non copiare una snapshot di un'istanza Amazon EMR esistente per creare un'AMI personalizzata, perché causa errori.
- Sono supportati solo il tipo di virtualizzazione HVM e le istanze compatibili con Amazon EMR. Assicurati di selezionare l'immagine HVM e un tipo di istanza compatibile con Amazon EMR durante il processo di personalizzazione dell'AMI. Per le istanze e i tipi di virtualizzazione compatibili, consulta Tipi di istanze supportati.
- Il tuo servizio ruolo deve disporre di autorizzazioni di avvio per l'AMI, perciò l'AMI deve essere pubblica oppure deve essere di tua proprietà o condivisa con te dal proprietario.
- Creare utenti dell'AMI con nome uguale a quello delle applicazioni causa errori (ad esempio, `hadoop`, `hdfs`, `yarn` o `spark`).
- I contenuti di `/tmp`, `/var` e `/emr` (se presenti sull'AMI) vengono spostati rispettivamente in `/mnt/tmp`, `/mnt/var` e `/mnt/emr` durante l'avvio. I file vengono conservati ma, in presenza di una grande quantità di dati, l'avvio potrebbe richiedere più tempo del previsto.
- Se si utilizza un'AMI Amazon Linux personalizzata basata su un'AMI Amazon Linux con una data di creazione 2018-08-11, il server Oozie non si avvia. Se si utilizza Oozie, è possibile creare un'AMI personalizzata basata su un ID AMI Amazon Linux con una data di creazione diversa. Puoi utilizzare il comando AWS CLI seguente per restituire un elenco di ID immagine per tutte le AMI Amazon Linux HVM con una versione 2018.03, insieme alla data di rilascio, in modo da poter scegliere un'AMI Linux appropriata come base. Sostituisci MyRegion con l'identificatore della Regione, ad esempio 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
```

- Nei casi in cui utilizzi un VPC con un nome di dominio non standard e AmazonProvidedDNS, non devi utilizzare l'opzione `rotate` nella configurazione DNS dei sistemi operativi.

Per maggiori informazioni consulta Creazione di AMI Linux supportate da Amazon EBS nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Modifica del rilascio di Amazon Linux durante la creazione di un cluster

Quando avvii un cluster utilizzando Amazon EMR versione 6.6.0 o successive, viene utilizzata automaticamente la versione più recente di Amazon Linux 2 convalidata per l'AMI Amazon EMR predefinita. Puoi specificare un rilascio di Amazon Linux diverso per il cluster con la console Amazon EMR o la AWS CLI.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Modifica del rilascio di Amazon Linux durante la creazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. Per Versione di EMR, scegli emr-6.6.0 o versione successiva.
4. In Operating system options (Opzioni del sistema operativo), scegli Amazon Linux version (Versione di Amazon Linux) e seleziona la casella di controllo Automatically apply latest Amazon Linux updates (Applica automaticamente gli ultimi aggiornamenti di Amazon Linux).
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Modifica del rilascio di Amazon Linux durante la creazione di un cluster con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce.

2. Seleziona Create cluster (Crea cluster) e vai a Advanced options (Opzioni avanzate).
3. In Configurazione software, per Rilascio, scegli emr-6.6.0 o successivi e quindi scegli le altre opzioni appropriate per l'applicazione. Seleziona Successivo.
4. In Hardware Configuration (Configurazione hardware), seleziona i valori appropriati per l'applicazione e scegli Next (Avanti).
5. In Additional Options (Opzioni aggiuntive), seleziona Amazon Linux Release (Rilascio di Amazon Linux), quindi seleziona la versione del rilascio di Amazon Linux per il tuo cluster.
6. Per avviare il cluster, scegli Next (Avanti) e completa le altre opzioni di configurazione.

AWS CLI

Modifica del rilascio di Amazon Linux durante la creazione di un cluster con la AWS CLI

- Utilizzo del parametro `--os-release-label` per specificare il parametro Versione Amazon Linux quando esegui il 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
```

Specifica della dimensione del volume del dispositivo di root Amazon EBS

Questa opzione è disponibile solo con Amazon EMR rilasci 4.x e successivi. È possibile specificare la dimensione del volume da 10 GiB (impostazione predefinita) fino a 100 GiB, al momento della creazione di un cluster utilizzando la AWS Management Console, la AWS CLI o l'API di Amazon EMR. Il dimensionamento si applica solo al volume dispositivo root EBS e a tutte le istanze del cluster. Non si applica ai volumi di storage, che vanno specificati separatamente per ogni tipo di istanza al momento della creazione di un cluster.

Per ulteriori informazioni su Amazon EBS, consulta Volume del dispositivo di root di Amazon EC2.

> **Note**
>
> Se utilizzi l'AMI di default, Amazon EMR collega gp2 (General Purpose SSD) come volume dispositivo di root. Un'AMI personalizzata può avere un altro tipo di volume dispositivo root.

Tuttavia, anche la dimensione minima del volume di root per un'AMI personalizzata è 10 GiB. Per ulteriori informazioni, consulta Utilizzo di un'AMI personalizzata.

Il costo del volume dispositivo root EBS viene calcolato all'ora in base alle tariffe EBS mensili per il tipo di volume nella Regione in cui viene eseguito il cluster. Lo stesso vale per i volumi di storage. I costi sono in GB, ma le dimensioni del volume root vanno specificate in GiB: tienine conto nella tua stima (1 GB corrisponde a 0,931323 GiB). Per stimare i costi associati ai volumi dispositivo root EBS nel cluster, utilizza la seguente formula:

($EBS in GB al mese)×0,931323÷30÷24×EMR_EBSRootGiB×InstanceCount

Prendiamo come esempio un cluster con un nodo master e un nodo principale che utilizzi l'AMI Amazon Linux di base, con l'impostazione predefinita di 10 GiB per il volume dispositivo di root. Se il costo EBS nella Regione è di $0,10/GB al mese, il totale è di circa $0,00129 per istanza all'ora e $0,00258 all'ora per il cluster ($0,10/GB al mese diviso per 30 giorni, diviso per 24 ore, moltiplicato per 10 GB, moltiplicato per 2 istanze del cluster).

Note

Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Specifica della dimensione del volume del dispositivo root di Amazon EBS con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Cluster configuration (Configurazione del cluster), vai alla sezione Cluster scaling and provisioning option (Opzioni di dimensionamento e provisioning del cluster. Espandi la freccia per il volume principale di EBS e inserisci un valore compreso tra 10 GiB e 100 GiB.
4. Scegli qualsiasi altra opzione applicabile al cluster.

5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica della dimensione del volume del dispositivo root di Amazon EBS con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate).
4. Per Configurazione software, in Rilascio, scegli Amazon EMR versione 4.x o successive. Scegli altre opzioni come appropriato per l'applicazione, quindi seleziona Next (Avanti).
5. In Hardware Configuration (Configurazione hardware), per Root device EBS volume size (Dimensioni volume EBS dispositivo root), immetti un valore compreso tra 10 GiB e 100 GiB.

CLI

Specifica della dimensione del volume del dispositivo root di Amazon EBS con la AWS CLI

- Utilizza il parametro `--ebs-root-volume-size` del comando create-cluster, come mostrato nell'esempio seguente.

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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
```

Configurazione del software cluster

Quando selezioni un rilascio del software, Amazon EMR utilizza un'Amazon Machine Image (AMI) con Amazon Linux per installare il software che scegli quando avvii il cluster, ad esempio Hadoop, Spark e Hive. Amazon EMR fornisce regolarmente nuove versioni, aggiungendo nuove caratteristiche, nuove applicazioni e aggiornamenti generali. Se possibile, ti consigliamo di utilizzare la versione più recente per avviare il cluster. La versione più recente è l'opzione predefinita nel caso di avvio di un cluster dalla console.

Per ulteriori informazioni sui rilasci Amazon EMR e sulle versioni dei software disponibili con ciascun rilascio, consulta Guida ai rilasci di Amazon EMR. Per ulteriori informazioni su come modificare le configurazioni predefinite di applicazioni e software installati nel cluster, passa alla sezione relativa alla Configurazione delle applicazioni nella Guida ai rilasci di Amazon EMR. Alcune versioni dei componenti dell'ecosistema Hadoop e Spark open source incluse nei rilasci Amazon EMR dispongono di patch e miglioramenti che sono descritti nella Guida ai rilasci di Amazon EMR.

Oltre al software e alle applicazioni standard che sono disponibili per l'installazione sul cluster, puoi utilizzare operazioni di bootstrap per installare software personalizzato. Le operazioni di bootstrap sono script che vengono eseguiti sulle istanze all'avvio del cluster e su nuovi nodi che vengono aggiunti al cluster quando vengono creati. Le operazioni di bootstrap sono utili anche per richiamare comandi di AWS CLI su ogni nodo per copiare oggetti da Amazon S3 in ogni nodo nel cluster.

> **Note**
>
> Le operazioni di bootstrap vengono utilizzate in maniera diversa in Amazon EMR versione 4.x e successive. Per ulteriori informazioni su queste differenze rispetto alle versioni 2.x e 3.x delle AMI Amazon EMR, consulta Differenze nella versione 4.x nella Guida ai rilasci di Amazon EMR.

Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo

Puoi utilizzare un'operazione di bootstrap per installare software aggiuntivo o personalizzare la configurazione delle istanze del cluster. Le operazioni di bootstrap sono script eseguiti sul cluster dopo che Amazon EMR ha avviato l'istanza utilizzando l'Amazon Machine Image (AMI) Amazon Linux. Le operazioni di bootstrap vengono eseguite prima che Amazon EMR installi le applicazioni specificate alla creazione del cluster e prima che i nodi del cluster inizino l'elaborazione dei dati.

Aggiungendo nodi a un cluster in esecuzione, le operazioni di bootstrap vengono eseguite allo stesso modo anche su questi nodi. È possibile creare e specificare operazioni di bootstrap personalizzate al momento della creazione del cluster.

La maggior parte delle operazioni di bootstrap predefinite per le versioni 2.x e 3.x dell'AMI Amazon EMR non sono supportate nelle versioni 4.x di Amazon EMR. Ad esempio, `configure-Hadoop` e `configure-daemons` non sono supportati nel rilascio 4.x di Amazon EMR. Al contrario, il rilascio 4.x di Amazon EMR offre questa caratteristica in modo nativo. Per ulteriori informazioni su come eseguire la migrazione delle operazioni di bootstrap dalle versioni 2.x e 3.x dell'AMI Amazon EMR al rilascio 4.x di Amazon EMR, consulta Personalizzazione della configurazione di cluster e applicazioni con versioni AMI precedenti di Amazon EMR nella Guida ai rilasci di Amazon EMR.

Nozioni di base sulle operazioni di bootstrap

Per impostazione predefinita, le operazioni di bootstrap vengono eseguite come utente Hadoop. Puoi eseguire un'operazione di bootstrap con privilegi root utilizzando `sudo`.

Tutte le interfacce di gestione Amazon EMR supportano le operazioni di bootstrap. Puoi specificare fino a 16 operazioni di bootstrap per cluster fornendo più parametri `bootstrap-actions` dalla console, da AWS CLI o dall'API.

Dalla console Amazon EMR, è possibile anche specificare un'operazione di bootstrap durante la creazione di un cluster.

Con l'interfaccia a riga di comando (CLI), puoi trasferire ad Amazon EMR i riferimenti agli script delle operazioni di bootstrap aggiungendo il parametro `--bootstrap-actions` quando crei il cluster con il comando `create-cluster`.

```
--bootstrap-actions Path="s3://mybucket/filename",Args=[arg1,arg2]
```

Se l'operazione di bootstrap restituisce un codice errore diverso da zero, Amazon EMR lo considera un errore e termina l'istanza. Se si verificano errori nelle operazioni di bootstrap su troppe istanze, Amazon EMR termina il cluster. Se gli errori coinvolgono solo poche istanze, Amazon EMR cerca di riassegnarle e continuare. Utilizza il codice di errore `lastStateChangeReason` del cluster per identificare gli errori causati da un'operazione di bootstrap.

Esecuzione di un'operazione di bootstrap in modo condizionale

Per eseguire solo operazioni di bootstrap sul nodo principale, puoi utilizzare un'operazione di bootstrap personalizzata con una logica per determinare se il nodo è quello principale.

```
#!/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
```

Il seguente output verrà stampato da un nodo principale.

```
This is not master node, do nothing, exiting
```

Il seguente output verrà stampato da un nodo principale.

```
This is master, continuing to execute script
```

Per utilizzare questa logica, carica l'operazione di bootstrap, incluso il codice sopra, nel bucket Amazon S3. Sulla AWS CLI, aggiungi il parametro `--bootstrap-actions` per la chiamata API `aws emr create-cluster` e specifica la posizione dello script bootstrap come valore di `Path`.

Operazioni di arresto

Lo script di un'operazione di bootstrap può creare una o più operazioni di arresto scrivendo script nella directory `/mnt/var/lib/instance-controller/public/shutdown-actions/`. Alla chiusura di un cluster, tutti gli script della directory vengono eseguiti in parallelo. Ogni script deve essere eseguito e completato entro 60 secondi.

Se il nodo viene chiuso con un errore, l'esecuzione degli script per le operazioni di arresto non è garantita.

> **Note**
>
> Con le versioni 4.0 e successive di Amazon EMR, è necessario creare manualmente la directory `/mnt/var/lib/instance-controller/public/shutdown-actions/` nel nodo master. Non esiste di default ma, dopo la sua creazione, gli script nella directory vengono comunque eseguiti prima dell'arresto. Per ulteriori informazioni sulla connessione al nodo master per creare directory, consulta Connessione al nodo primario tramite SSH.

Utilizzo di operazioni di bootstrap personalizzate

Puoi creare uno script personalizzato per eseguire un'operazione di bootstrap personalizzata. Alle operazioni di bootstrap personalizzate può fare riferimento qualsiasi interfaccia Amazon EMR.

> **Note**
>
> Per ottenere prestazioni ottimali, ti consigliamo di archiviare operazioni di bootstrap personalizzate, script e altri file che desideri utilizzare con Amazon EMR in un bucket Amazon S3 che si trova nella stessa Regione AWS del tuo cluster.

Indice

Aggiunta di operazioni di bootstrap personalizzate

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Creazione di un cluster con un'operazione di bootstrap con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Bootstrap actions (Operazioni di bootstrap), scegli Add (Aggiungi) per specificare un nome, la posizione dello script e gli argomenti opzionali per l'operazione. Seleziona Add bootstrap action (Aggiungi operazione di bootstrap).

4. Se lo desideri, puoi aggiungere altre operazioni di bootstrap.
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Creazione di un cluster con un'operazione di bootstrap personalizzata con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Fai clic su Go to advanced options (Vai alle opzioni avanzate).
4. Nella fase di creazione del cluster (opzioni avanzate, fase 1 e 2), scegli le opzioni desiderate e passa alla voce Step 3: General Cluster Settings (Fase 3: impostazioni generali del cluster).
5. In Bootstrap Actions (Operazioni di bootstrap), seleziona Configure and add (Configura e aggiungi) per specificare il nome, la posizione JAR e gli argomenti per l'operazione di bootstrap. Scegliere Add (Aggiungi).
6. Facoltativamente, è possibile aggiungere altre operazioni di bootstrap.
7. Procedi alla creazione del cluster. Le operazioni di bootstrap saranno eseguite dopo che il cluster sarà stato assegnato e inizializzato.

   Mentre il nodo primario del cluster è in esecuzione, è possibile connettersi al nodo primario e visualizzare i file di log generati dallo script dell'operazione di bootstrap nella directory `/mnt/var/log/bootstrap-actions/1`.

CLI

Creazione di un cluster con un'operazione di bootstrap personalizzata con la AWS CLI

Quando utilizzi AWS CLI per includere un'operazione di bootstrap, specifica `Path` e `Args` come elenco separato da virgole. L'esempio seguente non utilizza un elenco di argomenti.

- Per avviare un cluster con un'operazione di bootstrap personalizzata, digita il comando seguente e sostituisci *myKey* con il nome della coppia di chiavi EC2. Includi `--bootstrap-`

`actions` come parametro e specifica la posizione dello script bootstrap come valore di `Path`.

- Utenti Linux, UNIX e 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"
```

- Utenti 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"
```

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi core. Tutti i nodi utilizzeranno il tipo di istanza specificato nel comando.

> **Note**
>
> Se in precedenza non sono stati creati il ruolo del servizio Amazon EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Utilizzo di un'operazione di bootstrap personalizzata per la copia di un oggetto da Amazon S3 su ogni nodo

Puoi utilizzare un'operazione di bootstrap per copiare oggetti da Amazon S3 su ogni nodo di un cluster prima di installare le tue applicazioni. AWS CLI è installato su ciascun nodo di un cluster, in modo che l'operazione di bootstrap possa chiamare i comandi di AWS CLI.

L'esempio seguente mostra il semplice script di un'operazione di bootstrap che copia un file, `myfile.jar`, da Amazon S3 in una cartella locale, `/mnt1/myfolder`, su ogni nodo del cluster. Lo script viene salvato su Amazon S3 con il nome di file `copymyfile.sh` e con i seguenti contenuti.

```
#!/bin/bash
aws s3 cp s3://mybucket/myfilefolder/myfile.jar /mnt1/myfolder
```

All'avvio del cluster, è necessario specificare lo script. Il seguente esempio di AWS CLI ne è la dimostrazione:

```
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"
```

Configurazione di hardware e reti cluster

Quando crei un cluster Amazon EMR è importante considerare come configurare istanze Amazon EC2 e opzioni di rete. In questo capitolo vengono descritte le seguenti opzioni e vengono illustrate le best practice e linee guida per tutte queste opzioni.

- Tipi di nodi: le istanze Amazon EC2 in un cluster EMR sono organizzate in tipi di nodi. Esistono tre tipi di nodi: nodi primari, nodi core e nodi attività. Ogni tipo di nodo esegue un set di ruoli definiti dalle applicazioni distribuite installate sul cluster. Durante un processo Hadoop MapReduce o Spark, ad esempio, i componenti sui nodi attività e core elaborano i dati, trasferiscono l'output in Amazon S3 o HDFS e restituiscono i metadati di stato al nodo primario. Con un cluster a nodo singolo, tutti i componenti vengono eseguiti sul nodo primario. Per ulteriori informazioni, consulta Informazioni sui tipi di nodi: nodi primari, core e attività.
- Istanze EC2: quando si crea un cluster, è possibile effettuare delle scelte sulle istanze Amazon EC2 su cui verrà eseguito ogni tipo di nodo. Il tipo di istanza EC2 determina l'elaborazione e il profilo di archiviazione del nodo. La scelta dell'istanza Amazon EC2 per i nodi è importante perché determina il profilo delle prestazioni dei singoli tipi di nodo nel cluster. Per ulteriori informazioni, consulta Configurazione delle istanze Amazon EC2.
- Reti: è possibile avviare il cluster Amazon EMR in un VPC utilizzando una sottorete pubblica, una sottorete privata o una sottorete condivisa. La configurazione di rete determina il modo in cui i clienti e i servizi possono connettersi ai cluster per eseguire il lavoro, il modo in cui i cluster si

connettono agli archivi dati e ad altre risorse AWS e le opzioni disponibili per controllare il traffico su tali connessioni. Per ulteriori informazioni, consulta Configurazione delle reti.

- Raggruppamento di istanze: la raccolta di istanze EC2 che ospita ogni tipo di nodo è denominata parco istanze o gruppo di istanze uniforme. La scelta se configurare o meno i gruppi di istanze viene fatta quando si crea un cluster, Questa scelta determina il modo in cui è possibile aggiungere nodi al cluster mentre è in esecuzione. La configurazione si applica a tutti i tipi di nodo. In seguito non può più essere modificata. Per ulteriori informazioni, consulta Creazione di un cluster con parchi istanze o gruppi di istanze uniformi.

> **Note**
>
> La configurazione dei parchi istanze è disponibile solo in Amazon EMR rilasci 4.8.0 e successivi, esclusi i rilasci 5.0.0 e 5.0.3.

Informazioni sui tipi di nodi: nodi primari, core e attività

Utilizza questa sezione per scoprire il modo in cui Amazon EMR utilizza ognuno di questi tipi di nodo e come base per la pianificazione della capacità del cluster.

Nodo primario

Il nodo primario gestisce il cluster ed esegue in genere i componenti primari delle applicazioni distribuite. Ad esempio, il nodo primario esegue il servizio YARN ResourceManager per gestire le risorse per le applicazioni. Inoltre, esegue il servizio HDFS NameNode, tiene traccia dello stato di processi inviati al cluster e monitora l'integrità dei gruppi di istanze.

Per monitorare lo stato di avanzamento di un cluster e interagire direttamente con le applicazioni, puoi connetterti al nodo primario su SSH come utente Hadoop. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH. La connessione al nodo primario consente di accedere direttamente a directory e file, ad esempio i file di log Hadoop. Per ulteriori informazioni, consulta Visualizzare file di log di . Puoi anche visualizzare le interfacce utente pubblicate dalle applicazioni come siti Web in esecuzione sul nodo primario. Per ulteriori informazioni, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

> **Note**
>
> Con Amazon EMR rilascio 5.23.0 e successivi, puoi avviare un cluster con tre nodi primari per supportare applicazioni a elevata disponibilità come YARN Resource Manager, HDFS Name Node, Spark, Hive e Ganglia. Con questa caratteristica, il nodo primario non rappresenta più un potenziale singolo punto di errore. Se uno dei nodi primari ha esito negativo, Amazon EMR esegue automaticamente il failover in un nodo primario in standby e sostituisce il nodo primario guasto con uno nuovo con le medesime operazioni di configurazione e di bootstrap. Per ulteriori informazioni, consulta la sezione Plan and Configure Primary Nodes (Pianificazione e configurazione dei nodi primari).

Nodi principali

I nodi core sono gestiti dal nodo primario. I nodi principali eseguono il daemon Data Node per coordinare lo storage dei dati come parte di Hadoop Distributed File System (HDFS). Inoltre, eseguono il daemon Task Tracker e altre attività di calcolo parallelo sui dati richieste dalle applicazioni installate. Ad esempio, un nodo principale esegue i daemon YARN NodeManager, le attività Hadoop MapReduce e gli esecutori Spark.

Per ogni cluster, è disponibile un solo gruppo di istanze o parco istanze, ma possono essere presenti più nodi in esecuzione su più istanze Amazon EC2 nel gruppo di istanze o parco istanze. Con i gruppi di istanze, puoi aggiungere o rimuovere istanze Amazon EC2 mentre il cluster è in esecuzione. È inoltre possibile impostare la scalabilità automatica per aggiungere istanze in base al valore di un parametro. Per ulteriori informazioni sull'aggiunta e la rimozione di istanze Amazon EC2 con la configurazione dei gruppi di istanze, consulta Uso del dimensionamento del cluster.

Con i parchi istanze, puoi aggiungere e rimuovere agevolmente istanze modificando le capacità target del parco istanze su on demand e Spot di conseguenza. Per ulteriori informazioni sulle capacità target, consulta Opzioni del parco istanze.

> **Warning**
>
> La rimozione dei daemon HDFS da un nodo principale in esecuzione o la terminazione di nodi principali comporta il rischio di perdita dei dati. Fai attenzione quando configuri i nodi principali per l'utilizzo delle istanze Spot. Per ulteriori informazioni, consulta Quando occorre utilizzare le istanze Spot?.

Nodi attività

Puoi utilizzare i nodi attività per aggiungere potenza di elaborazione per eseguire attività di calcolo parallelo sui dati, ad esempio attività Hadoop MapReduce ed executor Spark. I nodi di task non eseguono il daemon Data Node, né archiviano dati in HDFS. Analogamente ai nodi principali, puoi aggiungere nodi attività a un cluster aggiungendo istanze Amazon EC2 a un gruppo di istanze uniforme esistente o modificando le capacità target per un parco istanze dell'attività.

Con la configurazione del gruppo di istanze uniforme puoi avere un totale di 48 gruppi di istanze attività. La possibilità di aggiungere gruppi di istanze in questo modo consente di combinare tipi di istanze Amazon EC2 e opzioni di prezzo, ad esempio istanze On Demand e istanze Spot. Questo consente di rispondere ai requisiti di carico di lavoro in modo conveniente.

Con la configurazione del parco istanze, la possibilità di combinare tipi di istanze e opzioni di acquisto è integrata, perciò esiste un solo parco istanze attività.

Poiché le istanze Spot vengono spesso utilizzate per eseguire nodi attività, Amazon EMR dispone delle caratteristiche predefinite per la pianificazione dei processi YARN in modo che i processi in esecuzione non abbiano esito negativo quando i nodi attività in esecuzione su istanze Spot vengono terminati. Amazon EMR esegue questa operazione consentendo ai processi master delle applicazioni di funzionare solo sui nodi principali. Il processo master dell'applicazione controlla i processi in esecuzione e deve rimanere attivo per tutta la durata del processo.

Amazon EMR rilascio 5.19.0 e successivi utilizzano la caratteristica integrata etichette nodo YARN per questo scopo. (Le versioni precedenti utilizzavano una patch di codice). Le proprietà nelle classificazioni di configurazione `yarn-site` e `capacity-scheduler` sono configurate per impostazione predefinita in modo che capacity-scheduler e fair-scheduler YARN sfruttino le etichette dei nodi. Amazon EMR etichetta in automatico i nodi principali con l'etichetta `CORE` e imposta le proprietà in modo che i master dell'applicazione siano pianificati solo sui nodi con l'etichetta CORE. La modifica manuale delle proprietà correlate nelle classificazioni di configurazione del sito di YARN e del pianificatore di capacità o direttamente nei file XML associati potrebbe interrompere o alterare questa funzionalità.

A partire dalla serie di rilascio Amazon EMR 6.x, la funzione etichette nodo YARN è disabilitata per impostazione predefinita. Per impostazione predefinita, i processi primari dell'applicazione possono essere eseguiti sia sui nodi core sia su quelli attività. È possibile abilitare la caratteristica etichette nodo YARN configurando le seguenti proprietà:

- `yarn.node-labels.enabled: true`

- `yarn.node-labels.am.default-node-label-expression: 'CORE'`

Per informazioni su proprietà specifiche, consulta Impostazioni di Amazon EMR per impedire gli errori nei processi a causa dell'interruzione delle istanze Spot nei nodi attività.

Configurazione delle istanze Amazon EC2

Le istanze EC2 sono disponibili in diverse configurazioni, note come tipi di istanze. I tipi di istanza dispongono di diverse capacità di CPU, input/output e archiviazione. Oltre al tipo di istanza, puoi scegliere diverse opzioni di acquisto per istanze Amazon EC2. Puoi specificare diversi tipi di istanze e opzioni di acquisto all'interno di gruppi di istanze o parchi istanze uniformi. Per ulteriori informazioni, consulta Creazione di un cluster con parchi istanze o gruppi di istanze uniformi. Per indicazioni sulla scelta di tipi di istanza e opzioni di acquisto per l'applicazione, consulta Best practice per la configurazione dei cluster.

> **Important**
>
> Quando si sceglie un tipo di istanza utilizzando la AWS Management Console, il numero di VPCU per ogni tipo di istanza è il numero di vcore YARN per quel tipo di istanza, non il numero di vCPU EC2 per quel tipo di istanza. Per ulteriori informazioni sul numero di vCPU per ogni tipo di istanza, consulta Tipi di istanza di Amazon EC2.

Argomenti

Tipi di istanze supportati

Questa sezione descrive i tipi di istanza supportati da Amazon EMR, organizzate per Regione AWS. Per ulteriori informazioni sui tipi di istanza, consulta Istanze Amazon EC2 e Matrice dei tipi di istanza AMI Amazon Linux.

Non tutti i tipi di istanza sono disponibili in tutte le Regioni e la disponibilità dell'istanza è soggetta alla disponibilità e alla domanda nella Regione e nella zona di disponibilità specificate. La zona di disponibilità di un'istanza è determinata dalla sottorete utilizzata per avviare il cluster.

Considerazioni

Considera quanto segue quando scegli i tipi di istanza per il tuo cluster Amazon EMR.

> **Important**
>
> Quando si sceglie un tipo di istanza utilizzando la AWS Management Console, il numero di VPCU per ogni tipo di istanza è il numero di vcore YARN per quel tipo di istanza, non il numero di vCPU EC2 per quel tipo di istanza. Per ulteriori informazioni sul numero di vCPU per ogni tipo di istanza, consulta Tipi di istanza di Amazon EC2.

- Se crei un cluster utilizzando un tipo di istanza che non è disponibile nella regione specificata in una zona di disponibilità, è possibile che il cluster non riesca a effettuare il provisioning o che il provisioning si blocchi. Per informazioni sulla disponibilità delle istanze, consulta Pagina dei prezzi di Amazon EMR o consulta le tabelle Tipi di istanze supportati da Regione AWS in questa pagina.
- A partire dal rilascio di Amazon EMR versione 5.13.0, tutte le istanze utilizzano la virtualizzazione HVM e l'archiviazione supportata da EBS per i volumi root. Quando si utilizzano versioni del rilascio di Amazon EMR precedenti a 5.13.0, alcune istanze di generazione precedenti utilizzano la virtualizzazione PVM. Per ulteriori informazioni, consulta Tipi di virtualizzazione delle AMI Linux.
- Alcuni tipi di istanze supportano caratteristiche di rete avanzate. Per ulteriori informazioni, consulta Reti avanzate su Linux.
- I driver NVIDIA e CUDA sono installati su tipi di istanze GPU per impostazione predefinita.

Tipi di istanze supportati da Regione AWS

Nella seguente tabella vengono descritti i tipi di istanze supportati da Amazon EMR, organizzati per Regione AWS. Le tabelle elencano anche i primi rilasci di Amazon EMR nelle serie 4.x, 5.x o 6.x che supportano ciascun tipo di istanza. Ad esempio, EMR supporta le istanze m4.16xlarge nella regione Stati Uniti Orientali (Virginia settentrionale) nelle versioni 4.8.3 e successive, 5.2.1 e successive e 6.0.0 e successive.

Stati Uniti orientali (Virginia settentrionale), us-east-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Stati Uniti orientali (Ohio): us-east-2

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6a.24xlarge	emr-5.36.1, emr-6.8.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6i.16xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m7i.2xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	x2gd.16xlarge	emr-5.36.1, emr-6.10.0
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Stati Uniti occidentali (California settentrionale) - us-west-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Stati Uniti occidentali (Oregon): us-west-2

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m5d.4xlarge	emr-5.13.0, emr-6.0.0
	m5d.8xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m5zn.2xlarge	emr-5.33.0, emr-6.3.0
	m5zn.3xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6gd.4xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6id.32xlarge	emr-5.36.1, emr-6.8.0
	m6idn.xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m7g.8xlarge	emr-5.36.1, emr-6.10.0
	m7g.12xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c5d.9xlarge	emr-5.13.0, emr-6.0.0
	c5d.12xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6g.4xlarge	emr-5.31.0, emr-6.1.0
	c6g.8xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6i.16xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c7g.xlarge	emr-5.36.1, emr-6.7.0
	c7g.2xlarge	emr-5.36.1, emr-6.7.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	p3.8xlarge	emr-5.10.0, emr-6.0.0
	p3.16xlarge	emr-5.10.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	p5.48xlarge	emr-6.14.0
Ottimizzazione per la 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
	r5ad.12xlarge	emr-5.20.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5dn.8xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6g.xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6id.xlarge	emr-5.36.1, emr-6.8.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6in.12xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	x1e.4xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	z1d.xlarge	emr-5.17.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	i4g.8xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

AWS GovCloud (Stati Uniti occidentali): us-gov-west-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
Ottimizzazione per il calcolo	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.12xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

AWS GovCloud (Stati Uniti orientali): us-gov-east-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	g4dn.12xlarge	emr-5.30.0, emr-6.0.0
	g4dn.16xlarge	emr-5.30.0, emr-6.0.0
Ottimizzazione per la 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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	x2idn.16xlarge	emr-5.36.1, emr-6.9.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Africa (Città del Capo) - af-south-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia Pacifico (Hong Kong) - ap-east-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	m5.xlarge	emr-5.20.0, emr-6.0.0
	m5.2xlarge	emr-5.20.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia Pacifico (Giacarta) - ap-southeast-3

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	m5.xlarge	emr-5.30.2, emr-6.0.1
	m5.2xlarge	emr-5.30.2, emr-6.0.1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Ottimizzazione per il calcolo	c5.xlarge	emr-5.30.2, emr-6.0.1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Elaborazione accelerata	g5.xlarge	emr-5.36.1, emr-6.9.0
	g5.2xlarge	emr-5.36.1, emr-6.9.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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 Pacifico (Mumbai) - ap-south-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6i.4xlarge	emr-5.35.0, emr-6.6.0
	m6i.8xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c7g.xlarge	emr-5.36.1, emr-6.7.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	p2.8xlarge	emr-5.10.0, emr-6.0.0
	p2.16xlarge	emr-5.10.0, emr-6.0.0
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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 Pacifico (Hyderabad) (ap-south-2)

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Ottimizzazione per il calcolo	c5.xlarge	emr-5.36.0, emr-6.7.0
	c5.2xlarge	emr-5.36.0, emr-6.7.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Asia Pacifico (Osaka) - ap-northeast-3

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Ottimizzazione per il calcolo	c5.xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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
Ottimizzazione per la 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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	x2iedn.24xlarge	emr-5.36.1, emr-6.9.0
	x2iedn.32xlarge	emr-5.36.1, emr-6.9.0
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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 Pacifico (Seoul) - ap-northeast-2

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6i.32xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Ottimizzazione per il calcolo	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.12xlarge	emr-5.13.0, emr-6.0.0
	c5d.18xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5b.4xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5n.2xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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 Pacifico (Singapore) - ap-southeast-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	c5.xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c5ad.24xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6a.32xlarge	emr-5.36.1, emr-6.8.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6i.xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	c7g.16xlarge	emr-5.36.1, emr-6.7.0
Elaborazione accelerata	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5b.4xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5n.2xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Storage ottimizzato	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Asia Pacifico (Sydney): ap-southeast-2

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	m7g.16xlarge	emr-5.36.1, emr-6.10.0
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5n.2xlarge	emr-5.31.0, emr-6.1.0
	r5n.4xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	z1d.6xlarge	emr-5.17.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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 Pacifico (Tokyo), ap-northeast-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5b.12xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	z1d.6xlarge	emr-5.17.0, emr-6.0.0
	z1d.12xlarge	emr-5.17.0, emr-6.0.0
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Canada (Centrale) - ca-central-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6in.12xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5n.xlarge	emr-5.31.0, emr-6.1.0
	r5n.2xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6i.2xlarge	emr-5.35.0, emr-6.6.0
	r6i.4xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	im4gn.8xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Cina (Ningxia) - cn-nordovest-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Storage ottimizzato	d2.xlarge	emr-4.9.1, emr-5.5.3, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Cina (Pechino) - cn-north-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	m5.xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m5d.16xlarge	emr-5.13.0, emr-6.0.0
	m5d.24xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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
	r5d.xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	d2.xlarge	emr-4.0.0, emr-5.0.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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 (Francoforte) - eu-central-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	m5.xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m5ad.16xlarge	emr-5.33.0, emr-6.3.0
	m5ad.24xlarge	emr-5.20.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m5n.12xlarge	emr-5.31.0, emr-6.1.0
	m5n.16xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6g.8xlarge	emr-5.30.0, emr-6.1.0
	m6g.12xlarge	emr-5.30.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6id.12xlarge	emr-5.36.1, emr-6.8.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6in.32xlarge	emr-5.36.1, emr-6.10.0
	m7g.xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5b.24xlarge	emr-5.33.0, emr-6.3.0
	r5d.xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5n.16xlarge	emr-5.31.0, emr-6.1.0
	r5n.24xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6gd.8xlarge	emr-5.33.0, emr-6.3.0
	r6gd.12xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6idn.2xlarge	emr-5.36.1, emr-6.10.0
	r6idn.4xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r7g.16xlarge	emr-5.36.1, emr-6.10.0
	r7gd.xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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 (Zurigo) (eu-central-2)

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	c7gn.12xlarge	emr-5.36.1, emr-6.10.0
	c7gn.16xlarge	emr-5.36.1, emr-6.10.0
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5d.16xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6a.12xlarge	emr-5.36.1, emr-6.8.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6i.8xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6idn.32xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r7gd.16xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	x2iedn.4xlarge	emr-5.36.1, emr-6.9.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	i3en.12xlarge	emr-5.25.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	is4gen.4xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	is4gen.8xlarge	emr-5.35.0, emr-6.6.0

Europa (Londra) - eu-west-2

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	m6i.16xlarge	emr-5.35.0, emr-6.6.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	m6i.24xlarge	emr-5.35.0, emr-6.6.0
	m6i.32xlarge	emr-5.35.0, emr-6.6.0
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6g.4xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r5b.24xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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 (Milano) (eu-south-1)

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Europa (Spagna) (eu-south-2)

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Ottimizzazione per il calcolo	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Ottimizzazione per la 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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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 (Parigi) - eu-west-3

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	m5.xlarge	emr-5.13.0, emr-6.0.0
	m5.2xlarge	emr-5.13.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6gd.2xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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 (Stoccolma) - eu-north-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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 (Bahrein) - me-south-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	m5.xlarge	emr-5.24.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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-5.24.0, 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Storage ottimizzato	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	i4i.32xlarge	emr-5.36.1, emr-6.8.0

Regione Medio Oriente (EAU) (me-central-1)

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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
Ottimizzazione per il calcolo	c5.xlarge	emr-5.36.0, emr-6.7.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	x2idn.24xlarge	emr-5.36.1, emr-6.9.0
	x2idn.32xlarge	emr-5.36.1, emr-6.9.0
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Sud America (San Paolo) - sa-east-1

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
Uso generale	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per il calcolo	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6a.4xlarge	emr-5.36.1, emr-6.8.0
	c6a.8xlarge	emr-5.36.1, emr-6.8.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	c6gn.2xlarge	emr-5.33.0, emr-6.3.0
	c6gn.4xlarge	emr-5.33.0, emr-6.3.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Elaborazione accelerata	g4dn.xlarge	emr-5.30.0, emr-6.0.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Ottimizzazione per la 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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	r6g.8xlarge	emr-5.31.0, emr-6.1.0
	r6g.12xlarge	emr-5.31.0, emr-6.1.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
	x1e.2xlarge	emr-5.36.1, emr-6.10.0
	x1e.4xlarge	emr-5.36.1, emr-6.10.0

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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
Storage ottimizzato	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

Classe istanza	Tipo di istanza	Versione minima di Amazon EMR supportata
	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

Istanze di generazioni precedenti

Amazon EMR supporta le istanze di generazioni precedenti per supportare le applicazioni ottimizzate per tali istanze e non ancora aggiornate. Per ulteriori informazioni su questi tipi di istanze e percorsi di aggiornamento, consulta Istanze di generazioni precedenti.

Classe istanza	Tipi di istanza
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

Classe istanza	Tipi di istanza
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] Utilizza l'AMI di virtualizzazione PVM con Amazon EMR in versione precedente alla 5.13.0. Per ulteriori informazioni, consulta Tipi di virtualizzazione delle AMI Linux.

[2] Non supportata nella versione di rilascio 5.15.0.

Opzioni di acquisto delle istanze

Durante la configurazione di un cluster, scegliere un'opzione di acquisto per le istanze Amazon EC2. Puoi scegliere istanze on demand, istanze Spot o entrambe. I prezzi variano in base al tipo di istanza e alla Regione. Il prezzo di Amazon EMR si aggiunge al prezzo di Amazon EC2 (il prezzo per i server sottostanti) e al prezzo di Amazon EBS (se si allegano volumi Amazon EBS). Per i prezzi correnti, consulta Prezzi di Amazon EMR.

La scelta di utilizzare gruppi di istanze o parchi istanze nel cluster determina la modalità di modifica delle opzioni di acquisto dell'istanza mentre un cluster è in esecuzione. Se scegli gruppi di istanze uniformi, puoi specificare l'opzione di acquisto solo per un gruppo di istanze al momento della creazione e il tipo di istanza e l'opzione di acquisto si applicano a tutte le istanze Amazon EC2 in ogni gruppo di istanze. Se scegli parchi istanze, puoi modificare le opzioni di acquisto dopo aver creato il parco istanze e combinare le opzioni di acquisto per raggiungere la capacità target specificata. Per ulteriori informazioni su queste configurazioni, consulta Creazione di un cluster con parchi istanze o gruppi di istanze uniformi.

Istanze on demand

Con istanze On-Demand, paghi per capacità di elaborazione a ore. In alternativa, puoi configurare le istanze on demand in modo da utilizzare le opzioni di acquisto Istanza riservata o Istanza dedicata. Con le istanze riservate, puoi effettuare un pagamento una tantum per un'istanza per prenotare la capacità. Le istanze dedicate sono fisicamente isolate a livello di hardware dell'host dalle istanze

che appartengono ad altri account AWS. Per ulteriori informazioni sulle opzioni di acquisto, consulta Opzioni di acquisto delle istanze nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Uso di istanze riservate

Per utilizzare istanze riservate in Amazon EMR, utilizza Amazon EC2 per acquistare l'istanza riservata e specifica i parametri della prenotazione, incluso l'ambito della prenotazione in riferimento a una Regione o a una zona di disponibilità. Per ulteriori informazioni, consulta Istanze riservate di Amazon EC2 e Acquisto di istanze riservate nella Guida per l'utente di Amazon EC2 per le istanze Linux. Quando acquisti un'istanza riservata, se tutte le condizioni seguenti sono vere, Amazon EMR utilizza l'istanza riservata all'avvio di un cluster:

- Un'istanza on demand viene specificata nella configurazione cluster che corrisponde alla specifica dell'istanza riservata.
- Il cluster viene avviato nell'ambito della prenotazione dell'istanza (la zona di disponibilità o Regione).
- La capacità dell'istanza riservata è ancora disponibile

Ad esempio, supponiamo di acquistare un'istanza riservata `m5.xlarge` con la prenotazione dell'istanza calibrata sulla Regione Stati Uniti orientali. Viene quindi avviato un cluster Amazon EMR negli Stati Uniti orientali che utilizza due istanze `m5.xlarge`. La prima istanza viene fatturata alla tariffa dell'istanza riservata e l'altra viene fatturata alla tariffa on demand. La capacità dell'istanza riservata viene utilizzata prima della creazione di qualsiasi istanza on demand.

Uso di istanze dedicate

Per utilizzare istanze dedicate, acquistale utilizzando Amazon EC2 e crea un VPC con l'attributo di tenancy Dedicated (Dedicato). All'interno di Amazon EMR, specifica quindi che un cluster deve essere avviato in questo VPC. Qualsiasi istanza on demand nel cluster che corrisponde alle specifiche dell'istanza dedicata utilizza istanze dedicate disponibili all'avvio del cluster.

> **Note**
>
> Amazon EMR non supporta l'impostazione dell'attributo `dedicated` su singole istanze.

Spot Instances

Le istanze Spot in Amazon EMR forniscono un'opzione per acquistare capacità di istanze Amazon EC2 a un costo ridotto rispetto all'acquisto di istanze on demand. Lo svantaggio dell'utilizzo di istanze Spot è che le istanze possono terminare se la capacità Spot diventa indisponibile per il tipo di istanza in esecuzione. Per ulteriori informazioni su quando l'utilizzo di istanze Spot può risultate appropriato per l'applicazione, consulta Quando occorre utilizzare le istanze Spot?.

Quando Amazon EC2 dispone di capacità inutilizzata, offre istanze EC2 a un costo ridotto, denominato prezzo Spot. Questo prezzo varia in base alla disponibilità e alla domanda e viene stabilito in base alla Regione e alla zona di disponibilità. Quando scegli le istanze Spot, specifica il prezzo Spot massimo che sei disposto a pagare per ogni tipo di istanza EC2. Quando il prezzo Spot nella zona di disponibilità del cluster è inferiore al prezzo Spot massimo specificato per tale tipo di istanza, le istanze vengono avviate. Mentre le istanze sono in esecuzione, ti viene addebitato il prezzo Spot corrente non il prezzo Spot massimo.

> **Note**
>
> Le istanze spot con una durata definita (note anche come blocchi Spot) non sono più disponibili per i nuovi clienti a partire dal 1° luglio 2021. Per i clienti che hanno già utilizzato la funzione, continueremo a supportare le istanze spot con una durata definita fino al 31 dicembre 2022.

Per i prezzi correnti, consulta Prezzi delle istanze Spot Amazon EC2. Per ulteriori informazioni, consulta Istanze Spot nella Guida per l'utente di Amazon EC2 per le istanze Linux. Quando crei e configuri un cluster, devi specificare le opzioni di rete che determinano in ultima analisi la zona di disponibilità in cui il cluster viene avviato. Per ulteriori informazioni, consulta Configurazione delle reti.

> **Tip**
>
> Puoi visualizzare il prezzo Spot in tempo reale nella console passando il mouse sul suggerimento informazioni accanto all'opzione di acquisto Spot quando crei un cluster utilizzando Advanced Options (Opzioni avanzate). Vengono visualizzati i prezzi per ogni zona di disponibilità nella Regione selezionata. I prezzi più bassi si trovano nelle righe di colore verde. A causa della fluttuazione dei prezzi Spot tra zone di disponibilità, è possibile che selezionando la zona di disponibilità con il prezzo iniziale più basso tale prezzo non venga mantenuto per l'intera durata del cluster. Per risultati ottimali, occorre studiare la cronologia

dei prezzi della zona di disponibilità prima di scegliere. Per ulteriori informazioni, consulta Cronologia dei prezzi dell'istanza Spot nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Le opzioni Istanza Spot variano a seconda che nella configurazione del cluster si utilizzano gruppi di istanze o parchi istanze uniformi.

Istanze Spot in gruppi di istanze uniformi

Quando utilizzi istanze Spot in un gruppo di istanze uniformi, tutte le istanze in un gruppo di istanze devono essere istanze Spot. Specifica una sottorete o zona di disponibilità singola per il cluster. Per ogni gruppo di istanze, specifica un'istanza Spot singola e un prezzo Spot massimo. Le istanze Spot di questo tipo vengono avviate se il prezzo Spot nella Regione e nella zona di disponibilità del cluster è inferiore al prezzo Spot massimo. Le istanze terminano se il prezzo Spot è superiore al prezzo Spot massimo. Puoi importare il prezzo Spot massimo solo quando configuri un gruppo di istanze. In seguito non può più essere modificata. Per ulteriori informazioni, consulta Creazione di un cluster con parchi istanze o gruppi di istanze uniformi.

Istanze Spot in parchi istanze

Quando utilizzi la configurazione dei parchi istanze, opzioni aggiuntive offrono un maggiore controllo sulla modalità di avvio e di terminazione delle istanze Spot. Sostanzialmente, i parchi istanze utilizzano un metodo diverso rispetto a gruppi di istanze uniformi per avviare le istanze. In pratica, definisci una capacità target per istanze Spot (e istanze on demand) e un massimo di cinque tipi di istanze. Puoi anche specificare una capacità ponderata per ogni tipo di istanza o utilizzare il vCPU (vcore YARN) del tipo di istanza come capacità ponderata. Questa capacità ponderata viene conteggiata per il raggiungimento della capacità target quando si esegue il provisioning di un'istanza di tale tipo. Amazon EMR esegue il provisioning di istanze con entrambe le opzioni di acquisto finché non viene raggiunta la capacità target per ogni target. Inoltre, puoi definire un intervallo di zone di disponibilità per Amazon EMR da cui scegliere quando si avviano istanze. Puoi anche fornire opzioni Spot aggiuntive per ciascun parco istanze, incluso un timeout di provisioning. Per ulteriori informazioni, consulta Configurazione di parchi istanze.

Archiviazione dell'istanza

Archivio istanza e archiviazione del volume Amazon EBS vengono utilizzati per dati HDFS, nonché buffer, cache, dati Scratch e altri contenuti temporanei che alcune applicazioni possono "riversare" nel file system locale.

Amazon EBS funziona in modo diverso all'interno di Amazon EMR in confronto alle normali istanze Amazon EC2. I volumi Amazon EBS collegati a cluster Amazon EMR sono temporanei: i volumi vengono eliminati al termine del cluster e dell'istanza (ad esempio, durante la riduzione di gruppi di istanze), pertanto è importante non aspettarsi che i dati siano persistenti. Anche se sono temporanei, i dati in HDFS possono essere replicati a seconda del numero e della specializzazione dei nodi nel cluster. Quando si aggiungono volumi di stogare Amazon EBS, questi vengono montati come volumi aggiuntivi. Non fanno parte del volume di avvio. YARN è configurato per utilizzare tutti i volumi aggiuntivi, ma l'utente è responsabile dell'allocazione di volumi aggiuntivi come storage locale (ad esempio, per file di log locali).

Di seguito sono riportate altre avvertenze per l'utilizzo di Amazon EBS con cluster Amazon EMR:

- Non puoi eseguire lo snapshot di un volume Amazon EBS per poi ripristinarlo all'interno di Amazon EMR. Per creare configurazioni personalizzate riutilizzabili, utilizza un'AMI personalizzata (disponibile in Amazon EMR versione 5.7.0 e successive). Per ulteriori informazioni, consulta Utilizzo di un'AMI personalizzata.
- Un volume dispositivo root Amazon EBS crittografato è supportato solo utilizzando un'AMI personalizzata. Per ulteriori informazioni, consulta Creazione di un'AMI personalizzata con un volume del dispositivo di root Amazon EBS crittografato.
- Se applichi tag utilizzando l'API di Amazon EMR, tali operazioni vengono applicate a volumi EBS.
- Esiste un limite di 25 volumi per istanza.
- I volumi Amazon EBS sui nodi principali non possono essere inferiori a 5 GB.

Archiviazione Amazon EBS di default per istanze

Amazon EMR collega automaticamente un volume Amazon EBS di 10 GB General Purpose SSD (gp2) come dispositivo di root per le AMI per migliorare le prestazioni. Inoltre, per istanze EC2 con archiviazione solo su EBS, Amazon EMR assegna volumi di archiviazione gp2 di Amazon EBS alle istanze. Quando crei un cluster utilizzando la versione 5.22.0 e successive di Amazon EMR, la quantità predefinita di spazio di archiviazione di Amazon EBS aumenta in base alle dimensioni dell'istanza. Dividiamo lo spazio di archiviazione aumentato su più volumi, offrendo migliori prestazioni IOPS e, di conseguenza, migliori prestazioni per alcuni carichi di lavoro standardizzati. Se desideri utilizzare una diversa configurazione dell'archiviazione delle istanze gp2 di Amazon EBS, puoi specificarla al momento della creazione di un cluster Amazon EMR o quando aggiungi nodi a un cluster esistente. Al momento, i volumi gp3 di Amazon EBS non possono essere utilizzati come volumi root su un cluster Amazon EMR. Puoi utilizzare solo volumi gp2 di Amazon EBS come volumi root e aggiungere volumi gp3 come volumi aggiuntivi. La tabella seguente identifica il numero

predefinito di volumi di archiviazione gp2 di Amazon EBS, le dimensioni e le dimensioni totali per tipo di istanza.

I costi di Amazon EBS vengono calcolati all'ora in base alle tariffe mensili per i volumi gp2 nella Regione AWS in cui viene eseguito il cluster. Ad esempio, il costo Amazon EBS per ora per il volume root su ogni nodo di cluster in una Regione che addebita $0,10/GB/mese è di circa $0,00139 all'ora ($0,10/GB/mese diviso per 30 giorni diviso per 24h moltiplicato per 10 GB).

Volumi e dimensioni di archiviazione Amazon EBS gp2 predefiniti per tipo di istanza per Amazon EMR 5.22.0 e successive

Dimensioni istanza	Numero di volumi	Dimensioni del volume (GiB)	Dimensione totale (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	1.024
18xlarge	4	288	1152

Dimensioni istanza	Numero di volumi	Dimensioni del volume (GiB)	Dimensione totale (GiB)
24xlarge	4	384	1536

Specifica di volumi di archiviazione EBS aggiuntivi

Quando configuri tipi di istanze in Amazon EMR, puoi specificare volumi EBS aggiuntivi per aggiungere capacità oltre l'archivio istanza (se presente) e il volume EBS predefinito. Amazon EBS fornisce i seguenti tipi di volume: per scopo generico (SSD), IOPS con provisioning (SSD), velocità effettiva ottimizzata (HDD), Cold (HDD) e magnetici. Si differenziano per caratteristiche di prestazioni e prezzo, perciò puoi personalizzare il tuo storage in base alle esigenze analitiche e aziendali delle applicazioni. Ad esempio, per alcune applicazioni potrebbe essere necessario riversare su disco, mentre altre possono funzionare in modo sicuro in memoria o utilizzando Amazon S3.

Puoi collegare volumi Amazon EBS a istanze solo al momento dell'avvio del cluster e quando aggiungi un gruppo di istanze del nodo attività aggiuntivo. Se un'istanza in un cluster Amazon EMR non va a buon fine, l'istanza e i volumi Amazon EBS collegati vengono sostituiti con nuovi volumi. Di conseguenza, se si scollega manualmente un volume Amazon EBS, questa operazione viene considerata da Amazon EMR come un errore e sostituisce l'archiviazione dell'istanza (se applicabile) e gli store di volumi.

Amazon EMR non consente di modificare il tipo di volume da gp2 a gp3 per un cluster EMR esistente. Per utilizzare gp3 per i tuoi carichi di lavoro/casi d'uso, devi avviare un nuovo cluster EMR. Inoltre, non è consigliabile aggiornare la velocità di trasmissione effettiva e gli IOPS su un cluster in uso o allocazione, poiché Amazon EMR utilizza i valori di velocità di trasmissione effettiva e IOPS specificati al momento dell'avvio del cluster per ogni nuova istanza aggiunta durante l'aumento del cluster. Consultare Confronto tra i tipi di volume gp2 e gp3 di Amazon EBS e Selezione di IOPS e velocità di trasmissione effettiva durante la migrazione a gp3.

> **Important**
>
> Per utilizzare un volume gp3 con il cluster EMR, avvia un nuovo cluster EMR utilizzando l'API, l'SDK o la CLI.

Confronto tra i tipi di volume gp2 e gp3 di Amazon EBS

Ecco un confronto tra i costi dei volumi gp2 e gp3 nella regione us-east-1 (Virginia settentrionale)

Tipo di volume	gp3	gp2
Volume size (Dimensione dei volumi)	Da 1 GiB a 16 TiB	Da 1 GiB a 16 TiB
IOPS predefiniti/di base	3000	3 IOPS/GiB (minimo 100 IOPS) fino a un massimo di 16.000 IOPS. I volumi inferiori a 1 TiB possono anche espandersi fino a 3.000 IOPS.
IOPS massimi per volume	16,000	16,000
Velocità di trasmissione effettiva predefinita/base	125 MiB/s	*Il limite di velocità di trasmissione effettiva è compreso tra 128 MiB/s e 250 MiB/s, a seconda delle dimensioni del volume.
Velocità di trasmissione effettiva massima per volume	1,000 MiB/s	250 MiB/s
Prezzo	0,08 USD per GiB al mese, 3.000 IOPS gratuiti e 0,005 USD per capacità di IOPS allocata al mese oltre i 3.000; 125 MiB/s gratuiti e 0,04 USD per MiB/s al mese allocato oltre i 125 MiB/s	0,10 USD per GiB al mese

Selezione di IOPS e velocità di trasmissione effettiva durante la migrazione a gp3

Per allocare un volume gp2, è necessario calcolare le dimensioni del volume al fine di ottenere IOPS e velocità di trasmissione effettiva proporzionali. Con gp3, non è necessario allocare un volume di dimensioni maggiori per ottenere prestazioni più elevate. Puoi scegliere le dimensioni e le prestazioni desiderate in base alle esigenze dell'applicazione. La selezione delle dimensioni e dei parametri

prestazionali corretti (IOPS e velocità di trasmissione effettiva) consente di ridurre al massimo i costi senza influire sulle prestazioni.

Ecco una tabella per aiutarti a selezionare le opzioni di configurazione di gp3:

Volume size (Dimensione dei volumi)	IOPS	Prestazioni
Da 1 a 170 GiB	3000	125 MiB/s
Da 170 a 334 GiB	3000	125 MiB/s se il tipo di istanza EC2 scelto supporta 125 MiB/s o meno, utilizza un valore superiore in base all'utilizzo, massimo 250 MiB/s*.
Da 334 a 1.000 GiB	3000	125 MiB/s se il tipo di istanza EC2 scelto supporta 125 MiB/s o meno, utilizza un valore superiore in base all'utilizzo, massimo 250 MiB/s*.
1.000 GiB e oltre	Corrisponde agli IOPS di gp2 (dimensioni in GiB x 3) o agli IOPS massimi forniti dal volume gp2 corrente	125 MiB/s se il tipo di istanza EC2 scelto supporta 125 MiB/s o meno, utilizza un valore superiore in base all'utilizzo, massimo 250 MiB/s*.

*Gp3 ha la capacità di fornire una velocità di trasmissione effettiva fino a 1.000 MiB/s. Poiché gp2 fornisce una velocità di trasmissione effettiva massima di 250 MiB/s, potrebbe non essere necessario andare oltre questo limite quando si utilizza gp3.

Configurazione delle reti

La maggior parte dei cluster viene avviata in una rete virtuale utilizzando Amazon Virtual Private Cloud (Amazon VPC). Un VPC è una rete virtuale isolata all'interno di AWS che è logicamente isolata all'interno dell'account AWS. È possibile configurare aspetti quali gli intervalli di indirizzi IP privati,

le sottoreti, le tabelle di routing e i gateway di rete. Per ulteriori informazioni, consulta la Guida per l'utente di Amazon VPC.

VPC offre le seguenti caratteristiche:

- Elaborazione di dati sensibili

  L'avvio di un cluster in un VPC è simile all'avvio del cluster in una rete privata con strumenti aggiuntivi, ad esempio tabelle di routing e liste di controllo degli accessi di rete, per definire chi può accedere alla rete. Se si stanno elaborando dati sensibili nel cluster, potrebbe essere necessario il controllo degli accessi aggiuntivo fornito dall'avvio del cluster in un VPC. Inoltre, puoi scegliere di avviare le risorse in una sottorete privata in cui nessuna di tali risorse dispone di una connessione a Internet diretta.
- Accesso alle risorse su una rete interna

  Se l'origine dati si trova in una rete privata, potrebbe essere difficile o non opportuno caricare tali dati in AWS per l'importazione in Amazon EMR, a causa della quantità di dati da trasferire o della natura riservata dei dati. Invece, puoi avviare il cluster in un VPC e collegare il data center al VPC tramite una connessione VPN, consentendo al cluster di accedere a risorse sulla rete interna. Ad esempio, se nel data center è disponibile un database Oracle, l'avvio del cluster in un VPC connesso a tale rete tramite VPN consente al cluster di accedere al database Oracle.

Sottoreti pubbliche e private

Puoi avviare cluster Amazon EMR in sottoreti VPC pubbliche e private. Ciò significa che non occorre una connessione a Internet per eseguire un cluster Amazon EMR; tuttavia, potrebbe essere necessario configurare Network Address Translation (NAT) e gateway VPN per accedere a servizi o risorse che si trovano all'esterno del VPC, ad esempio in una intranet aziendale o endpoint dei servizi AWS pubblici, come AWS Key Management Service.

> **Important**
>
> Amazon EMR supporta solo l'avvio di cluster in sottoreti private nelle versioni 4.2 e successive.

Per ulteriori informazioni su Amazon VPC, consulta la Guida per l'utente di Amazon VPC.

Argomenti

Opzioni di Amazon VPC

Quando avvii un cluster Amazon EMR all'interno di un VPC, è possibile avviarlo all'interno di una sottorete pubblica, privata o condivisa. A seconda del tipo di sottorete scelta per un cluster, esistono leggere ma significative differenze di configurazione.

Sottoreti pubbliche

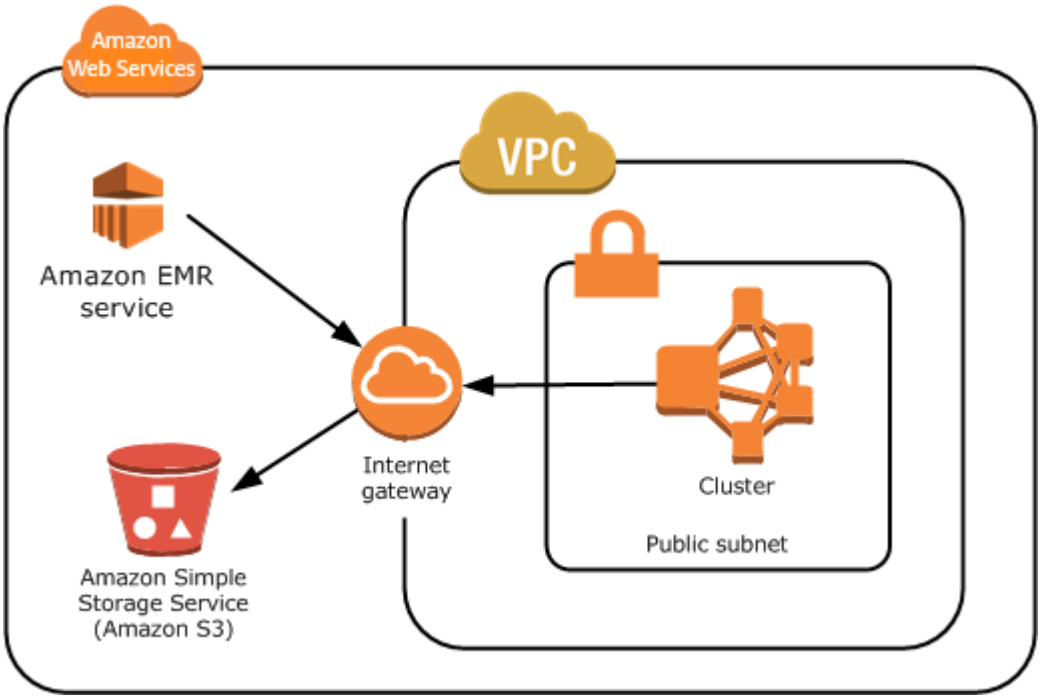
I cluster EMR in una sottorete pubblica richiedono un gateway Internet connesso. Questo accade perché i cluster Amazon EMR devono accedere ai servizi AWS e ad Amazon EMR. Se un servizio, ad esempio Amazon S3, offre la possibilità di creare un endpoint VPC, puoi accedere a tali servizi utilizzando l'endpoint anziché accedere a un endpoint pubblico tramite un gateway Internet. Inoltre, Amazon EMR non è in grado di comunicare con i cluster in sottoreti pubbliche tramite un dispositivo NAT (Network Address Translation). A questo scopo è richiesto un gateway Internet, ma puoi continuare a utilizzare un'istanza NAT o gateway per altro traffico in scenari più complessi.

Tutte le istanze in un cluster si connettono ad Amazon S3 tramite un endpoint VPC o un gateway Internet. Altri servizi AWS che non supportano attualmente endpoint VPC utilizzano solo un gateway Internet.

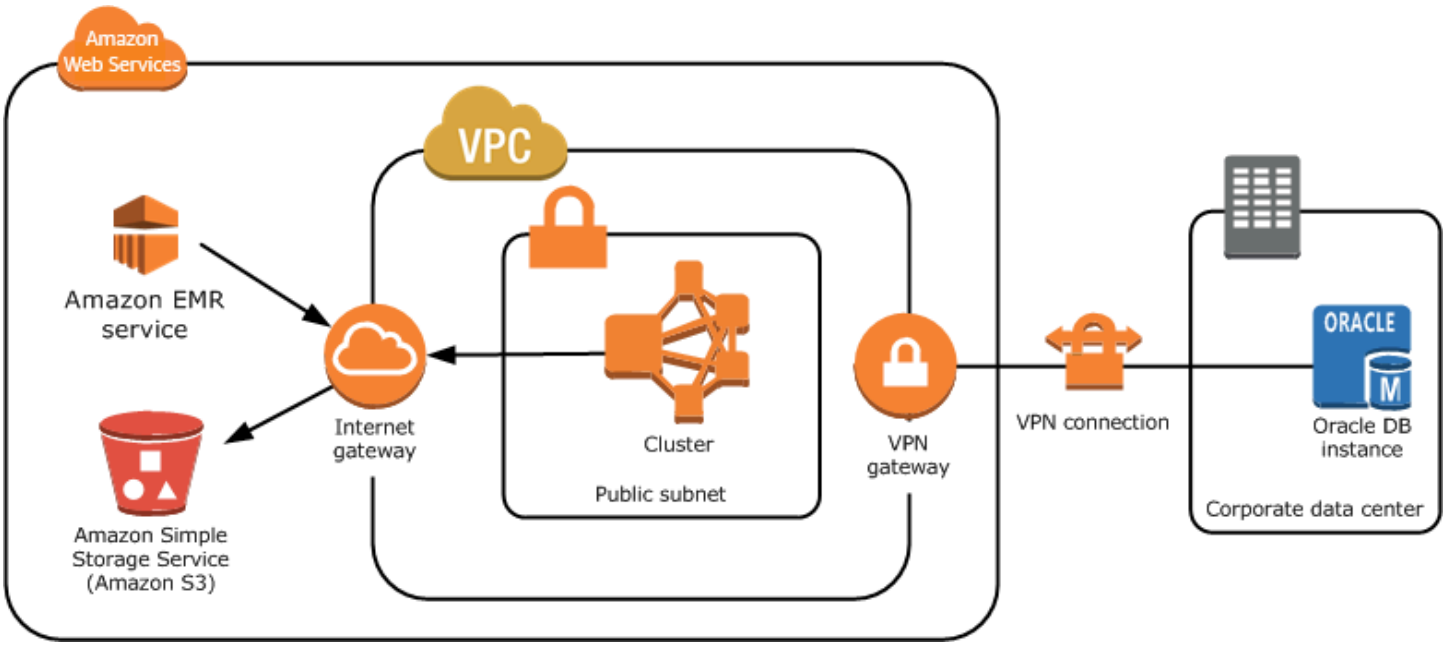
Se disponi di ulteriori risorse AWS che non desideri connettere al gateway Internet, puoi avviare questi componenti in una sottorete privata che crei all'interno di VPC.

I cluster in esecuzione in una sottorete pubblica utilizzano due gruppi di sicurezza: uno per il nodo primario e l'altro per i nodi core e attività. Per ulteriori informazioni, consulta Controllo del traffico di rete con gruppi di sicurezza.

Nel seguente diagramma viene mostrato in che modo un cluster Amazon EMR viene eseguito in un VPC utilizzando una sottorete pubblica. Il cluster è in grado di connettersi ad altre risorse AWS, ad esempio bucket Amazon S3, tramite il gateway Internet.



Nel seguente diagramma viene mostrato come configurare un VPC per consentire a un cluster nel VPC di accedere a risorse nella propria rete, ad esempio un database Oracle.



Sottoreti private

Una sottorete privata consente di avviare risorse AWS senza che sia necessario collegare un gateway Internet alla sottorete. Amazon EMR supporta l'avvio di cluster in sottoreti private con versioni 4.2.0 o successive.

> **Note**
>
> Quando si configura un cluster Amazon EMR in una sottorete privata, si consiglia di configurare anche gli endpoint VPC per Simple Storage Service (Amazon S3). Se il cluster EMR si trova in una sottorete privata senza endpoint VPC per Simple Storage Service (Amazon S3), verranno addebitati costi aggiuntivi del gateway NAT associati al traffico S3 perché il traffico tra il cluster EMR e S3 non rimarrà all'interno del VPC.

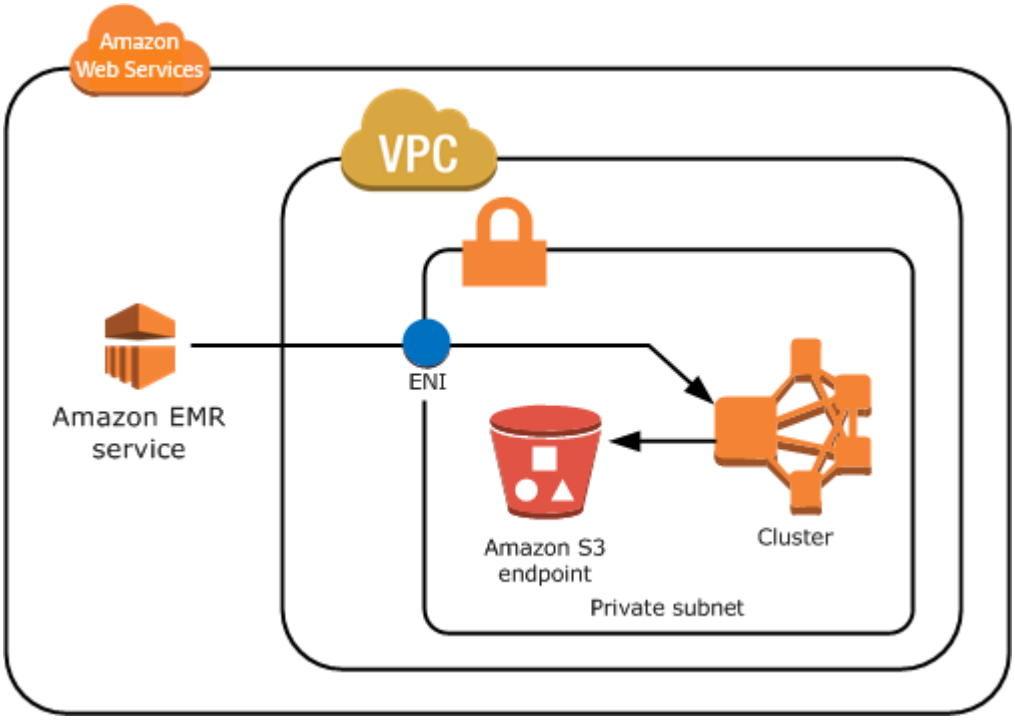
Le sottoreti private differiscono dalle sottoreti pubbliche nei modi seguenti:

- Per accedere ai servizi AWS che non forniscono un endpoint VPC, devi comunque utilizzare un'istanza NAT o un gateway Internet.
- Ti consigliamo di fornire almeno un percorso ai bucket di log del servizio Amazon EMR e al repository Amazon Linux in Amazon S3. Per ulteriori informazioni, consulta Policy Amazon S3 minima per sottorete privata
- Se utilizzi caratteristiche EMRFS, devi disporre di un endpoint VPC Amazon S3 e un percorso dalla sottorete privata a DynamoDB.
- Il debug funziona solo se fornisci un percorso dalla sottorete privata a un endpoint Amazon SQS pubblico.
- La creazione di una configurazione della sottorete privata con un'istanza NAT o un gateway in una sottorete pubblica è supportata solo utilizzando AWS Management Console. Il modo più semplice per aggiungere e configurare istanze NAT ed endpoint VPC Amazon S3 per cluster Amazon EMR è utilizzare la pagina VPC Subnets List (Elenco di sottoreti VPC) nella console Amazon EMR. Per configurare gateway NAT, consulta Gateway NAT nella Guida per l'utente di Amazon VPC.
- Non puoi modificare una sottorete con un cluster Amazon EMR esistente da pubblica a privata o viceversa. Per individuare un cluster Amazon EMR all'interno di una sottorete privata, il cluster deve essere avviato in tale sottorete privata.

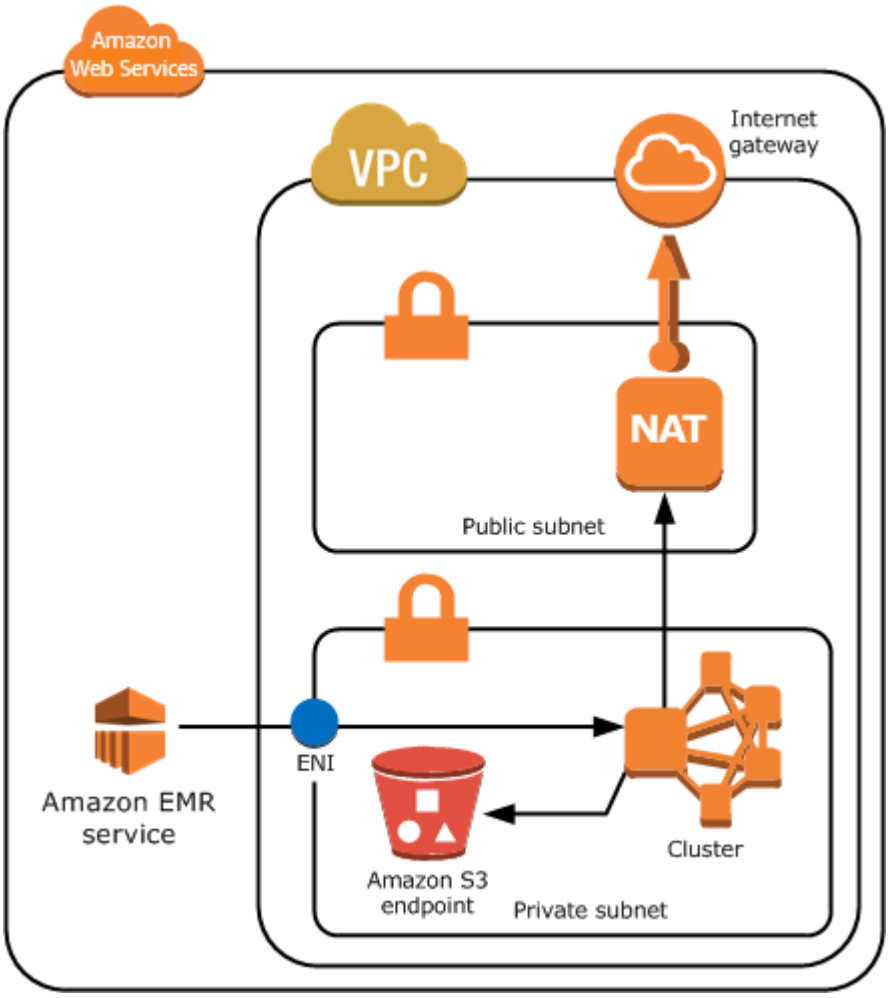
Amazon EMR crea e utilizza gruppi di sicurezza diversi per i cluster in una sottorete privata: ElasticMapReduce-Master-Private, ElasticMapReduce-Slave-Private ed ElasticMapReduce-ServiceAccess. Per ulteriori informazioni, consulta Controllo del traffico di rete con gruppi di sicurezza.

Per un elenco completo di NACL del tuo cluster, scegli Security groups for Primary (Gruppi di sicurezza per primario) e Security groups for Core & Task(Gruppi di sicurezza per principale e attività) nella pagina Cluster Details (Dettagli del cluster) della console di Amazon EMR.

Nella seguente immagine viene mostrato in che modo un cluster Amazon EMR viene configurato all'interno di una sottorete privata. La sola comunicazione all'esterno della sottorete è verso Amazon EMR.



Nell'immagine seguente viene mostrata una configurazione di esempio per un cluster Amazon EMR all'interno di una sottorete privata connessa a un'istanza NAT che si trova in una sottorete pubblica.



Sottoreti condivise

La condivisione VPC consente ai clienti di condividere sottoreti con altri account AWS all'interno della stessa organizzazione AWS. Puoi avviare i cluster Amazon EMR in sottoreti condivise pubbliche e private, con le seguenti avvertenze.

Il proprietario della sottorete deve condividere una sottorete prima di poter avviare un cluster Amazon EMR. Tuttavia, le sottoreti condivise possono essere successivamente non condivise. Per ulteriori informazioni, consulta Utilizzo dei VPC condivisi. Quando un cluster viene avviato in una sottorete condivisa e tale sottorete condivisa viene quindi non condivisa, è possibile osservare comportamenti specifici basati sullo stato del cluster Amazon EMR quando la sottorete non è condivisa.

- La condivisione della sottorete viene annullata prima che il cluster venga avviato: se il proprietario interrompe la condivisione dell'Amazon VPC o della sottorete mentre il partecipante avvia un cluster, il cluster potrebbe non avviarsi o essere parzialmente inizializzato senza eseguire il provisioning di tutte le istanze richieste.

- La condivisione della sottorete viene annullata dopo il cluster è stato lanciato: quando il proprietario interrompe la condivisione di una sottorete o dell'Amazon VPC con il partecipante, i cluster del partecipante non saranno in grado di eseguire il dimensionamento per aggiungere nuove istanze o sostituire istanze non integre.

Quando avvii un cluster Amazon EMR, vengono creati più i gruppi di sicurezza. In una sottorete condivisa, il partecipante controlla questi gruppi di sicurezza. Il proprietario della sottorete può visualizzare i gruppi di sicurezza, ma non è in grado di eseguire azioni. Se il proprietario della sottorete desidera rimuovere o modificare il gruppo di sicurezza, il partecipante che ha creato il gruppo di sicurezza deve eseguire l'azione.

Controllo delle autorizzazioni VPC con IAM

Per impostazione predefinita, tutti gli utenti possono visualizzare la totalità di sottoreti per l'account e qualsiasi utente può avviare un cluster in qualsiasi sottorete.

Quando avvii un cluster in un VPC, puoi utilizzare AWS Identity and Access Management (IAM) per controllare l'accesso ai cluster e limitare le operazioni utilizzando policy, proprio come nel caso dei cluster avviati in Amazon EC2 Classic. Per ulteriori informazioni su IAM, consulta la Guida per l'utente IAM.

Puoi anche utilizzare IAM per controllare chi può creare e amministrare le sottoreti. Ad esempio, puoi creare un account per amministrare sottoreti e un secondo account che può avviare cluster, ma non può modificare le impostazioni di Amazon VPC. Per ulteriori informazioni sull'amministrazione di policy e operazioni in Amazon EC2 e Amazon VPC, consulta Policy IAM per Amazon EC2 nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Configurazione di un VPC a cluster host

Prima di poter avviare i cluster in un VPC, devi creare un VPC e una sottorete. Per sottoreti pubbliche, devi creare un gateway Internet e collegarlo alla sottorete. Le seguenti istruzioni descrivono come creare un VPC in grado di ospitare cluster Amazon EMR.

Creazione di un VPC con sottoretI per un cluster Amazon EMR

1. Accedere alla console Amazon VPC all'indirizzo https://console.aws.amazon.com/vpc/.
2. Nella parte in alto a destra della pagina, scegli la Regione AWS per il VPC.
3. Seleziona Crea VPC.

4. Nella pagina VPC settings (Impostazioni VPC), scegli VPC and more (VPC e altro).
5. In Name tag auto-generation (Generazione automatica del tag nome), abilita Auto-generate (Generazione automatica) e inserisci un nome per il VPC. Ciò consente di identificare il VPC e la sottorete nella console Amazon VPC dopo che sono stati creati.
6. Nel campo IPv4 CIDR block (Blocco CIDR IPv4), inserisci uno spazio di indirizzi IP privati per il VPC per assicurare la corretta risoluzione del nome host DNS; in caso contrario, si potrebbero verificare errori del cluster Amazon EMR. Ciò include i seguenti intervalli di indirizzi 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. Per Number of Availability Zones (AZs) (Numero di zone di disponibilità [AZ]), scegli il numero di zone di disponibilità in cui desideri avviare le sottoreti.
8. In Number of public subnets (Numero di sottoreti pubbliche), scegli una singola sottorete pubblica che vuoi aggiungere al tuo VPC. Se i dati utilizzati nel cluster sono disponibili su Internet (ad esempio, in Amazon S3 o Amazon RDS), ti basta utilizzare una sottorete pubblica senza bisogno di aggiungere una sottorete privata.
9. Per Number of private subnets (Numero di sottoreti private), scegli il numero di sottoreti private che vuoi aggiungere al tuo VPC. Selezionane una o più se i dati per l'applicazione sono archiviati nella rete in uso (ad esempio, in un database Oracle). Per un VPC in una sottorete privata, tutte le istanze Amazon EC2 devono avere almeno un percorso ad Amazon EMR tramite l'interfaccia di rete elastica. Nella console, questo viene configurato automaticamente.
10. In NAT gateways (Gateway NAT), puoi facoltativamente scegliere di aggiungere gateway NAT. Sono necessari solo se disponi di sottoreti private che devono comunicare con Internet.
11. In VPC endpoints (Endpoint VPC), puoi facoltativamente scegliere di aggiungere endpoint per Amazon S3 alle tue sottoreti.
12. Verifica che i campi Enable DNS hostnames (Abilita hostname DNS) e Enable DNS resolution (Abilita risoluzione DNS) siano selezionati. Per ulteriori informazioni, consulta Utilizzo del DNS con il tuo VPC.
13. Seleziona Crea VPC.
14. Una finestra di stato mostra lo stato di avanzamento del lavoro. Al termine, scegli View VPC (Visualizza VPC) per navigare alla pagina Your VPCs (I tuoi VPC), che mostrerà il VPC predefinito e il VPC appena creato. Il VPC creato è un VPC non predefinito, di conseguenza la colonna Default VPC (VPC predefinito) visualizza No.

15. Se desideri associare il VPC a una voce DNS che non include un nome di dominio, vai su DHCP option sets (Set di opzioni DHCP), scegli Create DHCP options set (Crea set di opzioni DHCP) e ometti un nome di dominio. Dopo aver creato il set di opzioni, accedi al nuovo VPC, scegli Edit DHCP options set (Modifica set di opzioni DHCP) nel menu Actions (Operazioni) e seleziona il nuovo set di opzioni. Non puoi modificare il nome di dominio utilizzando la console dopo che il set di opzioni DNS è stato creato.

    Con Hadoop e applicazioni correlate è una best practice garantire la risoluzione del nome di dominio completo (FQDN) per i nodi. Per garantire la risoluzione DNS corretta, configura un VPC che include un set di opzioni DHCP i cui parametri sono impostati sui seguenti valori:

    - domain-name = **`ec2.internal`**

      Utilizza **`ec2.internal`** se la Regione è Stati Uniti orientali (Virginia settentrionale). Per le altre Regioni, utilizza *`region-name`***`.compute.internal`**. Per gli esempi in `us-west-2`, utilizza **`us-west-2.compute.internal`**. Per la Regione AWS GovCloud (Stati Uniti occidentali), utilizza **`us-gov-west-1.compute.internal`**.
    - domain-name-servers = **`AmazonProvidedDNS`**

    Per ulteriori informazioni, consulta Set opzioni DHCP nella Guida per l'utente di Amazon VPC.
16. Dopo che il VPC è stato creato, vai alla pagina Subnets (Sottoreti) e prendi nota del Subnet ID (ID sottorete) di una delle sottoreti del nuovo VPC. Utilizza queste informazioni quando avvii il cluster Amazon EMR nel VPC.

Avvio di cluster in un VPC

Quando disponi di una sottorete configurata per ospitare cluster Amazon EMR, avvia il cluster in tale sottorete specificando l'identificatore di sottorete associato durante la creazione del cluster.

> **Note**
>
> Amazon EMR supporta sottoreti private nelle versioni finali 4.2 e successive.

Quando il cluster viene avviato, Amazon EMR aggiunge gruppi di sicurezza a seconda che il cluster venga avviato in sottoreti private o pubbliche VPC. Tutti i gruppi di sicurezza consentono l'ingresso alla porta 8443 per la comunicazione con il servizio Amazon EMR, ma gli intervalli di indirizzi IP

variano per sottoreti pubbliche e private. Amazon EMR gestisce tutti questi gruppi di sicurezza e nel corso del tempo potrebbe essere necessario aggiungere altri indirizzi IP all'intervallo AWS. Per ulteriori informazioni, consulta Controllo del traffico di rete con gruppi di sicurezza.

Per gestire il cluster su un VPC, Amazon EMR collega un dispositivo di rete al nodo primario e lo gestisce tramite questo dispositivo. Puoi visualizzare questo dispositivo utilizzando l'operazione API Amazon EC2 `DescribeInstances`. Se modifichi questo dispositivo in qualsiasi modo, il cluster potrebbe non riuscire.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Avvio di un cluster in un VPC con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Networking (Reti), vai al campo Virtual private cloud (VPC) (Cloud privato virtuale [VPC]). Inserisci il nome del tuo VPC o scegli Browse (Sfoglia) per selezionarlo. In alternativa, scegli Create VPC (Crea VPC) per creare un VPC da utilizzare per il cluster.
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Avvio di un cluster in un VPC con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.

2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate).
4. Nella sezione Hardware Configuration (Configurazione hardware), per Network (Rete), seleziona l'ID di una rete VPC creata in precedenza.
5. Per EC2 Subnet (Sottorete EC2), seleziona l'ID di una sottorete creata in precedenza.

    a. Se la sottorete privata è configurata correttamente con opzioni NAT instance (Istanza NAT) e S3 endpoint (Endpoint S3), sopra i nomi sottorete e gli identificatori viene visualizzato (EMR Ready) (Pronto per EMR).

    b. Se la sottorete privata non dispone di un'istanza NAT e/o endpoint S3, puoi completare la configurazione scegliendo Add S3 endpoint and NAT instance (Aggiungi endpoint S3 e istanza NAT), Add S3 endpoint (Aggiungi endpoint S3) o Add NAT instance (Aggiungi istanza NAT). Seleziona le opzioni desiderate per l'istanza NAT e l'endpoint S3 e scegli Configure (Configura).

    > **Important**
    >
    > Per creare un'istanza NAT da Amazon EMR, sono necessarie le autorizzazioni ec2:CreateRoute, `ec2:RevokeSecurityGroupEgress`, `ec2:AuthorizeSecurityGroupEgress`, `cloudformation:DescribeStackEvents` e `cloudformation:CreateStack`.

    > **Note**
    >
    > Esiste un costo aggiuntivo per avviare un'istanza Amazon EC2 per il dispositivo NAT.

6. Procedere con la creazione del cluster.

AWS CLI

Avvio di un cluster in un VPC con la AWS CLI

> **Note**
>
> AWS CLI non fornisce un modo per creare un'istanza NAT automaticamente e collegarla alla sottorete privata. Tuttavia, per creare un endpoint S3 nella sottorete puoi utilizzare i comandi della CLI di Amazon VPC. Utilizza la console per creare istanze NAT e avviare cluster in una sottorete privata.

Dopo che il VPC è stato configurato, puoi avviare cluster Amazon EMR al suo interno utilizzando il sottocomando `create-cluster` con il parametro `--ec2-attributes`. Utilizza il parametro `--ec2-attributes` per specificare la sottorete VPC per il cluster.

- Per creare un cluster in una sottorete specifica, digita il comando seguente, sostituisci *myKey* con il nome della coppia di chiavi Amazon EC2 e sostituisci *77XXXX03* con l'ID sottorete.

```
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
```

  Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi core. Tutti i nodi utilizzano il tipo di istanza specificato nel comando.

  > **Note**
  >
  > Se in precedenza non sono stati creati il ruolo del servizio Amazon EMR predefinito e il profilo dell'istanza EC2, digita `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Policy Amazon S3 minima per sottorete privata

Per le sottoreti private, devi offrire ad Amazon EMR almeno la possibilità di accedere ai repository Amazon Linux. Questa policy della sottorete privata fa parte delle policy endpoint VPC per accedere

ad Amazon S3. Con Amazon EMR 5.25.0 o versioni successive, per abilitare l'accesso con un clic a Spark History Server persistente, devi consentire ad Amazon EMR di accedere al bucket di sistema che raccoglie i log di eventi Spark. Se abiliti la registrazione, fornisci le autorizzazioni PUT a un bucket `aws157-logs-*`. Per ulteriori informazioni, consulta Accesso con un clic a Spark History Server persistente.

Spetta a te determinare le restrizioni della policy che soddisfano le esigenze aziendali. Ad esempio, è possibile specificare la Regione `packages.us-east-1.amazonaws.com` per evitare un nome di bucket Amazon S3 ambiguo. La seguente policy di esempio fornisce le autorizzazioni per accedere ai repository Amazon Linux e al bucket di sistema Amazon EMR per la raccolta dei log di eventi Spark. Sostituisci *`MyRegion`* con la Regione in cui si trovano i bucket di log, ad esempio `us-east-1`.

Per ulteriori informazioni sull'utilizzo delle policy IAM con gli endpoint Amazon VPC, consulta Policy dell'endpoint per 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/*"
```

```
            ]
        }
    ]
}
```

La policy di seguito fornisce le autorizzazioni necessarie per accedere ai repository Amazon Linux 2. L'AMI Amazon Linux 2 è l'impostazione predefinita.

```
{
    "Statement": [
        {
            "Sid": "AmazonLinux2AMIRepositoryAccess",
            "Effect": "Allow",
            "Principal": "*",
            "Action": "s3:GetObject",
            "Resource": [
             "arn:aws:s3:::amazonlinux.MyRegion.amazonaws.com/*",
             "arn:aws:s3:::amazonlinux-2-repos-MyRegion/*"
            ]
        }
    ]
}
```

Più risorse per l'approfondimento della conoscenza di VPC

Utilizza gli argomenti seguenti per avere ulteriori informazioni sui VPC e sulle sottoreti.

- Sottoreti private in un VPC
  - Scenario 2: VPC con sottoreti pubbliche e private (NAT)
  - Istanze NAT
  - Elevata disponibilità per istanze NAT Amazon VPC: un esempio
- Sottoreti pubbliche in un VPC
  - Scenario 1: VPC con una sottorete pubblica singola
- Informazioni VPC generali
  - Guida per l'utente di Amazon VPC
  - Peering di VPC
  - Utilizzo di interfacce di rete elastiche con il tuo VPC
  - Connessione sicura alle istanze Linux in esecuzione in un VPC privato

Creazione di un cluster con parchi istanze o gruppi di istanze uniformi

Quando si crea un cluster e si specifica la configurazione del nodo primario, dei nodi core e dei nodi attività, sono disponibili due opzioni di configurazione. Puoi utilizzare parchi istanze o gruppi di istanze uniformi. L'opzione di configurazione scelta si applica a tutti i nodi, vale per la durata del cluster e i parchi istanze e gruppi di istanze non possono coesistere in un cluster. La configurazione di parchi istanze è disponibile in Amazon EMR versione 4.8.0 e successive, escluse le versioni 5.0.x.

Puoi utilizzare la console Amazon EMR, AWS CLI o l'API Amazon EMR per creare cluster con una delle due configurazioni. Quando utilizzi il comando `create-cluster` di AWS CLI, puoi utilizzare i parametri `--instance-fleets` per creare il cluster utilizzando parchi istanze o, in alternativa, puoi utilizzare i parametri `--instance-groups` per crearlo utilizzando gruppi di istanze uniformi.

Lo stesso vale utilizzando l'API Amazon EMR. Utilizza la configurazione `InstanceGroups` per specificare una matrice di oggetti `InstanceGroupConfig`, oppure utilizza la configurazione `InstanceFleets` per specificare una matrice di oggetti `InstanceFleetConfig`.

Nella nuova console Amazon EMR, puoi scegliere di utilizzare gruppi di istanze o parchi istanze quando crei un cluster e hai la possibilità di utilizzare le istanze spot con ciascuno di essi. Nella vecchia console Amazon EMR, se utilizzi le impostazioni Quick Options (Opzioni rapide) predefinite quando crei un cluster, Amazon EMR applica al cluster la configurazione dei gruppi di istanze uniformi e utilizza le istanze on demand. Per utilizzare istanze Spot con gruppi di istanze uniformi o per configurare parchi istanze e altre personalizzazioni, scegli Advanced Options (Opzioni avanzate).

Parchi istanze

La configurazione dei parchi istanze offre la più ampia varietà di opzioni di provisioning per istanze Amazon EC2. Ogni tipo di nodo dispone di un singolo parco istanze e l'utilizzo di un parco istanze dell'attività è opzionale. Puoi specificare fino a cinque tipi di istanze EC2 per parco istanze o 30 tipi di istanze EC2 per parco istanze al momento della creazione di un cluster utilizzando la AWS CLI o l'API di Amazon EMR e una strategia di allocazione per istanze on demand e Spot. Per i parchi istanze principali e attività, puoi assegnare una capacità target per istanze on demand e un'altra per istanze Spot. Amazon EMR sceglie qualsiasi combinazione dei tipi di istanze specificati per raggiungere le capacità target, eseguendo il provisioning sia delle istanze on demand che Spot.

Per il tipo di nodo primario, Amazon EMR consente di scegliere un singolo tipo di istanza dall'elenco e di specificare se è stato effettuato il provisioning come un'istanza on demand o spot. I parchi istanze offrono anche opzioni aggiuntive per gli acquisti di istanze Spot e on demand. Le opzioni di istanza Spot includono un timeout che specifica un'azione da eseguire se non è possibile assegnare

la capacità Spot e una strategia di allocazione preferita (ottimizzata per la capacità) per avviare i parchi di istanze Spot. I parchi di istanze on demand possono anche essere avviati utilizzando la strategia di allocazione (prezzo più basso). Se si utilizza un ruolo di servizio che non è il ruolo di servizio predefinito EMR o si utilizza una policy gestita da EMR nel ruolo del servizio, è necessario aggiungere autorizzazioni aggiuntive al ruolo del servizio cluster personalizzato per abilitare la strategia di allocazione. Per ulteriori informazioni, consulta Ruolo di servizio per Amazon EMR (ruolo EMR).

Per ulteriori informazioni sulla configurazione dei parchi istanze, consulta Configurazione di parchi istanze.

Gruppi di istanze uniformi

I gruppi di istanze uniformi offrono una configurazione più semplice rispetto ai parchi istanze. Ogni cluster Amazon EMR può includere fino a un massimo di 50 gruppi di istanze: un gruppo di istanze primarie che contiene un'unica istanza Amazon EC2, un gruppo di istanze primarie che contiene una o più istanze EC2 e fino a 48 gruppi di istanze attività opzionali. Ogni gruppo di istanze principale e dell'attività può contenere qualsiasi numero di istanze Amazon EC2. Puoi dimensionare ogni gruppo di istanze aggiungendo e rimuovendo istanze Amazon EC2 manualmente oppure puoi configurare la scalabilità automatica. Per informazioni sull'aggiunta e la rimozione di istanze, consulta Uso del dimensionamento del cluster.

Per ulteriori informazioni sulla configurazione di gruppi di istanze uniformi, consulta Configurazione di gruppi di istanze uniformi.

Utilizzo di parchi istanze e gruppi di istanze

Argomenti

- Configurazione di parchi istanze
- Utilizzo di prenotazioni della capacità con parchi istanze
- Configurazione di gruppi di istanze uniformi
- Best practice per la flessibilità delle istanze e delle zone di disponibilità
- Best practice per la configurazione dei cluster

Configurazione di parchi istanze

> **Note**
>
> La configurazione dei parchi istanze è disponibile solo in Amazon EMR rilasci 4.8.0 e successivi, esclusi i rilasci 5.0.0 e 5.0.3.

La configurazione del parco istanze per i cluster Amazon EMR ti consente di selezionare un'ampia gamma di opzioni di provisioning per le istanze Amazon EC2 e ti aiuta a sviluppare una strategia di risorse flessibile ed elastica per ogni tipo di nodo nel cluster.

Nella configurazione di un parco istanze, specifica una target capacity (capacità target) per le On-Demand Instances (istanze on demand) e le Spot Instances (istanze Spot) all'interno di ogni parco istanze. All'avvio del cluster, Amazon EMR effettua il provisioning di istanze fino al raggiungimento delle capacità target. Mentre un cluster è in esecuzione, se Amazon EC2 recupera un'istanza Spot a causa di un aumento del prezzo o dell'esito negativo di un'istanza, Amazon EMR prova a sostituire l'istanza con uno qualsiasi dei tipi di istanze specificati. In questo modo è più semplice riguadagnare la capacità durante un picco dei prezzi Spot.

Puoi specificare un massimo di cinque tipi di istanza Amazon EC2 per parco istanze per Amazon EMR da utilizzare per soddisfare i target, oppure un massimo di 30 tipi di istanza EC2 per parco istanze quando crei un cluster utilizzando la AWS CLI o l'API di Amazon EMR e una strategia di allocazione per istanze on demand e Spot.

Puoi anche selezionare più sottoreti per zone di disponibilità differenti. Quando Amazon EMR avvia il cluster, viene eseguita una ricerca in tali sottoreti per individuare le istanze e le opzioni di acquisto specificate. Se Amazon EMR rileva un evento AWS su larga scala in una o più zone di disponibilità, Amazon EMR tenta automaticamente di allontanare il traffico dalle zone di disponibilità interessate e tenta di avviare nuovi cluster creati in zone di disponibilità alternative in base alle selezioni effettuate. La selezione della zona di disponibilità del cluster avviene solo alla creazione del cluster. I nodi del cluster esistenti non vengono riavviati automaticamente in una nuova zona di disponibilità in caso di interruzione della zona di disponibilità.

Considerazioni

Considera gli elementi seguenti quando utilizzi il parco istanze con Amazon EMR.

- Puoi avere uno e un solo parco istanze per tipo di nodo (primario, principale, attività). Puoi specificare fino a cinque tipi di istanze Amazon EC2 per ogni parco istanze sulla AWS

Management Console (o un massimo di 30 tipi per parco istanze al momento della creazione di un cluster utilizzando la AWS CLI o l'API di Amazon EMR e una Strategia di allocazione per parchi istanze).

- Amazon EMR sceglie uno o tutti i tipi di istanze Amazon EC2 specificati per effettuare il provisioning con opzioni di acquisto Spot e on demand.
- Puoi definire capacità target per istanze Spot e on demand per il parco istanze principale e il parco istanze attività. Utilizza vCPU o un'unità generica assegnata a ogni istanza Amazon EC2 che esegue il conteggio per il raggiungimento del target. Amazon EMR effettua il provisioning di istanze finché ogni capacità target non viene completamente raggiunta. Per il parco istanze primarie, il target è sempre uno.
- Puoi scegliere una sottorete (zona di disponibilità) o un intervallo. Se selezioni un intervallo, Amazon EMR effettua il provisioning della capacità nella zona di disponibilità più adatta.
- Quando si specifica una capacità target per istanze Spot:
  - Per ogni tipo di istanza, specifica un prezzo Spot massimo. Amazon EMR effettua il provisioning delle istanze Spot se il prezzo Spot è inferiore al prezzo Spot massimo. L'utente paga il prezzo Spot, non necessariamente il prezzo Spot massimo.
  - Per ogni parco istanze, definire un periodo di timeout per il provisioning delle istanze Spot. Se Amazon EMR non è in grado di eseguire il provisioning della capacità Spot, è possibile terminare il cluster o passare al provisioning della capacità on demand. Ciò vale solo per il provisioning dei cluster, non per il loro ridimensionamento. Se il periodo di timeout termina durante il processo di ridimensionamento del cluster, le richieste Spot non sottoposte a provisioning verranno annullate senza il trasferimento alla capacità on demand.
- Per ogni parco istanze, puoi specificare una delle seguenti strategie di allocazione per le istanze Spot: ottimizzata per prezzo e capacità, ottimizzata per capacità, prezzo più basso o diversificata in tutti i pool.
- Per ogni parco istanze, puoi applicare la strategia di allocazione del prezzo più basso per le istanze On-Demand; non puoi personalizzare la strategia di allocazione per le istanze On-Demand.
- Per ogni parco istanze con `allocation strategy - lowest-price` on demand, è possibile scegliere di applicare le opzioni di prenotazione della capacità.
- Verifica le dimensioni della sottorete prima di avviare il cluster. Quando si esegue il provisioning di un cluster con un parco attività e non ci sono sufficienti indirizzi IP disponibili nella sottorete corrispondente, il parco istanze entrerà in uno stato sospeso anziché terminare il cluster con un errore. Per evitare questo problema, si consiglia di aumentare il numero di indirizzi IP nelle sottoreti.

Opzioni del parco istanze

Utilizza le seguenti linee guida per comprendere le opzioni del parco istanze.

Argomenti

Impostazione delle capacità target

Specifica le capacità target desiderate per il parco istanze core e il parco istanze di attività. Una volta fatto, ciò determina il numero di istanze on demand e istanze Spot di cui Amazon EMR effettua il provisioning. Durante la specifica di un'istanza è l'utente che decide la misura di conteggio di ogni istanza per il raggiungimento del target. Quando si effettua il provisioning di un'istanza on demand, esegue il conteggio fino al raggiungimento della capacità target on demand. Lo stesso vale per istanze Spot. A differenza dei parchi istanze core e attività, il parco istanze primarie è sempre un'unica istanza. Pertanto, la capacità target di questo parco istanze è sempre uno.

Quando utilizzi la console, il vCPU del tipo di istanza Amazon EC2 viene utilizzato come il conteggio per le capacità target per impostazione predefinita. È possibile modificare questo conteggio in unità generiche, quindi specificare il conteggio per ogni tipo di istanza EC2. Quando utilizzi AWS CLI, assegna manualmente unità generiche per ogni tipo di istanza.

> **Important**
>
> Quando si sceglie un tipo di istanza utilizzando la AWS Management Console, il numero di VPCU per ogni tipo di istanza è il numero di vcore YARN per quel tipo di istanza, non il numero di vCPU EC2 per quel tipo di istanza. Per ulteriori informazioni sul numero di vCPU per ogni tipo di istanza, consulta Tipi di istanza di Amazon EC2.

Per ogni parco istanze, specifica fino a cinque tipi di istanza Amazon EC2. Se utilizzi un Strategia di allocazione per parchi istanze e crei un cluster utilizzando la AWS CLI o l'API di Amazon EMR, puoi specificare fino a 30 tipi di istanze EC2 per parco istanze. Amazon EMR sceglie qualsiasi combinazione di questi tipi di istanza EC2 per raggiungere le capacità target. Poiché l'obiettivo di

Amazon EMR è raggiungere la capacità target completa, si possono verificare costi aggiuntivi. Ad esempio, se la capacità di due unità non viene raggiunta e Amazon EMR è in grado di effettuare il provisioning di un'istanza con un conteggio di cinque unità, il provisioning dell'istanza viene ancora effettuato, ossia la capacità target viene superata di tre unità.

Se riduci la capacità target per ridimensionare un cluster in esecuzione, Amazon EMR tenta di completare le attività dell'applicazione e termina le istanze per soddisfare il nuovo target. Per ulteriori informazioni, consulta Terminazione al completamento dell'attività.

Opzioni di avvio

Per le istanze Spot, puoi specificare un Maximum Spot price (Prezzo Spot massimo) per ogni tipo di istanza in un parco istanze. Puoi impostare questo prezzo come una percentuale del prezzo on demand o come un importo in dollari specifico. Amazon EMR effettua il provisioning di istanze Spot se il prezzo Spot corrente in una zona di disponibilità è inferiore al prezzo Spot massimo. L'utente paga il prezzo Spot, non necessariamente il prezzo Spot massimo.

> **Note**
>
> Le istanze spot con una durata definita (note anche come blocchi Spot) non sono più disponibili per i nuovi clienti a partire dal 1° luglio 2021. Per i clienti che hanno già utilizzato la funzione, continueremo a supportare le istanze spot con una durata definita fino al 31 dicembre 2022.

In Amazon EMR 5.12.1 e versioni successive hai la possibilità di avviare parchi di istanze Spot e on demand con allocazione della capacità ottimizzata. Questa opzione di strategia di allocazione può essere impostata nella vecchia AWS Management Console o con l'API `RunJobFlow`. Tieni presente che non è possibile personalizzare la strategia di allocazione nella nuova console. L'utilizzo della strategia di allocazione richiede autorizzazioni aggiuntive per il ruolo di servizio. Se utilizzi il ruolo di servizio Amazon EMR predefinito e la policy gestita (`EMR_DefaultRole` e `AmazonEMRServicePolicy_v2`) per il cluster, le autorizzazioni per l'opzione della strategia di allocazione sono già incluse. Se non si utilizza il ruolo di servizio Amazon EMR e la policy gestita predefiniti, è necessario aggiungerli per utilizzare questa opzione. Per informazioni, consultare Ruolo di servizio per Amazon EMR (ruolo EMR).

Per ulteriori informazioni sulle istanze Spot, consulta Istanze Spot nella Guida per l'utente di Amazon EC2 per le istanze Linux. Per maggiori informazioni sulle istanze Linux, consulta Istanze on demand nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Se si sceglie di avviare parchi istanze on demand con la strategia di allocazione del prezzo più basso, è possibile utilizzare le prenotazioni di capacità. Le opzioni di prenotazione della capacità possono essere impostate utilizzando l'API `RunJobFlow` di Amazon EMR. Le prenotazioni della capacità richiedono autorizzazioni aggiuntive per il ruolo di servizio che è necessario aggiungere per utilizzare queste opzioni. Per informazioni, consultare Autorizzazioni della strategia di allocazione. Tieni presente che non è possibile personalizzare le prenotazioni di capacità nella nuova console.

Opzioni di sottorete (zone di disponibilità) multiple

Quando utilizzi parchi istanze, puoi specificare più sottoreti Amazon EC2 all'interno di un VPC ciascuna corrispondente a una zona di disponibilità diversa. Se utilizzi EC2-Classic, specifica le zone di disponibilità in modo esplicito. Amazon EMR identifica la migliore zona di disponibilità per avviare istanze in base alle specifiche del parco istanze. Il provisioning delle istanze viene sempre effettuato in una sola zona di disponibilità. Puoi selezionare sottoreti private o pubbliche, ma non puoi combinare i due tipi. Inoltre, le sottoreti che specifichi devono trovarsi all'interno dello stesso VPC.

Configurazione del nodo master

Poiché il parco istanze primarie è solo un'istanza singola, la sua configurazione è leggermente diversa dai parchi istanze core e attività. Per il parco istanze primarie, puoi selezionare solo on demand o spot perché è costituito da una sola istanza. Se utilizzi la console per creare il parco istanze, la capacità target per l'opzione di acquisto selezionata è impostata su 1. Se utilizzi AWS CLI, imposta `TargetSpotCapacity` o `TargetOnDemandCapacity` su 1 a seconda delle necessità. È comunque possibile scegliere fino a 5 tipi di istanza per il parco istanze primarie (o un massimo di 30 quando si utilizza l'opzione di allocazione delle istanze on demand o spot). Tuttavia, a differenza del parco istanze core e attività, in cui Amazon EMR può effettuare il provisioning di più istanze di diversi tipi, Amazon EMR seleziona un singolo tipo di istanza per effettuare il provisioning per il parco istanze primarie.

Strategia di allocazione per parchi istanze

Con Amazon EMR versione 5.12.1 e successive, puoi utilizzare la strategia di allocazione con istanze on demand e Spot per ogni nodo del cluster. Quando crei un cluster utilizzando la AWS CLI, l'API Amazon EMR o la console Amazon EMR con una strategia di allocazione, puoi specificare fino a 30 tipi di istanze Amazon EC2 per parco istanze. Con la configurazione predefinita del parco istanze del cluster Amazon EMR, puoi avere fino a 5 tipi di istanze per parco istanze. Si consiglia di utilizzare la strategia di allocazione per un provisioning dei cluster più rapido, un'allocazione più accurata delle istanze Spot e un minor numero di interruzioni delle istanze Spot.

Argomenti

- Strategia di allocazione con istanze On-Demand
- Strategia di allocazione con istanze Spot
- Autorizzazioni della strategia di allocazione
- Autorizzazioni IAM necessarie per una strategia di allocazione

Strategia di allocazione con istanze On-Demand

Quando utilizzi la strategia di allocazione, le tue istanze On Demand utilizzano la strategia del prezzo più basso. In questo modo vengono avviate per prime le istanze con il prezzo più basso. Quando avvii istanze On-Demand, puoi utilizzare prenotazioni di capacità aperte o mirate nei tuoi account. È possibile utilizzare prenotazioni di capacità aperte per nodi primari, core e attività. Quando utilizzi istanze On-Demand con la strategia di allocazione per i parchi istanze, è possibile che si verifichi una condizione di capacità insufficiente. Si consiglia di specificare un numero maggiore di tipi di istanza per diversificare e ridurre la possibilità che si verifichi una condizione di capacità insufficiente. Per ulteriori informazioni, consulta Utilizzo di prenotazioni della capacità con parchi istanze.

Strategia di allocazione con istanze Spot

Per Istanze Spot puoi scegliere una delle seguenti strategie di allocazione:

`price-capacity-optimized` (consigliato)

La strategia di allocazione ottimizzata per prezzo e capacità avvia le istanze Spot dai pool di istanze Spot che hanno la capacità più elevata disponibile e il prezzo più basso per il numero di istanze in fase di avvio. Di conseguenza, la strategia ottimizzata per prezzo e capacità ha in genere maggiori possibilità di ottenere capacità Spot e offre tassi di interruzione inferiori.

`capacity-optimized`

La strategia di allocazione ottimizzata per capacità avvia le istanze Spot nei pool più disponibili con le minori possibilità di interruzione nel breve termine. Questa è una buona opzione per i carichi di lavoro che potrebbero avere un costo di interruzione più elevato associato al riavvio del lavoro. Questa è la strategia predefinita per Amazon EMR rilasci 6.9.0 e precedenti.

`diversified`

Con la strategia di allocazione diversificata, Amazon EC2 distribuisce le istanze Spot in tutti i pool di capacità Spot.

`lowest-price`

La strategia di allocazione del prezzo più basso avvia le istanze Spot dal pool con il prezzo più basso con capacità disponibile. Se il pool con il prezzo più basso non ha capacità disponibile, le istanze Spot provengono dal successivo pool con il prezzo più basso che ha capacità disponibile. Se un pool esaurisce la capacità prima di soddisfare la capacità desiderata, il parco istanze Amazon EC2 continua a soddisfare la richiesta attingendo dal successivo pool con il prezzo più basso. Per accertarti che la capacità desiderata sia soddisfatta, potresti ricevere istanze spot da vari pool. Poiché questa strategia considera solo il prezzo dell'istanza e non la capacità disponibile, potrebbe comportare tassi di interruzione elevati.

Autorizzazioni della strategia di allocazione

L'opzione della strategia di allocazione richiede diverse autorizzazioni IAM incluse automaticamente nel ruolo di servizio Amazon EMR predefinito e nella policy gestita da Amazon EMR (`EMR_DefaultRole` e `AmazonEMRServicePolicy_v2`). Se si utilizza un ruolo di servizio o una policy gestita personalizzati per il cluster, è necessario aggiungere queste autorizzazioni prima di creare il cluster. Per ulteriori informazioni, consulta Autorizzazioni della strategia di allocazione.

Quando si utilizza la strategia di allocazione on demand, sono disponibili delle prenotazioni della capacità on demand (ODCR). Le opzioni di prenotazione della capacità ti consentono di specificare una preferenza per utilizzare prima la capacità riservata per i cluster Amazon EMR. È possibile utilizzarla per garantire che i carichi di lavoro critici utilizzino la capacità già riservata utilizzando ODCR aperti o mirati. Per i carichi di lavoro non critici, le preferenze di prenotazione della capacità consentono di specificare se utilizzare la capacità riservata.

Le prenotazioni della capacità possono essere utilizzate solo da istanze che dispongono di attributi corrispondenti (tipo di istanza, piattaforma e zona di disponibilità). Per impostazione predefinita, le prenotazioni di capacità aperta vengono automaticamente utilizzate da Amazon EMR durante il provisioning di istanze on demand che corrispondono agli attributi dell'istanza. Se non disponi di istanze in esecuzione che corrispondono agli attributi delle prenotazioni di capacità, queste rimangono inutilizzate finché non avvii un'istanza con attributi corrispondenti. Se non si desidera utilizzare alcuna prenotazione di capacità durante l'avvio del cluster, è necessario impostare la preferenza di prenotazione di capacità su none (nessuna) nelle opzioni di avvio.

Tuttavia, puoi anche utilizzare una prenotazione di capacità per carichi di lavoro specifici. In questo modo puoi controllare in modo esplicito quali istanze possono essere eseguite in quella capacità

riservata. Per ulteriori informazioni sulle prenotazioni di capacità On-Demand, consulta Utilizzo di prenotazioni della capacità con parchi istanze.

Autorizzazioni IAM necessarie per una strategia di allocazione

Il Ruolo di servizio per Amazon EMR (ruolo EMR) richiede autorizzazioni aggiuntive per creare un cluster che utilizzi la strategia di allocazione per parchi istanze on demand o Spot.

Queste autorizzazioni sono incluse automaticamente nel ruolo di servizio Amazon EMR `EMR_DefaultRole` e nella policy gestita da Amazon EMR `AmazonEMRServicePolicy_v2`.

Se utilizzi un ruolo di servizio o una policy gestita personalizzati per il cluster, devi aggiungere le autorizzazioni seguenti:

```
{
 "Version": "2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Action": [
    "ec2:DeleteLaunchTemplate",
    "ec2:CreateLaunchTemplate",
    "ec2:DescribeLaunchTemplates",
    "ec2:CreateLaunchTemplateVersion",
    "ec2:CreateFleet"
   ],
   "Resource": "*"
  }
}
```

Di seguito sono riportate le autorizzazioni del ruolo di servizio necessarie per creare un cluster che utilizza prenotazioni di capacità aperte o mirate. È necessario includere queste autorizzazioni oltre alle autorizzazioni necessarie per l'utilizzo della strategia di allocazione.

Example Documento della policy per le prenotazioni di capacità del ruolo di servizio

Per utilizzare le prenotazioni di capacità aperte, è necessario includere le seguenti autorizzazioni aggiuntive.

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

```
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions"
            ],
            "Resource": "*"
        }
    ]
}
```

Example

Per utilizzare le prenotazioni di capacità mirate, è necessario includere le seguenti autorizzazioni aggiuntive.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeCapacityReservations",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DeleteLaunchTemplateVersions",
                "resource-groups:ListGroupResources"
            ],
            "Resource": "*"
        }
    ]
}
```

Configurazione di parchi istanze per il cluster

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Creazione di un cluster con parchi istanze con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster), quindi seleziona Create cluster (Crea cluster).
3. In Cluster configuration (Configurazione del cluster), scegli Instance fleets (Parchi istanze).
4. Per ogni Node group (Gruppo di nodi), seleziona Add instance type (Aggiungi tipo di istanza) e scegli fino a 5 tipi di istanze per i parchi istanze primarie e core e fino a 15 tipi di istanze per i parchi istanze attività. Amazon EMR può effettuare il provisioning di qualsiasi combinazione di questi tipi di istanze quando avvia il cluster.
5. In ogni tipo di gruppo di nodi, scegli il menu a discesa Actions (Operazioni) accanto a ciascuna istanza per modificare queste impostazioni:

   Aggiunta di volumi EBS

   Specifica i volumi EBS da collegare al tipo di istanza dopo che Amazon EMR ne effettua il provisioning.

   Modifica della capacità ponderata

   Per il gruppo di nodi core, modifica questo valore con un numero qualsiasi di unità adatto alle tue applicazioni. Il numero di vCore YARN per ogni tipo di istanza del parco istanze viene utilizzato come relativa capacità ponderata predefinita. Non è possibile modificare la capacità ponderata del nodo primario.

   Modifica del prezzo spot massimo

   Per ogni tipo di istanza all'interno di un parco, specifica un prezzo spot massimo. Puoi impostare questo prezzo come una percentuale del prezzo on demand o come un importo in dollari specifico. Se il prezzo spot corrente in una zona di disponibilità è inferiore al prezzo spot massimo, Amazon EMR effettua il provisioning delle istanze spot. L'utente paga il prezzo Spot, non necessariamente il prezzo Spot massimo.
6. Facoltativamente, per aggiungere gruppi di sicurezza per i tuoi nodi, espandi EC2 security groups (firewall) (Gruppi di sicurezza EC2 ([firewall]) nella sezione Networking (Reti) e seleziona il gruppo di sicurezza per ogni tipo di nodo.

7. Facoltativamente, seleziona la casella di controllo accanto a Applica strategia di allocazione se desideri utilizzare tale opzione e seleziona la strategia di allocazione che desideri specificare per le istanze Spot. Se il tuo ruolo di servizio Amazon EMR non dispone delle autorizzazioni richieste, è consigliabile non selezionare questa opzione. Per ulteriori informazioni, consulta Strategia di allocazione per parchi istanze.
8. Scegli qualsiasi altra opzione applicabile al cluster.
9. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Creazione di un cluster con parchi istanze con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Nella parte superiore della finestra della console, scegli Go to advanced options (Vai alle opzioni avanzate), immetti le opzioni Software Configuration (Configurazione software) e seleziona Next (Successivo).
4. In Cluster Composition (Composizione cluster), scegli Instance fleets (Parchi istanze). Quando selezioni l'opzione parco istanze, dovresti visualizzare le opzioni per specificare la Target capacity (Capacità di destinazione) di istanze on demand e Spot nella tabella Cluster Nodes and Instances (Nodi cluster e istanze).
5. Per Network (Rete), immetti un valore. Se scegli un VPC per Network (Rete), seleziona una EC2 Subnet (Sottorete EC2) singola o fai CTRL+clic per scegliere più sottoreti Amazon EC2. Le sottoreti selezionate devono essere dello stesso tipo (pubblica o privata). Se si sceglie solo una sottorete, il cluster viene avviato in tale sottorete. Se si sceglie un gruppo di sottoreti, la sottorete più adatta viene selezionata dal gruppo all'avvio del cluster.

> **Note**
>
> L'account e la Regione offrono la possibilità di scegliere Launch into EC2-Classic (Avvio in EC2-Classic) per Network (Rete). Se scegli questa opzione, effettua una o più selezioni da EC2 Availability Zones (Zone di disponibilità EC2) anziché da EC2

Subnets (Sottoreti EC2). Per ulteriori informazioni, consulta Amazon EC2 e Amazon VPC nella Guida per l'utente di Amazon EC2 per le istanze Linux.

6. In Allocation strategy (Strategia di allocazione), seleziona la casella di controllo per applicare le strategie di allocazione se desideri utilizzare tale opzione. Per ulteriori informazioni, consulta Strategia di allocazione per parchi istanze.
7. Per ciascun Node type (Tipo di nodo), se desideri modificare il nome predefinito di un parco istanze, scegli l'icona a forma di matita e immetti un nome descrittivo. Se desideri rimuovere il parco istanze Task (Attività), scegli l'icona X a destra della riga Task (Attività).
8. Scegli Add/remove instance types to fleet (Aggiungi/rimuovi tipi di istanza al/dal parco istanze) e scegli fino a 5 tipi di istanze dall'elenco di parchi istanze primarie e core e fino a 15 tipi di istanze dall'elenco di parchi istanze attività. Amazon EMR può scegliere di effettuare il provisioning di qualsiasi combinazione di questi tipi di istanze quando avvia il cluster.
9. Per ogni tipo di istanza principale e attività, scegli come definire la capacità ponderata (ogni istanza conta come X unità) per quell'istanza. Il numero di vCore YARN per ogni tipo di istanza del parco istanze viene utilizzato come unità di capacità ponderata predefinita, ma è possibile modificare il valore in qualsiasi unità utile per le applicazioni.
10. In Target capacity (Capacità di destinazione), definisci il numero totale di istanze on demand e Spot che desideri per ogni parco istanze. EMR assicura che le istanze del parco istanze soddisfino le unità richieste per la capacità target on demand e Spot. Se per un parco istanze non sono specificate unità on demand o Spot, non viene fornita alcuna capacità per tale parco istanze.
11. Se un parco istanze è configurato con una capacità target per Spot, è possibile inserire il prezzo Spot massimo come % of On-Demand (% del prezzo on demand) o immettere una quantità Dollars ($) (Dollari ($)) in USD.
12. Per avere volumi EBS collegati al tipo di istanza quando viene effettuato il provisioning, seleziona la matita accanto a EBS Storage (Archiviazione EBS), quindi immetti le opzioni di configurazione EBS.
13. Se hai definito un conteggio istantaneo per Spot units (Unità Spot), scegli Advanced Spot options (Opzioni Spot avanzate) in base alle seguenti linee guida:

    - Provisioning timeout (Timeout di provisioning): utilizza queste impostazioni per controllare il comportamento di Amazon EMR quando non è in grado di effettuare il provisioning di istanze Spot tra i Fleet instance types (Tipi di istanze del parco istanze) specificati. Inserisci un periodo di timeout in minuti, quindi scegli tra le opzioni Terminate the cluster (Termina il

cluster) o Switch to provisioning On-Demand Instances (Passa al provisioning di istanze on demand). Se scegli di passare al provisioning di istanze on demand, la capacità assegnata delle istanze on demand esegue il conteggio per il raggiungimento della capacità target per istanze Spot e Amazon EMR effettua il provisioning di istanze on demand finché non viene raggiunta la capacità target per istanze Spot.

14. Scegli Next (Successivo), modifica altre impostazioni cluster, quindi scegli Next (Successivo).
15. Se hai scelto di applicare la nuova strategia di allocazione, nelle impostazioni Security Options (Opzioni di sicurezza), seleziona un EMR role (Ruolo EMR) e un EC2 instance profile (Profilo dell'istanza EC2) che contengano le autorizzazioni necessarie per la strategia di allocazione. In caso contrario, la creazione del cluster avrà esito negativo.
16. Scegli Crea cluster.

AWS CLI

Per creare e avviare un cluster con parchi istanze con la AWS CLI, segui queste linee guida:

- Per creare e avviare un cluster con parchi istanze, utilizza il comando `create-cluster` insieme ai parametri `--instance-fleet`.
- Per ottenere i dettagli sulla configurazione dei parchi istanze in un cluster, utilizza il comando `list-instance-fleets`.
- Per aggiungere AMI Amazon Linux personalizzate multiple a un cluster che stai creando, usa l'opzione `CustomAmiId` con ogni specifica `InstanceType`. È possibile configurare i nodi della flotta di istanze con più tipi di istanza e più AMI personalizzate in base alle proprie esigenze. Per informazioni, consultare Esempi: creazione di un cluster con la configurazione di parchi istanze.
- Per apportare modifiche alla capacità target per un parco istanze, utilizza il comando `modify-instance-fleet`.
- Per aggiungere un parco istanze dell'attività a un cluster che non ne possiede già uno, utilizza il comando `add-instance-fleet`.
- È possibile aggiungere più AMI personalizzate al parco istanze di attività utilizzando l'argomento CustomAmiId con il comando add-instance-fleet. Per informazioni, consultare Esempi: creazione di un cluster con la configurazione di parchi istanze.
- Per utilizzare l'opzione di strategia di allocazione durante la creazione di un parco istanze, aggiornare il ruolo del servizio per includere il documento del policy di esempio nella sezione seguente.

- Per utilizzare le opzioni di riserva di capacità durante la creazione di un parco istanze con strategia di allocazione On-Demand, aggiornare il ruolo del servizio per includere il documento del policy di esempio nella sezione seguente.
- I parchi istanze vengono inclusi automaticamente nel ruolo di servizio EMR e nella policy gestita da Amazon EMR predefiniti (`EMR_DefaultRole` e `AmazonEMRServicePolicy_v2`). Se si utilizza un ruolo di servizio personalizzato o una policy gestita personalizzata per il cluster, è necessario aggiungere le nuove autorizzazioni per la strategia di allocazione nella sezione seguente.

Esempi: creazione di un cluster con la configurazione di parchi istanze

Negli esempi seguenti vengono illustrati i comandi `create-cluster` con opzioni diverse che è possibile combinare.

> **Note**
>
> Se in precedenza non sono stati creati il ruolo del servizio Amazon EMR predefinito e il profilo dell'istanza EC2, digitare `aws emr create-default-roles` per crearli prima di utilizzare il comando `create-cluster`.

Example Esempio: primaria on demand, core on demand con tipo di istanza singola, VPC predefinito

```
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 Esempio: primaria spot, core spot con tipo di istanza singola, VPC predefinito

```
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 Esempio: primaria on demand, core mista con tipo di istanza singola, sottorete EC2 singola

```
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 Esempio: primaria on demand, core spot con più tipi di istanze ponderate, durata definita e timeout per spot, intervallo di sottoreti EC2

```
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 Esempio: primaria on demand, core mista e attività con più tipi di istanze ponderate, timeout per istanze spot core, intervallo di sottoreti EC2

```
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 Esempio: primaria spot, nessuna core o attività, configurazione Amazon EBS, VPC predefinito

```
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 Esempio: AMI personalizzate multiple, tipi di istanza multipli, primario on-demand, principale on-demand

```
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 Esempio: aggiunta di un nodo attività a un cluster in esecuzione con più tipi di istanza e più AMI personalizzate

```
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 Esempio: utilizzo di un file di configurazione JSON

Puoi configurare i parametri del parco istanze in un file JSON, quindi fare riferimento al file JSON come l'unico parametro per i parchi istanze. Ad esempio, il comando seguente fa riferimento a un file di configurazione 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
```

Il file *my-fleet-config.json* specifica parchi istanze primarie, core e attività come mostrato nel seguente esempio. Il parco istanze core utilizza un prezzo spot massimo (`BidPrice`) come una percentuale di on demand, mentre i parchi istanze attività e primarie utilizzano il prezzo spot massimo (BidPriceAsPercentageofOnDemandPrice) come una stringa in 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"
            }
        ]
    }
```

```
]
```

Modifica delle capacità target per un parco istanze

Per specificare nuove capacità target per un parco istanze, utilizza il comando `modify-instance-fleet`. Devi specificare l'ID cluster e l'ID del parco istanze. Utilizza il comando `list-instance-fleets` per recuperare ID del parco istanze.

```
aws emr modify-instance-fleet --cluster-id <cluster-id> \
  --instance-fleet \
    InstanceFleetId='<instance-fleet-id>',TargetOnDemandCapacity=1,TargetSpotCapacity=1
```

Aggiunta di un parco istanze attività a un cluster

Se un cluster contiene solo parchi istanze primarie e core, puoi utilizzare il comando `add-instance-fleet` per aggiungere un parco istanze attività. Puoi utilizzare questo comando solo per aggiungere parchi istanze dell'attività.

```
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}']
```

Ottenimento dei dettagli di configurazione per parchi istanze in un cluster

Per ottenere i dettagli di configurazione dei parchi istanze in un cluster, utilizza il comando `list-instance-fleets`. Il comando richiede un ID cluster come input. Nell'esempio seguente viene illustrato il comando e il relativo output per un cluster che contiene un gruppo di istanze attività primarie e un gruppo di istanze attività core. Per la sintassi di risposta completa, consulta ListInstanceFleets nella Guida di riferimento alle API di 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"
        }
    ]
}
```

Utilizzo di prenotazioni della capacità con parchi istanze

Per avviare parchi istanze on demand con opzioni di prenotazione della capacità, allega autorizzazioni aggiuntive per il ruolo di servizio necessarie per utilizzare le opzioni di prenotazione della capacità. Poiché le opzioni di prenotazione della capacità devono essere utilizzate insieme alla strategia di allocazione on demand, è necessario includere anche le autorizzazioni necessarie per la strategia di allocazione nel ruolo di servizio e nella policy gestita. Per ulteriori informazioni, consulta Autorizzazioni della strategia di allocazione.

Amazon EMR supporta prenotazioni della capacità aperte e mirate. Negli argomenti seguenti vengono illustrate le configurazioni dei gruppi di istanze che è possibile utilizzare con l'operazione `RunJobFlow` o il comando `create-cluster` per avviare parchi istanze mediante prenotazioni della capacità on demand.

Utilizzo delle prenotazioni della capacità aperte in base al miglior tentativo

Se le istanze on demand del cluster corrispondono agli attributi delle prenotazioni di capacità aperta (tipo di istanza, piattaforma, tenancy e zona di disponibilità) disponibili nell'account, le prenotazioni di capacità vengono applicate automaticamente. Tuttavia, non è garantito che le prenotazioni di capacità verranno utilizzate. Per il provisioning del cluster, Amazon EMR valuta tutti i pool di istanze specificati nella richiesta di avvio e utilizza quello con il prezzo più basso che dispone di capacità sufficienti per avviare tutti i nodi principali richiesti. Le prenotazioni della capacità aperte disponibili che corrispondono al pool di istanze vengono applicate automaticamente. Se le prenotazioni della capacità aperte disponibili non corrispondono al pool di istanze, rimangono inutilizzate.

Una volta eseguito il provisioning dei nodi principali, la zona di disponibilità viene selezionata e risolta. Amazon EMR effettua il provisioning dei nodi attività in pool di istanze, a partire da quelli con il prezzo più basso, nella zona di disponibilità selezionata fino a quando non viene eseguito il provisioning

di tutti i nodi attività. Le prenotazioni della capacità aperte disponibili che corrispondono al pool di istanze vengono applicate automaticamente.

Di seguito sono riportati i casi d'uso della logica di allocazione della capacità di Amazon EMR per l'utilizzo delle prenotazioni della capacità aperte in base al miglior tentativo.

Esempio 1: il pool di istanze a prezzo più basso nella richiesta di avvio dispone di prenotazioni della capacità aperte disponibili

In questo caso, Amazon EMR avvia la capacità nel pool di istanze a prezzo più basso con istanze on demand. Le prenotazioni della capacità aperte disponibili in quel pool di istanze vengono utilizzate automaticamente.

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	$	$$	$$$
Istanze con provisioning	100	-	-
Prenotazione della capacità aperta utilizzata	100	-	-
Prenotazioni della capacità aperte disponibili	50	100	100

Dopo l'avvio del parco istanze, puoi eseguire `describe-capacity-reservations` per vedere quante prenotazioni di capacità sono rimaste inutilizzate.

Esempio 2: il pool di istanze a prezzo più basso nella richiesta di avvio non dispone di prenotazioni della capacità aperte disponibili

In questo caso, Amazon EMR avvia la capacità nel pool di istanze a prezzo più basso con istanze on demand. Tuttavia, le prenotazioni della capacità aperte rimangono inutilizzate.

On-Demand Strategy	lowest-price		
Requested Capacity	100		
Instance Type	c5.xlarge	m5.xlarge	r5.xlarge
Prenotazioni della capacità aperte disponibili	-	-	100
On-Demand Price	$	$$	$$$
Istanze con provisioning	100	-	-
Prenotazione della capacità aperta utilizzata	-	-	-
Prenotazioni della capacità aperte disponibili	-	-	100

Configurazione dei parchi istanze per utilizzare le prenotazioni della capacità aperte in base al miglior tentativo

Quando utilizzi l'operazione `RunJobFlow` per creare un cluster basato sul parco istanze, imposta la strategia di allocazione on demand `lowest-price` e `CapacityReservationPreference` per le opzioni di prenotazione di capacità su `open`. In alternativa, se lasci vuoto questo campo, Amazon EMR configura come impostazione predefinita la preferenza di prenotazione della capacità dell'istanza on demand su `open`.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
```

```
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "open"
         }
        }
    }
```

È inoltre possibile utilizzare l'interfaccia CLI di Amazon EMR per creare un cluster basato sul parco istanze utilizzando prenotazioni della capacità aperte.

```
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}}'}
```

Dove,

- `open-ODCR-cluster` viene sostituito con il nome del cluster che utilizza le prenotazioni della capacità aperte.
- `subnet-22XXXX01` viene sostituito con l'ID della sottorete.

Usa prima le prenotazioni della capacità aperte

Puoi scegliere di sovrascrivere la strategia di allocazione del prezzo più basso e assegnare priorità utilizzando le prenotazioni della capacità aperte disponibili prima durante il provisioning di un cluster Amazon EMR. In questo caso, Amazon EMR valuta tutti i pool di istanze con prenotazioni della capacità specificati nella richiesta di avvio e utilizza quello con il prezzo più basso che dispone di capacità sufficienti per avviare tutti i nodi principali richiesti. Se nessuno dei pool di istanze con prenotazioni della capacità dispone di capacità sufficiente per i nodi principali richiesti, Amazon EMR ritorna al caso sulla base del miglior tentativo descritto nell'argomento precedente. In altre parole, Amazon EMR valuta nuovamente tutti i pool di istanze specificati nella richiesta di avvio

e utilizza quello con il prezzo più basso che dispone di capacità sufficiente per avviare tutti i nodi principali richiesti. Le prenotazioni della capacità aperte disponibili che corrispondono al pool di istanze vengono applicate automaticamente. Se le prenotazioni della capacità aperte disponibili non corrispondono al pool di istanze, rimangono inutilizzate.

Una volta eseguito il provisioning dei nodi principali, la zona di disponibilità viene selezionata e risolta. Amazon EMR effettua il provisioning dei nodi attività in pool di istanze con prenotazioni della capacità, a partire da quelli con il prezzo più basso, nella zona di disponibilità selezionata fino a quando non viene eseguito il provisioning di tutti i nodi attività. Amazon EMR utilizza prima le prenotazioni della capacità aperte disponibili per ogni pool di istanze nella zona di disponibilità selezionata e, solo se necessario, utilizza la strategia di prezzo più basso per eseguire il provisioning di eventuali nodi attività rimanenti.

Di seguito sono riportati i casi d'uso della logica di allocazione della capacità di Amazon EMR per l'utilizzo delle prenotazioni della capacità aperte per prime.

Esempio 1: il pool di istanze con prenotazioni della capacità aperte disponibili nella richiesta di avvio ha una capacità sufficiente per i nodi principali

In questo caso, Amazon EMR avvia la capacità nel pool di istanze con prenotazioni della capacità aperte disponibili indipendentemente dal prezzo del pool di istanze. Di conseguenza, le prenotazioni della capacità aperte vengono utilizzate quando possibile, fino a quando non viene eseguito il provisioning di tutti i nodi principali.

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	$	$$	$$$
Istanze con provisioning	-	-	100

Prenotazione della capacità aperta utilizzata	-	-	100
Prenotazioni della capacità aperte disponibili	-	-	50

Esempio 2: il pool di istanze con prenotazioni della capacità aperte disponibili nella richiesta di avvio non dispone di capacità sufficiente per i nodi principali

In questo caso, Amazon EMR torna ad avviare nodi principali utilizzando una strategia di prezzo più basso in base al miglior tentativo per utilizzare le prenotazioni di capacità.

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	$	$$	$$$
Istanze con provisioning	100	-	-
Prenotazione della capacità aperta utilizzata	10	-	-
Prenotazioni della capacità aperte disponibili	-	50	50

Dopo l'avvio del parco istanze, puoi eseguire `describe-capacity-reservations` per vedere quante prenotazioni di capacità sono rimaste inutilizzate.

Configurazione dei parchi istanze per utilizzare prima le prenotazioni della capacità aperte

Quando utilizzi l'operazione `RunJobFlow` per creare un cluster basato sul parco istanze, imposta la strategia di allocazione on-demand su `lowest-price` e `UsageStrategy` per `CapacityReservationOptions` su `use-capacity-reservations-first`.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first"
         }
       }
    }
```

È inoltre possibile utilizzare l'interfaccia CLI Amazon EMR per creare un cluster basato sul parco istanze utilizzando prima prenotazioni della capacità.

```
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}}'}
```

Dove,

- `use-CR-first-cluster` viene sostituito con il nome del cluster che utilizza le prenotazioni della capacità aperte.

- `subnet-22XXXX01` viene sostituito con l'ID della sottorete.

Usa prima le prenotazioni della capacità mirate

Quando si esegue il provisioning di un cluster Amazon EMR, è possibile scegliere di sostituire la strategia di allocazione a prezzo più basso e assegnare priorità utilizzando innanzitutto le prenotazioni della capacità mirate disponibili. In questo caso, Amazon EMR valuta tutti i pool di istanze con prenotazioni della capacità mirate specificate nella richiesta di avvio e seleziona quello con il prezzo più basso che dispone di capacità sufficiente per avviare tutti i nodi principali richiesti. Se nessuno dei pool di istanze con prenotazioni della capacità mirate dispone di capacità sufficiente per i nodi principali, Amazon EMR ritorna al caso sulla base del miglior tentativo descritto in precedenza. In altre parole, Amazon EMR valuta nuovamente tutti i pool di istanze specificati nella richiesta di avvio e seleziona quello con il prezzo più basso che dispone di capacità sufficiente per avviare tutti i nodi principali richiesti. Le prenotazioni della capacità aperte disponibili che corrispondono al pool di istanze vengono applicate automaticamente. Tuttavia, le prenotazioni della capacità mirate rimangono inutilizzate.

Una volta eseguito il provisioning dei nodi principali, la zona di disponibilità viene selezionata e risolta. Amazon EMR effettua il provisioning dei nodi attività in pool di istanze con prenotazioni della capacità mirate, a partire da quelli con il prezzo più basso, nella zona di disponibilità selezionata fino a quando non viene eseguito il provisioning di tutti i nodi attività. Amazon EMR tenta di utilizzare prima le prenotazioni della capacità mirate disponibili in ogni pool di istanze nella zona di disponibilità selezionata. Quindi, solo se necessario, Amazon EMR utilizza la strategia di prezzo più basso per effettuare il provisioning di eventuali nodi attività rimanenti.

Di seguito sono riportati i casi d'uso della logica di allocazione della capacità di Amazon EMR per l'utilizzo delle prenotazioni della capacità mirate per prime.

Esempio 1: il pool di istanze con prenotazioni della capacità mirate disponibili nella richiesta di avvio ha una capacità sufficiente per i nodi principali

In questo caso, Amazon EMR avvia la capacità nel pool di istanze con prenotazioni della capacità mirate disponibili indipendentemente dal prezzo del pool di istanze. Di conseguenza, le prenotazioni della capacità mirate vengono utilizzate quando possibile, fino a quando non viene eseguito il provisioning di tutti i nodi principali.

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	$	$$	$$$
Istanze con provisioning	-	-	100
Prenotazione della capacità mirata utilizzata	-	-	100
Prenotazioni della capacità mirate disponibili	-	-	50

Example Esempio 2: il pool di istanze con prenotazioni della capacità mirate disponibili nella richiesta di avvio non dispone di capacità sufficiente per i nodi principali

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	$	$$	$$$

Istanze con provisioning	100	-	-
Prenotazioni della capacità mirate utilizzate	-	-	-
Prenotazioni della capacità mirate disponibili	10	50	50

Dopo l'avvio del parco istanze, puoi eseguire `describe-capacity-reservations` per vedere quante prenotazioni di capacità sono rimaste inutilizzate.

Configurazione dei parchi istanze per utilizzare prima le prenotazioni della capacità mirate

Quando utilizzi l'operazione `RunJobFlow` per creare un cluster basato sul parco istanze, imposta la strategia di allocazione on demand su `lowest-price`, `UsageStrategy` per `CapacityReservationOptions` su `use-capacity-reservations-first` e `CapacityReservationResourceGroupArn` per `CapacityReservationOptions` su `<your resource group ARN>`. Per ulteriori informazioni, consulta Utilizzo di prenotazioni della capacità nella Guida per l'utente di Amazon EC2 per le istanze Linux.

```
"LaunchSpecifications":
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "UsageStrategy": "use-capacity-reservations-first",
            "CapacityReservationResourceGroupArn": "arn:aws:resource-groups:sa-
east-1:123456789012:group/MyCRGroup"
         }
       }
    }
```

Dove `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup` viene sostituito con l'ARN del gruppo di risorse.

È inoltre possibile utilizzare la CLI Amazon EMR per creare un cluster basato sul parco istanze utilizzando prima prenotazioni della capacità mirate.

```
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}}'}
```

Dove,

- `targeted-CR-cluster` viene sostituito con il nome del cluster che utilizza le prenotazioni della capacità mirate.
- `subnet-22XXXX01` viene sostituito con l'ID della sottorete.
- `arn:aws:resource-groups:sa-east-1:123456789012:group/MyCRGroup` viene sostituito con l'ARN del gruppo di risorse.

Come evitare l'utilizzo delle prenotazioni della capacità aperte disponibili

Example

Se desideri evitare di utilizzare una qualsiasi delle prenotazioni della capacità aperte durante l'avvio di un cluster Amazon EMR, imposta la strategia di allocazione on demand su `lowest-price` e `CapacityReservationPreference` per `CapacityReservationOptions` su `none`. In caso contrario, Amazon EMR imposta come impostazione predefinita la preferenza di prenotazione della capacità dell'istanza on demand su `open` e prova a utilizzare le prenotazioni della capacità aperte disponibili sulla base della migliore ipotesi.

```
"LaunchSpecifications":
```

```
    {"OnDemandSpecification": {
        "AllocationStrategy": "lowest-price",
        "CapacityReservationOptions":
         {
            "CapacityReservationPreference": "none"
         }
       }
    }
```

È inoltre possibile utilizzare l'interfaccia CLI di Amazon EMR per creare un cluster basato sul parco istanze senza utilizzare prenotazioni della capacità aperte.

```
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}}'}
```

Dove,

- `none-CR-cluster` viene sostituito con il nome del cluster che non utilizza le prenotazioni della capacità aperte.
- `subnet-22XXXX01` viene sostituito con l'ID della sottorete.

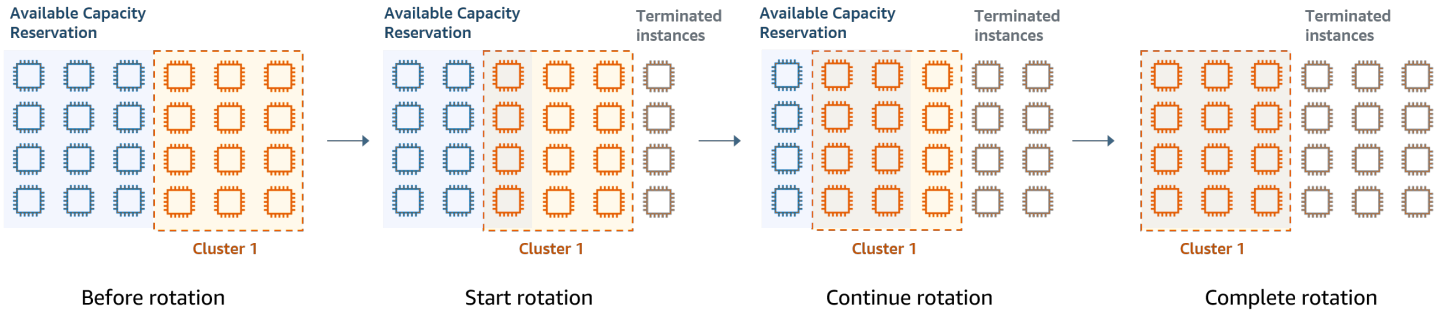
Scenari per l'utilizzo delle prenotazioni della capacità

È possibile utilizzare le prenotazioni della capacità nei seguenti scenari.

Scenario 1: Rotazione di un cluster a esecuzione prolungata utilizzando le prenotazioni della capacità

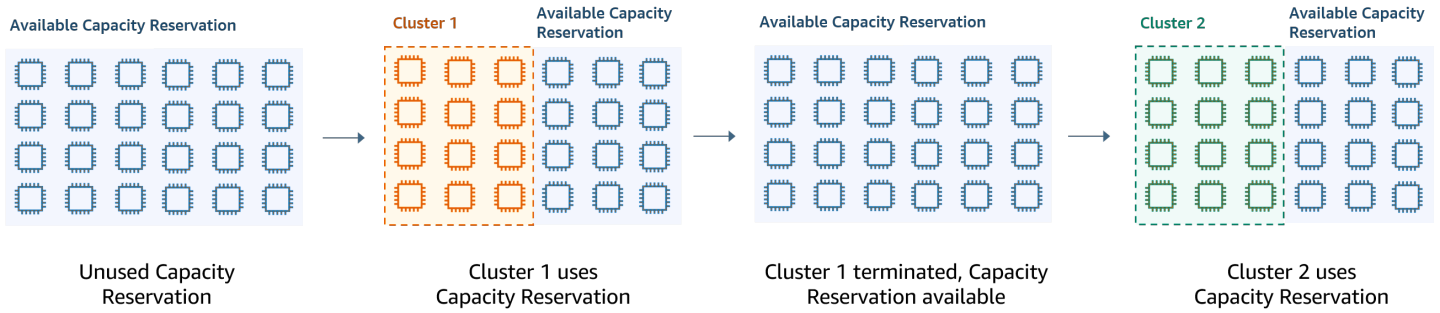
Quando si ruota un cluster a esecuzione prolungata, è possibile che siano richiesti requisiti rigorosi per i tipi di istanza e le zone di disponibilità per le nuove istanze di cui si esegue il provisioning. Con

le prenotazioni della capacità, è possibile utilizzare la garanzia della capacità per completare la rotazione del cluster senza interruzioni.



Scenario 2: Provisioning di cluster di breve durata successivi utilizzando le prenotazioni della capacità

È possibile utilizzare le prenotazioni della capacità anche per eseguire il provisioning di un gruppo di cluster successivi e di breve durata per singoli carichi di lavoro in modo che quando si chiude un cluster, il cluster successivo possa utilizzare le prenotazioni della capacità. È possibile utilizzare le prenotazioni della capacità mirate per assicurarsi che solo i cluster previsti utilizzino le prenotazioni della capacità.



Configurazione di gruppi di istanze uniformi

Con la configurazione dei gruppi di istanze, ogni tipo di nodo (master, principale o di task) è costituito dallo stesso tipo di istanza e dalla stessa opzione di acquisto per istanze: on demand o Spot. Queste impostazioni vengono specificate al momento della creazione di un gruppo di istanze. Le impostazioni non possono essere modificate in seguito. Tuttavia, puoi aggiungere istanze dello stesso tipo e opzioni di acquisto a gruppi di istanze principali e dell'attività Puoi anche rimuovere istanze.

Se le istanze on demand del cluster corrispondono agli attributi delle prenotazioni di capacità aperta (tipo di istanza, piattaforma, tenancy e zona di disponibilità) disponibili nell'account, le prenotazioni di capacità vengono applicate automaticamente. È possibile utilizzare prenotazioni di capacità aperte per nodi primari, core e attività. Tuttavia, non è possibile utilizzare prenotazioni della capacità mirate o impedire l'avvio di istanze in prenotazioni della capacità aperte con attributi corrispondenti quando

si esegue il provisioning di cluster utilizzando gruppi di istanze. Se si desidera utilizzare prenotazioni della capacità mirate o impedire l'avvio di istanze in prenotazioni della capacità aperte, utilizza piuttosto parchi istanze. Per ulteriori informazioni, consulta Utilizzo di prenotazioni della capacità con parchi istanze.

Per aggiungere tipi di istanze differenti dopo che un cluster è stato creato, puoi aggiungere gruppi di istanze di attività aggiuntivi. Puoi scegliere tipi di istanze e opzioni di acquisto differenti per ogni gruppo di istanze. Per ulteriori informazioni, consulta Uso del dimensionamento del cluster.

Quando si avviano delle istanze, la preferenza della prenotazione di capacità dell'istanza on demand è impostata in modo predefinito su `open`, che consente di eseguirla in qualsiasi prenotazione di capacità aperta che abbia gli attributi corrispondenti (tipo di istanza, piattaforma, zona di disponibilità). Per informazioni sulla condivisione delle prenotazioni di capacità on demand, consulta Utilizzo di prenotazioni della capacità con parchi istanze.

In questa sezione viene descritta la creazione di un cluster con gruppi di istanze uniformi. Per ulteriori informazioni sulla modifica di un gruppo di istanze esistente aggiungendo o rimuovendo istanze manualmente o con scalabilità automatica, consulta Gestione di cluster.

Utilizzo della console per configurare gruppi di istanze uniformi

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Creazione di un cluster con gruppi di istanze con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster), quindi seleziona Create cluster (Crea cluster).
3. In Cluster configuration (Configurazione del cluster), scegli Instance groups (Gruppi di istanze).

4. In Node groups (Gruppi di nodi) è presente una sezione per ogni tipo di gruppo di nodi. Per il gruppo di nodi primari, seleziona la casella di controllo Use multiple primary nodes (Usa più nodi primari) se desideri avere 3 nodi primari. Seleziona la casella di controllo Use Spot purchasing option (Utilizza l'opzione di acquisto spot) se desideri utilizzare gli acquisti spot.
5. Per i gruppi di nodi primari e core, seleziona Add instance type (Aggiungi tipo di istanza) e scegli fino a 5 tipi di istanza. Per il gruppo attività, seleziona Add instance type (Aggiungi tipo di istanza) e scegli fino a 15 tipi di istanza. Amazon EMR può effettuare il provisioning di qualsiasi combinazione di questi tipi di istanze quando avvia il cluster.
6. In ogni tipo di gruppo di nodi, scegli il menu a discesa Actions (Operazioni) accanto a ciascuna istanza per modificare queste impostazioni:

   Aggiunta di volumi EBS

   Specifica i volumi EBS da collegare al tipo di istanza dopo che Amazon EMR ne effettua il provisioning.

   Modifica del prezzo spot massimo

   Per ogni tipo di istanza all'interno di un parco, specifica un prezzo spot massimo. Puoi impostare questo prezzo come una percentuale del prezzo on demand o come un importo in dollari specifico. Se il prezzo spot corrente in una zona di disponibilità è inferiore al prezzo spot massimo, Amazon EMR effettua il provisioning delle istanze spot. L'utente paga il prezzo Spot, non necessariamente il prezzo Spot massimo.
7. Facoltativamente, espandi Configurazione nodo per inserire una configurazione JSON o per caricare JSON da Amazon S3.
8. Scegli qualsiasi altra opzione applicabile al cluster.
9. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Nella procedura seguente vengono descritte Advanced options (Opzioni avanzate) quando si crea un cluster. L'utilizzo di Quick options (Opzioni rapide) consente inoltre di creare un cluster con la configurazione dei gruppi di istanze.

Creazione di un cluster con gruppi di istanze uniformi con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate), immetti le opzioni Software Configuration (Configurazione software) e seleziona Next (Successivo).
4. Nella schermata Hardware Configuration (Configurazione hardware), lascia Uniform instance groups (Gruppi di istanze uniformi) selezionato.
5. Scegli la Network (Rete), quindi seleziona la EC2 Subnet (Sottorete EC2) per il cluster. La sottorete selezionata è associata a un gruppo di disponibilità, che viene elencato con ciascuna sottorete. Per ulteriori informazioni, consulta Configurazione delle reti.

   > **Note**
   >
   > L'account e la Regione offrono la possibilità di scegliere Launch into EC2-Classic (Avvio in EC2-Classic) per Network (Rete). Se scegli questa opzione, seleziona una EC2 Availability Zone (Zona di disponibilità EC2) anziché una EC2 Subnet (Sottorete EC2). Per ulteriori informazioni, consulta Amazon EC2 e Amazon VPC nella Guida per l'utente di Amazon EC2 per le istanze Linux.

6. All'interno di ogni riga Node type (Tipo di nodo):

   - In Node type (Tipo di nodo), se desideri modificare il nome predefinito del gruppo di istanze, scegli l'icona a forma di matita e immetti un nome descrittivo. Se desideri rimuovere il gruppo di istanze di Task (Attività), seleziona l'icona X. Scegli Add task instance group (Aggiungi gruppo di istanze di attività) per aggiungere ulteriori gruppi di istanze di Task (Attività).
   - In Instance type (Tipo di istanza), scegli l'icona a forma di matita e scegli il tipo di istanza da utilizzare per tale tipo di nodo.

     > **Important**
     >
     > Quando si sceglie un tipo di istanza utilizzando la AWS Management Console, il numero di VPCU per ogni tipo di istanza è il numero di vcore YARN per quel tipo di istanza, non il numero di vCPU EC2 per quel tipo di istanza. Per ulteriori

informazioni sul numero di vCPU per ogni tipo di istanza, consulta Tipi di istanza di Amazon EC2.

- In Instance type (Tipo di istanza), seleziona l'icona a forma di matita per Configurations (Configurazioni) e modifica le configurazioni per applicazioni per ogni gruppo di istanze.
- In Instance count (Conteggio istanze), immetti il numero di istanze da utilizzare per ogni tipo di nodo.
- In Purchasing option (Opzione di acquisto), scegli On-Demand (On demand) o Spot. Se scegli Spot, seleziona un'opzione per il prezzo massimo per le istanze Spot. Per impostazione predefinita, è selezionata l'opzione Use on-demand as max price (Usa on demand come prezzo massimo). Puoi selezionare Set max $/hr (Imposta max $/hr) e immettere il prezzo massimo. La zona di disponibilità della sottorete EC2 scelta è al di sotto del prezzo Spot massimo.

  > **Tip**
  >
  > Posiziona il mouse sul suggerimento informazioni per Spot per visualizzare il prezzo Spot corrente per le zone di disponibilità nella Regione corrente. Il prezzo Spot più basso è scritto in verde. È possibile utilizzare queste informazioni per modificate la selezione EC2 Subnet (Sottorete EC2).

- In Auto Scaling for Core and Task node types (Dimensionamento automatico per tipi di nodo principale e di task), scegli l'icona a forma di matita, quindi configura le opzioni di scalabilità automatica. Per ulteriori informazioni, consulta Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze .

7. Scegli Add task instance group (Aggiungi gruppo di istanze dell'attività) come desiderato e configura le impostazioni come descritto nella fase precedente.
8. Scegli Next (Successivo), modifica altre impostazioni cluster, quindi avvia il cluster.

Utilizzo della AWS CLI per creare un cluster con gruppi di istanze uniformi

Per specificare la configurazione di gruppi di istanze per un cluster utilizzando AWS CLI, utilizza il comando `create-cluster` insieme al parametro `--instance-groups`. Amazon EMR presuppone l'opzione On-Demand Instance (Istanza on demand) a meno che non venga specificato l'argomento `BidPrice` per un gruppo di istanze. Per esempi di comandi `create-cluster` che avviano gruppi di istanze uniformi con istanze on demand e opzioni di cluster diverse, digita `aws`

`emr create-cluster help` dalla riga di comando oppure consulta create-cluster nella Guida di riferimento ai comandi della AWS CLI.

Puoi utilizzare AWS CLI per creare gruppi di istanze uniformi in un cluster che utilizza istanze Spot. Il prezzo Spot offerto dipende dalla zona di disponibilità. Quando utilizzi la CLI o l'API, puoi specificare la zona di disponibilità con l'argomento `AvailabilityZone` (se stai utilizzando una rete EC2-Classic) o l'argomento `SubnetID` del parametro `--ec2-attributes` . La zona di disponibilità o la sottorete selezionata vale per il cluster, pertanto viene utilizzata per tutti i gruppi di istanze. Se non specifichi una zona di disponibilità o sottorete in maniera esplicita, Amazon EMR seleziona la zona di disponibilità con il prezzo Spot più basso quando avvia il cluster.

Nel seguente esempio viene illustrato un comando `create-cluster` che crea un gruppo di istanze primarie, uno di core e due di attività, tutti i quali utilizzano istanze spot. Sostituisci *myKey* con il nome della coppia di chiavi Amazon EC2.

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuovili o sostituiscili con un accento circonflesso (^).

```
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
```

Utilizzando la CLI, è possibile creare cluster di gruppi di istanze uniformi che specificano un'AMI personalizzata univoca per ogni tipo di istanza nel gruppo di istanze. Ciò consente di utilizzare architetture di istanze diverse nello stesso gruppo di istanze. Ogni tipo di istanza deve utilizzare un'AMI personalizzata con un'architettura corrispondente. Ad esempio, si dovrebbe configurare un tipo di istanza m5.xlarge con un'AMI personalizzata dell'architettura x86_64 e un tipo di istanza m6g.xlarge con una corrispondente (ARM) architettura AMI personalizzata `AWS AARCH64`.

L'esempio seguente mostra un cluster di gruppi di istanze uniforme creato con due tipi di istanza, ognuno con una propria AMI personalizzata. Notare che le AMI personalizzate sono specificate solo a livello di tipo di istanza, non a livello di cluster. Questo per evitare conflitti tra le AMI del tipo di istanza e un'AMI a livello di cluster, il che causerebbe il fallimento dell'avvio del cluster.

```
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
```

È possibile aggiungere AMI personalizzate multiple a un gruppo di istanze aggiunto a un cluster in esecuzione. L'argomento `CustomAmiId` può essere utilizzato con il comando `add-instance-groups` come mostrato nell'esempio seguente.

```
aws emr add-instance-groups --cluster-id j-123456 \
  --instance-groups \

 InstanceGroupType=Task,InstanceType=m5.xlarge,InstanceCount=1,CustomAmiId=ami-123456
```

Utilizzo di Java SDK per creare un gruppo di istanze

Crea un'istanza di un oggetto `InstanceGroupConfig` che specifica la configurazione di un gruppo di istanze per un cluster. Per utilizzare istanze Spot, imposta le proprietà `withBidPrice` e `withMarket` sull'oggetto `InstanceGroupConfig`. Nel codice seguente viene mostrato come definire gruppi di istanze primarie, core e attività che eseguono istanze 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");
```

Best practice per la flessibilità delle istanze e delle zone di disponibilità

Ciascuna Regione AWS presenta più località isolate, chiamate zone di disponibilità. Quando avvii un'istanza, puoi decidere di specificare una zona di disponibilità (AZ) nella Regione AWS utilizzata. La flessibilità della zona di disponibilità consiste nella distribuzione delle istanze su più AZ. Se un'istanza fallisce, è possibile progettare l'applicazione in modo che un'istanza in un'altra zona dAZ possa gestire le richieste. Per maggiori informazioni sulle zone di disponibilità, consulta la documentazione relativa a Region and zones (Regioni e zone di disponibilità) nella Guida per l'utente di Amazon EC2.

La flessibilità delle istanze consiste nell'utilizzare più tipi di istanze per soddisfare i requisiti di capacità. Quando esprimi la flessibilità con le istanze, puoi utilizzare la capacità aggregata attingendo da dimensioni, famiglie e generazioni di istanze diverse. Una maggiore flessibilità consente di migliorare la possibilità di trovare e allocare la quantità di capacità di elaborazione richiesta rispetto a un cluster che utilizza un singolo tipo di istanza.

La flessibilità delle istanze e delle zone di disponibilità riduce gli errori di capacità insufficiente (ICE) e le interruzioni Spot rispetto a un cluster con un singolo tipo di istanza o AZ. Utilizza le best practice qui descritte per determinare quali istanze diversificare dopo avere stabilito la famiglia e le dimensioni iniziali delle istanze. Questo approccio massimizza la disponibilità dei pool di capacità di Amazon EC2 con una variazione minima delle prestazioni e dei costi.

Flessibilità nella scelta delle zone di disponibilità

Ti consigliamo di configurare tutte le zone di disponibilità per l'uso nel tuo cloud privato virtuale (VPC) e di selezionarle per il cluster EMR. I cluster devono esistere in una sola zona di disponibilità, ma con i parchi istanze Amazon EMR è possibile selezionare più sottoreti per diverse zone di disponibilità. Quando Amazon EMR avvia il cluster, viene eseguita una ricerca in tali sottoreti per individuare le

istanze e le opzioni di acquisto che hai specificato. Quando esegui il provisioning di un cluster EMR per più sottoreti, il cluster può accedere a un pool di capacità Amazon EC2 più profondo rispetto ai cluster in una singola sottorete.

Se devi dare priorità a un certo numero di zone di disponibilità da utilizzare nel tuo cloud privato virtuale (VPC) per il tuo cluster EMR, puoi sfruttare la funzionalità di punteggio di posizionamento Spot con Amazon EC2. Con il punteggio di posizionamento Spot, specifichi i requisiti di calcolo per le tue istanze Spot, quindi EC2 restituisce le prime dieci Regioni AWS o zone di disponibilità con punteggio su una scala da 1 a 10. Un punteggio pari a 10 indica che è molto probabile che la richiesta Spot abbia successo; un punteggio pari a 1 indica che è improbabile che la richiesta Spot abbia successo. Per ulteriori informazioni su come utilizzare il punteggio di posizionamento Spot, consulta Punteggio di posizionamento Spot nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Flessibilità nella scelta dei tipi di istanza

La flessibilità delle istanze consiste nell'utilizzare più tipi di istanze per soddisfare i requisiti di capacità. La flessibilità delle istanze offre vantaggi nell'utilizzo sia delle istanze spot sia delle istanze on demand di Amazon EC2. Con le istanze spot, la flessibilità delle istanze consente ad Amazon EC2 di avviare istanze da pool di capacità più profondi utilizzando dati sulla capacità in tempo reale. Inoltre, prevede quali sono le istanze più disponibili. Ciò garantisce un numero minore di interruzioni e può ridurre il costo complessivo di un carico di lavoro. Con le istanze on demand, la flessibilità delle istanze riduce gli errori di capacità insufficiente (ICE) quando la capacità totale viene fornita da un numero maggiore di pool di istanze.

Per i cluster Instance Group (Gruppi di istanze) puoi specificare fino a 50 tipi di istanza EC2. Per gli Instance Fleets (Parchi istanze) con strategia di allocazione, puoi specificare fino a 30 tipi di istanza EC2 per ogni gruppo di nodi primari, core e attività. Una gamma più ampia di istanze migliora i vantaggi della flessibilità delle istanze.

Espressione della flessibilità delle istanze

Tieni in considerazione le seguenti best practice per usufruire della flessibilità delle istanze per la tua applicazione.

Argomenti

- Determinazione della famiglia e delle dimensioni delle istanze
- Inclusione di istanze aggiuntive

Determinazione della famiglia e delle dimensioni delle istanze

Amazon EMR supporta diversi tipi di istanza per casi d'uso differenti. Questi tipi di istanze sono elencati nella documentazione di Tipi di istanze supportati. Ogni tipo di istanza appartiene a una famiglia di istanze che descrive l'applicazione per cui il tipo è ottimizzato.

Per i nuovi carichi di lavoro, è necessario eseguire un benchmark con i tipi di istanze della famiglia per uso generico, ad esempio `m5` o `c5`. Quindi, osserva i parametri del sistema operativo e YARN di Ganglia e Amazon CloudWatch per determinare i colli di bottiglia del sistema al massimo carico. I colli di bottiglia includono le operazioni di CPU, memoria, archiviazione e I/O. Dopo avere identificato i colli di bottiglia, scegli una famiglia di istanze ottimizzate per l'elaborazione, per la memoria, per l'archiviazione o un'altra famiglia appropriata per i tuoi tipi di istanze. Per maggiori dettagli, consulta la sezione Determine right infrastructure for your Spark workloads (Determinare l'infrastruttura giusta per i carichi di lavoro Spark) nella guida alle best practice di Amazon EMR su GitHub.

Quindi, identifica il container YARN o l'esecutore Spark più piccolo richiesto dall'applicazione. Questa è la dimensione dell'istanza più piccola adatta al container e la dimensione minima dell'istanza per il cluster. Utilizza questo parametro per determinare le istanze con cui puoi diversificare ulteriormente. Un'istanza più piccola consentirà una maggiore flessibilità delle istanze.

Per la massima flessibilità delle istanze, dovresti sfruttare il maggior numero possibile di istanze. Ti consigliamo di diversificare con istanze dotate di specifiche hardware simili. In questo modo, massimizzi l'accesso dei pool di capacità di EC2 con una variazione minima delle prestazioni e dei costi. Diversifica in base alle dimensioni. Per farlo, dai prima priorità ad AWS Graviton e alle generazioni precedenti. Una buona regola è quella di dimostrare flessibilità su almeno 15 tipi di istanza per ogni carico di lavoro. Ti consigliamo di iniziare con istanze per uso generico, ottimizzate per il calcolo oppure ottimizzate per la memoria. Questi tipi di istanze offriranno la massima flessibilità.

Inclusione di istanze aggiuntive

Per garantire la massima diversità, includi tipi di istanze aggiuntivi. Dai priorità alla flessibilità di dimensioni delle istanze, Graviton e di generazione. Ciò consente l'accesso a pool di capacità EC2 aggiuntivi con profili di costi e prestazioni simili. Se hai bisogno di ulteriore flessibilità a causa degli ICE o delle interruzioni spot, prendi in considerazione la flessibilità delle varianti e delle famiglie. Ogni approccio presenta dei compromessi che dipendono dal caso d'uso e dai requisiti.

- Flessibilità di dimensione: in primo luogo, diversifica scegliendo istanze di dimensioni diverse all'interno della stessa famiglia. Le istanze della medesima famiglia offrono lo stesso costo e le

stesse prestazioni, ma possono avviare un numero diverso di container su ogni host. Ad esempio, se la dimensione minima dell'esecutore necessaria è 2 vCPU con 8 Gb di memoria, la dimensione minima dell'istanza è `m5.xlarge`. Per garantire la flessibilità delle dimensioni, includi `m5.xlarge`, `m5.2xlarge`, `m5.4xlarge`, `m5.8xlarge`, `m5.12xlarge`, `m5.16xlarge` e `m5.24xlarge`.

- Flessibilità di Graviton: oltre alle dimensioni, puoi diversificare con le istanze Graviton. Le istanze Graviton sono alimentate da processori AWS Graviton2 che offrono il rapporto prezzo/prestazioni migliore per i carichi di lavoro cloud in Amazon EC2. Ad esempio, con la dimensione minima dell'istanza di `m5.xlarge`, è possibile includere `m6g.xlarge`, `m6g.2xlarge`, `m6g.4xlarge`, `m6g.8xlarge` e `m6g.16xlarge` per la flessibilità di Graviton.
- Flessibilità di generazione: analogamente alla flessibilità di dimensione e di Graviton, le istanze delle famiglie della generazione precedente condividono le stesse specifiche hardware. Ciò si traduce in un profilo di costi e prestazioni simile con un aumento del pool totale a cui Amazon EC2 può accedere. Per garantire la flessibilità di generazione, includi `m4.xlarge`, `m4.2xlarge`, `m4.10xlarge` e `m4.16xlarge`.
- Flessibilità per famiglie e varianti
  - Capacità: per ottimizzare la capacità, consigliamo la flessibilità delle istanze scegliendo tra famiglie di istanze diverse. Le istanze comuni di famiglie di istanze diverse dispongono di pool di istanze più profondi che possono contribuire a soddisfare i requisiti di capacità. Tuttavia, istanze di famiglie diverse hanno rapporti tra vCPU e memoria differenti. Ciò comporta un sottoutilizzo se il container dell'applicazione previsto è dimensionato per un'istanza diversa. Ad esempio, con `m5.xlarge`, includi istanze ottimizzate per il calcolo come `c5` oppure istanze ottimizzate per la memoria come `r5` per la flessibilità della famiglia di istanze.
  - Costo: per ottimizzare i costi, consigliamo la flessibilità delle istanze scegliendo tra le varianti. Queste istanze hanno lo stesso rapporto tra memoria e vCPU dell'istanza iniziale. Il compromesso con la flessibilità delle varianti è che queste istanze hanno pool di capacità più piccoli, il che potrebbe comportare una capacità aggiuntiva limitata o interruzioni spot più numerose. Ad esempio, con `m5.xlarge`, includi istanze basate su AMD (`m5a`), istanze basate su SSD (`m5d`) o istanze ottimizzate per la rete (`m5n`) per garantire la flessibilità delle varianti.

Best practice per la configurazione dei cluster

Usa le linee guida in questa sezione per determinare i tipi di istanza, le opzioni di acquisto e la quantità di storage di cui effettuare il provisioning per ogni tipo di nodo in un cluster EMR.

Che tipo di istanza utilizzare?

Esistono diversi modi per aggiungere istanze Amazon EC2 a un cluster. Il metodo da scegliere dipende dal fatto che si utilizzi la configurazione dei gruppi di istanze o la configurazione dei parchi istanze per il cluster.

- Gruppi di istanze
  - Aggiungi manualmente istanze dello stesso tipo a gruppi di istanze principali e dell'attività esistenti.
  - Aggiungi manualmente un gruppo di istanze dell'attività, che può utilizzare un tipo di istanza diverso.
  - Configura la scalabilità automatica in Amazon EMR per un gruppo di istanze, aggiungendo e rimuovendo istanze automaticamente in base al valore di un parametro Amazon CloudWatch specificato. Per ulteriori informazioni, consulta Uso del dimensionamento del cluster.
- Parchi istanze
  - Aggiungi un singolo parco istanze attività.
  - Cambia la capacità target per istanze on demand e Spot per parchi istanze principale e dell'attività. Per ulteriori informazioni, consulta Configurazione di parchi istanze.

Un modo per pianificare le istanze del cluster è eseguire un cluster di test con un set di dati di esempio rappresentativo e monitorare l'utilizzo dei nodi del cluster. Per ulteriori informazioni, consulta Visualizzazione e monitoraggio di un cluster. Un altro modo è calcolare la capacità delle istanze che si stanno valutando e confrontare tale valore rispetto alle dimensioni dei dati.

In generale: il tipo di nodo primario, che assegna le attività, non richiede un'istanza EC2 con tanta potenza di elaborazione; le istanze Amazon EC2 per il tipo di nodo core, che elaborano le attività e archiviano i dati in HDFS, richiedono sia potenza di elaborazione sia capacità di archiviazione; le istanze Amazon EC2 per il tipo di nodo attività, che non archiviano dati, richiedono solo potenza di elaborazione. Per le linee guida sulle istanze Amazon EC2 disponibili e la relativa configurazione, consulta Configurazione delle istanze Amazon EC2.

Le seguenti linee guida si applicano alla maggior parte dei cluster Amazon EMR.

- Esiste un limite di vCPU per il numero totale di istanze Amazon EC2 on demand eseguite su un account AWS per Regione AWS. Per ulteriori informazioni sul limite per vCPU e su come richiedere un aumento del limite per l'account, consulta Istanze on demand nella Guida per l'utente per istanze Linux di Amazon EC2.

- Generalmente, il nodo primario non ha requisiti di calcolo elevati. Per i cluster con un numero elevato di nodi o per i cluster con applicazioni implementate in modo specifico sul nodo primario (JupyterHub, Hue e così via), potrebbe rendersi necessario un nodo primario più grande, che contribuirebbe a migliorare le prestazioni del cluster. Ad esempio, potresti prendere in considerazione l'utilizzo di un'istanza m5.xlarge per cluster piccoli (50 nodi o meno) e l'aumento a un tipo di istanza più grande per cluster più grandi.
- Le esigenze di calcolo dei nodi principali e di task dipendono dal tipo di elaborazione eseguita dall'applicazione. Molti processi possono essere eseguiti su tipi di istanze per uso generico, che offrono prestazioni bilanciate in termini di CPU, spazio su disco e input/output. Cluster che richiedono molti calcoli possono trarre beneficio dall'esecuzione su istanze con elevata CPU, che hanno proporzionalmente più CPU che RAM. Applicazioni di database e di caching nella memoria possono trarre vantaggio dall'esecuzione su istanze a memoria elevata. Applicazioni che prevedono un uso intensivo della rete e della CPU, ad esempio operazioni di analisi, procedimenti NLP e Machine Learning, possono trarre vantaggio dall'esecuzione su istanze Cluster Compute, che forniscono risorse di CPU proporzionalmente elevate e prestazioni di rete avanzate.
- Se fasi diverse del cluster hanno diverse esigenze di capacità, puoi iniziare con un numero ridotto di nodi principali e aumentare o diminuire il numero di nodi di task per soddisfare i diversi requisiti di capacità del flusso di elaborazione.
- La quantità di dati che puoi elaborare dipende dalla capacità dei nodi principali e dalla dimensione dei dati come input, durante l'elaborazione, e come output. I dataset di input, intermedio e di output risiedono tutti sul cluster durante l'elaborazione.

Quando occorre utilizzare le istanze Spot?

Quando avvii un cluster in Amazon EMR, puoi scegliere di avviare istanze primarie, core o attività su istanze spot. Poiché ogni tipo di gruppo di istanze gioca un ruolo diverso nel cluster, esistono implicazioni legate all'avvio di ciascun tipo di nodo sulle istanze Spot. Non è possibile modificare un'opzione di acquisto di un'istanza quando un cluster è in esecuzione. Per passare dalle istanze on demand alle istanze spot o viceversa per i nodi primari e core è necessario terminare il cluster e avviarne uno nuovo. Per i nodi di task, puoi avviare una nuovo gruppo di istanze di task o un parco istanze e rimuovere quella precedente.

Argomenti

- Impostazioni di Amazon EMR per impedire gli errori nei processi a causa dell'interruzione delle istanze Spot nei nodi attività
- Nodo primario su un'istanza spot

- Nodi principali sulle istanze Spot
- Nodi attività sulle istanze Spot
- Configurazioni di istanze per scenari applicativi

Impostazioni di Amazon EMR per impedire gli errori nei processi a causa dell'interruzione delle istanze Spot nei nodi attività

Poiché le istanze Spot vengono spesso utilizzate per eseguire nodi attività, Amazon EMR dispone delle funzionalità predefinite per la pianificazione dei processi YARN in modo che i processi in esecuzione non abbiano esito negativo quando i nodi attività in esecuzione su istanze Spot vengono terminati. Amazon EMR esegue questa operazione consentendo ai processi master delle applicazioni di funzionare solo sui nodi principali. Il processo master dell'applicazione controlla i processi in esecuzione e deve rimanere attivo per tutta la durata del processo.

Amazon EMR rilascio 5.19.0 e successivi utilizzano la caratteristica integrata etichette nodo YARN per questo scopo. (Le versioni precedenti utilizzavano una patch di codice). Le proprietà nelle classificazioni di configurazione `yarn-site` e `capacity-scheduler` sono configurate per impostazione predefinita in modo che capacity-scheduler e fair-scheduler YARN sfruttino le etichette dei nodi. Amazon EMR etichetta in automatico i nodi principali con l'etichetta `CORE` e imposta le proprietà in modo che i master dell'applicazione siano pianificati solo sui nodi con l'etichetta CORE. La modifica manuale delle proprietà correlate nelle classificazioni di configurazione del sito di YARN e del pianificatore di capacità o direttamente nei file XML associati potrebbe interrompere o alterare questa funzionalità.

Per impostazione predefinita, Amazon EMR configura le proprietà e i valori riportati di seguito. Fai attenzione quando configuri queste proprietà.

- yarn-site (yarn-site.xml) Su tutti i nodi
  - `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) su nodi primari e core
  - `yarn.nodemanager.node-labels.provider: 'config'`
  - `yarn.nodemanager.node-labels.provider.configured-node-partition: 'CORE'`
- capacity-scheduler (capacity-scheduler.xml) Su tutti i nodi

- `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 partire dalla serie di rilascio Amazon EMR 6.x, la funzione etichette nodo YARN è disabilitata per impostazione predefinita. Per impostazione predefinita, i processi primari dell'applicazione possono essere eseguiti sia sui nodi core sia su quelli attività. È possibile abilitare la caratteristica etichette nodo YARN configurando le seguenti proprietà:
>
> - `yarn.node-labels.enabled: true`
> - `yarn.node-labels.am.default-node-label-expression: 'CORE'`

Nodo primario su un'istanza spot

Il nodo primario controlla e gestisce il cluster. Quando termina, termina anche il cluster, pertanto devi avviare il nodo primario come istanza spot solo se stai eseguendo un cluster in cui la terminazione improvvisa è accettabile. Ciò è possibile se stai eseguendo il test di una nuova applicazione, disponi di un cluster che conserva periodicamente i dati in uno store esterno come Amazon S3 o stai eseguendo un cluster in cui il costo è più importante che garantire il completamento del cluster.

Quando avvii il gruppo di istanze primarie come istanza spot, il cluster non si avvia finché tale richiesta di istanza spot non viene soddisfatta. È importare tenere presente ciò quando si seleziona il prezzo Spot massimo.

Puoi aggiungere un nodo primario di istanza spot solo quando avvii il cluster. I nodi primari non possono essere aggiunti o rimossi da un cluster in esecuzione.

In genere, devi eseguire il nodo primario come istanza spot solo se stai eseguendo l'intero cluster (tutti i gruppi di istanze) come istanze spot.

Nodi principali sulle istanze Spot

I nodi principali elaborano dati e memorizzano informazioni utilizzando HDFS. La terminazione di un'istanza principale comporta la perdita dei dati. Per questo motivo, è consigliabile eseguire solo i nodi principali sulle istanze Spot quando la perdita dei dati HDFS parziale è tollerabile.

Quando avvii il gruppo di istanze principale come istanze Spot, Amazon EMR attende finché non puoi effettuare il provisioning di tutte le istanze principali richieste prima di avviare il gruppo di istanze. In altre parole, se richiedi sei istanze Amazon EC2 e solo cinque sono disponibili al prezzo Spot massimo o al di sotto di tale prezzo, il gruppo di istanze non verrà avviato. Amazon EMR continua ad attendere finché tutte e sei le istanze Amazon EC2 non sono disponibili o finché non termini il cluster. È possibile modificare il numero di istanze Spot in un gruppo di istanze principali per aggiungere capacità a un cluster in esecuzione. Per ulteriori informazioni sulle operazioni con i gruppi di istanze e sul funzionamento delle istanze Spot con il parco istanze, consulta the section called “Configurazione di parchi istanze o gruppi di istanze”.

Nodi attività sulle istanze Spot

I nodi di task elaborano i dati, ma non mantengono dati persistenti in HDFS. Se terminano perché il prezzo Spot ha superato il prezzo Spot massimo, nessun dato viene perso e l'effetto sul cluster è minimo.

Quando avvii uno o più gruppi di istanze attività come istanze Spot, Amazon EMR esegue il provisioning di quanti più nodi attività possibili, utilizzando il prezzo Spot massimo. Questo significa che se richiedi un gruppo di istanze attività con sei nodi e solo cinque istanze Spot sono disponibili al o sotto il prezzo Spot massimo, Amazon EMR avvia il gruppo di istanze con cinque nodi, aggiungendo il sesto in seguito, se possibile.

Avviare gruppi di istanze dell'attività come istanze Spot è un modo strategico per espandere la capacità del cluster riducendo i costi. Se avvii i gruppi di istanze primarie e core come istanze on demand, la loro capacità è garantita per l'esecuzione del cluster. Puoi aggiungere istanze dell'attività ai gruppi di istanze dell'attività in base alle esigenze, per gestire picchi di traffico o velocizzare l'elaborazione dei dati.

Puoi aggiungere o rimuovere i nodi di task utilizzando la console, AWS CLI o API. Puoi anche aggiungere altri gruppi di attività, ma non puoi rimuovere un gruppo di attività dopo che è stato creato.

Configurazioni di istanze per scenari applicativi

La tabella seguente è un riferimento rapido per le opzioni di acquisto e le configurazioni dei tipi di nodo in genere appropriate per vari scenari applicativi. Scegli il link per visualizzare ulteriori informazioni su ciascun tipo di scenario.

Scenario applicativo	Opzione di acquisto per i nodi primari	Opzione di acquisto per i nodi principali	Opzione di acquisto per i nodi attività
Cluster di lunga durata e data warehouse	On demand	Combinazione di on demand o parco istanze	Combinazione di Spot o parco istanze
Carichi di lavoro basati sui costi	Spot	Spot	Spot
Carichi di lavoro critici	On demand	On demand	Combinazione di Spot o parco istanze
Test dell'applicazione	Spot	Spot	Spot

Esistono diversi scenari in cui le istanze Spot sono utili per l'esecuzione di un cluster Amazon EMR.

Cluster di lunga durata e data warehouse

Se stai eseguendo un cluster Amazon EMR persistente caratterizzato da una variazione prevedibile della capacità di elaborazione, ad esempio un data warehouse, puoi gestire picchi di domanda a costi inferiori con istanze Spot. Puoi avviare gruppi di istanze primarie e core come istanze on demand per gestire la normale capacità e avviare il gruppo di istanze attività come istanze spot per gestire i requisiti di carico di picco.

Carichi di lavoro basati sui costi

Se stai eseguendo cluster transitori per i quali un costo inferiore è più importante del tempo necessario per il completamento e la perdita parziale del lavoro è accettabile, puoi eseguire l'intero cluster (gruppi di istanze primarie, core e attività) come istanze spot per trarre vantaggio dal maggiore risparmio sui costi.

Carichi di lavoro critici

Se stai eseguendo un cluster per i quale un costo inferiore è più importante del tempo necessario per il completamento, ma la perdita parziale del lavoro non è accettabile, avvia i gruppi di istanze primarie e core come istanze on demand e integra con uno o più gruppi di istanze attività di istanze spot. L'esecuzione di gruppi di istanze primarie e core come istanze on demand garantisce che i dati vengono mantenuti in HDFS e che il cluster sia protetto dalla terminazione dovuta a fluttuazioni del mercato spot, fornendo al contempo risparmi sui costi derivanti dall'esecuzione di gruppi di istanze attività come istanze spot.

Test dell'applicazione

Quando esegui il test di una nuova applicazione come preparazione all'avvio in un ambiente di produzione, puoi eseguire l'intero cluster (gruppi di istanze primarie, core e attività) come istanze spot per ridurre i costi di test.

Calcolo della capacità HDFS richiesta di un cluster

La quantità di storage HDFS disponibile per il cluster dipende dai seguenti fattori:

- Il numero di istanze Amazon EC2 usate per nodi principali.
- La capacità dell'archivio istanza Amazon EC2 per il tipo di istanza usato. Per ulteriori informazioni sui volumi dell'archivio istanza, consulta Archivio istanza Amazon EC2 nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Il numero e la dimensione dei volumi Amazon EBS collegati ai nodi principali.
- Un fattore di replica, che tiene conto del modo in cui ogni blocco di dati viene archiviato in HDFS per ridondanza di tipo RAID. Per impostazione predefinita, il fattore di replica è tre per un cluster di almeno 10 nodi principali, due per un cluster di 4-9 nodi principali e uno per un cluster di massimo tre nodi.

Per calcolare la capacità HDFS di un cluster, per ogni nodo principale aggiungi la capacità del volume instance store alla capacità di archiviazione Amazon EBS (se utilizzata). Moltiplica il risultato per il numero di nodi principali, quindi dividi il totale per il fattore di replica in base al numero di nodi principali. Ad esempio, un cluster con 10 nodi principali di tipo i2.xlarge, che dispone di 800 GB di storage dell'istanza senza alcun volume Amazon EBS collegato, dispone di un totale di circa 2.666 GB disponibili per HDFS (10 nodi x 800 GB ÷ fattore di replica 3).

Se il valore di capacità HDFS calcolato è inferiore ai dati, puoi aumentare la quantità di storage HDFS nei seguenti modi:

- Creando un cluster con volumi Amazon EBS aggiuntivi o aggiungendo gruppi di istanze con volumi Amazon EBS collegati a un cluster esistente
- Aggiungendo più nodi principali
- Scegliendo un tipo di istanza Amazon EC2 con maggiore capacità di archiviazione
- Utilizzando la compressione dati
- Modificando le impostazioni di configurazione Hadoop per ridurre il fattore di replica

La riduzione del fattore di replica deve essere utilizzata con attenzione poiché riduce la ridondanza dei dati HDFS e la capacità del cluster di recuperare da blocchi HDFS persi o danneggiati.

Configurazione della registrazione e del debug di cluster

Quando pianifichi il cluster, devi determinare la quantità di supporto di debug che vuoi rendere disponibile. Durante lo sviluppo iniziale dell'applicazione di elaborazione dati, ti consigliamo di testare l'applicazione su un cluster che elabora un piccolo, ma rappresentativo, sottoinsieme dei tuoi dati. È probabile che per eseguire questa operazione tu intenda utilizzare tutti gli strumenti di debug forniti da Amazon EMR, come l'archiviazione dei file di log in Amazon S3.

Una volta terminata la fase di sviluppo dell'applicazione di elaborazione dati e avviata quella di produzione, puoi scegliere di ridurre il debug. In questo modo, azzeri il costo relativo all'archiviazione degli archivi di file di log in Amazon S3 e riduci il carico di elaborazione sul cluster in quanto non è più necessario scrivere lo stato su Amazon S3. L'inconveniente di questa scelta è che in caso di problemi avrai a disposizione un minor numero di strumenti per gestirli.

File di log di default

Per impostazione predefinita, ogni cluster scrive file di log sul nodo primario. Questi file sono scritti nella directory `/mnt/var/log/`. Puoi accedervi utilizzando SSH per la connessione al nodo primario come descritto in Connessione al nodo primario tramite SSH.

> **Note**
>
> Se utilizzi Amazon EMR versione 6.8.0 o precedenti, i file di log vengono salvati in Amazon S3 durante la terminazione del cluster, quindi non puoi accedere ai file di log una volta terminato il nodo primario. Amazon EMR versione 6.9.0 e successive archiviano i log

in Amazon S3 durante la riduzione del cluster, cosicché i file di log generati sul cluster persistono anche dopo la terminazione del nodo.

Non è necessario attivare alcuna opzione affinché i file di log siano scritti sul nodo primario. In effetti, questo è il comportamento predefinito di Amazon EMR e Hadoop.

Un cluster genera vari tipi di file di log, tra cui:

- Step logs (Log di fase): questi log sono generati dal servizio Amazon EMR e contengono informazioni sul cluster e i risultati di ogni fase. I file di log sono archiviati nella directory `/mnt/var/log/hadoop/steps/` nel nodo primario. Ogni fase registra i risultati a essa relativi in una sottodirectory numerata distinta: `/mnt/var/log/hadoop/steps/s-`*`stepId1`*`/` per la prima fase, `/mnt/var/log/hadoop/steps/s-`*`stepId2`*`/`, per la seconda e così via. Gli identificatori di fase a 13 caratteri (ad esempio stepId1, stepId2) sono esclusivi di un cluster.
- Hadoop and YARN component logs (Registro dei componenti Hadoop e YARN): i registri per i componenti associati a Apache YARN e MapReduce, ad esempio, sono contenuti in cartelle distinte in `/mnt/var/log`. Le posizioni dei file di log per i componenti Hadoop in `/mnt/var/log` sono le seguenti: hadoop-hdfs, hadoop-mapreduce, hadoop-httpfs e hadoop-yarn. La directory hadoop-state-pusher è per l'output del processo pusher dello stato di Hadoop.
- Bootstrap action logs (Log delle operazioni di bootstrap): se il processo utilizza operazioni di bootstrap, i risultati di queste operazioni sono registrati. I file di log sono archiviati in /mnt/var/log/bootstrap-actions/ nel nodo primario. Ogni operazione di bootstrap registra i risultati a essa relativi in una sottodirectory numerata distinta: `/mnt/var/log/bootstrap-actions/1/` per la prima operazione di bootstrap, `/mnt/var/log/bootstrap-actions/2/` per la seconda e così via.
- Instance state logs (Log di stato dell'istanza): forniscono informazioni su CPU, stato della memoria e thread del garbage collector del nodo. I file di log sono archiviati in `/mnt/var/log/instance-state/` nel nodo primario.

Archiviazione di file di log in Amazon S3

> **Note**
>
> Non è al momento possibile utilizzare l'aggregazione dei log in Amazon S3 con l'utility `yarn logs`.

Amazon EMR rilascio 6.9.0 e successivi archiviano i log in Amazon S3 durante la riduzione del cluster, cosicché i file di log generati sul cluster persistono anche dopo la terminazione del nodo. Questo comportamento è abilitato automaticamente, quindi non devi fare nulla per attivarlo.
Per Amazon EMR rilascio 6.8.0 e precedenti, è possibile configurare un cluster per archiviare periodicamente i file di log presenti nel nodo primario in Amazon S3. In questo modo, i file di log saranno disponibili anche dopo la terminazione del cluster, indipendentemente dalla causa della stessa (arresto normale, errore, ecc.). Amazon EMR archivia i file di log in Amazon S3 ogni 5 minuti.

Per Amazon EMR rilascio 6.8.0 e successivi, per archiviare i file di log in Amazon S3 devi abilitare tale caratteristica all'avvio del cluster. Puoi eseguire tale operazione mediante la console, la CLI o l'API. Per impostazione predefinita, l'archiviazione dei log è abilitata per i cluster avviati utilizzando la console. Per i cluster avviati mediante la CLI o l'API, la registrazione in Amazon S3 deve essere abilitata manualmente.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Archiviazione dei file di log in Amazon S3 con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Cluster logs (Log del cluster), seleziona la casella di controllo Publish cluster-specific logs to Amazon S3 (Pubblica log specifici del cluster su Amazon S3).
4. Nel campo Amazon S3 location (Posizione Amazon S3), digita o cerca il percorso Amazon S3 in cui archiviare i log. Se digiti il nome di una cartella che non esiste nel bucket, Amazon S3 crea tale cartella.

   Quando imposti questo valore, Amazon EMR copia i file di log dalle istanze EC2 nel cluster in Amazon S3. Questa operazione evita che i file di log vadano persi quando il cluster termina

ed EC2 termina le istanze EC2 che ospitano il cluster. Questi log sono utili per la risoluzione dei problemi. Per ulteriori informazioni, consulta Visualizzazione dei file di log.

5. Facoltativamente, seleziona la casella di controllo Encrypt cluster-specific logs (Crittografa log specifici del cluster). Quindi, seleziona una chiave AWS KMS dall'elenco, inserisci un ARN della chiave oppure crea una nuova chiave. Questa opzione è disponibile solo con Amazon EMR versione 5.30.0 e successive, esclusa la versione 6.0.0. Per utilizzare questa opzione, aggiungi l'autorizzazione AWS KMS per il profilo dell'istanza EC2 e il ruolo Amazon EMR. Per ulteriori informazioni, consulta Crittografia dei file di log archiviati in Amazon S3 con una chiave gestita dal cliente di AWS KMS.
6. Scegli qualsiasi altra opzione applicabile al cluster.
7. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Archiviazione dei file di log in Amazon S3 con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate).
4. Nella sezione Opzioni generali, nel campo Registrazione, accetta l'opzione predefinita: Abilitato.

   Questo determina se Amazon EMR cattura i dati di log dettagliati su Amazon S3. È possibile effettuare questa impostazione solo quando si crea il cluster. Per ulteriori informazioni, consulta Visualizzare file di log di .
5. Nel campo S3 folder (Cartella S3), digita (o passa a) un percorso Amazon S3 per archiviare i log. Puoi anche permettere alla console di generare un percorso Amazon S3 al posto tuo. Se digita il nome di una cartella che non esiste nel bucket, questa viene creata.

   Quando questo valore è impostato, Amazon EMR copia i file di log dalle istanze EC2 nel cluster in Amazon S3. Questa azione evita che i file di log vengano persi quando il cluster termina e le istanze EC2 che ospitano il cluster sono terminate. Questi log sono utili per la risoluzione dei problemi.

Per ulteriori informazioni, consulta Visualizzazione dei file di log.

6. Nel campo Log encryption (Crittografa di log), seleziona Encrypt logs stored in S3 with an AWS KMS customer managed key (Crittografa i log archiviati in S3 con una chiave gestita dal cliente di AWS KMS). Quindi, seleziona una chiave AWS KMS dall'elenco o immetti un ARN della chiave. È anche possibile creare una nuova chiave AWS KMS.

   Questa opzione è disponibile solo con Amazon EMR versione 5.30.0 e successive, esclusa la versione 6.0.0. Per utilizzare questa opzione, aggiungi l'autorizzazione AWS KMS per il profilo dell'istanza EC2 e il ruolo Amazon EMR. Per ulteriori informazioni, consulta Crittografia dei file di log archiviati in Amazon S3 con una chiave gestita dal cliente di AWS KMS.
7. Procedere con la creazione del cluster come descritto in Pianificazione e configurazione di cluster.

CLI

Archiviazione dei file di log in Amazon S3 con la AWS CLI

Per archiviare file di log in Amazon S3 mediante la AWS CLI, digita il comando `create-cluster` e specifica il percorso dei log in Amazon S3 utilizzando il parametro `--log-uri`.

1. Per registrare file di log in Amazon S3, digita il comando seguente e sostituisci *myKey* con il nome della coppia di chiavi 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. Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi core. Tutti i nodi utilizzeranno il tipo di istanza specificato nel comando.

> **Note**
>
> Se in precedenza non sono stati creati il ruolo di servizio Amazon EMR predefinito e il profilo dell'istanza EC2, inserisci `aws emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Crittografia dei file di log archiviati in Amazon S3 con una chiave gestita dal cliente di AWS KMS

Con Amazon EMR versione 5.30.0 e successive (tranne Amazon EMR 6.0.0), puoi crittografare i file di log archiviati in Amazon S3 con una chiave gestita dal cliente di AWS KMS. Per abilitare questa opzione nella console, segui le fasi in Archiviazione di file di log in Amazon S3. Il profilo dell'istanza Amazon EC2 e il ruolo Amazon EMR devono soddisfare i seguenti prerequisiti:

- Il profilo dell'istanza Amazon EC2 utilizzato per il cluster deve disporre dell'autorizzazione per utilizzare `kms:GenerateDataKey`.
- Il ruolo Amazon EMR utilizzato per il cluster deve disporre dell'autorizzazione per utilizzare `kms:DescribeKey`.
- Il profilo dell'istanza Amazon EC2 e il ruolo Amazon EMR devono essere aggiunti all'elenco degli utenti chiave per la chiave gestita dal cliente di AWS KMS specificata, come illustrato nelle fasi seguenti:
  1. Apri la console AWS Key Management Service (AWS KMS) all'indirizzo https://console.aws.amazon.com/kms.
  2. Per modificare la Regione AWS, utilizza il Selettore di regione nell'angolo in alto a destra della pagina.
  3. Seleziona l'alias della chiave KMS da modificare.
  4. Nella pagina dei dettagli della chiave, in Key Users (Utenti di chiavi), scegli Add (Aggiungi).
  5. Nella finestra di dialogo Add key users (Aggiungi utenti delle chiavi), seleziona il profilo dell'istanza Amazon EC2 e il ruolo Amazon EMR.
  6. Scegliere Add (Aggiungi).

Per ulteriori informazioni, consulta Ruoli del servizio IAM utilizzati da Amazon EMR e Utilizzo delle policy delle chiavi nella Guida per gli sviluppatori di AWS Key Management Service.

Aggregazione di log in Amazon S3 mediante la AWS CLI

> **Note**
>
> Non è al momento possibile utilizzare l'aggregazione dei log con l'utility `yarn logs`. È possibile utilizzare soltanto l'aggregazione supportata da questa procedura.

L'aggregazione di log (Hadoop 2.x) compila i log di tutti i container di una singola applicazione in un unico file. Per abilitare l'aggregazione di log in Amazon S3 mediante la AWS CLI, è possibile utilizzare un'operazione di bootstrap all'avvio del cluster per abilitare l'aggregazione di log e specificare il bucket in cui archiviare i log.

- Per abilitare l'aggregazione di log, crea il file di configurazione denominato `myConfig.json`, che contiene quanto segue:

```
[
  {
    "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"
    }
  }
]
```

  Digita il comando seguente e sostituisci *myKey* con il nome della coppia di chiavi EC2. Puoi inoltre sostituire il testo in rosso con le configurazioni desiderate.

```
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
```

  Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi core. Tutti i nodi utilizzeranno il tipo di istanza specificato nel comando.

> **Note**
>
> Se in precedenza non sono stati creati il ruolo di servizio EMR predefinito e il profilo dell'istanza EC2, esegui `aws emr create-default-roles` per crearli prima di eseguire il sottocomando `create-cluster`.

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta la Guida di riferimento ai comandi della AWS CLI.

Posizione dei log

L'elenco seguente include tutti i tipi di log e le relative posizioni in Amazon S3. Puoi utilizzarli per risolvere i problemi di Amazon EMR.

Log delle fasi

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/steps/<step-id>/
```

Log di applicazioni

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/containers/
```

Questa posizione include i log del container `stderr` e `stdout`, `directory.info`, `prelaunch.out` e `launch_container.sh`.

Log del gestore delle risorse

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
applications/hadoop-yarn/
```

Hadoop HDFS

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/
applications/hadoop-hdfs/
```

Questa posizione include i log NameNode, DataNode e YARN TimelineServer.

Log del gestore dei nodi

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/
applications/hadoop-yarn/
```

Log degli stati istanza

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<all-instance-id>/daemons/
instance-state/
```

Log di provisioning di Amazon EMR

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
provision-node/*
```

Log Hive

```
s3://DOC-EXAMPLE-LOG-BUCKET/<cluster-id>/node/<leader-instance-id>/
applications/hive/*
```

- Per trovare i log di Hive sul tuo cluster, rimuovi l'asterisco (`*`) e aggiungi `/var/log/hive/` al link precedente.
- Per trovare i log di HiveServer2, rimuovi l'asterisco (`*`) e aggiungi `var/log/hive/hiveserver2.log` al link precedente.
- Per trovare i log HiveCLI, rimuovi l'asterisco (`*`) e aggiungi `/var/log/hive/user/hadoop/hive.log` al link precedente.
- Per trovare i log di Hive Metastore Server, rimuovi l'asterisco (`*`) e aggiungi `/var/log/hive/user/hive/hive.log` al link precedente.

Se l'errore si trova nel nodo primario o nel nodo attività dell'applicazione Tez, fornisci i log del container Hadoop appropriato.

Abilitare lo strumento di debug

Lo strumento di debug ti consente di accedere più facilmente ai file di log dalla console Amazon EMR. Per ulteriori informazioni, consulta Visualizzazione dei file di log nello strumento di debug. Quando abiliti lo strumento di debug su un cluster, Amazon EMR archivia i file di log in Amazon S3 e quindi li indicizza. Puoi quindi utilizzare la console per individuare log di fasi, processi, attività e tentativi di attività per un cluster in modo intuitivo.

Per utilizzare lo strumento di debug nella console Amazon EMR, devi abilitare il debug all'avvio del cluster mediante la console, la CLI o l'API. Tieni presente che la nuova console Amazon EMR non offre lo strumento di debug.

Old console

Attivazione dello strumento di debug con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Scegli Go to advanced options (Vai alle opzioni avanzate).
4. Nella sezione Cluster Configuration (Configurazione cluster), nel campo Logging (Registrazione), scegli Enabled (Abilitata). Non è possibile abilitare il debug senza abilitare la registrazione.
5. Nel campo Log folder S3 location (Percorso cartella di log S3), digita un percorso Amazon S3 per archiviare i log.
6. Nel campo Debugging (Debug), scegli Enabled (Abilitato). L'opzione di debug crea uno scambio Amazon SQS per pubblicare messaggi di debug sul back-end del servizio Amazon EMR. È possibile che la pubblicazione di messaggi comporti dei costi. Per ulteriori informazioni, consulta la pagina del prodotto di Amazon SQS.
7. Procedere con la creazione del cluster come descritto in Pianificazione e configurazione di cluster.

AWS CLI

Attivazione dello strumento di debug con la AWS CLI

Per abilitare il debug mediante AWS CLI, digita il `create-cluster` sottocomando con il parametro `--enable-debugging`. È inoltre necessario specificare il parametro `--log-uri` durante l'abilitazione del debug.

- Per abilitare il debug mediante AWS CLI, digita il comando seguente e sostituisci *myKey* con il nome della coppia di chiavi 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
```

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo primario e le istanze rimanenti vengono avviate come nodi core. Tutti i nodi utilizzeranno il tipo di istanza specificato nel comando.

> **Note**
>
> Se in precedenza non sono stati creati il ruolo del servizio EMR predefinito e il profilo dell'istanza EC2, digitare aws `emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

API

Attivazione dello strumento di debug con l'API di Amazon EMR

- Abilita il debug utilizzando la seguente configurazione dell'SDK Java.

```
    StepFactory stepFactory = new StepFactory();
StepConfig enabledebugging = new StepConfig()
    .withName("Enable debugging")
    .withActionOnFailure("TERMINATE_JOB_FLOW")
    .withHadoopJarStep(stepFactory.newEnableDebuggingStep());
```

In questo esempio, `new StepFactory()` utilizza `us-east-1` come regione predefinita. Se il cluster viene avviato in una regione diversa, è necessario specificarla utilizzando `new StepFactory("region.elasticmapreduce")`, ad esempio `new StepFactory("ap-northeast-2.elasticmapreduce")`.

Informazioni sulle opzioni di debug

I rilasci da 4.1.0 a 5.27.0 di Amazon EMR supportano il debug in tutte le Regioni. Altre versioni di Amazon EMR non supportano l'opzione di debug. A partire dal 23 gennaio 2023, Amazon EMR interromperà lo strumento di debug per tutte le versioni.

Amazon EMR crea una coda Amazon SQS per elaborare i dati di debug. È possibile che per i messaggi siano applicati costi aggiuntivi. Amazon SQS include tuttavia un piano gratuito con 1.000.000 di richieste. Per ulteriori informazioni, consulta https://aws.amazon.com/https://aws.amazon.com/sqs.

Il debug richiede l'utilizzo di ruoli. Il tuo ruolo di servizio e il tuo profilo dell'istanza devono consentirti di utilizzare tutte le operazioni API di Amazon SQS. Se i tuoi ruoli sono collegati a policy gestite di Amazon EMR, non devi eseguire alcuna operazione per modificare i ruoli. Se disponi di ruoli personalizzati, devi aggiungere le autorizzazioni `sqs:*`. Per ulteriori informazioni, consulta Configurazione dei ruoli di servizio IAM per le autorizzazioni di Amazon EMR per i servizi e risorse AWS.

Aggiunta di tag ai cluster

Può rivelarsi utile classificare le risorse AWS in vari modi, ad esempio, per scopo, proprietario o ambiente. Puoi eseguire questa operazione in Amazon EMR assegnando metadati personalizzati ai cluster Amazon EMR mediante tag. Un tag consiste di una chiave e di un valore che definisci. Per Amazon EMR, il cluster è il livello di risorse che puoi taggare. Ad esempio, puoi definire un set di tag per i cluster del tuo account per tenere traccia del proprietario di ogni cluster o per identificare un cluster di produzione rispetto a un cluster di testing. Ti consigliamo di creare un set di tag coerente per soddisfare i requisiti dell'organizzazione.

Quando aggiungi un tag a un cluster Amazon EMR, il tag viene propagato anche a ogni istanza Amazon EC2 attiva associata al cluster. Analogamente, quando rimuovi un tag da un cluster Amazon EMR, tale tag viene rimosso da ciascuna istanza Amazon EC2 attiva associata.

> **Important**
>
> Utilizza la console o la CLI di Amazon EMR per gestire i tag sulle istanze Amazon EC2 che fanno parte di un cluster anziché la console o la CLI di Amazon EC2, in quanto le modifiche effettuate in Amazon EC2 non sono sincronizzate nel sistema di assegnazione di tag di Amazon EMR.

Puoi identificare un'istanza Amazon EC2 che fa parte di un cluster Amazon EMR cercando i tag di sistema riportati di seguito. In questo esempio, *`CORE`* è il valore del ruolo del gruppo di istanze e *`j-12345678`* è un esempio di valore di identificatore di un flusso di elaborazione (cluster):

- aws:elasticmapreduce:instance-group-role=*CORE*
- aws:elasticmapreduce:job-flow-id=*j-12345678*

> **Note**
>
> Amazon EMR e Amazon EC2 interpretano i tag come una stringa di caratteri senza significato semantico.

Puoi utilizzare i tag utilizzando la AWS Management Console, l'CLI e l'API.

Puoi aggiungere tag durante la creazione di un nuovo cluster Amazon EMR nonché aggiungere, modificare o rimuovere tag da un cluster Amazon EMR in esecuzione. La modifica di un tag è un concetto che si applica alla console di Amazon EMR. Tuttavia, se utilizzi la CLI e l'API, la modifica di un tag consiste nel rimuovere il vecchio tag e aggiungerne un nuovo. Puoi modificare chiavi e valori di tag e rimuovere tag da una risorsa in qualsiasi momento durante l'esecuzione di un cluster. Tuttavia, non puoi aggiungere, modificare o rimuovere tag da un cluster o da istanze terminati precedentemente associati a un cluster che è ancora attivo. Puoi inoltre impostare il valore di un tag su una stringa vuota, ma non su null.

Se utilizzi AWS Identity and Access Management (IAM) con le istanze Amazon EC2 per le autorizzazioni basate sulle risorse in base ai tag, le policy IAM vengono applicate ai tag che Amazon EMR propaga alle istanze Amazon EC2 di un cluster. Per propagare i tag di Amazon EMR alle istanze Amazon EC2, la tua policy IAM per Amazon EC2 deve concedere le autorizzazioni per chiamare le API CreateTags e DeleteTags di Amazon EC2. Inoltre, i tag propagati possono avere un'incidenza sulle autorizzazioni basate sulle risorse di Amazon EC2. I tag propagati ad Amazon EC2 possono essere letti come condizioni nella policy IAM, esattamente come avviene con gli altri tag Amazon EC2. Fai riferimento alla policy IAM quando aggiungi tag ai cluster Amazon EMR, in modo da evitare che gli utenti dispongano di autorizzazioni non corrette per un cluster. Per evitare problemi, assicurati che le policy IAM non includano condizioni sui tag che intendi utilizzare anche sui cluster Amazon EMR. Per ulteriori informazioni, consulta Controllo dell'accesso alle risorse di Amazon EC2.

Limitazioni applicate ai tag

Ai tag si applicano le seguenti limitazioni di base:

- Le limitazioni applicabili alle risorse Amazon EC2 lo sono anche per Amazon EMR. Per ulteriori informazioni, consulta https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/Using_Tags.html#tag-restrictions.
- Non utilizzare il prefisso `aws:` nei nomi e nei valori di tag in quanto è riservato per l'uso in AWS. Inoltre, non puoi modificare né eliminare nomi o valori di tag con tale prefisso.
- Non puoi cambiare o modificare i tag su un cluster terminato.
- Un valore di tag può essere una stringa vuota, ma non nullo. Inoltre, una chiave di tag non può essere una stringa vuota.
- Chiavi e valori possono contenere qualsiasi carattere alfabetico in qualsiasi lingua, qualsiasi carattere numerico, spazi bianchi, separatori sottointesi e i seguenti simboli: _ . : / = + - @

Per ulteriori informazioni sull'assegnazione di tag utilizzando la AWS Management Console, consulta Utilizzo dei tag nella console nella Guida per l'utente di Amazon EC2 per le istanze Linux. Per ulteriori informazioni sull'assegnazione di tag mediante Amazon EC2API o la riga di comando, consulta Panoramica su API e CLI nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Aggiunta di tag alle risorse per la fatturazione

Puoi utilizzare i tag per organizzare la fattura AWS affinché rifletta la tua struttura dei costi. Per eseguire questa operazione, registrati per far sì che la fattura del tuo account AWS includa i valori di chiave di tag. Puoi quindi organizzare le informazioni di fatturazione in base ai valori di chiave di tag per visualizzare il costo delle risorse combinate. Sebbene Amazon EMR e Amazon EC2 abbiano estratti conti differenti, i tag applicati a ogni cluster sono propagati anche a ogni istanza associata in modo che sia possibile utilizzare i tag per collegare i costi di Amazon EMR e Amazon EC2 correlati.

Puoi ad esempio applicare tag a numerose risorse con un nome di applicazione specifico, quindi organizzare le informazioni di fatturazione per visualizzare il costo totale dell'applicazione in più servizi. Per ulteriori informazioni, consulta Assegnazione di tag e allocazione dei costi nella Guida per l'utente di AWS Billing.

Aggiunta di tag a un cluster

Puoi aggiungere tag ai cluster al momento della loro creazione.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Aggiunta di tag durante la creazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Tags (Tag), seleziona Add new tag (Aggiungi nuovo tag). Specifica un tag nel campo Key (Chiave). Facoltativamente, puoi specificare un tag nel campo Value (Valore).
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Aggiunta di tag durante la creazione di un cluster con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Nella pagina Step 3: General Cluster Settings (Fase 3: Impostazioni cluster generali), nella sezione Tags (Tag), digitare un valore in Key (Chiave) per il tag.

   Quando si inizia a digitare il valore Key (Chiave), un'altra riga viene visualizzata automaticamente per il nuovo tag successivo.
4. Facoltativamente, in Value (Valore) digitare un valore per il tag.
5. Ripeti i passaggi precedenti per ogni coppia chiave/valore di tag da aggiungere al cluster. All'avvio del cluster, tutti i tag inseriti vengono automaticamente associati al cluster.

AWS CLI

Aggiunta di tag durante la creazione di un cluster con la AWS CLI

L'esempio seguente illustra come aggiungere un tag a un nuovo cluster mediante l'AWS CLI. Per aggiungere dei tag quando si crea un cluster, digita il sottocomando `create-cluster` con il parametro `--tags`.

- Per aggiungere un tag denominato *`costCenter`* con il valore chiave *`marketing`* alla creazione di un cluster, digita il comando seguente e sostituisci *`myKey`* con il nome della coppia di chiavi 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
```

Quando si specifica il numero di istanze senza utilizzare il parametro `--instance-groups`, viene avviato un singolo nodo master e le istanze rimanenti vengono avviate come nodi principali. Tutti i nodi utilizzeranno il tipo di istanza specificato nel comando.

> **Note**
>
> Se in precedenza non sono stati creati il ruolo del servizio EMR predefinito e il profilo dell'istanza EC2, digitare aws `emr create-default-roles` per crearli prima di digitare il sottocomando `create-cluster`.

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Puoi anche aggiungere tag a un cluster esistente.

New console

Aggiunta di tag a un cluster esistente con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.

2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Nella scheda Tags (Tag) nella pagina dei dettagli del cluster, seleziona Manage tags (Gestisci tag). Specifica un tag nel campo Key (Chiave). Facoltativamente, puoi specificare un tag nel campo Value (Valore).
4. Selezionare Save changes (Salva modifiche). La scheda Tags (Tag) si aggiorna con il nuovo numero di tag presenti nel cluster. Ad esempio, se ora hai due tag, l'etichetta della tua scheda è Tags (2).

Old console

Aggiunta di tag a un cluster esistente con la vecchia console

1. Nella console di Amazon EMR, seleziona la pagina Cluster List (Elenco cluster) e fai clic su un cluster a cui aggiungere i tag.
2. Nella pagina Cluster Details (Dettagli del cluster), nel campo Tags (Tag), fai clic su View All/Edit (Visualizza tutto/Modifica).
3. Nella pagina View All/Edit (Visualizza tutto/Modifica), fai clic su Add (Aggiungi).
4. Fai clic sul campo vuoto nella colonna Key (Chiave) e digita il nome della chiave.
5. Eventualmente, fai clic sul campo vuoto nella colonna Value (Valore) e digita il nome del valore.
6. Quando si inizia a inserire un nuovo tag, un'altra riga di tag vuota appare sotto il tag in corso di modifica. Ripeti i passaggi precedenti sulla nuova riga di tag per ogni tag da aggiungere.

AWS CLI

Aggiunta di tag a un cluster in esecuzione con la AWS CLI

- Inserisci il sottocomando `add-tags` con il parametro `--tag` per assegnare tag a un ID del cluster. Puoi recuperare l'ID del cluster tramite la console o il comando `list-clusters`. Il sottocomando `add-tags` accetta attualmente un solo ID di risorsa.

  Ad esempio, per aggiungere due tag a un cluster in esecuzione (uno con una chiave denominata *`costCenter`* con un valore di *`marketing`* e un altro denominato *`other`* con un valore di *`accounting`*), inserisci il comando seguente e sostituisci *`j-KT4XXXXXXXX1NM`* con il l'ID del cluster.

```
aws emr add-tags --resource-id j-KT4XXXXXXXX1NM --tag "costCenter=marketing" --
tag "other=accounting"
```

Nota: quando i tag vengono aggiunti mediante la CLI AWS, il comando non genera output. Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Visualizzazione di tag su un cluster

Se desideri visualizzare tutti i tag associati a un cluster, puoi farlo nella console o nella AWS CLI.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Visualizzazione dei tag di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Per visualizzare tutti i tag, seleziona la scheda Tags (Tag) nella pagina dei dettagli del cluster.

Old console

Visualizzazione dei tag di un cluster con la vecchia console

1. Nella console di Amazon EMR, seleziona la pagina Cluster List (Elenco cluster) e fai clic su un cluster per visualizzare i tag.

2. Alcuni tag sono visualizzati nella pagina Cluster Details (Dettagli del cluster), nel campo Tags (Tag). Fai clic su View All/Edit (Visualizza tutto/Modifica) per visualizzare tutti i tag disponibili sul cluster.

AWS CLI

Visualizzazione dei tag di un cluster con la AWS CLI

Per visualizzare i tag su un cluster mediante AWS CLI, digita il sottocomando `describe-cluster` con il parametro `--query`.

- Per visualizzare i tag di un cluster, digita il comando seguente e sostituisci *j-KT4XXXXXXXX1NM* con l'ID del cluster.

```
aws emr describe-cluster --cluster-id j-KT4XXXXXX1NM --query Cluster.Tags
```

L'output visualizza tutte le informazioni sui tag del cluster in un formato simile al seguente:

```
Value: accounting     Value: marketing
Key: other            Key: costCenter
```

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Rimozione di tag da un cluster

Se un tag non è più necessario, puoi rimuoverlo dal cluster.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Rimozione dei tag di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Nella scheda Tags (Tag) nella pagina dei dettagli del cluster, seleziona Manage tags (Gestisci tag).
4. Scegli Remove (Rimuovi) per ogni coppia chiave-valore che desideri rimuovere.
5. Seleziona Salva modifiche.

Old console

Rimozione dei tag di un cluster con la vecchia console

1. Nella console di Amazon EMR, seleziona la pagina Cluster List (Elenco cluster) e fai clic sul cluster da cui rimuovere i tag.
2. Nella pagina Cluster Details (Dettagli del cluster), nel campo Tags (Tag), fai clic su View All/Edit (Visualizza tutto/Modifica).
3. Nella finestra di dialogo View All/Edit (Visualizza tutto/Modifica), fai clic sull'icona X accanto al tag da eliminare e fai clic su Save (Salva).
4. (Facoltativo) Ripeti il passaggio precedente per ogni coppia chiave-valore di tag da rimuovere dal cluster.

AWS CLI

Rimozione dei tag di un cluster con la AWS CLI

Digita il sottocomando `remove-tags` con il parametro `--tag-keys`. Quando si rimuove un tag, è richiesto solo il nome di chiave.

- Per rimuovere un tag da un cluster, digita il comando seguente e sostituisci *j-KT4XXXXXXXX1NM* con l'ID del cluster.

```
aws emr remove-tags --resource-id j-KT4XXXXXX1NM --tag-keys "costCenter"
```

> **Note**
>
> Non è attualmente possibile rimuovere più tag utilizzando un unico comando.

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Driver e integrazione delle applicazioni di terze parti

Puoi eseguire diverse popolari applicazioni per i Big Data su Amazon EMR con prezzi di tipo utility: ciò significa che pagherai una tariffa oraria aggiuntiva nominale per l'applicazione di terze parti mentre il cluster è in esecuzione. In questo modo puoi utilizzare l'applicazione senza dover acquistare una licenza annuale. Le seguenti sezioni descrivono alcuni degli strumenti che è possibile utilizzare con EMR.

Argomenti

Utilizzo degli strumenti di Business Intelligence con Amazon EMR

Puoi utilizzare gli strumenti di Business Intelligence più diffusi, quali Microsoft Excel, MicroStrategy, QlikView e Tableau con Amazon EMR per esplorare e visualizzare i dati. Molti di questi strumenti richiedono un driver ODBC (Open Database Connectivity) o JDBC (Open Database Connectivity). Per scaricare e installare i driver più recenti, consulta la pagina http://awssupportdatasvcs.com/bootstrap-actions/Simba/latest/.

Per trovare le versioni precedenti dei driver, consulta la pagina http://awssupportdatasvcs.com/bootstrap-actions/Simba/.

Sicurezza in Amazon EMR

Per AWS, la sicurezza del cloud ha la massima priorità. In quanto cliente AWS, è possibile trarre vantaggio da un'architettura di data center e di rete progettata per soddisfare i requisiti delle organizzazioni più esigenti a livello di sicurezza.

La sicurezza è una responsabilità condivisa tra te e AWS. Il modello di responsabilità condivisa descrive questo come sicurezza del cloud e sicurezza nel cloud:

- La sicurezza del cloud: AWS è responsabile della protezione dell'infrastruttura che esegue i servizi AWS nel cloud AWS. AWS fornisce inoltre servizi che puoi utilizzare in sicurezza. I revisori di terze parti testano e verificano regolarmente l'efficacia della sicurezza come parte dei programmi di conformità AWS. Per ulteriori informazioni sui programmi di conformità che si applicano ad Amazon EMR, consulta Servizi AWS coperti dal programma di conformità.
- Sicurezza nel cloud: la tua responsabilità è determinata dal servizio AWS che utilizzi. Inoltre, sei responsabile anche di altri fattori, tra cui la riservatezza dei dati, i requisiti dell'azienda e le leggi e le normative applicabili.

La presente documentazione aiuta a comprendere come applicare il modello di responsabilità condivisa quando si utilizza Amazon EMR. Quando sviluppi soluzioni su Amazon EMR, utilizza le seguenti tecnologie per proteggere le risorse e i dati del cluster in base ai requisiti aziendali. Gli argomenti in questo capitolo mostrano come configurare Amazon EMR e utilizzare altri servizi AWS per soddisfare gli obiettivi di sicurezza e conformità.

Configurazioni di sicurezza

Le configurazioni di sicurezza in Amazon EMR sono modelli per impostazioni di sicurezza diverse. Puoi creare una configurazione di sicurezza per riutilizzare in modo pratico un'impostazione di sicurezza ogni volta che crei un cluster. Per ulteriori informazioni, consulta Utilizzo delle configurazioni di sicurezza per impostare la sicurezza del cluster.

Protezione dei dati

Puoi implementare la crittografia dei dati per proteggere i dati a riposo in Amazon S3, i dati a riposo nell'archiviazione delle istanze del cluster e i dati in transito. Per ulteriori informazioni, consulta Crittografia dei dati a riposo e in transito.

AWS Identity and Access Management con Amazon EMR

AWS Identity and Access Management (IAM) è un servizio AWS che consente agli amministratori di controllare in modo sicuro l'accesso alle risorse AWS. Gli amministratori IAM controllano chi è authenticated (autenticato) (accesso effettuato) e authorized (autorizzato) (dispone di autorizzazioni) a utilizzare risorse Amazon EMR. IAM è un servizio AWS che è possibile utilizzare senza alcun costo aggiuntivo.

- Policy IAM basate su identità: le policy IAM concedono o negano a utenti e gruppi le autorizzazioni per eseguire operazioni. Le policy possono essere combinate con le funzionalità di tagging per controllare gli accessi cluster per cluster. Per ulteriori informazioni, consulta AWS Identity and Access Management per Amazon EMR.
- Ruoli IAM: il ruolo di servizio Amazon EMR, il profilo dell'istanza e il ruolo collegato ai servizi controllano come Amazon EMR può accedere ad altri servizi AWS. Per ulteriori informazioni, consulta Configurazione dei ruoli di servizio IAM per le autorizzazioni di Amazon EMR per i servizi e risorse AWS.
- Ruoli IAM per le richieste EMRFS ad Amazon S3: quando Amazon EMR accede ad Amazon S3, puoi specificare il ruolo IAM da utilizzare in funzione dell'utente, del gruppo o della posizione dei dati EMRFS in Amazon S3. Questo consente di controllare in modo preciso se gli utenti del cluster possono accedere ai file da Amazon EMR. Per ulteriori informazioni, consulta Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3.

Kerberos

Puoi configurare Kerberos per fornire un'autenticazione efficace tramite crittografia con chiave segreta. Per ulteriori informazioni, consulta Utilizzo di Kerberos per l'autenticazione con Amazon EMR.

Lake Formation

Puoi utilizzare le autorizzazioni Lake Formation insieme ad AWS Glue Data Catalog per fornire accesso granulare a livello di colonna a database e tabelle in AWS Glue Data Catalog. Lake Formation consente l'accesso federato single sign-on ai EMR Notebooks o Apache Zeppelin da un sistema di identità aziendale. Per ulteriori informazioni, consulta Integrazione di Amazon EMR con AWS Lake Formation.

Secure Socket Shell (SSH)

SSH consente di utilizzare un modo sicuro per connettersi alla riga di comando sulle istanze di cluster. Fornisce inoltre il tunneling per visualizzare le interfacce Web che le applicazioni ospitano sul nodo master. I clienti possono eseguire l'autenticazione utilizzando Kerberos o una coppia di chiavi Amazon EC2. Per ulteriori informazioni, consulta Utilizzo di una coppia di chiavi EC2 per le credenziali SSH e Connessione a un cluster.

Gruppi di sicurezza Amazon EC2

I gruppi di sicurezza agiscono come firewall virtuale per le istanze di cluster EMR, limitando il traffico di rete in entrata e in uscita. Per ulteriori informazioni, consulta Controllo del traffico di rete con gruppi di sicurezza.

Aggiornamenti per l'AMI predefinita di Amazon Linux per Amazon EMR

> **Important**
>
> I cluster Amazon EMR che eseguono le AMI (Amazon Linux Machine Images) Amazon Linux o Amazon Linux 2 utilizzano il comportamento predefinito di Amazon Linux e non scaricano e installano automaticamente aggiornamenti importanti e critici dei kernel che richiedono un riavvio. Si tratta dello stesso comportamento assunto da altre istanze Amazon EC2 che eseguono l'AMI predefinita di Amazon Linux. Se nuovi aggiornamenti software Amazon Linux che richiedono un riavvio (ad esempio, aggiornamenti del kernel, NVIDIA e CUDA) risultano disponibili dopo il rilascio di una versione di Amazon EMR, le istanze del cluster Amazon EMR che eseguono l'AMI predefinita non scaricano e installano automaticamente tali aggiornamenti. Per ottenere gli aggiornamenti del kernel, puoi personalizzare l'AMI di Amazon EMR per utilizzare l'AMI di Amazon Linux più recente.

A seconda dell'assetto di sicurezza dell'applicazione e la durata di esecuzione di un cluster, è possibile scegliere di riavviare periodicamente il cluster per applicare gli aggiornamenti di sicurezza, oppure creare un'operazione di bootstrap per personalizzare l'installazione dei pacchetti e degli aggiornamenti. È anche possibile scegliere di provare e quindi installare determinati aggiornamenti di sicurezza sulle istanze del cluster in esecuzione. Per ulteriori informazioni, consulta Utilizzo

dell'AMI Amazon Linux predefinita per Amazon EMR. Tieni presente che la tua configurazione di rete deve consentire l'uscita HTTP e HTTPS ai repository Amazon Linux in Amazon S3, altrimenti gli aggiornamenti della sicurezza non avranno esito positivo.

Utilizzo delle configurazioni di sicurezza per impostare la sicurezza del cluster

Con Amazon EMR versione 4.8.0 o versioni successive, puoi utilizzare le configurazioni di sicurezza per configurare la crittografia dei dati, l'autenticazione Kerberos (disponibile nella versione 5.10.0 e versioni successive) e l'autorizzazione Amazon S3 per EMRFS (disponibile nella versione 5.10.0 o versioni successive).

Dopo aver creato una configurazione di sicurezza, devi specificarla alla creazione di un cluster e puoi riutilizzarla per un numero qualsiasi di cluster.

Per creare una configurazione di sicurezza, puoi utilizzare la console AWS Command Line Interface (AWS CLI) oppure gli SDK AWS. Puoi anche usare un modello AWS CloudFormation per creare una configurazione di sicurezza. Per ulteriori informazioni, consulta Guida per l'utente di AWS CloudFormation e il modello di riferimento per AWS::EMR::SecurityConfiguration.

Argomenti

- Creazione di una configurazione di sicurezza
- Impostazione di una configurazione di sicurezza per un cluster

Creazione di una configurazione di sicurezza

Questo argomento descrive le procedure generali per la creazione di una configurazione di sicurezza mediante la console di EMR e AWS CLI. Include inoltre un riferimento per i parametri relativi a crittografia, autenticazione e ruoli IAM per EMRFS. Per ulteriori informazioni su queste caratteristiche, consulta i seguenti argomenti:

- Crittografia dei dati a riposo e in transito
- Utilizzo di Kerberos per l'autenticazione con Amazon EMR
- Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3

Per creare una configurazione di sicurezza mediante la console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. Nel riquadro di navigazione, scegliere Security Configurations (Configurazioni di sicurezza), quindi Create security configuration (Crea configurazione di sicurezza).
3. Digitare un nome in Name (Nome) per la configurazione di sicurezza.
4. Scegli le opzioni per Encryption (Crittografia) e Authentication (Autenticazione) come descritto nelle sezioni successive, quindi scegli Create (Crea).

Per creare una configurazione di sicurezza usando AWS CLI

- Usa il comando `create-security-configuration` come mostrato nell'esempio seguente.
  - Per *`SecConfigName`*, specifica il nome della configurazione di sicurezza. Si tratta del nome che hai specificato alla creazione di un cluster che utilizza questa configurazione di sicurezza.
  - Per *`SecConfigDef`*, specifica una struttura JSON inline o il percorso di un file JSON locale *`file://MySecConfig.json`*. I parametri JSON definiscono le opzioni per Encryption (Crittografia), IAM Roles for EMRFS access to Amazon S3 (Ruoli IAM per l'accesso EMRFS ad Amazon S3) e Authentication (Autenticazione) come descritto nelle sezioni successive.

```
aws emr create-security-configuration --name "SecConfigName" --security-
configuration SecConfigDef
```

Configurazione della crittografia di dati

Prima di configurare la crittografia in una configurazione di sicurezza, è necessario creare le chiavi e i certificati utilizzati per la crittografia. Per ulteriori informazioni, consulta Fornitura di chiavi per la crittografia di dati a riposo con Amazon EMR e Fornitura di certificati per la crittografia di dati in transito con Amazon EMR.

Quando crei una configurazione di sicurezza, specifichi due set di opzioni di crittografia: la crittografia di dati inattivi e la crittografia di dati in transito. Le opzioni per la crittografia di dati a riposo includono Amazon S3 con EMRFS e la crittografia per dischi locali. Le opzioni per la crittografia In transito abilitano le funzionalità di crittografia open source per determinate applicazioni che supportano il protocollo Transport Layer Security (TLS). Le opzioni per la crittografia inattiva e per quella in transito

possono essere attivate insieme o separatamente. Per ulteriori informazioni, consulta Crittografia dei dati a riposo e in transito.

> **Note**
>
> L'uso di AWS KMS comporta l'applicazione di costi per l'archiviazione e l'uso delle chiavi di crittografia. Per ulteriori informazioni, consultare Prezzi di AWS KMS.

Impostazione delle opzioni di crittografia mediante la console

Scegli le opzioni in Encryption (Crittografia) in base alle linee guida riportate di seguito.

- Scegli le opzioni in At rest encryption (Crittografia inattiva) per crittografare i dati archiviati nel file system.

  Puoi scegliere di crittografare i dati in Amazon S3, nei dischi locali o in entrambi.
- In S3 data encryption (Crittografia di dati S3), per Encryption mode (Modalità di crittografia), scegli un valore per determinare il modo in cui Amazon EMR crittografa i dati Amazon S3 con EMRFS.

  La fase successiva dipende dalla modalità di crittografia scelta:

  - SSE-S3

    Specifica la crittografia lato server con le chiavi di crittografia gestite da Amazon S3. Non è necessaria alcuna altra operazione in quanto Amazon S3 gestisce automaticamente le chiavi.
  - SSE-KMS o CSE-KMS

    Specifica la crittografia lato server con chiavi gestite da AWS KMS (KMS-CSE) o la crittografia lato client con chiavi gestite da AWS KMS (CSE-KMS). Per AWS KMS key, seleziona una chiave. La chiave deve esistere nella stessa regione del cluster EMR. Per i requisiti relativi alle chiavi, consulta Utilizzo di AWS KMS keys per la crittografia.
  - CSE-Custom

    Specifica la crittografia lato client mediante una chiave root lato client personalizzata (CSE-Custom). Per S3 object (Oggetto S3), immetti il percorso in Amazon S3 oppure l'ARN Amazon S3 del file JAR del provider di chiavi personalizzato. Quindi, per Key provider class (Classe provider di chiavi), immetti il nome di classe completo di una classe dichiarata nell'applicazione che implementa l'interfaccia EncryptionMaterialsProvider.

- In Local disk encryption (Crittografia per dischi locali), scegli un valore per Key provider type (Tipo di provider di chiavi).
  - AWS KMS key

    Selezionare questa opzione per specificare una AWS KMS key. Per AWS KMS key, seleziona una chiave. La chiave deve esistere nella stessa regione del cluster EMR. Per ulteriori informazioni sui requisiti relativi alle chiavi, consulta Utilizzo di AWS KMS keys per la crittografia.

    Crittografia EBS

    Quando specifichi AWS KMS come provider di chiavi, puoi abilitare la crittografia EBS per crittografare il dispositivo root EBS e i volumi di archiviazione. Per abilitare tale opzione, è necessario concedere al ruolo del servizio EMR `EMR_DefaultRole` le autorizzazioni per utilizzare la AWS KMS key specificata. Per ulteriori informazioni sui requisiti relativi alle chiavi, consulta Abilitazione della crittografia EBS fornendo autorizzazioni aggiuntive per le chiavi KMS.
  - Personalizza

    Seleziona questa opzione per specificare un provider di chiavi personalizzato. Per S3 object (Oggetto S3), immetti il percorso in Amazon S3 oppure l'ARN Amazon S3 del file JAR del provider di chiavi personalizzato. Per Key provider class (Classe provider di chiavi), immetti il nome di classe completo di una classe dichiarata nell'applicazione che implementa l'interfaccia EncryptionMaterialsProvider. Il nome di classe qui specificato deve essere diverso dal nome di classe fornito per CSE-Custom.
- Scegli In-transit encryption (Crittografia in transito) per abilitare le funzionalità della crittografia TLS open source per dati in transito. Scegli Certificate provider type (Tipo di provider di certificati) in base alle seguenti linee guida:
  - PEM

    Seleziona questa opzione per utilizzare i file PEM forniti in un file ZIP. Due artefatti devono essere inclusi nel file ZIP: privateKey.pem e certificateChain.pem. Un terzo file, trustedCertificates.pem, è facoltativo. Per informazioni dettagliate, consulta Fornitura di certificati per la crittografia di dati in transito con Amazon EMR. Per S3 object (Oggetto S3), specifica il percorso in Amazon S3 o l'ARN Amazon S3 del campo del file ZIP.
  - Personalizza

    Seleziona questa opzione per specificare un provider di certificati personalizzato, quindi, per S3 object (Oggetto S3), immetti il percorso in Amazon S3, o l'ARN Amazon S3, del file JAR del

provider di certificati personalizzato. Per Key provider class (Classe provider di chiavi), immettere il nome di classe completo di una classe dichiarata nell'applicazione che implementa l'interfaccia TLSArtifactsProvider.

Impostazione delle opzioni di crittografia mediante la AWS CLI

Le sezioni seguenti includono scenari di esempio per illustrare il JSON ben formato --security-configuration per differenti configurazioni e provider di chiavi nonché un riferimento per i parametri JSON e i valori appropriati.

Esempio di opzioni di crittografia di dati in transito

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati In transito è abilitata e la crittografia di dati inattivi è disabilitata.
- Un file ZIP contenente certificati in Amazon S3 è utilizzato come provider di chiavi (consulta Fornitura di certificati per la crittografia di dati in transito con Amazon EMR per i requisiti relativi ai certificati).

```
aws emr create-security-configuration --name "MySecConfig" --security-configuration '{
 "EncryptionConfiguration": {
  "EnableInTransitEncryption": true,
  "EnableAtRestEncryption": false,
  "InTransitEncryptionConfiguration": {
   "TLSCertificateConfiguration": {
    "CertificateProviderType": "PEM",
    "S3Object": "s3://MyConfigStore/artifacts/MyCerts.zip"
   }
  }
 }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati In transito è abilitata e la crittografia di dati inattivi è disabilitata.
- È utilizzato un provider di chiavi personalizzato (vedi Fornitura di certificati per la crittografia di dati in transito con Amazon EMR per i requisiti relativi ai certificati).

```
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"
   }
  }
  }
}'
```

Esempio di opzioni di crittografia di dati a riposo

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati In transito è disabilitata e la crittografia di dati inattivi è abilitata.
- SSE-S3 è utilizzato per la crittografia di Amazon S3.
- La crittografia per dischi locali utilizza AWS KMS come provider di chiavi.

```
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"
   }
  }
  }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati in transito è abilitata e fa riferimento a un file ZIP con certificati PEM in Amazon S3 mediante l'ARN.
- SSE-KMS è utilizzato per la crittografia di Amazon S3.
- La crittografia per dischi locali utilizza AWS KMS come provider di chiavi.

```
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"
   }
  }
 }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati in transito è abilitata e fa riferimento a un file ZIP con certificati PEM in Amazon S3.
- CSE-KMS è utilizzato per la crittografia di Amazon S3.
- La crittografia per dischi locali utilizza un provider di chiavi personalizzato a cui si fa riferimento con il relativo 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"
   }
  }
 }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati in transito è abilitata con un provider di chiavi personalizzato.
- CSE-Custom è utilizzato per i dati Amazon S3.
- La crittografia per dischi locali utilizza un provider di chiavi personalizzato.

```
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"
   }
  }
 }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati In transito è disabilitata e la crittografia di dati inattivi è abilitata.
- La crittografia Amazon S3 è abilitata con SSE-KMS.
- Vengono utilizzate più chiavi AWS KMS, una per ogni bucket S3 e le eccezioni di crittografia vengono applicate a questi singoli bucket S3.
- La crittografia del disco locale è disabilitata.

```
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
  }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati In transito è disabilitata e la crittografia di dati inattivi è abilitata.
- La crittografia Amazon S3 è abilitata con SSE-S3 e la crittografia dei dischi locali è disabilitata.

```
aws emr create-security-configuration --name "MyS3EncryptionConfig" --security-
configuration '{
    "EncryptionConfiguration": {
        "EnableInTransitEncryption": false,
        "EnableAtRestEncryption": true,
        "AtRestEncryptionConfiguration": {
            "S3EncryptionConfiguration": {
                "EncryptionMode": "SSE-S3"
            }
        }
     }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati In transito è disabilitata e la crittografia di dati inattivi è abilitata.
- La crittografia su disco locale è abilitata con AWS KMS come provider di chiavi e la crittografia Amazon S3 è disabilitata.

```
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"
            }
        }
     }
}'
```

L'esempio successivo illustra lo scenario seguente:

- La crittografia di dati In transito è disabilitata e la crittografia di dati inattivi è abilitata.
- La crittografia su disco locale è abilitata con AWS KMS come provider di chiavi e la crittografia Amazon S3 è disabilitata.
- La crittografia EBS è abilitata.

```
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"
            }
        }
     }
}'
```

Riferimento JSON per impostazioni di crittografia

La tabella seguente elenca i parametri JSON per le impostazioni di crittografia e fornisce una descrizione dei valori accettabili per ogni parametro.

Parametro	Descrizione
`"EnableInTransitEncryption" : true | false`	Specify `true` to enable in-transit encryption and `false` to disable it. If omitted, `false` is assumed, and in-transit encryption is disabled.
`"EnableAtRestEncryption": true | false`	Specify `true` to enable at-rest encryption and `false` to disable it. If omitted, `false` is assumed and at-rest encryption is disabled.
Parametri di crittografia in transito	
`"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 `Personalizza` 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` 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 `Personalizza` 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

Parametro	Descrizione
	named exactly `privateKey.pem` and `certificateChain.pem`. A file named `trustedCertificates.pem` is optional.
`"CertificateProviderClass" : "MyClassID "`	Required only if `Personalizza` 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`.
Parametri di crittografia inattiva	
`"AtRestEncryptionConfiguration" :`	Specifies a collection of values for at-rest encryption when `EnableAtRestEncryption` is `true`, including Amazon S3 encryption and local disk encryption.
Parametri di crittografia di Amazon S3	
`"S3EncryptionConfiguration" :`	Specifies a collection of values used for Amazon S3 encryption with the EMR File System (EMRFS).
`"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.

Parametro	Descrizione
"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* .
Parametri di crittografia per dischi locali	
"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 .

Parametro	Descrizione
`"EncryptionKeyProviderType": "AwsKms" | "Custom"`	Specifies the key provider. If `AwsKms` is specified, an KMS key ARN must be specified as the `AwsKmsKey` value. If **`Personalizza`** 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`).
`"EncryptionKeyProviderClass" : "MyLocalDiskKeyClassID "`	Required only when `Personalizza` 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`*.

Configurazione dell'autenticazione Kerberos

Una configurazione di sicurezza con impostazioni Kerberos può essere utilizzata da un cluster creato con attributi Kerberos, altrimenti si verifica un errore. Per ulteriori informazioni, consulta Utilizzo

di Kerberos per l'autenticazione con Amazon EMR. Kerberos è disponibile solo in Amazon EMR versione 5.10.0 e versioni successive.

Configurazione delle impostazioni di Kerberos mediante la console

Scegliere le opzioni in Kerberos authentication (Autenticazione Kerberos) in base alle linee guida seguenti.

Parametro		Descrizione
Kerberos		Specifica che Kerberos è abilitato per i cluster che utilizzano questa configurazione di protezione. Se un cluster utilizza questa configurazione di protezione, deve avere anche le impostazioni Kerberos specificate o, in caso contrario, genererà un errore.
Provider	KDC dedicato del cluster	Specifica che Amazon EMR crea un KDC sul nodo primario di qualsiasi cluster che utilizza questa configurazione di sicurezza. Quando crei il cluster, puoi specificare il nome del realm e la password di amministratore KDC. Se necessario, puoi fare riferimento a questo KDC da altri cluster. Crea tali cluster utilizzando una configura zione di sicurezza diversa, specifica un KDC esterno e utilizza il nome del realm e la password di amministr atore KDC specificati per il KDC dedicato al cluster.
	KDC esterno	Disponibili solo in Amazon EMR versione 5.20.0 e successive. Specifica che i cluster che utilizzan o questa configurazione di sicurezza autenticano le entità Kerberos principali utilizzando un server KDC esterno al cluster. Non viene creato un KDC nel cluster. Quando crei il cluster, specifichi il nome del realm e la password di amministratore KDC per il KDC esterno.

Parametro		Descrizione
Durata del ticket		Facoltativo. Specifica il periodo di validità di un ticket Kerberos emesso dal KDC nei cluster che utilizzano questa configurazione di sicurezza. La durata dei ticket è limitata per motivi di sicurezza . Le applicazioni e i servizi del cluster rinnovano automaticamente i ticket dopo la loro scadenza. Gli utenti che si connettono al cluster tramite SSH utilizzan do le credenziali Kerberos devono eseguire `kinit` dalla linea di comando del nodo primario per rinnovare un ticket dopo la relativa scadenza.
Fiducia tra realm		Specifica una relazione di fiducia tra realm tra un KDC dedicato al cluster in cluster che utilizzano questa configurazione di sicurezza e un KDC in un altro realm Kerberos. Le entità principali (in genere gli utenti) di un altro realm vengono autenticate nei cluster che utilizzano questa configurazione. È necessaria una configura zione aggiuntiva nell'altro realm Kerberos. Per ulteriori informazioni, consulta Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory.
Proprietà di fiducia tra realm	Realm (Dominio)	Specifica il nome di realm Kerberos dell'altro realm della relazione di fiducia. Per convenzione, i nomi del realm Kerberos sono uguali al nome di dominio, ma utilizzano solo lettere maiuscole.
	Dominio	Specifica il nome di dominio dell'altro realm della relazione di fiducia.

Parametro		Descrizione
	Server amministratore	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server di amministrazione nell'altro realm della relazione di fiducia. Generalmente, il server amministratore e il server KDC vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma comunicano su porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 749, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :749`).
	Server KDC	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server KDC nell'altro realm della relazione di fiducia. Generalmente, il server KDC e il server amministratore vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma utilizzano porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 88, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :88`).
KDC esterno		Specifica che i cluster KDC esterni vengono utilizzati dal cluster.

Parametro			Descrizione
Proprietà del KDC esterno	Server amministratore		Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server amministratore esterno. Generalmente, il server amministratore e il server KDC vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma comunicano su porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 749, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :749`).
	Server KDC		Specifica il nome di dominio completo (FQDN) del server KDC esterno. Generalmente, il server KDC e il server amministratore vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma utilizzano porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 88, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :88`).
	Integrazione con Active Directory		Specifica che l'autenticazione dell'entità Kerberos principale è integrata con un dominio Microsoft Active Directory.
	Proprietà di integrazione con Active Directory	Realm di Active Directory	Specifica il nome del realm Kerberos del dominio Active Directory. Per convenzione, i nomi del realm Kerberos sono uguali al nome di dominio, ma utilizzano solo lettere maiuscole.
		Dominio Active Directory	Specifica il nome del dominio di Active Directory.

Parametro			Descrizione
		Server Active Directory	Specifica il nome di dominio completo (FQDN) del controller di dominio Microsoft Active Directory.

Configurazione delle impostazioni di Kerberos mediante la AWS CLI

La seguente tabella mostra i parametri JSON per le impostazioni di Kerberos in una configurazione di sicurezza. Per gli esempi di configurazione, consulta Esempi di configurazione.

Parametro	Descrizione
`"AuthenticationConfiguration": {`	Obbligatorio per Kerberos. Specifica che una configurazione di autentica zione fa parte di questa configura zione di sicurezza.
`"KerberosConfiguration": {`	Obbligatorio per Kerberos. Specifica le proprietà di configurazione Kerberos.
`"Provider": "ClusterD edicatedKdc",` —oppure— `"Provider: "ExternalKdc",`	*ClusterDedicatedKdc* specifica che Amazon EMR crea un KDC sul nodo primario di qualsiasi cluster che utilizza questa configura zione di sicurezza. Quando crei il cluster, puoi specificare il nome del realm e la password di amministr atore KDC. Se necessario, puoi fare riferimento a questo KDC da altri cluster. Crea tali cluster utilizzan do una configurazione di sicurezza diversa, specifica un KDC esterno e utilizza il nome del realm e la password di amministratore KDC che

Parametro	Descrizione
	hai specificato quando hai creato il cluster con il KDC dedicato al cluster. *ExternalKdc* specifica che il cluster utilizza un KDC esterno. Amazon EMR non crea un KDC sul nodo primario. Un cluster che utilizza questa configurazione di sicurezza deve specificare il nome del realm e la password di amministratore KDC del KDC esterno.
`"ClusterDedicatedKdcConfiguration": {`	Opzione obbligatoria quando specifichi *ClusterDedicatedKdc* .
`"TicketLifetimeInHours": 24,`	Facoltativo. Specifica il periodo di validità di un ticket Kerberos emesso dal KDC nei cluster che utilizzano questa configurazione di sicurezza. La durata dei ticket è limitata per motivi di sicurezza. Le applicazioni e i servizi del cluster rinnovano automaticamente i ticket dopo la loro scadenza. Gli utenti che si connettono al cluster tramite SSH utilizzando le credenziali Kerberos devono eseguire `kinit` dalla linea di comando del nodo primario per rinnovare un ticket dopo la relativa scadenza.

Parametro	Descrizione
`"CrossRealmTrustConfiguration": {`	Specifica una relazione di fiducia tra realm tra un KDC dedicato al cluster in cluster che utilizzano questa configurazione di sicurezza e un KDC in un altro realm Kerberos. Le entità principali (in genere gli utenti) di un altro realm vengono autenticate nei cluster che utilizzano questa configurazione. È necessari a una configurazione aggiuntiva nell'altro realm Kerberos. Per ulteriori informazioni, consulta Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory.
`"Realm": "KDC2.COM",`	Specifica il nome di realm Kerberos dell'altro realm della relazione di fiducia. Per convenzione, i nomi del realm Kerberos sono uguali al nome di dominio, ma utilizzano solo lettere maiuscole.
`"Domain": "kdc2.com",`	Specifica il nome di dominio dell'altro realm della relazione di fiducia.

Parametro	Descrizione
"AdminServer": "*kdc.com:749* ",	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server di amministrazione nell'altro realm della relazione di fiducia. Generalmente, il server amministratore e il server KDC vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma comunicano su porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 749, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com` *:749*).
"KdcServer": "*kdc.com:88* "	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server KDC nell'altro realm della relazione di fiducia. Generalmente, il server KDC e il server amministratore vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma utilizzano porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 88, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com` *:88*).

Parametro	Descrizione
}	
}	
"ExternalKdcConfiguration": {	Opzione obbligatoria quando specifichi *ExternalKdc* .
"TicketLifetimeInHours": *24*,	Facoltativo. Specifica il periodo di validità di un ticket Kerberos emesso dal KDC nei cluster che utilizzano questa configurazione di sicurezza. La durata dei ticket è limitata per motivi di sicurezza. Le applicazioni e i servizi del cluster rinnovano automaticamente i ticket dopo la loro scadenza. Gli utenti che si connettono al cluster tramite SSH utilizzando le credenziali Kerberos devono eseguire `kinit` dalla linea di comando del nodo primario per rinnovare un ticket dopo la relativa scadenza.
"KdcServerType": "Single",	Specifica che viene fatto riferimento a un singolo server KDC. Attualmente, `Single` è l'unico valore supportato.

Parametro				Descrizione
			"AdminServer": "*kdc.com:749* ",	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server amministratore esterno. Generalmente, il server amministratore e il server KDC vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma comunicano su porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 749, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com` *:749*).
			"KdcServer": "*kdc.com:88* ",	Specifica il nome di dominio completo (FQDN) del server KDC esterno. Generalmente, il server KDC e il server amministratore vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma utilizzano porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 88, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com` *:88*).

Parametro	Descrizione
"AdIntegrationConfiguration": {	Specifica che l'autenticazione dell'entità Kerberos principale è integrata con un dominio Microsoft Active Directory.
"AdRealm": "*AD.DOMAIN.COM* ",	Specifica il nome del realm Kerberos del dominio Active Directory. Per convenzione, i nomi del realm Kerberos sono uguali al nome di dominio, ma utilizzano solo lettere maiuscole.
"AdDomain": "*ad.domain.com* "	Specifica il nome del dominio di Active Directory.
"AdServer": "*ad.domain.com* "	Specifica il nome di dominio completo (FQDN) del controller di dominio Microsoft Active Directory.
}	
}	
}	
}	

Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3

I ruoli IAM per EMRFS ti consentono di fornire differenti autorizzazioni per dati EMRFS in Amazon S3. A questo proposito, crei mappature che specificano un ruolo IAM utilizzato per le autorizzazioni quando una richiesta di accesso contiene un identificatore da te specificato. L'identificatore può essere un ruolo o un utente Hadoop oppure un prefisso Amazon S3.

Per ulteriori informazioni, consulta Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3.

Specifica dei ruoli IAM per EMRFS mediante la AWS CLI

Di seguito è riportato un esempio di frammento JSON per specificare ruoli IAM personalizzati per EMRFS all'interno di una configurazione di sicurezza. Vengono illustrate le mappature dei ruoli per i tre diversi tipi di identificatori, seguite da un riferimento ai parametri.

```
{
  "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" ]
      }]
    }
  }
}
```

Parametro	Descrizione
`"AuthorizationConfiguration":`	Campo obbligatorio.
`"EmrFsConfiguration":`	Campo obbligatorio. Contiene mappature dei ruoli.
`"RoleMappings":`	Campo obbligatorio. Contiene una o più definizioni di mappatura dei ruoli. Le mappature dei ruoli vengono valutate dall'alto verso il basso nell'ordine in cui vengono visualizzate. Se una mappatura dei ruoli viene valutata come true (vera) per una chiamata EMRFS per i dati in Amazon S3, non vengono valutate altre mappature dei ruoli ed EMRFS utilizza il ruolo

Parametro	Descrizione
	IAM specificato per la richiesta. Le mappature dei ruoli sono costituite dai parametri obbligatori seguenti:
`"Role":`	Specifica l'identificatore ARN di un ruolo IAM nel formato `arn:aws:iam::` *`account-id`* `:role/`*`role-name`* . Questo è il ruolo IAM che Amazon EMR assume se la richiesta EMRFS ad Amazon S3 corrisponde a uno degli `Identifiers` specificato.
`"IdentifierType":`	Il valore può essere uno dei seguenti: • `"User"` specifica che gli identificatori sono uno o più utenti Hadoop, i quali possono essere utenti di account Linux o entità Kerberos. Quando la richiesta EMRFS viene originata dall'utente o dagli utenti specificati, viene assunto il ruolo IAM. • `"Prefix"` specifica che l'identificatore è un percorso Amazon S3. Il ruolo IAM viene assunto per le chiamate alla posizione o alle posizioni con i prefissi specificati. Ad esempio, il prefisso `s3://mybucket/` corrisponde a `s3://mybucket/mydir` e `s3://mybucket/yetanotherdir` . • `"Group"` specifica che gli identificatori sono uno o più Gruppi Hadoop. Il ruolo IAM viene assunto se la richiesta proviene da un utente nel gruppo o nei gruppi specificati.
`"Identifiers":`	Specifica uno o più identificatori del tipo di identificatore appropriato. Separa più identificatori con virgole senza spazi.

Configurazione di richieste del servizio di metadati per istanze Amazon EC2

I metadati dell'istanza sono dati relativi all'istanza che puoi utilizzare per configurare o gestire un'istanza in esecuzione. Puoi accedere ai metadati dell'istanza da un'istanza in esecuzione utilizzando uno dei metodi seguenti:

- Servizio di metadati dell'istanza Versione 1 (IMDSv1): un metodo di richiesta/risposta
- Servizio di metadati dell'istanza Versione 2 (IMDSv2): un metodo orientato alla sessione

Mentre Amazon EC2 supporta sia IMDSv1 che IMDSv2, Amazon EMR supporta IMDSv2 in Amazon EMR 5.23.1, 5.27.1, 5.32 o versioni successive e 6.2 o versioni successive. In queste versioni, i componenti di Amazon EMR utilizzano IMDSv2 per tutte le chiamate IMDS. Per le chiamate IMDS nel codice dell'applicazione, è possibile utilizzare sia IMDSv1 che IMDSv2 oppure configurare IMDS per utilizzare solo IMDSv2 per una maggiore sicurezza. Quando specifichi l'utilizzo di IMDSv2, IMDSv1 non funziona più.

Per ulteriori informazioni, consulta Configurazione del servizio di metadati dell'istanza nella Guida per l'utente di Amazon EC2 per le istanze Linux.

> **Note**
>
> Nelle versioni precedenti di Amazon EMR 5.x o 6.x, la disattivazione di IMDSv1 causa un errore di avvio del cluster poiché i componenti di Amazon EMR utilizzano IMDSv1 per tutte le chiamate IMDS. Quando si disattiva IMDSv1, assicurati che qualsiasi software personalizzato che utilizza IMDSv1 venga aggiornato a IMDSv2.

Specifica della configurazione del servizio di metadati dell'istanza utilizzando la AWS CLI

Di seguito è riportato un esempio di frammento JSON per specificare il servizio di metadati dell'istanza Amazon EC2 (IMDS) all'interno di una configurazione di sicurezza.

```
{
  "InstanceMetadataServiceConfiguration" : {
      "MinimumInstanceMetadataServiceVersion": integer,
      "HttpPutResponseHopLimit": integer
   }
}
```

Parametro	Descrizione
`"InstanceMetadataServiceConfiguration":`	Campo obbligatorio.
`"MinimumInstanceMetadataServiceVersion":`	Campo obbligatorio. Specificare 1 o 2. Un valore di 1 consente IMDSv1 e IMDSv2. Un valore di 2 consente solo IMDSv2.
`"HttpPutResponseHopLimit":`	Campo obbligatorio. Il limite di hop della risposta HTTP PUT per le richieste di metadati dell'istanza. Maggiore è il numero, più è lungo il tragitto che le richieste di metadati dell'istanza possono percorrere. Default: 1. Specifica un numero intero da 1 a 64.

Specifica della configurazione del servizio di metadati dell'istanza utilizzando la console

Puoi configurare l'utilizzo di IMDS per un cluster quando lo avvii dalla console di Amazon EMR. Controlli delle configurazioni di sicurezza IMDS nella console di Amazon EMR

Per configurare l'utilizzo di IMDS tramite la console:

1. Durante la creazione di una nuova configurazione di sicurezza, nella pagina Security configurations (Configurazioni di sicurezza), seleziona Configure EC2 Instance metadata service (Configura servizio di metadati dell'istanza EC2) sotto l'impostazione EC2 Instance Metadata Service (Servizio di metadati dell'istanza EC2). Questa configurazione è supportata solo in Amazon EMR 5.23.1, 5.27.1, 5.32 o versioni successive e 6.2 o versioni successive.
2. Per l'opzione Minimum Instance Metadata Service Version (Versione del servizio di metadati dell'istanza minima), seleziona una delle seguenti opzioni:
   - Turn off IMDSv1 and only allow IMDSv2 (Disattiva IMDSv1 e consenti solo IMDSv2) se desideri consentire solo IMDSv2 in questo cluster. Consulta Passaggio all'utilizzo del servizio di metadati dell'istanza versione 2 nella Guida per l'utente di Amazon EC2 per le istanze Linux.
   - Allow both IMDSv1 and IMDSv2 on cluster (Consenti sia IMDSv1 che IMDSv2 sul cluster) se desideri consentire IMDSv1 e IMDSv2 orientato alla sessione su questo cluster.

3. Per IMDSv2, è inoltre possibile configurare il numero consentito di hop di rete per il token dei metadati impostando HTTP put response hop limit (Limite di hop di risposta HTTP put) su un numero intero compreso tra 1 e 64.

Per ulteriori informazioni, consulta Configurazione del servizio di metadati dell'istanza nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Consulta Configurazione dei dettagli dell'istanza e Configurazione del servizio di metadati dell'istanza nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Impostazione di una configurazione di sicurezza per un cluster

Puoi definire impostazioni di crittografia quando crei un cluster specificando la configurazione di sicurezza. Puoi utilizzare la AWS Management Console o l'AWS CLI.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Specifica di una configurazione di sicurezza con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Security configuration and permissions (Configurazione e autorizzazioni di sicurezza), cerca il campo Security configuration (Configurazione di sicurezza). Seleziona il menu a discesa o scegli Browse (Sfoglia) per selezionare il nome di una configurazione di sicurezza che hai creato in precedenza. In alternativa, scegli Create security configuration (Crea configurazione di sicurezza) per creare una configurazione da utilizzare per il cluster.
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica di una configurazione di sicurezza con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Nella schermata Step 1: Software and Steps (Fase 1: software e fasi), dall'elenco Release (Rilascio), scegli emr-4.8.0 o una versione più recente. Scegliere le impostazioni desiderate e quindi Next (Avanti).
4. Nella schermata Step 2: Hardware (Fase 2: hardware), scegliere le impostazioni desiderate e quindi Next (Avanti). Eseguire la stessa operazione per Step 3: General Cluster Settings (Fase 3: impostazioni generali per cluster).
5. Nella schermata Step 4: Security (Fase 4: sicurezza), in Encryption Options (Opzioni di crittografia), scegliere un valore per Security configuration (Configurazione di sicurezza).
6. Configurare altre opzioni di sicurezza come desiderato e scegliere Crea cluster (Crea cluster).

CLI

Specifica di una configurazione di sicurezza con la AWS CLI

- Utilizza `aws emr create-cluster` per applicare facoltativamente una configurazione di sicurezza con `--security-configuration` *`MySecConfig`*, dove *`MySecConfig`* è il nome della configurazione di sicurezza, come mostrato nell'esempio seguente. Il valore `--release-label` specificato deve essere 4.8.0 o successivo e il parametro `--instance-type` può essere qualsiasi tipo disponibile.

```
aws emr create-cluster --instance-type m5.xlarge --release-label emr-5.0.0 --
security-configuration mySecConfig
```

Protezione dei dati in Amazon EMR

Ulteriori informazioni su come il modello di responsabilità condivisa AWS si applica alla protezione dei dati in Amazon EMR. Come descritto in questo modello, AWS è responsabile della protezione dell'infrastruttura globale che esegue l'intero cloud AWS. L'utente è responsabile di mantenere il controllo sui contenuti ospitati su questa infrastruttura. Questo contenuto include la configurazione della protezione e le attività di gestione per i servizi AWS utilizzati. Per ulteriori informazioni sulla

privacy dei dati, consulta Domande frequenti sulla privacy dei dati. Per ulteriori informazioni sulla protezione dei dati, consulta il post the Amazon shared responsibility model and GDPR (Modello di responsabilità condivisa di Amazon e GDPR) del Blog sulla sicurezza AWS.

Per garantire la protezione dei dati, si consiglia di proteggere le credenziali dell'account AWS e di configurare singoli account con AWS Identity and Access Management. In questo modo, a ogni utente verranno assegnate solo le autorizzazioni necessarie per svolgere il proprio lavoro. Ti suggeriamo, inoltre, di proteggere i dati nei seguenti modi:

- Utilizza l'autenticazione a più fattori (MFA) con ogni account.
- Utilizza TLS per comunicare con risorse AWS. È richiesto TLS 1.2.
- Configura la creazione di logo delle attività di API e utenti con AWS CloudTrail.
- Utilizza le soluzioni di crittografia AWS, insieme a tutti i controlli di sicurezza predefiniti all'interno dei servizi AWS.
- Utilizza i servizi di sicurezza gestiti avanzati, ad esempio Amazon Macie, che aiutano a individuare e proteggere i dati personali archiviati in Amazon S3.
- Se si richiedono moduli crittografici convalidati FIPS 140-2 quando si accede ad AWS tramite una CLI o un'API, utilizzare un endpoint FIPS. Per ulteriori informazioni sugli endpoint FIPS disponibili, consulta il Federal Information Processing Standard (FIPS) 140-2.

Consigliamo di non inserire mai informazioni identificative sensibili, ad esempio i numeri di account dei clienti, in campi a formato libero come un campo Nome. Questo include il lavoro con Amazon EMR o altri servizi AWS tramite la console, l'API, la AWS CLI o gli SDK AWS. Gli eventuali dati immessi in Amazon EMR o altri servizi potrebbero essere prelevati per l'inserimento nei registri di diagnostica. Quando fornisci un URL a un server esterno, non includere informazioni sulle credenziali nell'URL per convalidare la tua richiesta a tale server.

Crittografia dei dati a riposo e in transito

La crittografia dei dati consente di impedire agli utenti non autorizzati di leggere dati su un cluster e sui sistemi di archiviazione di dati associati. Sono inclusi i dati salvati su supporti persistenti, noti come dati inattivi, e i dati che possono essere intercettati quando circolano in rete, noti come dati in transito.

A partire dalla versione Amazon EMR 4.8.0, puoi utilizzare le configurazioni di sicurezza di Amazon EMR per configurare più facilmente impostazioni di crittografia di dati per cluster. Le configurazioni di

sicurezza offrono impostazioni che assicurano la sicurezza dei dati in transito e a riposo nei volumi di Amazon Elastic Block Store (Amazon EBS) e in EMRFS su Amazon S3.

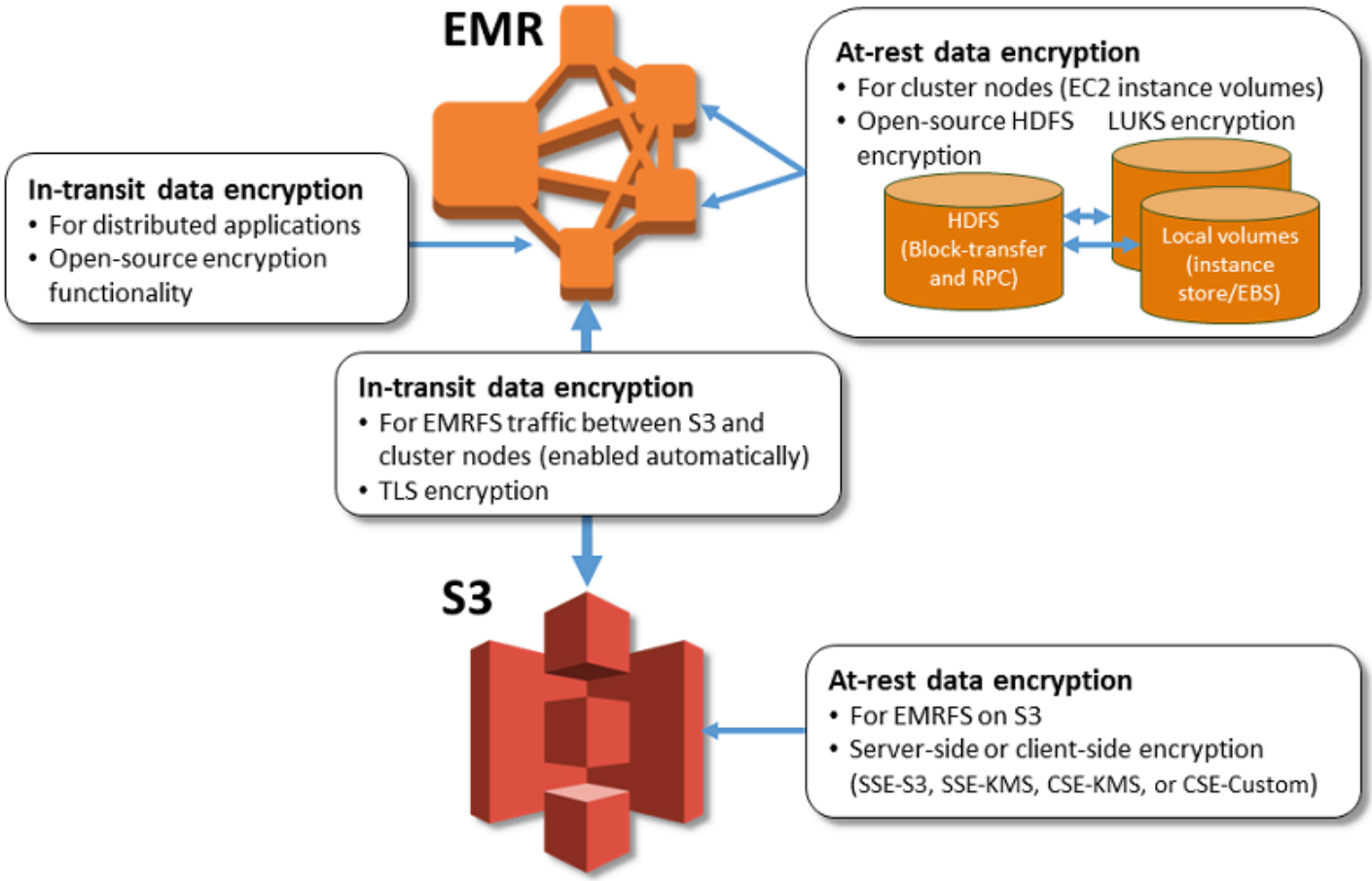
Facoltativamente, a partire da Amazon EMR versione 4.1.0 e successive, puoi scegliere di configurare la crittografia trasparente in HDFS, che non è configurata utilizzando le configurazioni di sicurezza. Per ulteriori informazioni, consulta Crittografia trasparente in HDFS su Amazon EMR nella Guida ai rilasci di Amazon EMR.

Argomenti

Opzioni di crittografia

Con Amazon EMR versione 4.8.0 e versioni successive, puoi utilizzare una configurazione di sicurezza per specificare impostazioni di crittografia di dati a riposo, in transito o entrambi. Quando abiliti la crittografia dei dati a riposo, puoi scegliere di crittografare i dati EMRFS in Amazon S3, i dati in dischi locali o entrambi. Ogni configurazione di sicurezza creata viene archiviata in Amazon EMR anziché nella configurazione del cluster. Di conseguenza, puoi facilmente riutilizzare una configurazione per specificare impostazioni di crittografia dei dati ogni volta che crei un cluster. Per ulteriori informazioni, consulta Creazione di una configurazione di sicurezza.

Il diagramma seguente mostra le differenti opzioni di crittografia dei dati disponibili con le configurazioni di sicurezza.



Anche le seguenti opzioni di crittografia sono disponibili e non sono configurate utilizzando una configurazione di sicurezza:

- Facoltativamente, con Amazon EMR versione 4.1.0 e versioni successive, puoi scegliere di configurare la crittografia trasparente in HDFS. Per ulteriori informazioni, consulta Crittografia trasparente in HDFS su Amazon EMR nella Guida ai rilasci di Amazon EMR.
- Se utilizzi una versione di Amazon EMR che non supporta configurazioni di sicurezza, puoi configurare manualmente la crittografia per i dati EMRFS in Amazon S3. Per ulteriori informazioni, consulta Specifica della crittografia Amazon S3 utilizzando le proprietà EMRFS.
- Se utilizzi una versione Amazon EMR precedente alla 5.24.0, un volume del dispositivo di root EBS crittografato è supportato solo quando utilizzi un'AMI personalizzata. Per ulteriori informazioni, consulta la sezione Creazione di un'AMI personalizzata con un volume del dispositivo di root Amazon EBS crittografato nella Guida alla gestione di Amazon EMR.

> **Note**
>
> A partire dalla versione 5.24.0 di Amazon EMR, puoi utilizzare un'opzione di configurazione di sicurezza per crittografare il dispositivo di root EBS e i volumi di archiviazione quando specifichi AWS KMS come provider di chiavi. Per ulteriori informazioni, consulta Crittografia del disco locale.

La crittografia dei dati richiede chiavi e certificati. Una configurazione di sicurezza offre la flessibilità necessaria per scegliere tra diverse opzioni, tra cui chiavi gestite da AWS Key Management Service, chiavi gestite da Amazon S3 e chiavi e certificati di provider personalizzati che specifichi. L'uso di AWS KMS come provider di chiavi comporta l'addebito dei costi relativi all'archiviazione e all'uso delle chiavi di crittografia. Per ulteriori informazioni, consulta Prezzi di AWS KMS.

Prima di specificare le opzioni di crittografia, scegli i sistemi di gestione di chiavi e certificati che intendi utilizzare, in modo da cominciare a creare chiavi e certificati o i provider personalizzati che specifichi durante l'impostazione dei parametri di crittografia.

Crittografia dei dati a riposo per dati EMRFS in Amazon S3

La crittografia di Amazon S3 funziona con gli oggetti del file system EMR (EMRFS) letti da e scritti su Amazon S3. Puoi specificare la crittografia lato server (SSE) o la crittografia lato client (CSE) di Amazon S3 come modalità di crittografia predefinita quando abiliti la crittografia dei dati a riposo. Facoltativamente, è possibile specificare diversi metodi di crittografia per singoli bucket utilizzando override di crittografia per bucket. Indipendentemente dall'abilitazione o meno della crittografia di Amazon S3, il protocollo Transport Layer Security (TLS) esegue la crittografia degli oggetti EMRFS in transito tra i nodi cluster EMR e Amazon S3. Per informazioni dettagliate sulla crittografia Amazon S3, consulta Protezione dei dati con la crittografia nella Guida per l'utente di Amazon Simple Storage Service.

> **Note**
>
> L'uso di AWS KMS comporta l'applicazione di costi per l'archiviazione e l'uso delle chiavi di crittografia. Per ulteriori informazioni, consultare Prezzi di AWS KMS.

Crittografia lato server di Amazon S3

Quando configuri la crittografia lato server Amazon S3, Amazon S3 esegue la crittografia dei dati a livello di oggetto, tramite la scrittura dei dati su disco, e ne esegue la decrittografia al momento dell'accesso. Per ulteriori informazioni sulla SEE, consulta Protezione dei dati con la crittografia lato server nella Guida per l'utente di Amazon Simple Storage Service.

Puoi scegliere tra due diversi sistemi di gestione delle chiavi quando specifichi la SSE in Amazon EMR:

- SSE-S3: Amazon S3 gestisce le chiavi per te.
- SSE-KMS: si utilizza una AWS KMS key configurata con policy adatte per Amazon EMR. Per ulteriori informazioni sui requisiti chiave per Amazon EMR, consulta Utilizzo di AWS KMS keys per crittografia.

La SSE con chiavi fornite dal cliente (SSE-C) non è disponibile per l'utilizzo con Amazon EMR.

File crittografato lato client Amazon S3

Con la crittografia lato client di Amazon S3, la crittografia e la decrittografia Amazon S3 avvengono nel client EMRFS nel tuo cluster. Gli oggetti vengono crittografati prima di essere caricati su Amazon S3 e decrittati dopo il loro download. Il provider specificato fornisce la chiave di crittografia utilizzata dal client. Il client può utilizzare le chiavi fornite da AWS KMS (CSE-KMS) o una classe Java personalizzata che fornisce la chiave principale lato client (CSE-C). Le specifiche di crittografia sono leggermente diverse tra CSE-KMS e CSE-C, a seconda del provider specificato e dei metadati dell'oggetto da decrittare o crittografare. Per ulteriori informazioni su queste differenze, consulta Protezione dei dati tramite la crittografia lato client nella Guida per l'utente di Amazon Simple Storage Service.

> **Note**
>
> Amazon S3 CSE garantisce solo che i dati EMRFS scambiati con Amazon S3 siano crittografati; non tutti i dati sui volumi delle istanze del cluster sono crittografati. Inoltre, poiché Hue non utilizza EMRFS, gli oggetti che il browser di file Hue S3 scrive su Amazon S3 non vengono crittografati.

Crittografia del disco locale

I seguenti meccanismi interagiscono per crittografare i dischi locali quando abiliti la crittografia dei dischi locali utilizzando una configurazione Amazon EMR di sicurezza.

Crittografia HDFS open source

HDFS scambia i dati tra le istanze del cluster durante l'elaborazione distribuita. Inoltre, legge e scrive i dati nei volumi instance store e nei volumi EBS collegati alle istanze. Le seguenti opzioni di crittografia Hadoop open source sono attivate quando abiliti la crittografia per dischi locali:

- Secure Hadoop RPC (RPC Hadoop protetto) è impostata su `Privacy`, che utilizza Simple Authentication Security Layer (SASL).
- Data encryption on HDFS block data transfer (Crittografia di dati su trasferimento di dati di blocco HDFS) è impostata su `true` ed è configurata per utilizzare la crittografia AES 256.

> **Note**
>
> Puoi attivare una crittografia Apache Hadoop supplementare abilitando la crittografia in transito. Per ulteriori informazioni, consulta Crittografia in transito. Queste impostazioni di crittografia non attivano la crittografia trasparente HDFS, che puoi configurare manualmente. Per ulteriori informazioni, consulta Crittografia trasparente in HDFS su Amazon EMR nella Guida ai rilasci di Amazon EMR.

Crittografia dell'archivio istanza

Per i tipi di istanza EC2 che utilizzano SSD basati su NVMe come volume di archivio istanza, la crittografia NVMe viene utilizzata indipendentemente dalle impostazioni di crittografia Amazon EMR. Per ulteriori informazioni, consulta Volumi SSD NVMe nella Guida per l'utente di Amazon EC2 per le istanze Linux. Per altri volumi di archivio istanza, Amazon EMR utilizza LUKS per crittografare il volume di archivio istanza quando la crittografia su disco locale è abilitata, indipendentemente dal fatto che i volumi EBS siano crittografati utilizzando la crittografia EBS o LUKS.

Crittografia dei volumi EBS

Se crei un cluster in una regione in cui la crittografia Amazon EC2 dei volumi EBS è abilitata per impostazione predefinita per il tuo account, i volumi EBS vengono crittografati anche se la

crittografia su disco locale non è abilitata. Per ulteriori informazioni, consulta la sezione Crittografia per impostazione predefinita nella Guida per l'utente di Amazon EC2 per le istanze Linux. Con la crittografia dei dischi locali abilitata in una configurazione di sicurezza, le impostazioni Amazon EMR hanno la precedenza sulle impostazioni di crittografia Amazon EC2 per impostazione predefinita per le istanze EC2 del cluster.

Sono disponibili le seguenti opzioni per crittografare i volumi EBS utilizzando una configurazione di sicurezza:

- Crittografia EBS: a partire da Amazon EMR versione 5.24.0, puoi scegliere di abilitare la crittografia EBS. L'opzione di crittografia EBS crittografa il volume dispositivo root EBS e i volumi di storage collegati. L'opzione di crittografia EBS è disponibile solo quando specifichi AWS Key Management Service come provider di chiavi. Ti consigliamo di utilizzare la crittografia EBS.
- Crittografia LUKS: se scegli di utilizzare la crittografia LUKS per i volumi Amazon EBS, la crittografia LUKS si applica solo ai volumi di archiviazione collegati, non al volume del dispositivo di root. Per ulteriori informazioni sulla crittografia LUKS, vedi la specifica di LUKS su disco.

  Per il tuo provider di chiavi, puoi utilizzare una AWS KMS key configurata con policy appropriate per Amazon EMR oppure una classe Java personalizzata che fornisce artefatti di crittografia. L'uso di AWS KMS comporta l'applicazione di costi per l'archiviazione e l'uso delle chiavi di crittografia. Per ulteriori informazioni, consulta Prezzi di AWS KMS.

> **Note**
>
> Per verificare se la crittografia EBS è abilitata sul cluster, è consigliabile utilizzare la chiamata API `DescribeVolumes`. Per ulteriori informazioni, consultare DescribeVolumes. L'esecuzione di `lsblk` sul cluster verificherà solo lo stato della crittografia LUKS anziché la crittografia EBS.

Crittografia in transito

Diversi meccanismi di crittografia sono abilitati con la crittografia in transito. Si tratta di funzionalità open source specifiche dell'applicazione che possono variare a seconda della versione Amazon EMR. Le seguenti caratteristiche di crittografia specifiche dell'applicazione possono essere abilitate utilizzando le configurazioni dell'applicazione Apache. Per ulteriori informazioni, consulta la sezione Configurazione delle applicazioni.

Hadoop

- Hadoop MapReduce Encrypted Shuffle (Shuffle crittografato MapReduce Hadoop) utilizza TLS.
- Secure Hadoop RPC (RPC Hadoop protetto) è impostata su "Privacy" e utilizza SASL (attivato in Amazon EMR quando la crittografia di dati a riposo è abilitata).
- Data encryption on HDFS block data transfer (Crittografia di dati su trasferimento di dati di blocco HDFS) utilizza AES 256 (attivata in Amazon EMR quando la crittografia di dati a riposo è abilitata nella configurazione di sicurezza).
- Per ulteriori informazioni, consulta Hadoop in secure mode (Hadoop in modalità protetta) nella documentazione di Apache Hadoop.

HBase

- Quando Kerberos è abilitato, la proprietà `hbase.rpc.protection` è impostata su `privacy` per la comunicazione crittografata.
- Per ulteriori informazioni, consulta la sezione Client-side configuration for secure operation (Configurazione lato client per funzionamento sicuro) nella documentazione di Apache HBase.
- Per ulteriori informazioni su Kerberos con Amazon EMR, consulta Utilizzo di Kerberos per l'autenticazione con Amazon EMR.

Hive

- La comunicazione del client JDBC/ODBC con HiveServer2 (HS2) è crittografata mediante configurazioni SSL nel rilascio di Amazon EMR 6.9.0 e successivi.
- Per ulteriori informazioni, consulta la sezione SSL encryption (Crittografia SSL) della documentazione di Apache Hive.

Spark

- Le comunicazioni RPC interne tra componenti Spark, ad esempio il servizio di trasferimento dei blocchi e il servizio shuffle esterno, sono crittografate utilizzando la cifratura AES-256 in Amazon EMR versione 5.9.0 e versioni successive. Nelle versioni precedenti, le comunicazioni RPC interne sono crittografate utilizzando la cifratura SASL con DIGEST-MD5.

- Le comunicazioni HTTP con interfacce utente come Spark History Server e file server HTTPS sono crittografate mediante la configurazione SSL di Spark. Per ulteriori informazioni, consulta SSL Configuration (Configurazione SSL) nella documentazione Spark.
- Per ulteriori informazioni, consulta la sezione Spark release notes (Note di rilascio di Spark) nella documentazione di Apache Spark.

Tez

- Tez Shuffle Handler (Gestore shuffle Tez) utilizza il protocollo TLS (`tez.runtime.ssl.enable`).

Presto

- Le comunicazioni interne tra nodi Presto utilizzano SSL/TLS (solo Amazon EMR versione 5.6.0 e versioni successive).

Puoi specificare gli artifact di crittografia utilizzati per la crittografia di dati in transito in due modi: fornendo un file zippato di certificati che hai caricato in Amazon S3 oppure facendo riferimento a una classe Java personalizzata che fornisce artifact di crittografia. Per ulteriori informazioni, consulta Fornitura di certificati per la crittografia di dati in transito con Amazon EMR.

Creazione di chiavi e certificati per la crittografia di dati

Prima di specificare le opzioni di crittografia utilizzando una configurazione di sicurezza, scegli il provider che desideri utilizzare per le chiavi e gli artefatti di crittografia. Ad esempio, puoi usare AWS KMS o un provider personalizzato che hai creato. Quindi, crea le chiavi o il provider di chiavi come descritto in questa sezione.

Fornitura di chiavi per la crittografia di dati a riposo con Amazon EMR

Puoi utilizzare AWS Key Management Service (AWS KMS) o un provider di chiavi personalizzato per la crittografia di dati a riposo in Amazon EMR. L'uso di AWS KMS comporta l'applicazione di costi per l'archiviazione e l'uso delle chiavi di crittografia. Per ulteriori informazioni, consulta Prezzi di AWS KMS.

Questo argomento fornisce dettagli sulla policy delle chiavi per una chiave KMS da utilizzare con Amazon EMR, nonché linee guida ed esempi di codice per scrivere una classe di provider di chiavi

personalizzato per la crittografia di Amazon S3. Per ulteriori informazioni sulla creazione di chiavi, consulta Creazione di chiavi nella Guida per gli sviluppatori di AWS Key Management Service.

Utilizzo di AWS KMS keys per la crittografia

La chiave di crittografia AWS KMS deve essere stata creata nella stessa Regione dell'istanza cluster Amazon EMR e dei bucket Amazon S3 utilizzati con EMRFS. Se la chiave specificata si trova in un account diverso da quello utilizzato per configurare un cluster, è necessario specificare la chiave utilizzando il relativo ARN.

Il ruolo del profilo dell'istanza Amazon EC2 deve disporre delle autorizzazioni per utilizzare la chiave KMS specificata. Il ruolo predefinito per il profilo dell'istanza in Amazon EMR è `EMR_EC2_DefaultRole`. Se utilizzi un ruolo diverso per il profilo dell'istanza o utilizzi ruoli IAM per le richieste EMRFS ad Amazon S3, assicurati che ogni ruolo venga aggiunto come utente chiave a seconda dei casi. Ciò concede al ruolo l'autorizzazione a utilizzare la chiave KMS. Per ulteriori informazioni, consulta Utilizzo di policy delle chiavi nella Guida per gli sviluppatori di AWS Key Management Service e Configurazione dei ruoli IAM per richieste EMRFS ad Amazon S3.

Puoi utilizzare la AWS Management Console per aggiungere il tuo profilo dell'istanza o il profilo dell'istanza EC2 all'elenco degli utenti delle chiavi per la chiave KMS specificata oppure puoi utilizzare la AWS CLI o un SDK AWS per collegare una policy delle chiavi appropriata.

Nota che Amazon EMR supporta solo Chiavi KMS simmetriche. Non è possibile utilizzare una chiave KMS asimmetrica per crittografare i dati a riposo in un cluster Amazon EMR. Per informazioni su come determinare se una chiave KMS è simmetrica o asimmetrica, consulta Identificazione di chiavi KMS simmetriche e asimmetriche.

La procedura seguente descrive come aggiungere il profilo dell'istanza Amazon EMR predefinito, `EMR_EC2_DefaultRole` come utente di chiavi utilizzando la AWS Management Console. Presuppone che tu abbia già creato una chiave KMS. Per creare una nuova chiave KMS, consulta Creazione di chiavi nella Guida per gli sviluppatori di AWS Key Management Service.

Aggiunta del profilo dell'istanza EC2 per Amazon EMR all'elenco degli utenti della chiave di crittografia

1. Accedi alla AWS Management Console e apri la console AWS Key Management Service (AWS KMS) all'indirizzo https://console.aws.amazon.com/kms.
2. Per modificare la Regione AWS, utilizza il Selettore di regione nell'angolo in alto a destra della pagina.

3. Seleziona l'alias della chiave KMS da modificare.
4. Nella pagina dei dettagli della chiave, in Key Users (Utenti di chiavi), scegli Add (Aggiungi).
5. Nella finestra di dialogo Add key users (Aggiungi utenti chiave) selezionare il ruolo appropriato. Il nome del ruolo di default è `EMR_EC2_DefaultRole`.
6. Scegliere Add (Aggiungi).

Abilitazione della crittografia EBS fornendo autorizzazioni aggiuntive per le chiavi KMS

A partire da Amazon EMR versione 5.24.0, puoi crittografare il dispositivo di root EBS e i volumi di archiviazione utilizzando un'opzione di configurazione di sicurezza. Per abilitare tale opzione, è necessario specificare AWS KMS come provider di chiavi. Inoltre, è necessario concedere al ruolo del servizio EMR `EMR_DefaultRole` le autorizzazioni per utilizzare la AWS KMS key specificata.

Puoi utilizzare la AWS Management Console per aggiungere il ruolo del servizio EMR all'elenco degli utenti delle chiavi per la chiave KMS specificata oppure puoi utilizzare la AWS CLI o un SDK AWS per collegare una policy delle chiavi appropriata.

La procedura seguente descrive come aggiungere il profilo dell'istanza EMR predefinita, `EMR_DefaultRole` come utente di chiavi utilizzando la AWS Management Console. Presuppone che tu abbia già creato una chiave KMS. Per creare una nuova chiave KMS, consulta Creazione di chiavi nella Guida per gli sviluppatori di AWS Key Management Service.

Per aggiungere il ruolo del servizio EMR all'elenco degli utenti delle chiavi di crittografia

1. Accedi alla AWS Management Console e apri la console AWS Key Management Service (AWS KMS) all'indirizzo https://console.aws.amazon.com/kms.
2. Per modificare la Regione AWS, utilizza il Selettore di regione nell'angolo in alto a destra della pagina.
3. Nella barra laterale sinistra, scegli Customer managed keys (Chiavi gestite dal cliente).
4. Seleziona l'alias della chiave KMS da modificare.
5. Nella pagina dei dettagli della chiave, in Key Users (Utenti di chiavi), scegli Add (Aggiungi).
6. Nella finestra di dialogo Add key users (Aggiungi utenti chiave) selezionare il ruolo appropriato. Il nome del ruolo del servizio EMR predefinito è `EMR_DefaultRole`.
7. Scegliere Add (Aggiungi).

Creazione di un provider di chiavi personalizzato

Quando utilizzi una configurazione di sicurezza, devi specificare un nome di classe di provider differente per la crittografia per dischi locali e la crittografia di Amazon S3.

Quando si crea un provider di chiavi personalizzato, l'applicazione dovrebbe implementare l'interfaccia EncryptionMaterialsProvider, che è disponibile in AWS SDK for Java 1.11.0 e versioni successive. L'implementazione può utilizzare qualsiasi strategia per fornire materiali di crittografia. È possibile, ad esempio, scegliere di fornire materiali di crittografia statici o di integrare un sistema di gestione delle chiavi più complesso.

L'algoritmo di crittografia utilizzato per i materiali di crittografia personalizzati deve essere AES/GCM/NoPadding.

La classe EncryptionMaterialsProvider ottiene materiali di crittografia suddivisi in base al contesto. Amazon EMR popola le informazioni sul contesto di crittografia al runtime per aiutare il chiamante a determinare i materiali di crittografia corretti da restituire.

Example Esempio: utilizzo di un provider di chiavi personalizzato per la crittografia Amazon S3 con EMRFS

Quando Amazon EMR recupera i materiali di crittografia dalla classe EncryptionMaterialsProvider per eseguire la crittografia, EMRFS inserisce facoltativamente l'argomento materialsDescription con due campi: l'URI Amazon S3 per l'oggetto e il JobFlowId del cluster, che può essere utilizzato dalla classe EncryptionMaterialsProvider per restituire i materiali di crittografia in modo selettivo.

Ad esempio, il provider potrebbe restituire chiavi diverse per diversi prefissi URI di Amazon S3. Sarà la descrizione dei materiali di crittografia restituiti a essere memorizzata con l'oggetto Amazon S3 anziché il valore materialsDescription generato da EMRFS e passato al provider. Durante la decrittografia di un oggetto Amazon S3, la descrizione dei materiali di crittografia viene passata alla classe EncryptionMaterialsProvider in modo che possa, ancora una volta, restituire selettivamente la chiave corrispondente per decrittare l'oggetto.

Di seguito viene fornita un'implementazione di riferimento di EncryptionMaterialsProvider. Un altro provider personalizzato, EMRFSRSAEncryptionMaterialsProvider, è disponibile da 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;
  }
```

```
}
```

Fornitura di certificati per la crittografia di dati in transito con Amazon EMR

Con Amazon EMR versione 4.8.0 o successive, sono disponibili due opzioni per specificare artifact di crittografia dei dati in transito utilizzando una configurazione di sicurezza:

- Puoi creare manualmente certificati PEM, includerli in un file .zip e quindi fare riferimento al file .zip in Amazon S3.
- Puoi implementare un provider di certificati personalizzato come classe Java, specificare il file JAR dell'applicazione in Amazon S3 e infine fornire il nome di classe completo del provider come dichiarato nell'applicazione. La classe deve implementare l'interfaccia TLSArtifactsProvider disponibile a partire dalla versione 1.11.0 di AWS SDK for Java.

Amazon EMR scarica automaticamente artifact in ogni nodo nel cluster e successivamente li utilizza per implementare le funzionalità di crittografia in transito open source. Per ulteriori informazioni sulle opzioni disponibili, consulta Crittografia in transito.

Utilizzo di certificati PEM

Quando specifichi un file ZIP per la crittografia in transito, per la configurazione di sicurezza i file PEM nel file ZIP devono essere denominati esattamente come illustrato di seguito:

Certificati di crittografia in transito

Nome file	Obbligatorio/facoltativo	Informazioni
privateKey.pem	Obbligatorio	Chiave privata
certificateChain.pem	Obbligatorio	Catena di certificati
trustedCertificates.pem	Facoltativo	Obbligatorio se il certificato fornito non è firmato dall'autorità di certificazione principale attendibile Java di default o da un'autorità di certificazione intermedia a essa collegata. Le autorità di certificazione

Nome file	Obbligatorio/facoltativo	Informazioni
		principali attendibili Java di default sono disponibili in `jre/lib/security/cacerts` .

È probabile che tu intenda configurare il file PEM della chiave privata affinché sia un certificato generico che consente l'accesso al dominio Amazon VPC in cui si trovano le istanze di cluster. Ad esempio, se il cluster risiede in us-east-1 (Virginia settentrionale), puoi scegliere di specificare un nome comune nella configurazione del certificato che autorizza l'accesso al cluster specificando `CN=*.ec2.internal` nella definizione di oggetto del certificato. Se il cluster risiede in us-west-2 (Oregon), puoi specificare `CN=*.us-west-2.compute.internal`.

Se il file PEM fornito nell'artifact di crittografia non dispone di un carattere jolly nel CN per il dominio, è necessario modificare il valore di `hadoop.ssl.hostname.verifier` in `ALLOW_ALL`. Ciò può essere realizzato con la classificazione `core-site` quando si inviano configurazioni a un cluster o aggiungendo questo valore nel file `core-site.xml`. Questa modifica è necessaria perché il verificatore predefinito del nome host non accetta un nome host senza il carattere jolly, generando così un errore. Per ulteriori informazioni sulla configurazione di cluster EMR in Amazon VPC, consulta Configurazione delle reti.

L'esempio seguente spiega come utilizzare OpenSSL per generare un certificato autofirmato X.509 con una chiave privata RSA a 1024 bit. La chiave consente di accedere alle istanze del cluster Amazon EMR dell'emittente nella Regione `us-west-2` (Oregon) come specificato dal nome di dominio *`*.us-west-2.compute.internal`* come nome comune.

Sono specificati altri elementi di oggetto facoltativi, come paese (C), stato (S) e lingua (L). Poiché viene generato un certificato autofirmato, il secondo comando dell'esempio copia il file `certificateChain.pem` nel file `trustedCertificates.pem`. Il terzo comando usa `zip` per creare il file `my-certs.zip` che contiene i certificati.

> **Important**
>
> Questo esempio è soltanto un proof of concept. L'utilizzo di certificati autofirmati non è consigliato e presenta un potenziale rischio per la sicurezza. Per i sistemi di produzione, utilizza un'autorità di certificazione attendibile per emettere certificati.

```
$ 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 per Amazon EMR

AWS Identity and Access Management (IAM) è un Servizio AWS che consente agli amministratori di controllare in modo sicuro l'accesso alle risorse AWS. Gli amministratori IAM controllano chi è authenticated (autenticato) (accesso effettuato) e authorized (autorizzato) (dispone di autorizzazioni) a utilizzare risorse Amazon EMR. IAM è un Servizio AWS il cui uso non comporta costi aggiuntivi.

Argomenti

Destinatari

Il modo in cui AWS Identity and Access Management (IAM) viene utilizzato cambia a seconda delle operazioni da eseguire in Amazon EMR.

Utente del servizio: se utilizzi il servizio Amazon EMR per eseguire il tuo lavoro, l'amministratore fornisce le credenziali e le autorizzazioni necessarie. All'aumentare del numero di caratteristiche Amazon EMR utilizzate per il lavoro, potrebbero essere necessarie ulteriori autorizzazioni. La comprensione della gestione dell'accesso ti consente di richiedere le autorizzazioni corrette all'amministratore. Se non riesci ad accedere a una funzionalità in Amazon EMR, consulta Risoluzione dei problemi relativi all'identità e all'accesso di Amazon EMR.

Amministratore del servizio: se sei il responsabile delle risorse Amazon EMR presso la tua azienda, probabilmente disponi dell'accesso completo ad Amazon EMR. Il tuo compito è determinare le funzionalità e le risorse di Amazon EMR a cui gli utenti del servizio devono accedere. Devi inviare le richieste all'amministratore IAM per cambiare le autorizzazioni degli utenti del servizio. Esamina le informazioni contenute in questa pagina per comprendere i concetti di base relativi a IAM. Per ulteriori informazioni su come la tua azienda può utilizzare IAM con Amazon EMR, consulta Funzionamento di Amazon EMR con IAM.

Amministratore IAM: se sei un amministratore IAM, potresti essere interessato a ottenere informazioni su come scrivere policy per gestire l'accesso ad Amazon EMR. Per visualizzare policy basate su identità Amazon EMR di esempio che possono essere utilizzate in IAM, consulta Esempi di policy basate su identità per Amazon EMR.

Autenticazione con identità

L'autenticazione è la procedura di accesso ad AWS con le credenziali di identità. Devi essere autenticato (connesso a AWS) come utente root Utente root dell'account AWS, come utente IAM o assumere un ruolo IAM.

Puoi accedere ad AWS come identità federata utilizzando le credenziali fornite attraverso un'origine di identità. AWS IAM Identity Center Gli esempi di identità federate comprendono gli utenti del centro identità IAM, l'autenticazione Single Sign-On (SSO) dell'azienda e le credenziali di Google o Facebook. Se accedi come identità federata, l'amministratore ha configurato in precedenza la federazione delle identità utilizzando i ruoli IAM. Se accedi ad AWS tramite la federazione, assumi indirettamente un ruolo.

A seconda del tipo di utente, puoi accedere alla AWS Management Console o al portale di accesso AWS. Per ulteriori informazioni sull'accesso ad AWS, consulta la sezione Come accedere al tuo Account AWS nella Guida per l'utente di Accedi ad AWS.

Se accedi ad AWS in modo programmatico, AWS fornisce un Software Development Kit (SDK) e un'interfaccia della linea di comando (CLI) per firmare crittograficamente le richieste utilizzando le tue credenziali. Se non utilizzi gli strumenti AWS, devi firmare le richieste personalmente. Per ulteriori informazioni sulla firma delle richieste, consultare Firma delle richieste AWS nella Guida per l'utente IAM.

A prescindere dal metodo di autenticazione utilizzato, potrebbe essere necessario specificare ulteriori informazioni sulla sicurezza. AWS consiglia ad esempio di utilizzare l'autenticazione a più fattori (MFA) per aumentare la sicurezza dell'account. Per ulteriori informazioni, consulta Autenticazione

a più fattori nella Guida per l'utente di AWS IAM Identity Center e Utilizzo dell'autenticazione a più fattori (MFA) in AWS nella Guida per l'utente IAM.

Utente root di un Account AWS

Quando crei un Account AWS, inizi con una singola identità di accesso che ha accesso completo a tutti i Servizi AWS e le risorse nell'account. Tale identità è detta utente root Account AWS ed è possibile accedervi con l'indirizzo e-mail e la password utilizzati per creare l'account. Si consiglia vivamente di non utilizzare l'utente root per le attività quotidiane. Conserva le credenziali dell'utente root e utilizzale per eseguire le operazioni che solo l'utente root può eseguire. Per un elenco completo delle attività che richiedono l'accesso come utente root, consulta la sezione Tasks that require root user credentials (Attività che richiedono le credenziali dell'utente root) nella Guida per l'utente IAM.

Identità federata

Come best practice, richiedere agli utenti umani, compresi quelli che richiedono l'accesso di amministratore, di utilizzare la federazione con un provider di identità per accedere a Servizi AWS utilizzando credenziali temporanee.

Un'identità federata è un utente della directory degli utenti aziendali, un provider di identità Web, AWS Directory Service, la directory Identity Center o qualsiasi utente che accede ad Servizi AWS utilizzando le credenziali fornite tramite un'origine di identità. Quando le identità federate accedono ad Account AWS, assumono ruoli e i ruoli forniscono credenziali temporanee.

Per la gestione centralizzata degli accessi, consigliamo di utilizzare AWS IAM Identity Center. È possibile creare utenti e gruppi in IAM Identity Center oppure connettersi e sincronizzarsi con un gruppo di utenti e gruppi nell'origine di identità per utilizzarli in tutte le applicazioni e gli Account AWS. Per ulteriori informazioni su IAM Identity Center, consulta Cos'è IAM Identity Center? nella Guida per l'utente di AWS IAM Identity Center.

Utenti e gruppi IAM

Un utente IAM è una identità all'interno del tuo Account AWS che dispone di autorizzazioni specifiche per una singola persona o applicazione. Ove possibile, consigliamo di fare affidamento a credenziali temporanee invece di creare utenti IAM con credenziali a lungo termine come le password e le chiavi di accesso. Tuttavia, se si hanno casi d'uso specifici che richiedono credenziali a lungo termine con utenti IAM, si consiglia di ruotare le chiavi di accesso. Per ulteriori informazioni, consulta la pagina Rotazione periodica delle chiavi di accesso per casi d'uso che richiedono credenziali a lungo termine nella Guida per l'utente IAM.

Un gruppo IAM è un'identità che specifica un insieme di utenti IAM. Non è possibile eseguire l'accesso come gruppo. È possibile utilizzare gruppi per specificare le autorizzazioni per più utenti alla volta. I gruppi semplificano la gestione delle autorizzazioni per set di utenti di grandi dimensioni. Ad esempio, è possibile avere un gruppo denominato Amministratori IAM e concedere a tale gruppo le autorizzazioni per amministrare le risorse IAM.

Gli utenti sono diversi dai ruoli. Un utente è associato in modo univoco a una persona o un'applicazione, mentre un ruolo è destinato a essere assunto da chiunque ne abbia bisogno. Gli utenti dispongono di credenziali a lungo termine permanenti, mentre i ruoli forniscono credenziali temporanee. Per ulteriori informazioni, consulta Quando creare un utente IAM (invece di un ruolo) nella Guida per l'utente IAM.

Ruoli IAM

Un ruolo IAM è un'identità all'interno di Account AWS che dispone di autorizzazioni specifiche. È simile a un utente IAM, ma non è associato a una persona specifica. È possibile assumere temporaneamente un ruolo IAM nella AWS Management Console mediante lo scambio di ruoli. È possibile assumere un ruolo chiamando un'operazione AWS CLI o API AWS oppure utilizzando un URL personalizzato. Per ulteriori informazioni sui metodi per l'utilizzo dei ruoli, consulta Utilizzo di ruoli IAM nella Guida per l'utente IAM.

I ruoli IAM con credenziali temporanee sono utili nelle seguenti situazioni:

- Accesso utente federato: per assegnare le autorizzazioni a una identità federata, è possibile creare un ruolo e definire le autorizzazioni per il ruolo. Quando un'identità federata viene autenticata, l'identità viene associata al ruolo e ottiene le autorizzazioni da esso definite. Per ulteriori informazioni sulla federazione dei ruoli, consulta Creazione di un ruolo per un provider di identità di terza parte nella Guida per l'utente IAM. Se utilizzi IAM Identity Center, configura un set di autorizzazioni. IAM Identity Center mette in correlazione il set di autorizzazioni con un ruolo in IAM per controllare a cosa possono accedere le identità dopo l'autenticazione. Per informazioni sui set di autorizzazioni, consulta Set di autorizzazioni nella Guida per l'utente di AWS IAM Identity Center.
- Autorizzazioni utente IAM temporanee: un utente IAM o un ruolo può assumere un ruolo IAM per ottenere temporaneamente autorizzazioni diverse per un'attività specifica.
- Accesso multi-account: è possibile utilizzare un ruolo IAM per permettere a un utente (un principale affidabile) con un account diverso di accedere alle risorse nell'account. I ruoli sono lo strumento principale per concedere l'accesso multi-account. Tuttavia, per alcuni dei Servizi AWS, è possibile collegare una policy direttamente a una risorsa (anziché utilizzare un ruolo come proxy). Per

informazioni sulle differenze tra ruoli e policy basate su risorse per l'accesso multi-account, consulta Differenza tra i ruoli IAM e le policy basate su risorse nella Guida per l'utente IAM.

- Accesso multi-servizio: alcuni Servizi AWS utilizzano funzionalità che in altri Servizi AWS. Ad esempio, quando effettui una chiamata in un servizio, è comune che tale servizio esegua applicazioni in Amazon EC2 o archivi oggetti in Amazon S3. Un servizio può eseguire questa operazione utilizzando le autorizzazioni dell'entità chiamante, utilizzando un ruolo di servizio o utilizzando un ruolo collegato al servizio.
  - Autorizzazioni principale: quando si utilizza un utente o un ruolo IAM per eseguire operazioni in AWS, si viene considerati un principale. Le policy concedono autorizzazioni a un'entità. Quando si utilizzano alcuni servizi, è possibile eseguire un'azione che attiva un'altra azione in un servizio diverso. In questo caso è necessario disporre delle autorizzazioni per eseguire entrambe le operazioni. Per verificare se un'operazione richiede ulteriori operazioni dipendenti in una policy, consulta Operazioni, risorse e chiavi di condizione per Amazon EMR nella Guida di riferimento per l'autorizzazione al servizio.
  - Ruolo di servizio: un ruolo di servizio è un ruolo IAM che un servizio assume per eseguire operazioni per tuo conto. Un amministratore IAM può creare, modificare ed eliminare un ruolo di servizio dall'interno di IAM. Per ulteriori informazioni, consulta la sezione Creazione di un ruolo per delegare le autorizzazioni a un Servizio AWS nella Guida per l'utente IAM.
  - Ruolo collegato al servizio: un ruolo collegato al servizio è un tipo di ruolo di servizio collegato a un Servizio AWS. Il servizio può assumere il ruolo per eseguire un'azione per tuo conto. I ruoli collegati ai servizi sono visualizzati nell'account Account AWS e sono di proprietà del servizio. Un amministratore IAM può visualizzare le autorizzazioni per i ruoli collegati ai servizi, ma non modificarle.
- Applicazioni in esecuzione su Amazon EC2: è possibile utilizzare un ruolo IAM per gestire credenziali temporanee per le applicazioni in esecuzione su un'istanza EC2 che eseguono richieste di AWS CLI o dell'API AWS. Ciò è preferibile all'archiviazione delle chiavi di accesso nell'istanza EC2. Per assegnare un ruolo AWS a un'istanza EC2, affinché sia disponibile per tutte le relative applicazioni, puoi creare un profilo dell'istanza collegato all'istanza. Un profilo dell'istanza contiene il ruolo e consente ai programmi in esecuzione sull'istanza EC2 di ottenere le credenziali temporanee. Per ulteriori informazioni, consulta Utilizzo di un ruolo IAM per concedere autorizzazioni ad applicazioni in esecuzione su istanze di Amazon EC2 nella Guida per l'utente IAM.

Per informazioni sull'utilizzo dei ruoli IAM, consulta Quando creare un ruolo IAM (invece di un utente) nella Guida per l'utente IAM.

Gestione dell'accesso con policy

Per controllare l'accesso a AWS è possibile creare policy e collegarle a identità o risorse AWS. Una policy è un oggetto in AWS che, quando associato a un'identità o a una risorsa, ne definisce le autorizzazioni. AWS valuta queste policy quando un principale IAM (utente, utente root o sessione ruolo) effettua una richiesta. Le autorizzazioni nelle policy determinano l'approvazione o il rifiuto della richiesta. La maggior parte delle policy viene archiviata in AWS sotto forma di documenti JSON. Per ulteriori informazioni sulla struttura e sui contenuti dei documenti delle policy JSON, consulta Panoramica delle policy JSON nella Guida per l'utente IAM.

Gli amministratori possono utilizzare le policy AWS JSON per specificare l'accesso ai diversi elementi. In altre parole, quale principale può eseguire azioni su quali risorse e in quali condizioni.

Per impostazione predefinita, utenti e ruoli non dispongono di autorizzazioni. Per concedere agli utenti l'autorizzazione a eseguire operazioni sulle risorse di cui hanno bisogno, un amministratore IAM può creare policy IAM. Successivamente l'amministratore può aggiungere le policy IAM ai ruoli e gli utenti possono assumere i ruoli.

Le policy IAM definiscono le autorizzazioni relative a un'operazione, a prescindere dal metodo utilizzato per eseguirla. Ad esempio, supponiamo di disporre di una policy che consente l'operazione `iam:GetRole`. Un utente con tale policy può ottenere informazioni sul ruolo dalla AWS Management Console, la AWS CLI o l'API AWS.

Policy basate su identità

Le policy basate su identità sono documenti di policy di autorizzazione JSON che è possibile allegare a un'identità (utente, gruppo di utenti o ruolo IAM). Tali policy definiscono le operazioni che utenti e ruoli possono eseguire, su quali risorse e in quali condizioni. Per informazioni su come creare una policy basata su identità, consulta Creazione di policy IAM nella Guida per l'utente IAM.

Le policy basate su identità possono essere ulteriormente classificate come policy inline o policy gestite. Le policy inline sono incorporate direttamente in un singolo utente, gruppo o ruolo. Le policy gestite sono policy autonome che possono essere collegate a più utenti, gruppi e ruoli in Account AWS. Le policy gestite includono le policy gestite da AWS e le policy gestite dal cliente. Per informazioni su come scegliere tra una policy gestita o una policy inline, consulta Scelta fra policy gestite e policy inline nella Guida per l'utente IAM.

Policy basate su risorse

Le policy basate su risorse sono documenti di policy JSON che è possibile allegare a una risorsa. Gli esempi più comuni di policy basate su risorse sono le policy di trust dei ruoli IAM e le policy dei bucket Amazon S3. Nei servizi che supportano policy basate sulle risorse, gli amministratori dei servizi possono utilizzarli per controllare l'accesso a una risorsa specifica. Quando è allegata a una risorsa, una policy definisce le azioni che un principale può eseguire su tale risorsa e a quali condizioni. È necessario specificare un principale in una policy basata sulle risorse. I principali possono includere account, utenti, ruoli, utenti federati o Servizi AWS.

Le policy basate sulle risorse sono policy inline che si trovano in tale servizio. Non è possibile utilizzare le policy gestite da AWS da IAM in una policy basata su risorse.

Liste di controllo accessi (ACL)

Le liste di controllo degli accessi (ACL) controllano quali principali (membri, utenti o ruoli dell'account) hanno le autorizzazioni per accedere a una risorsa. Le ACL sono simili alle policy basate sulle risorse, sebbene non utilizzino il formato del documento di policy JSON.

Amazon S3, AWS WAF e Amazon VPC sono esempi di servizi che supportano le ACL. Per maggiori informazioni sulle ACL, consulta Panoramica delle liste di controllo degli accessi (ACL) nella Guida per gli sviluppatori di Amazon Simple Storage Service.

Altri tipi di policy

AWS supporta altri tipi di policy meno comuni. Questi tipi di policy possono impostare il numero massimo di autorizzazioni concesse dai tipi di policy più comuni.

- Limiti delle autorizzazioni: un limite delle autorizzazioni è una funzione avanzata nella quale si imposta il numero massimo di autorizzazioni che una policy basata su identità può concedere a un'entità IAM (utente o ruolo IAM). È possibile impostare un limite delle autorizzazioni per un'entità. Le autorizzazioni risultanti sono l'intersezione delle policy basate su identità dell'entità e i relativi limiti delle autorizzazioni. Le policy basate su risorse che specificano l'utente o il ruolo nel campo `Principal` sono condizionate dal limite delle autorizzazioni. Un rifiuto esplicito in una qualsiasi di queste policy sostituisce l'autorizzazione. Per ulteriori informazioni sui limiti delle autorizzazioni, consulta Limiti delle autorizzazioni per le entità IAM nella Guida per l'utente IAM.
- Policy di controllo dei servizi (SCP): le SCP sono policy JSON che specificano il numero massimo di autorizzazioni per un'organizzazione o unità organizzativa (OU) in AWS Organizations. AWS

Organizations è un servizio per il raggruppamento e la gestione centralizzata degli Account AWS multipli di proprietà dell'azienda. Se abiliti tutte le funzionalità in un'organizzazione, puoi applicare le policy di controllo dei servizi (SCP) a uno o tutti i tuoi account. La SCP limita le autorizzazioni per le entità negli account membri, compreso ogni Utente root dell'account AWS. Per ulteriori informazioni su organizzazioni e policy SCP, consulta la pagina sulle Policy di controllo dei servizi nella Guida per l'utente di AWS Organizations.

- Policy di sessione: le policy di sessione sono policy avanzate che vengono trasmesse come parametro quando si crea in modo programmatico una sessione temporanea per un ruolo o un utente federato. Le autorizzazioni della sessione risultante sono l'intersezione delle policy basate su identità del ruolo o dell'utente e le policy di sessione. Le autorizzazioni possono anche provenire da una policy basata su risorse. Un rifiuto esplicito in una qualsiasi di queste policy sostituisce l'autorizzazione. Per ulteriori informazioni, consulta Policy di sessione nella Guida per l'utente IAM.

Più tipi di policy

Quando più tipi di policy si applicano a una richiesta, le autorizzazioni risultanti sono più complicate da comprendere. Per informazioni su come AWS determina se consentire una richiesta quando sono coinvolti più tipi di policy, consulta Logica di valutazione delle policy nella Guida per l'utente IAM.

Funzionamento di Amazon EMR con IAM

Prima di utilizzare IAM per gestire l'accesso ad Amazon EMR, scopri quali funzionalità IAM sono disponibili per l'uso con Amazon EMR.

Funzionalità IAM utilizzabili con Amazon EMR

Funzionalità IAM	Supporto per Amazon EMR
Policy basate su identità	Sì
Policy basate su risorse	Sì
Operazioni di policy	Sì
Risorse relative alle policy	Sì
Chiavi di condizione delle policy	Sì
Liste di controllo degli accessi	No

Funzionalità IAM	Supporto per Amazon EMR
ABAC (tag nelle policy)	Sì
Credenziali temporanee	Sì
Autorizzazioni del principale	Sì
Ruoli di servizio	No
Ruoli collegati al servizio	Sì

Per ottenere un quadro generale del funzionamento di Amazon EMR e altri servizi AWS con la maggior parte delle funzionalità di IAM, consulta Servizi AWS supportati da IAM nella Guida per l'utente IAM.

Policy basate su identità per Amazon EMR

Supporta le policy basate su identità	Sì

Le policy basate su identità sono documenti di policy di autorizzazione JSON che è possibile allegare a un'identità (utente, gruppo di utenti o ruolo IAM). Tali policy definiscono le operazioni che utenti e ruoli possono eseguire, su quali risorse e in quali condizioni. Per informazioni su come creare una policy basata su identità, consulta Creazione di policy IAM nella Guida per l'utente di IAM.

Con le policy basate su identità di IAM, è possibile specificare quali operazioni e risorse sono consentite o rifiutate, nonché le condizioni in base alle quali le operazioni sono consentite o rifiutate. Non è possibile specificare l'entità principale in una policy basata sull'identità perché si applica all'utente o al ruolo a cui è associato. Per informazioni su tutti gli elementi utilizzabili in una policy JSON, consulta Guida di riferimento agli elementi delle policy JSON IAM nella Guida per l'utente IAM.

Esempi di policy basate su identità per Amazon EMR

Per visualizzare esempi di policy basate su identità Amazon EMR, consulta Esempi di policy basate su identità per Amazon EMR.

Policy basate su risorse all'interno di Amazon EMR

Supporta le policy basate su risorse	Sì

Le policy basate su risorse sono documenti di policy JSON che è possibile allegare a una risorsa. Gli esempi più comuni di policy basate su risorse sono le policy di trust dei ruoli IAM e le policy dei bucket Amazon S3. Nei servizi che supportano policy basate sulle risorse, gli amministratori dei servizi possono utilizzarli per controllare l'accesso a una risorsa specifica. Quando è allegata a una risorsa, una policy definisce le azioni che un principale può eseguire su tale risorsa e a quali condizioni. È necessario specificare un principale in una policy basata sulle risorse. I principali possono includere account, utenti, ruoli, utenti federati o Servizi AWS.

Per consentire l'accesso multi-account, puoi specificare un intero account o entità IAM in un altro account come principale in una policy basata sulle risorse. L'aggiunta di un principale multi-account a una policy basata sulle risorse rappresenta solo una parte della relazione di trust. Quando l'entità principale e la risorsa si trovano in diversi Account AWS, un amministratore IAM nell'account attendibile deve concedere all'entità principale (utente o ruolo) anche l'autorizzazione per accedere alla risorsa. L'autorizzazione viene concessa collegando all'entità una policy basata sull'identità. Tuttavia, se una policy basata su risorse concede l'accesso a un principale nello stesso account, non sono richieste ulteriori policy basate su identità. Per ulteriori informazioni, consulta Differenza tra i ruoli IAM e le policy basate su risorse nella Guida per l'utente IAM.

Operazioni di policy per Amazon EMR

Supporta le azioni di policy	Sì

Gli amministratori possono utilizzare le policy JSON AWS per specificare gli accessi ai diversi elementi. Cioè, quale principale può eseguire azioni su quali risorse, e in quali condizioni.

L'elemento `Action` di una policy JSON descrive le azioni che è possibile utilizzare per consentire o negare l'accesso a un criterio. Le azioni di policy hanno spesso lo stesso nome dell'operazione API AWS. Ci sono alcune eccezioni, ad esempio le azioni di sola autorizzazione che non hanno un'operazione API corrispondente. Esistono anche alcune operazioni che richiedono più azioni in una policy. Queste azioni aggiuntive sono denominate azioni dipendenti.

Includi le azioni in una policy per concedere le autorizzazioni a eseguire l'operazione associata.

Per visualizzare un elenco di operazioni Amazon EMR, consulta Operazioni, risorse e chiavi di condizione per Amazon EMR in Guida di riferimento all'autorizzazione del servizio.

Le operazioni delle policy in Amazon EMR utilizzano il seguente prefisso prima dell'operazione:

```
EMR
```

Per specificare più azioni in una sola istruzione, occorre separarle con la virgola.

```
"Action": [
      "EMR:action1",
      "EMR:action2"
         ]
```

Per visualizzare esempi di policy basate su identità Amazon EMR, consulta Esempi di policy basate su identità per Amazon EMR.

Risorse delle policy per Amazon EMR

Supporta le risorse di policy	Sì

Gli amministratori possono utilizzare le policy JSON AWS per specificare gli accessi ai diversi elementi. Cioè, quale principale può eseguire azioni su quali risorse, e in quali condizioni.

L'elemento JSON `Resource` della policy specifica l'oggetto o gli oggetti ai quali si applica l'operazione. Le istruzioni devono includere un elemento `Resource` o un elemento `NotResource`. Come best practice, specifica una risorsa utilizzando il suo nome della risorsa Amazon (ARN). Puoi eseguire questa operazione per azioni che supportano un tipo di risorsa specifico, note come autorizzazioni a livello di risorsa.

Per le azioni che non supportano le autorizzazioni a livello di risorsa, ad esempio le operazioni di elenco, utilizza un carattere jolly (*) per indicare che l'istruzione si applica a tutte le risorse.

```
"Resource": "*"
```

Per visualizzare un elenco di tipi di risorse di Amazon EMR, consulta Risorse definite da Amazon EMR in Guida di riferimento all'autorizzazione del servizio. Per informazioni sulle operazioni per cui è

possibile specificare l'ARN di ciascuna risorsa, consulta Operazioni, risorse e chiavi di condizione per Amazon EMR.

Per visualizzare esempi di policy basate su identità Amazon EMR, consulta Esempi di policy basate su identità per Amazon EMR.

Chiavi di condizione delle policy per Amazon EMR

Supporta le chiavi di condizione delle policy specifiche del servizio	Sì

Gli amministratori possono utilizzare le policy JSON AWS per specificare gli accessi ai diversi elementi. Cioè, quale principale può eseguire azioni su quali risorse, e in quali condizioni.

L'elemento `Condition` (o blocco `Condition`) consente di specificare le condizioni in cui un'istruzione è in vigore. L'elemento `Condition` è facoltativo. Puoi compilare espressioni condizionali che utilizzano operatori di condizione, ad esempio uguale a o minore di, per soddisfare la condizione nella policy con i valori nella richiesta.

Se specifichi più elementi `Condition` in un'istruzione o più chiavi in un singolo elemento `Condition`, questi vengono valutati da AWS utilizzando un'operazione `AND` logica. Se specifichi più valori per una singola chiave di condizione, AWS valuta la condizione utilizzando un'operazione `OR` logica. Tutte le condizioni devono essere soddisfatte prima che le autorizzazioni dell'istruzione vengano concesse.

Puoi anche utilizzare variabili segnaposto quando specifichi le condizioni. Ad esempio, puoi autorizzare un utente IAM ad accedere a una risorsa solo se è stata taggata con il relativo nome utente IAM. Per ulteriori informazioni, consulta Elementi delle policy IAM: variabili e tag nella Guida per l'utente IAM.

AWS supporta chiavi di condizione globali e chiavi di condizione specifiche per il servizio. Per visualizzare tutte le chiavi di condizione globali di AWS, consulta Chiavi di contesto delle condizioni globali di AWS nella Guida per l'utente IAM.

Per visualizzare un elenco di chiavi di condizione Amazon EMR e scoprire quali operazioni e risorse è possibile utilizzare con una chiave di condizione, consulta Operazioni, risorse e chiavi di condizione per Amazon EMR in Guida di riferimento all'autorizzazione del servizio.

Per visualizzare esempi di policy basate su identità Amazon EMR, consulta Esempi di policy basate su identità per Amazon EMR.

Liste di controllo degli accessi (ACL) in Amazon EMR

Supporta le ACL	No

Le policy di controllo degli accessi (ACL) controllano quali principali (membri, utenti o ruoli dell'account) hanno le autorizzazioni ad accedere a una risorsa. Le ACL sono simili alle policy basate su risorse, sebbene non utilizzino il formato del documento di policy JSON.

Controllo degli accessi basato su attributi (ABAC) con Amazon EMR

Supporta ABAC (tag nelle policy)	Sì

Il controllo dell'accesso basato su attributi (ABAC) è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. In AWS, tali attributi sono denominati tag. È possibile collegare dei tag alle entità IAM (utenti o ruoli) e a numerose risorse AWS. L'assegnazione di tag alle entità e alle risorse è il primo passaggio di ABAC. In seguito, vengono progettate policy ABAC per consentire operazioni quando il tag dell'entità principale corrisponde al tag sulla risorsa a cui si sta provando ad accedere.

La strategia ABAC è utile in ambienti soggetti a una rapida crescita e aiuta in situazioni in cui la gestione delle policy diventa impegnativa.

Per controllare l'accesso basato su tag, fornisci informazioni sui tag nell'elemento condizione di una policy utilizzando le chiavi di condizione `aws:ResourceTag/`*`key-name`*, `aws:RequestTag/`*`key-name`* o `aws:TagKeys`.

Se un servizio supporta tutte e tre le chiavi di condizione per ogni tipo di risorsa, il valore per il servizio è Yes (Sì). Se un servizio supporta tutte e tre le chiavi di condizione solo per alcuni tipi di risorsa, allora il valore sarà Partial (Parziale).

Per ulteriori informazioni su ABAC, consulta Che cos'è ABAC? nella Guida per l'utente IAM. Per visualizzare un tutorial con i passaggi per l'impostazione di ABAC, consulta Utilizzo del controllo degli accessi basato su attributi (ABAC) nella Guida per l'utente IAM.

Utilizzo di credenziali temporanee con Amazon EMR

Supporta le credenziali temporanee	Sì

Alcuni Servizi AWS non funzionano quando si accede utilizzando credenziali temporanee. Per ulteriori informazioni, inclusi i Servizi AWS che funzionano con le credenziali temporanee, consulta Servizi AWS supportati da IAM nella Guida per l'utente IAM.

Le credenziali temporanee sono utilizzate se si accede alla AWS Management Console utilizzando qualsiasi metodo che non sia la combinazione di nome utente e password. Ad esempio, quando accedi alla AWS utilizzando il collegamento Single Sign-On (SSO) della tua azienda, tale processo crea in automatico credenziali temporanee. Le credenziali temporanee vengono create in automatico anche quando accedi alla console come utente e poi cambi ruolo. Per ulteriori informazioni sullo scambio dei ruoli, consulta Cambio di un ruolo (console) nella Guida per l'utente IAM.

È possibile creare manualmente credenziali temporanee utilizzando la AWS CLI o l'API AWS. È quindi possibile utilizzare tali credenziali temporanee per accedere ad AWS. AWS consiglia di generare le credenziali temporanee dinamicamente anziché utilizzare chiavi di accesso a lungo termine. Per ulteriori informazioni, consulta Credenziali di sicurezza provvisorie in IAM.

Autorizzazioni delle entità principali tra servizi per Amazon EMR

Supporta le autorizzazioni delle entità principali	Sì

Quando si utilizza un utente o un ruolo IAM per eseguire operazioni in AWS, si viene considerati un'entità principale. Le policy concedono autorizzazioni al principale. Quando si utilizzano alcuni servizi, è possibile eseguire un'azione che attiva un'altra azione in un servizio diverso. In questo caso è necessario disporre delle autorizzazioni per eseguire entrambe le operazioni. Per verificare se un'operazione richiede ulteriori operazioni dipendenti in una policy, consulta Operazioni, risorse e chiavi di condizione per Amazon EMR nella Guida di riferimento per l'autorizzazione al servizio.

Ruoli di servizio per Amazon EMR

Supporta i ruoli di servizio	No

Ruoli collegati ai servizi per Amazon EMR

Supporta i ruoli collegati ai servizi	Sì

Per ulteriori informazioni su come creare e gestire i ruoli collegati ai servizi, consulta Servizi AWS supportati da IAM. Trova un servizio nella tabella che include un `Yes` nella colonna Service-linked role (Ruolo collegato ai servizi). Scegli il collegamento Sì per visualizzare la documentazione relativa al ruolo collegato ai servizi per tale servizio.

Utilizzo di tag di cluster e notebook con policy IAM per il controllo degli accessi

Le autorizzazioni per operazioni Amazon EMR associate a EMR Notebooks e cluster EMR possono essere ottimizzate utilizzando il controllo degli accessi basato su tag con policy IAM basate su identità. Puoi utilizzare chiavi di condizione all'interno di un elemento `Condition` (chiamato anche blocco `Condition`) per consentire determinate operazioni solo quando un notebook, un cluster o entrambi dispongono di una determinata combinazione di chiave di tag o di chiave-valore. Puoi anche limitare l'operazione `CreateEditor` (che crea un notebook EMR) e l'operazione `RunJobFlow` (che crea un cluster) in modo che una richiesta per un tag deve essere inviata quando la risorsa viene creata.

In Amazon EMR, le chiavi di condizioni che possono essere utilizzate in un elemento `Condition` si applicano solo a tali operazioni API Amazon EMR in cui `ClusterID` o `NotebookID` è un parametro di richiesta obbligatorio. Ad esempio, l'operazione ModifyInstanceGroups non supporta le chiavi di contesto perché `ClusterID` è un parametro opzionale.

Quando crei un notebook EMR, viene applicato un tag predefinito con una stringa di chiavi `creatorUserId` impostata sul valore dell'ID utente IAM che ha creato il notebook. Ciò è utile per limitare le operazioni consentite per il notebook al solo creatore.

Le seguenti chiavi di condizione sono disponibili in Amazon EMR:

- Utilizza la chiave di contesto di condizione `elasticmapreduce:ResourceTag/`*`TagKeyString`* per concedere o negare agli utenti operazioni su cluster p notebook con tag che hanno *`TagKeyString`* specificato. Se un'operazione trasferisce `ClusterID` e `NotebookID`, la condizione si applica al cluster e al notebook. Questo significa che entrambe le risorse devono avere la stringa della chiave di tag o la combinazione chiave-valore specificata. Puoi utilizzare l'elemento `Resource` per limitare l'istruzione in modo che si applichi solo ai cluster o ai notebook

in base alle esigenze. Per ulteriori informazioni, consulta Esempi di policy basate su identità per Amazon EMR.

- Utilizza la chiave di contesto di condizione `elasticmapreduce:RequestTag/`*`TagKeyString`* per richiedere un determinato tag con operazioni/chiamate API. Ad esempio, puoi utilizzare questa chiave di contesto di condizione insieme all'operazione `CreateEditor` per richiedere che una chiave con *`TagKeyString`* sia applicata a un notebook quando viene creato.

Esempi

Per visualizzare un elenco delle operazioni Amazon EMR, consulta Actions Defined by Amazon EMR (Operazioni definite da Amazon EMR) nella IAM User Guide (Guida per l'utente IAM).

Ruoli di runtime per le fasi di Amazon EMR

Un ruolo di runtime è un ruolo (IAM) AWS Identity and Access Management che puoi specificare quando invii un processo o una query a un cluster Amazon EMR. Il processo o la query inviata al cluster Amazon EMR utilizza il ruolo di runtime per accedere alle risorse AWS, ad esempio agli oggetti in Amazon S3. Puoi specificare i ruoli di runtime con Amazon EMR per i processi Spark e Hive.

Puoi anche specificare i ruoli di runtime quando ti connetti a cluster Amazon EMR in Amazon SageMaker e quando colleghi un WorkSpace Amazon EMR Studio a un cluster EMR. Per ulteriori informazioni, consulta Connessione a un cluster Amazon EMR da Studio e Esecuzione di un WorkSpace EMR Studio con un ruolo di runtime.

In precedenza, i cluster Amazon EMR eseguivano processi o query Amazon EMR con autorizzazioni basate sulla policy IAM associata al profilo dell'istanza utilizzato per avviare il cluster. Ciò significava che le policy dovevano contenere l'unione di tutte le autorizzazioni per tutti i processi e le query eseguiti su un cluster Amazon EMR. Con i ruoli di runtime, ora puoi gestire il controllo degli accessi per ogni singolo processo o query, invece di condividere il profilo dell'istanza Amazon EMR del cluster.

Nei cluster Amazon EMR con ruoli di runtime, puoi inoltre applicare il controllo degli accessi basato su AWS Lake Formation ai processi Spark, Hive e Presto e alle query sui tuoi data lake. Per maggiori informazioni su come eseguire l'integrazione con AWS Lake Formation, consulta la sezione Integrazione di Amazon EMR con AWS Lake Formation.

> **Note**
>
> Quando specifichi un ruolo di runtime per una fase di Amazon EMR, i processi o le query inviate possono accedere solo alle risorse AWS consentite dalle policy associate al ruolo di runtime. Questi processi e queste query non possono accedere al servizio metadati dell'istanza sulle istanze EC2 del cluster né utilizzare il profilo dell'istanza EC2 del cluster per accedere alle risorse AWS.

Prerequisiti per l'avvio di un cluster Amazon EMR con un ruolo di runtime

Argomenti

Fase 1: impostazione delle configurazioni di sicurezza in Amazon EMR

Utilizza la struttura JSON seguente per creare una configurazione di sicurezza nella AWS Command Line Interface (AWS CLI) e impostare `EnableApplicationScopedIAMRole` su `true`. Per ulteriori informazioni sulle configurazioni della sicurezza, consulta Utilizzo delle configurazioni di sicurezza per impostare la sicurezza del cluster.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true
        }
    }
}
```

Ti consigliamo di abilitare sempre le opzioni di crittografia in transito nella configurazione di sicurezza, in modo che i dati trasferiti su Internet siano crittografati anziché in testo semplice. Puoi ignorare queste opzioni se non desideri connetterti ai cluster Amazon EMR con ruoli di runtime di SageMaker Studio o EMR Studio. Per configurare la crittografia dei dati, consulta Configurazione della crittografia di dati.

In alternativa, puoi creare una configurazione di sicurezza con impostazioni personalizzate con la AWS Management Console.

Passaggio 2: Configurazione di un profilo dell'istanza EC2 per il cluster Amazon EMR

I cluster Amazon EMR utilizzano il ruolo del profilo dell'istanza Amazon EC2 per assumere i ruoli di runtime. Per utilizzare i ruoli di runtime con le fasi di Amazon EMR, aggiungi le seguenti policy al ruolo IAM che intendi utilizzare come ruolo del profilo dell'istanza. Per aggiungere le policy a un ruolo IAM o modificare una policy inline o gestita esistente, consulta Aggiunta e rimozione di autorizzazioni per identità IAM.

```
{
    "Version":"2012-10-17",
    "Statement":[
        {
            "Sid":"AllowRuntimeRoleUsage",
            "Effect":"Allow",
            "Action":[
                "sts:AssumeRole",
                "sts:TagSession"
            ],
            "Resource":[
                <runtime-role-ARN>
            ]
        }
    ]
}
```

Passaggio 3: Configurazione di una policy di attendibilità

Per ogni ruolo IAM che intendi utilizzare come ruolo di runtime, imposta la seguente policy di attendibilità, sostituendo `EMR_EC2_DefaultRole` con il ruolo del profilo dell'istanza. Per modificare la policy di attendibilità di un ruolo IAM, consulta Modifica di una policy di attendibilità del ruolo.

```
{
    "Sid":"AllowAssumeRole",
    "Effect":"Allow",
    "Principal":{
        "AWS":"arn:aws:iam::<AWS_ACCOUNT_ID>:role/EMR_EC2_DefaultRole"
    },
    "Action":"sts:AssumeRole"
}
```

Avvio di un cluster Amazon EMR con controllo degli accessi basato su ruoli

Dopo aver impostato le configurazioni, puoi avviare un cluster Amazon EMR con la configurazione di sicurezza da Fase 1: impostazione delle configurazioni di sicurezza in Amazon EMR. Per utilizzare i ruoli di runtime con le fasi di Amazon EMR, utilizza l'etichetta di rilascio `emr-6.7.0` o versioni successive e seleziona Hive, Spark o entrambi come applicazione del cluster. Per connetterti da SageMaker Studio, utilizza la versione `emr-6.9.0` o successive e seleziona Livy, Spark, Hive o Presto come applicazione del cluster. Per istruzioni sull'avvio del cluster, consulta la sezione Impostazione di una configurazione di sicurezza per un cluster.

Invio di processi Spark tramite la procedura di Amazon EMR

Di seguito è riportato un esempio di esecuzione dell'oggetto HdfsTest incluso con Apache Spark. Questa chiamata API ha esito positivo solo se il ruolo di runtime per Amazon EMR fornito dispone dell'accesso 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**
>
> Ti consigliamo di disattivare l'accesso al cluster Amazon EMR di SSH e di consentirne l'accesso solo all'API `AddJobFlowSteps` di Amazon EMR.

Invio di processi Hive tramite la procedura di Amazon EMR

L'esempio seguente utilizza Apache Hive con fasi Amazon EMR per inviare un processo per l'esecuzione del file `QUERY_FILE.hql`. Questa query ha esito positivo solo se il ruolo di runtime fornito può accedere al percorso Amazon S3 del file di query.

```
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
```

Connessione ai cluster Amazon EMR con ruoli di runtime da un notebook SageMaker Studio

Puoi applicare i ruoli di runtime di Amazon EMR alle query eseguite nei cluster Amazon EMR da SageMaker Studio. A tale scopo, segui la procedura seguente.

1. Segui le istruzioni in Launch Amazon SageMaker Studio (Avvio di Amazon SageMaker Studio) per creare un SageMaker Studio.
2. Nell'interfaccia utente di Sagemaker Studio, avvia un notebook con kernel supportati. Ad esempio, avvia un'immagine SparkMagic con un kernel PySpark.
3. Scegli un cluster Amazon EMR in SageMaker Studio, quindi scegli Connect (Connetti).
4. Scegli un ruolo di runtime, quindi seleziona Connect (Connetti).

Questa operazione crea una cella del notebook SageMaker con comandi magic per connettersi al cluster Amazon EMR con il ruolo di runtime Amazon EMR scelto. Nella cella del notebook, puoi inserire ed eseguire query con il ruolo di runtime e il controllo degli accessi basato su Lake Formation. Per un esempio più dettagliato, consulta la sezione Applicare controlli granulari degli accessi ai dati con AWS Lake Formation e Amazon EMR da Amazon SageMaker Studio.

Controllo dell'accesso al ruolo di runtime di Amazon EMR

Puoi controllare l'accesso al ruolo di runtime con la chiave di condizione `elasticmapreduce:ExecutionRoleArn`. La seguente policy consente a un principale IAM di utilizzare un ruolo IAM denominato `Caller` o qualsiasi ruolo IAM che inizia con la stringa `CallerTeamRole`, come il ruolo di runtime.

> **Important**
>
> Quando concedi l'accesso a un chiamante al fine di effettuare la chiamata all'API `AddJobFlowSteps` o `GetClusterSessionCredentials`, devi creare una condizione basata sulla chiave di contesto `elasticmapreduce:ExecutionRoleArn`, come mostrato nell'esempio seguente.

```
{
    "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*"
            ]
        }
    }
}
```

Configurazione di un rapporto di attendibilità tra i ruoli di runtime e i cluster Amazon EMR

Amazon EMR genera un identificatore univoco `ExternalId` per ogni configurazione di sicurezza con autorizzazione del ruolo di runtime attivata. Questa autorizzazione consente a ogni utente di possedere una serie di ruoli di runtime da utilizzare sui cluster di sua proprietà. Ad esempio, ogni reparto di un'azienda può utilizzare il proprio ID esterno per aggiornare la policy di attendibilità sulla propria serie di ruoli di runtime.

Puoi trovare l'ID esterno con l'API `DescribeSecurityConfiguration` di Amazon EMR, come mostrato nell'esempio seguente.

```
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"
}
```

Per ulteriori informazioni su come utilizzare un ID esterno, vedi l'argomento Come utilizzare un ID esterno quando si concede a una terza parte l'accesso alle proprie risorse AWS.

Verifica

Per monitorare e controllare le azioni che gli utenti finali compiono con i ruoli IAM, è possibile attivare la funzione di identità di origine. Per ulteriori informazioni sull'origine dell'identità, consulta la sezione Monitoraggio e controllo delle operazioni intraprese con i ruoli assunti.

Per tenere traccia dell'identità dell'origine, imposta `ApplicationScopedIAMRoleConfiguration/PropagateSourceIdentity` su `true` nella configurazione di sicurezza, come indicato di seguito.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true,
            "ApplicationScopedIAMRoleConfiguration":{
                "PropagateSourceIdentity":true
            }
        }
    }
}
```

Quando imposti `PropagateSourceIdentity` su `true`, Amazon EMR applica l'identità dell'origine dalle credenziali di chiamata a una sessione di processo o query creata con il ruolo di runtime. Se nelle credenziali di chiamata non è presente alcuna identità di origine, Amazon EMR non imposta l'identità di origine.

Per utilizzare questa proprietà, fornisci le autorizzazioni `sts:SetSourceIdentity` al tuo profilo dell'istanza, come indicato di seguito.

```
{ // PropagateSourceIdentity statement
    "Sid":"PropagateSourceIdentity",
    "Effect":"Allow",
    "Action":"sts:SetSourceIdentity",
    "Resource":[
        <runtime-role-ARN>
    ],
    "Condition":{
        "StringEquals":{
            "sts:SourceIdentity":<source-identity>
        }
    }
}
```

È necessario anche aggiungere l'istruzione `AllowSetSourceIdentity` alla policy di attendibilità dei ruoli di runtime.

```
{ // 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>
        }
    }
}
```

Ulteriori considerazioni

> **Note**
>
> Con la versione `emr-6.9.0` di Amazon EMR, potresti riscontrare errori intermittenti quando ti connetti ai cluster Amazon EMR da SageMaker Studio. Per risolvere questo problema, puoi installare la patch con un'operazione di bootstrap all'avvio del cluster. Per informazioni dettagliate sulle patch, consulta la sezione Amazon EMR release 6.9.0 known issues (Problemi noti di Amazon EMR versione 6.9.0).

Inoltre, quando configuri i ruoli di runtime per Amazon EMR, tieni presente quanto segue.

- Amazon EMR supporta i ruoli di runtime in tutte le Regioni AWS commerciali.
- Le fasi di Amazon EMR supportano i processi Apache Spark e Apache Hive con i ruoli di runtime quando utilizzi la versione `emr-6.7.0` o successive.
- SageMaker Studio supporta le query Spark, Hive e Presto con ruoli di runtime quando utilizzi la versione `emr-6.9.0` o successive.
- I seguenti kernel del notebook in SageMaker supportano i ruoli di runtime:
  - DataScience — Kernel Python 3
  - DataScience 2.0 — Kernel Python 3
  - DataScience 3.0 — Kernel Python 3
  - SparkAnalytics 1.0 — Kernel SparkMagic e PySpark
  - SparkAnalytics 2.0 — Kernel SparkMagic e PySpark
  - SparkMagic — Kernel PySpark
- Amazon EMR supporta le fasi che utilizzano `RunJobFlow` solo al momento della creazione del cluster. Questa API non supporta i ruoli di runtime.
- Amazon EMR non supporta i ruoli di runtime su cluster configurati per la disponibilità elevata.
- I ruoli di runtime non forniscono supporto per il controllo dell'accesso alle risorse del cluster, come HDFS e HMS.

Configurazione dei ruoli di servizio IAM per le autorizzazioni di Amazon EMR per i servizi e risorse AWS

Amazon EMR e applicazioni come Hadoop e Spark hanno bisogno di autorizzazioni per accedere ad altre risorse AWS ed eseguire operazioni quando sono in esecuzione. Ogni cluster in Amazon EMR deve avere un ruolo di servizio e un ruolo per il profilo dell'istanza Amazon EC2. Per ulteriori informazioni, consulta Ruoli IAM e Utilizzo dei profili dell'istanza nella Guida per l'utente IAM. Le policy IAM allegate a questi ruoli forniscono al cluster le autorizzazioni per interagire con altri servizi AWS per conto di un utente.

Un ruolo aggiuntivo, il ruolo Auto Scaling, è necessario se il cluster utilizza la scalabilità automatica in Amazon EMR. Il ruolo di servizio AWS per EMR Notebooks è necessario se si utilizza EMR Notebooks.

Amazon EMR fornisce ruoli e policy gestite predefiniti che determinano le autorizzazioni per ciascun ruolo. Le policy gestite sono create e mantenute da AWS, quindi vengono aggiornate automaticamente in caso di modifica dei requisiti di servizio. Per ulteriori informazioni, consulta Policy gestite da AWS nella Guida per l'utente IAM.

Se si sta creando un cluster o un notebook per la prima volta in un account, i ruoli per Amazon EMR non esistono ancora. Una volta creati, è possibile visualizzare i ruoli, le policy allegate e le autorizzazioni consentite o negate dalle policy nella console IAM (https://console.aws.amazon.com/iam/). È possibile specificare ruoli predefiniti per Amazon EMR da creare e utilizzare, è possibile creare ruoli propri e specificarli individualmente al momento della creazione di un cluster per personalizzare le autorizzazioni, infine è possibile specificare ruoli predefiniti da utilizzare al momento della creazione di un cluster utilizzando la AWS CLI. Per ulteriori informazioni, consulta Personalizzazione dei ruoli IAM.

Modifica di policy basate sull'identità per le autorizzazioni a passare i ruoli del servizio per Amazon EMR

Le policy gestite predefinite con autorizzazioni complete di Amazon EMR incorporano configurazioni di sicurezza `iam:PassRole`, tra cui:

- Autorizzazioni `iam:PassRole` solo per specifici ruoli Amazon EMR predefiniti.
- Condizioni `iam:PassedToService` che consentono di utilizzare la policy solo con servizi AWS specificati, quali `elasticmapreduce.amazonaws.com` e `ec2.amazonaws.com`.

È possibile visualizzare la versione JSON delle policy AmazonEMRFullAccessPolicy_v2 e AmazonEMRServicePolicy_v2 nella console IAM. Ti consigliamo di creare i nuovi cluster con le policy gestite v2.

Riepilogo del ruolo di servizio

La tabella seguente elenca i ruoli del servizio IAM associati ad Amazon EMR per riferimento rapido.

Funzione	Ruolo predefinito	Descrizione	Policy gestita predefinita
Ruolo di servizio per Amazon EMR (ruolo EMR)	`EMR_DefaultRole_V2`	Permette ad Amazon EMR di chiamare altri servizi AWS per conto dell'utente quando effettua il provisioning delle risorse ed esegue operazioni a livello di servizio. Questo ruolo è obbligatorio per tutti i cluster.	`AmazonEMRServicePolicy_v2` **Important** Per richiedere le Istanze spot è necessario un ruolo collegato al servizio. Se questo ruolo non esiste, il ruolo di servizio Amazon EMR deve avere l'autorizzazione per crearlo, altrimenti si verifica un errore di autorizzazione. Se si prevede di richiedere istanze Spot, è necessari

Funzione	Ruolo predefinito	Descrizione	Policy gestita predefinita
			o aggiornare questa policy per includere un'istruzione che consenta la creazione di questo ruolo collegato al servizio. Per ulteriori informazioni, consulta Ruolo di servizio per Amazon EMR (ruolo EMR) e Ruolo collegato ai servizi per le richieste di istanza Spot nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Funzione	Ruolo predefinito	Descrizione	Policy gestita predefinita
Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2)	`EMR_EC2_DefaultRole`	I processi di applicazioni eseguiti supportate dall'ecosistema Hadoop sulle istanze del cluster utilizzano questo ruolo durante la chiamata di altri servizi AWS. Per accedere ai dati in Amazon S3 utilizzando EMRFS, è possibile specificare diversi ruoli da assumere in base alla posizione dei dati in Amazon S3. Ad esempio, più team possono accedere a un singolo "account di archiviazione" dei dati di Amazon S3. Per ulteriori informazioni, consulta Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3. Questo ruolo è obbligatorio per tutti i cluster.	`AmazonElasticMapReduceforEC2Role`. Per ulteriori informazioni, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2).

Funzione	Ruolo predefinito	Descrizione	Policy gestita predefinita
Ruolo di servizio per il dimensionamento automatico in Amazon EMR (ruolo Auto Scaling)	`EMR_AutoScaling_DefaultRole`	Consente di eseguire azioni aggiuntive per ambienti con dimensionamento dinamico. Necessario solo per i cluster che utilizzano la scalabilità automatica in Amazon EMR. Per ulteriori informazioni, consulta Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze .	`AmazonElasticMapReduceforAutoScalingRole` . Per ulteriori informazioni, consulta Ruolo di servizio per il dimensionamento automatico in Amazon EMR (ruolo Auto Scaling).

Funzione	Ruolo predefinito	Descrizione	Policy gestita predefinita
Ruolo di servizio per EMR Notebooks	`EMR_Notebooks_DefaultRole`	Fornisce le autorizzazioni richieste da un notebook EMR per accedere ad altre risorse AWS ed eseguire operazioni. Richiesto solo se si utilizza EMR Notebooks.	`AmazonElasticMapReduceEditorsRole` . Per ulteriori informazioni, consulta Ruolo di servizio per EMR Notebooks. `S3FullAccessPolicy` è inoltre collegato per impostazione predefinita. Il contenuto di questa policy è mostrato di seguito. `{ "Version": "2012-10-17", "Statement": [ { "Effect": "Allow", "Action": "s3:*", "Resource": "*" } ] }`

Funzione	Ruolo predefinito	Descrizione	Policy gestita predefinita
Ruolo collegato ai servizi	`AWSServiceRoleForEMRCleanup`	Amazon EMR crea automaticamente un ruolo collegato ai servizi. Se il servizio per Amazon EMR ha perso la capacità di eliminare risorse Amazon EC2, Amazon EMR può utilizzare questo ruolo per farlo. Se un cluster usa le istanze Spot, le policy di autorizzazione associate a Ruolo di servizio per Amazon EMR (ruolo EMR) devono consentire la creazione di un ruolo collegato al servizio. Per ulteriori informazioni, consulta Autorizzazioni del ruolo collegato ai servizi per Amazon EMR.	`AmazonEMRCleanupPolicy`

Argomenti

- Ruoli di servizio IAM utilizzati da Amazon EMR
- Personalizzazione dei ruoli IAM
- Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3
- Utilizzo di policy basate su risorse per l'accesso Amazon EMR ad AWS Glue Data Catalog

- I ruoli IAM possono essere utilizzati con le applicazioni che richiamano direttamente i servizi AWS
- Consentire a utenti e gruppi di creare e modificare i ruoli

Ruoli di servizio IAM utilizzati da Amazon EMR

Amazon EMR impiega i ruoli di servizio IAM per eseguire operazioni per conto dell'utente durante il provisioning delle risorse cluster, l'esecuzione di applicazioni, il dimensionamento dinamico delle risorse e la creazione ed esecuzione di EMR Notebooks. Amazon EMR utilizza i seguenti ruoli durante l'interazione con altri servizi AWS. Ogni ruolo dispone di una funzione univoca all'interno di Amazon EMR. Gli argomenti in questa sezione descrivono la funzione del ruolo e forniscono i ruoli e le policy di autorizzazioni predefiniti per ciascun ruolo.

Se sul cluster è presente un codice applicativo che chiama direttamente i servizi AWS, potrebbe essere necessario utilizzare l'SDK per specificare i ruoli. Per ulteriori informazioni, consulta I ruoli IAM possono essere utilizzati con le applicazioni che richiamano direttamente i servizi AWS.

Argomenti

- Ruolo di servizio per Amazon EMR (ruolo EMR)
- Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2)
- Ruolo di servizio per il dimensionamento automatico in Amazon EMR (ruolo Auto Scaling)
- Ruolo di servizio per EMR Notebooks
- Utilizzo del ruolo collegato ai servizi per Amazon EMR

Ruolo di servizio per Amazon EMR (ruolo EMR)

Il ruolo di servizio di Amazon EMR definisce le azioni consentite ad Amazon EMR quando effettua il provisioning delle risorse ed esegue attività a livello di servizio che non vengono eseguite nel contesto di un'istanza Amazon EC2 in esecuzione in un cluster. Ad esempio, il ruolo del servizio viene utilizzato per effettuare il provisioning di istanze EC2 quando viene avviato un cluster.

- Il nome del ruolo predefinito è `EMR_DefaultRole_V2`.
- La policy gestita predefinita necessaria per Amazon EMR e associata a `EMR_DefaultRole_V2` è `AmazonEMRServicePolicy_v2`. Questa policy v2 sostituisce la policy gestita predefinita e obsoleta `AmazonElasticMapReduceRole`.

`AmazonEMRServicePolicy_v2` dipende dall'accesso ridotto alle risorse che Amazon EMR fornisce o utilizza. Quando si utilizza questa policy, è necessario passare il tag utente `for-use-with-amazon-emr-managed-policies = true` durante il provisioning del cluster. Amazon EMR propaga automaticamente tali tag. Inoltre, potrebbe essere necessario aggiungere manualmente un tag utente a tipi specifici di risorse, ad esempio gruppi di sicurezza EC2 che non sono stati creati da Amazon EMR. Per informazioni, consultare Assegnazione di tag alle risorse per l'utilizzo delle policy gestite.

> **Important**
>
> Amazon EMR utilizza questo ruolo di servizio Amazon EMR e il ruolo `AWSServiceRoleForEMRCleanup` per pulire le risorse del cluster del tuo account che non usi più, come le istanze Amazon EC2. È necessario includere azioni relative alle policy del ruolo per eliminare o terminare le risorse. Altrimenti, Amazon EMR non può eseguire queste azioni di pulizia e potresti incorrere in costi per le risorse inutilizzate che rimangono nel cluster.

Nell'esempio seguente viene mostrato il contenuto della policy `AmazonEMRServicePolicy_v2` corrente. È anche possibile visualizzare il contenuto corrente della policy gestita `AmazonEMRServicePolicy_v2` sulla console 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*"
     }
    }
   }
  ]
```

```
}
```

Il tuo ruolo di servizio dovrebbe utilizzare la seguente policy di attendibilità:

> **Important**
>
> La policy di attendibilità seguente include le chiavi di condizione globale `aws:SourceArn` e `aws:SourceAccount` per limitare le autorizzazioni concesse ad Amazon EMR per determinate risorse dell'account. Il loro utilizzo può proteggerti dal problema del "confused deputy".

```
{
    "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>:*"
                }
            }
        }
    ]
}
```

Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2)

Il ruolo di servizio per le istanze EC2 del cluster (detto anche profilo dell'istanza EC2 per Amazon EMR) è un tipo speciale di ruolo di servizio assegnato a ogni istanza EC2 in un cluster Amazon EMR quando l'istanza viene avviata. I processi delle applicazioni eseguite sull'ecosistema Hadoop presuppongono che questo ruolo per le autorizzazioni interagisca con altri servizi AWS.

Per ulteriori informazioni sui ruoli di servizio per le istanze EC2, consulta Utilizzo di un ruolo IAM per concedere autorizzazioni ad applicazioni in esecuzione su istanze di Amazon EC2 nella Guida per l'utente IAM.

> **Important**
>
> Il ruolo di servizio predefinito per le istanze EC2 del cluster e la relativa policy gestita predefinita AWS, `AmazonElasticMapReduceforEC2Role`, diventeranno presto obsoleti senza prevedere alcuna policy gestita AWS sostitutiva. Sarà necessario creare e specificare un profilo di istanza per sostituire il ruolo obsoleto e la policy predefinita.

Ruolo predefinito e policy gestita

- Il nome del ruolo predefinito è `EMR_EC2_DefaultRole`.
- Il supporto per la policy gestita da `EMR_EC2_DefaultRole` predefinita (`AmazonElasticMapReduceforEC2Role`) è quasi al termine. Invece di utilizzare una policy gestita predefinita per il profilo dell'istanza EC2, applica policy basate sulle risorse ai bucket S3 e ad altre risorse di cui Amazon EMR necessita, oppure utilizza la policy gestita dal cliente con un ruolo IAM come profilo dell'istanza. Per ulteriori informazioni, consulta Creazione di un ruolo di servizio per le istanze EC2 del cluster con le autorizzazioni con privilegi minimi.

Di seguito viene mostrato il contenuto della versione 3 di `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"
            ]
        }
    ]
}
```

Il tuo ruolo di servizio dovrebbe utilizzare la seguente policy di attendibilità:

```
{
    "Version": "2008-10-17",
    "Statement": [
        {
            "Sid": "",
            "Effect": "Allow",
            "Principal": {
                "Service": "ec2.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

Creazione di un ruolo di servizio per le istanze EC2 del cluster con le autorizzazioni con privilegi minimi

Come best practice, è fortemente consigliato creare un ruolo di servizio per le istanze EC2 del cluster e una policy di autorizzazione che abbia le autorizzazioni minime ad altri servizi AWS richiesti dalla tua applicazione.

La policy gestita predefinita `AmazonElasticMapReduceforEC2Role` offre le autorizzazioni che facilitano l'avvio del primo cluster. Tuttavia, `AmazonElasticMapReduceforEC2Role` diventerà presto obsoleta e Amazon EMR non fornirà una policy gestita predefinita AWS sostitutiva per il ruolo divenuto obsoleto. Per avviare un cluster iniziale, è necessario fornire una policy basata sulle risorse gestita dal cliente o basata su ID.

Le seguenti dichiarazioni di policy forniscono esempi di autorizzazioni richieste per differenti caratteristiche di Amazon EMR. È consigliabile utilizzare queste autorizzazioni per la creazione di una policy di autorizzazioni che limita l'accesso alle sole caratteristiche e risorse che richiede il cluster. Tutte le dichiarazioni delle policy di esempio utilizzano la Regione *`us-west-2`* e l'ID account AWS fittizio *`123456789012`*. Sostituirli nel modo appropriato per il cluster.

Per ulteriori informazioni sulla creazione e sula specifica di ruoli personalizzati, consulta Personalizzazione dei ruoli IAM.

> **Note**
>
> Se crei un ruolo EMR personalizzato per EC2, segui il flusso di lavoro di base che crea automaticamente un profilo di istanza con lo stesso nome. Amazon EC2 consente di creare profili di istanza e ruoli con nomi diversi, ma Amazon EMR non supporta questa

configurazione e genera l'errore "invalid instance profile (profilo di istanza non valido)" quando crei il cluster.

Lettura e scrittura di dati su Amazon S3 utilizzando EMRFS

Quando un'applicazione in esecuzione su un cluster Amazon EMR riferisce i dati utilizzando il formato `s3://mydata`, Amazon EMR utilizza il profilo dell'istanza EC2 per effettuare la richiesta. In genere, i cluster leggono e scrivono dati su Amazon S3 in questo modo e Amazon EMR utilizza le autorizzazioni associate al ruolo di servizio per le istanze EC2 del cluster per impostazione predefinita. Per ulteriori informazioni, consulta Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3.

Poiché i ruoli IAM per EMRFS verranno ripristinati alle autorizzazioni associate al ruolo di servizio per le istanze EC2 del cluster, come best practice ti consigliamo di utilizzare i ruoli IAM per EMRFS e limitare le autorizzazioni EMRFS e Amazon S3 associate al ruolo di servizio per le istanze EC2 del cluster.

L'istruzione di esempio di seguito mostra le autorizzazioni che EMRFS richiede per effettuare le richieste ad Amazon S3.

- *my-data-bucket-in-s3-for-emrfs-reads-and-writes* specifica il bucket in Amazon S3 dove il cluster legge e scrive i dati e tutte le sottocartelle utilizzando */**. Aggiungere solo i bucket e le cartelle necessari per la propria applicazione.
- La dichiarazione di policy che consente le azioni dynamodb è necessaria solo se è abilitata la visualizzazione coerente di EMRFS. *EmrFSMetadata* specifica la cartella predefinita per la visualizzazione coerente di 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-*"
        }
    ]
}
```

Archiviazione di file di log in Amazon S3

La seguente istruzione di policy consente al cluster Amazon EMR di archiviare i file di log nel percorso Amazon S3 specificato. Nell'esempio sottostante, quando il cluster è stato creato, *s3://MyLoggingBucket/MyEMRClusterLogs* è stato specificato utilizzando il Percorso cartella di log S3 nella console, utilizzando l'opzione `--log-uri` dalla AWS CLI oppure utilizzando il parametro `LogUri` nel comando `RunJobFlow`. Per ulteriori informazioni, consulta Archiviazione di file di log in Amazon S3.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:PutObject",
            "Resource": "arn:aws:s3:::MyLoggingBucket/MyEMRClusterLogs/*"
        }
    ]
}
```

Utilizzo degli strumenti di debug

L'istruzione di policy seguente consente operazioni necessarie se si abilita lo strumento di debug Amazon EMR. L'archiviazione dei file di log Amazon S3 e le autorizzazioni associate mostrate nell'esempio in alto sono necessarie per il debug. Per ulteriori informazioni, consulta Abilitare lo strumento di debug.

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "sqs:GetQueueUrl",
                "sqs:SendMessage"
            ],
            "Resource": "arn:aws:sqs:us-west-2:123456789012:AWS-ElasticMapReduce-*"
        }
    ]
}
```

Utilizzo di AWS Glue Data Catalog

L'istruzione di policy seguente consente operazioni che sono necessarie se si utilizza AWS Glue Data Catalog come metastore per le applicazioni. Per ulteriori informazioni, consulta Utilizzo di AWS Glue Data Catalog come metastore per Spark SQL, Utilizzo di AWS Glue Data Catalog come metastore per Hive e Utlizzo di Presto con AWS Glue Data Catalog nella Guida ai rilasci di 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": "*",
        }
    ]
}
```

Ruolo di servizio per il dimensionamento automatico in Amazon EMR (ruolo Auto Scaling)

Il ruolo Auto Scaling per Amazon EMR svolge una funzione simile a quella del ruolo di servizio, ma consente azioni aggiuntive per ambienti di scalabilità dinamica.

- Il nome del ruolo predefinito è `EMR_AutoScaling_DefaultRole`.
- La policy gestita predefinita associata a `EMR_AutoScaling_DefaultRole` è `AmazonElasticMapReduceforAutoScalingRole`.

I contenuti della versione 1 di `AmazonElasticMapReduceforAutoScalingRole` sono elencati di seguito.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "cloudwatch:DescribeAlarms",
                "elasticmapreduce:ListInstanceGroups",
                "elasticmapreduce:ModifyInstanceGroups"
            ],
            "Effect": "Allow",
            "Resource": "*"
        }
    ]
}
```

Il tuo ruolo di servizio dovrebbe utilizzare la seguente policy di attendibilità:

> **Important**
>
> La policy di attendibilità seguente include le chiavi di condizione globale `aws:SourceArn` e `aws:SourceAccount` per limitare le autorizzazioni concesse ad Amazon EMR per determinate risorse dell'account. Il loro utilizzo può proteggerti dal problema del "confused deputy".

```
{
    "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>:*"
                }
            }
        }
    ]
}
```

Ruolo di servizio per EMR Notebooks

Ogni notebook EMR necessita di autorizzazioni per accedere ad altre risorse AWS ed eseguire operazioni. Le policy IAM collegate a questo ruolo di servizio forniscono al notebook le autorizzazioni per interagire con altri servizi AWS. Quando crei un notebook utilizzando la AWS Management Console, specifichi un Ruolo di servizio AWS. È possibile utilizzare il ruolo predefinito, `EMR_Notebooks_DefaultRole`, oppure specificare un ruolo creato. Se un notebook non è stato creato in precedenza, è possibile scegliere di creare il ruolo predefinito.

- Il nome del ruolo predefinito è `EMR_Notebooks_DefaultRole`.
- Le policy gestite di default allegate a `EMR_Notebooks_DefaultRole` sono `AmazonElasticMapReduceEditorsRole` e `S3FullAccessPolicy`.

Il tuo ruolo di servizio dovrebbe utilizzare la seguente policy di attendibilità:

> **Important**
>
> La policy di attendibilità seguente include le chiavi di condizione globale `aws:SourceArn` e `aws:SourceAccount` per limitare le autorizzazioni concesse ad Amazon EMR per determinate risorse dell'account. Il loro utilizzo può proteggerti dal problema del "confused deputy".

```
{
    "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>:*"
                }
            }
        }
    ]
}
```

Il contenuto della versione 1 di `AmazonElasticMapReduceEditorsRole` è il seguente.

```
{
    "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"
                    ]
                }
            }
        }
    ]
}
```

Di seguito sono riportati il contenuto di `S3FullAccessPolicy`. La `S3FullAccessPolicy` consente al tuo ruolo di servizio per i EMR Notebooks di eseguire tutte le operazioni di Amazon S3

sugli oggetti nel Account AWS. Quando crei un ruolo di servizio personalizzato per i EMR Notebooks, devi assegnare al tuo ruolo di servizio le autorizzazioni Amazon S3.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "s3:*",
            "Resource": "*"
        }
    ]
}
```

È possibile individuare l'accesso in lettura e scrittura per il ruolo di servizio nella posizione Amazon S3 in cui si desidera salvare i file del notebook. Utilizza il seguente set minimo di autorizzazioni Amazon S3.

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

Se il bucket Amazon S3 è crittografato, devi includere le seguenti autorizzazioni per AWS Key Management Service.

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

Quando si collegano i repository Git al notebook e si deve creare un segreto per il repository, è necessario aggiungere l'autorizzazione `secretsmanager:GetSecretValue` nella policy IAM collegata al ruolo di servizio per i notebooks Amazon EMR. Di seguito è illustrato un esempio di policy:

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

```
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": "secretsmanager:GetSecretValue",
            "Resource": "*"
        }
    ]
}
```

Autorizzazioni del ruolo di servizio EMR Notebooks

In questa tabella sono elencate le azioni eseguite da EMR Notebooks utilizzando il ruolo di servizio, assieme alle autorizzazioni necessarie per ogni azione.

Azione	Autorizzazioni
Stabilisci un canale di rete protetto tra un notebook e un cluster Amazon EMR ed esegui le azioni di pulizia necessarie.	`"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"`
Utilizza le credenziali Git memorizzate in AWS Secrets Manager per collegare i repository Git a un notebook.	`"secretsmanager:GetSecretValue"`

Azione	Autorizzazioni
Applica tag AWS all'interfaccia di rete e ai gruppi di sicurezza predefiniti creati da EMR Notebooks durante la configurazione del canale di rete sicuro. Per ulteriori informazioni, consulta Assegnazione di tag alle risorse AWS.	`"ec2:CreateTags"`

Accedi o carica i file notebook e i metadati in Amazon S3.	`"s3:PutObject",`
	`"s3:GetObject",`
	`"s3:GetEncryptionConfiguration",`
	`"s3:ListBucket",`
	`"s3:DeleteObject"`
	
	Le seguenti autorizzazioni sono necessarie solo se utilizzi un bucket Amazon S3 crittografato.
	
	`"kms:Decrypt",`
	`"kms:GenerateDataKey",`
	`"kms:ReEncryptFrom",`
	`"kms:ReEncryptTo",`
	`"kms:DescribeKey"`

Aggiornamenti di EMR Notebooks sulle policy gestite da AWS

Visualizza i dettagli sugli aggiornamenti alle policy gestite da AWS per EMR Notebook dal 1° marzo 2021.

Modifica	Descrizione	Data
`AmazonElasticMapReduceEditorsRole - Added permissions`	EMR Notebooks ha aggiunto le autorizzazioni `ec2:describeVPCs` e `elastmicmapreduce:ListSteps`	8 febbraio 2023

Modifica	Descrizione	Data
	a `AmazonElasticMapReduceEditorsRole` .	
EMR Notebooks ha avviato il tracciamento delle modifiche	EMR Notebook ha avviato il tracciamento delle modifiche per le sue policy gestite da AWS.	8 febbraio 2023

Utilizzo del ruolo collegato ai servizi per Amazon EMR

Amazon EMR utilizza ruoli collegati ai servizi AWS Identity and Access Management (IAM)https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_terms-and-concepts.html#iam-term-service-linked-role. Un ruolo collegato ai servizi è un tipo di ruolo IAM univoco collegato direttamente ad Amazon EMR. Il ruolo collegato ai servizi è predefinito da Amazon EMR e include le autorizzazioni necessarie ad Amazon EMR per chiamare automaticamente Amazon EC2 ed eliminare le risorse del cluster dopo che non sono più in uso. Il ruolo collegato ai servizi viene utilizzato in combinazione con il ruolo di servizio Amazon EMR e il profilo dell'istanza Amazon EC2 per Amazon EMR. Per ulteriori informazioni sul ruolo del servizio e il profilo dell'istanza, consulta Configurazione dei ruoli di servizio IAM per le autorizzazioni di Amazon EMR per i servizi e risorse AWS.

Amazon EMR definisce le autorizzazioni di questo ruolo relativo ai servizi e, salvo diversamente definito, solo Amazon EMR potrà assumere tale ruolo. Le autorizzazioni definite includono la policy di attendibilità e la policy delle autorizzazioni che non può essere collegata a nessun'altra entità IAM. È possibile eliminare il ruolo solo dopo aver chiuso tutti i cluster EMR nell'account.

Per informazioni sugli altri servizi che supportano i ruoli collegati ai servizi, consultare Servizi AWS che funzionano con IAM e cercare i servizi che riportano Yes (Sì) nella colonna Service-Linked Role (Ruoli collegati ai servizi). Scegli Yes (Sì) in corrispondenza di un link per visualizzare la documentazione relativa al ruolo collegato ai servizi per tale servizio.

Autorizzazioni del ruolo collegato ai servizi per Amazon EMR

Amazon EMR utilizza il ruolo AWSServiceRoleForEMRCleanup, ossia un ruolo basato sul servizio che consente ad Amazon EMR di terminare ed eliminare le risorse Amazon EC2 per conto dell'utente se il ruolo del servizio Amazon EMR ha perso tale capacità. Amazon EMR crea automaticamente il ruolo durante la creazione del cluster, se non esiste già.

Ai fini dell'assunzione del ruolo, il ruolo collegato ai servizi AWSServiceRoleForEMRCleanup considera attendibili i seguenti servizi:

- `elasticmapreduce.amazonaws.com`

La policy delle autorizzazioni del ruolo collegato ai servizi AWSServiceRoleForEMRCleanup consente ad Amazon EMR di eseguire le seguenti operazioni sulle risorse specificate:

- Operazione: `DescribeInstances` su `ec2`
- Operazione: `DescribeSpotInstanceRequests` su `ec2`
- Operazione: `ModifyInstanceAttribute` su `ec2`
- Operazione: `TerminateInstances` su `ec2`
- Operazione: `CancelSpotInstanceRequests` su `ec2`
- Operazione: `DeleteNetworkInterface` su `ec2`
- Operazione: `DescribeInstanceAttribute` su `ec2`
- Operazione: `DescribeVolumeStatus` su `ec2`
- Operazione: `DescribeVolumes` su `ec2`
- Operazione: `DetachVolume` su `ec2`
- Operazione: `DeleteVolume` su `ec2`

Per consentire a un'entità IAM (come un utente, un gruppo o un ruolo) di creare, modificare o eliminare un ruolo collegato ai servizi devi configurare le relative autorizzazioni.

Consentire a un'entità IAM di creare il ruolo collegato ai servizi AWSServiceRoleForEMRCleanup

Aggiungi la seguente istruzione alla policy delle autorizzazioni per l'entità IAM che deve creare il ruolo collegato ai servizi:

```
{
    "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"
            ]
        }
    }
}
```

Consentire a un'entità IAM di modificare la descrizione del ruolo collegato ai servizi AWSServiceRoleForEMRCleanup

Aggiungi la seguente istruzione alla policy delle autorizzazioni per l'entità IAM che deve modificare la descrizione di un ruolo collegato ai servizi:

```
{
    "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"
            ]
        }
    }
}
```

Consentire a un'entità IAM di eliminare il ruolo collegato ai servizi AWSServiceRoleForEMRCleanup

Aggiungi la seguente istruzione alla policy delle autorizzazioni per l'entità IAM che deve eliminare un ruolo collegato ai servizi:

```
{
    "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"
            ]
        }
    }
}
```

Creazione di un ruolo collegato ai servizi per Amazon EMR

Non è necessario creare manualmente il ruolo AWSServiceRoleForEMRCleanup. Quando si avvia un cluster, per la prima volta o quando non è presente un ruolo collegato ai servizi, Amazon EMR crea il ruolo collegato ai servizi per conto dell'utente. Per creare il ruolo collegato ai servizi, devi disporre delle autorizzazioni . Per un esempio di istruzione che aggiunge questa funzionalità alla policy di autorizzazione di un'entità IAM (ad esempio un utente, gruppo o ruolo), consulta Autorizzazioni del ruolo collegato ai servizi per Amazon EMR.

> **Important**
>
> Se utilizzavi Amazon EMR prima del 24 ottobre 2017, data da cui è disponibile il supporto dei ruoli collegati ai servizi, Amazon EMR ha creato il ruolo AWSServiceRoleForEMRCleanup nel tuo account. Per ulteriori informazioni, consulta Comparsa di un nuovo ruolo nell'account IAM.

Modifica di un ruolo collegato ai servizi per Amazon EMR

Amazon EMR non consente di modificare il ruolo collegato ai servizi AWSServiceRoleForEMRCleanup. Dopo aver creato un ruolo collegato ai servizi, non potrai modificarne il nome perché varie entità potrebbero farvi riferimento. È possibile tuttavia modificarne la descrizione utilizzando IAM.

Modifica della descrizione di un ruolo collegato ai servizi (console IAM)

È possibile utilizzare la console IAM per modificare la descrizione di un ruolo collegato ai servizi.

Per modificare la descrizione di un ruolo collegato ai servizi (console)

1. Nel pannello di navigazione della console IAM seleziona Roles (Ruoli).
2. Scegliere il nome del ruolo da modificare.
3. Nella parte destra di Role description (Descrizione ruolo), scegli Edit (Modifica).
4. Digita una nuova descrizione nella casella e scegli Save changes (Salva modifiche).

Modifica della descrizione di un ruolo collegato ai servizi (CLI IAM)

Puoi utilizzare i comandi IAM dalla AWS Command Line Interface per modificare la descrizione di un ruolo collegato ai servizi.

Per modificare la descrizione di un ruolo collegato ai servizi (CLI)

1. (Facoltativo) Per visualizzare la descrizione attuale di un ruolo, utilizza i seguenti comandi:

   ```
   $ aws iam get-role --role-name role-name
   ```

   Per fare riferimento ai ruoli con i comandi della CLI utilizza il nome del ruolo, non l'ARN. Ad esempio, per fare riferimento a un ruolo il cui ARN è `arn:aws:iam::123456789012:role/myrole`, puoi usare **`myrole`**.
2. Per aggiornare la descrizione di un ruolo collegato ai servizi, utilizza uno dei seguenti comandi:

   ```
   $ aws iam update-role-description --role-name role-name --description description
   ```

Modifica della descrizione di un ruolo collegato ai servizi (API IAM)

È possibile utilizzare l'API IAM per modificare la descrizione di un ruolo collegato ai servizi.

Per modificare la descrizione di un ruolo collegato ai servizi (API)

1. (Facoltativo) Per visualizzare la descrizione attuale di un ruolo, utilizza il seguente comando:

   API IAM: GetRole
2. Per aggiornare la descrizione di un ruolo, utilizza il seguente comando:

   API IAM: UpdateRoleDescription

Eliminazione di un ruolo collegato ai servizi per Amazon EMR

Se non è più necessario utilizzare una funzionalità o un servizio che richiede un ruolo collegato ai servizi, ti consigliamo di eliminare il ruolo. In questo modo non sarà più presente un'entità non utilizzata che non viene monitorata e gestita attivamente. Tuttavia, è necessario effettuare la pulizia delle risorse associate al ruolo collegato ai servizi prima di poterlo eliminare.

Pulizia di un ruolo collegato ai servizi

Prima di utilizzare IAM per eliminare un ruolo collegato ai servizi, devi innanzitutto verificare che il ruolo non abbia sessioni attive ed eliminare tutte le risorse utilizzate dal ruolo.

Per verificare se il ruolo collegato ai servizi dispone di una sessione attiva nella console IAM

1. Apri la console IAM all'indirizzo https://console.aws.amazon.com/iam/.
2. Nel pannello di navigazione, seleziona Ruoli. Seleziona il nome (non la casella di controllo) del ruolo AWSServiceRoleForEMRCleanup.
3. Nella pagina Summary (Riepilogo) per il ruolo selezionato, scegli Access Advisor (Consulente accessi).
4. Nella scheda Access Advisor (Consulente accessi) esaminare l'attività recente per il ruolo collegato ai servizi.

> **Note**
>
> Se non hai la certezza che Amazon EMR stia utilizzando il ruolo AWSServiceRoleForEMRCleanup, puoi provare a eliminarlo. Se il servizio sta utilizzando il ruolo, l'eliminazione non andrà a buon fine e potrai visualizzare le regioni in cui il ruolo viene utilizzato. Se il ruolo è in uso, prima di poterlo eliminare dovrai attendere il termine della sessione. Non puoi revocare la sessione per un ruolo collegato ai servizi.

Rimozione delle risorse Amazon EMR utilizzate da AWSServiceRoleForEMRCleanup

- Chiudi tutti i cluster del tuo account. Per ulteriori informazioni, consulta Terminazione di un cluster.

Eliminazione di un ruolo collegato ai servizi (console IAM)

È possibile utilizzare la console IAM per eliminare un ruolo collegato ai servizi.

Per eliminare un ruolo collegato ai servizi (console)

1. Apri la console IAM all'indirizzo https://console.aws.amazon.com/iam/.
2. Nel pannello di navigazione, seleziona Ruoli. Seleziona la casella di controllo accanto ad AWSServiceRoleForEMRCleanup, non il nome o la riga.
3. In operazioni Role (Ruolo) nella parte superiore della pagina, seleziona Delete (Elimina) ruolo.
4. Nella finestra di dialogo di conferma controlla i dati relativi all'ultimo accesso ai servizi, che indicano quando ognuno dei ruoli selezionati ha effettuato l'accesso a un servizio AWS. In questo modo potrai verificare se il ruolo è attualmente attivo. Per procedere, seleziona Yes, Delete (Sì, elimina).
5. Controlla le notifiche della console IAM per monitorare lo stato dell'eliminazione del ruolo collegato ai servizi. Poiché l'eliminazione del ruolo collegato ai servizi IAM è asincrona, una volta richiesta l'eliminazione del ruolo, il task di eliminazione può essere eseguito correttamente o meno. Se il task non viene eseguito correttamente, puoi scegliere View details (Visualizza dettagli) o View Resources (Visualizza risorse) dalle notifiche per capire perché l'eliminazione non è stata effettuata. Se l'eliminazione non viene eseguita perché vi sono risorse nel servizio che sono usate dal ruolo, il motivo dell'errore include un elenco di risorse.

Eliminazione di un ruolo collegato ai servizi (CLI IAM)

Puoi utilizzare i comandi IAM dalla AWS Command Line Interface per eliminare un ruolo collegato ai servizi. Poiché un ruolo collegato ai servizi non può essere eliminato se è in uso o se a esso sono associate delle risorse, occorre inviare una richiesta di eliminazione. Se queste condizioni non sono soddisfatte, la richiesta può essere rifiutata.

Per eliminare un ruolo collegato ai servizi (CLI)

1. Per controllare lo stato dell'attività di eliminazione, devi acquisire il `deletion-task-id` dalla risposta. Per inviare una richiesta di eliminazione per un ruolo collegato ai servizi, digita il seguente comando:

```
$ aws iam delete-service-linked-role --role-name AWSServiceRoleForEMRCleanup
```

2. Digita il seguente comando per verificare lo stato del task di eliminazione:

```
$ aws iam get-service-linked-role-deletion-status --deletion-task-id deletion-task-id
```

Lo stato di un task di eliminazione può essere `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED` o `FAILED`. Se l'eliminazione non viene eseguita correttamente, la chiamata restituisce il motivo dell'errore per consentire all'utente di risolvere il problema.

Eliminazione di un ruolo collegato ai servizi (API IAM)

È possibile utilizzare l'API IAM per eliminare un ruolo collegato ai servizi. Poiché un ruolo collegato ai servizi non può essere eliminato se è in uso o se a esso sono associate delle risorse, occorre inviare una richiesta di eliminazione. Se queste condizioni non sono soddisfatte, la richiesta può essere rifiutata.

Per eliminare un ruolo collegato ai servizi (API)

1. Per inviare una richiesta di eliminazione per un ruolo collegato ai servizi, chiama DeleteServiceLinkedRole. Nella richiesta, specifica il nome del ruolo AWSServiceRoleForEMRCleanup.

   Per controllare lo stato dell'attività di eliminazione, devi acquisire il `DeletionTaskId` dalla risposta.
2. Per controllare lo stato dell'eliminazione, chiama GetServiceLinkedRoleDeletionStatus. Nella richiesta, specificare il `DeletionTaskId`.

   Lo stato di un task di eliminazione può essere `NOT_STARTED`, `IN_PROGRESS`, `SUCCEEDED` o `FAILED`. Se l'eliminazione non viene eseguita correttamente, la chiamata restituisce il motivo dell'errore per consentire all'utente di risolvere il problema.

Regioni supportate per i ruoli collegati ai servizi di Amazon EMR

Amazon EMR supporta l'utilizzo di ruoli collegati ai servizi nelle seguenti Regioni.

Nome Regione	Identità della regione	Supporto in Amazon EMR
Stati Uniti orientali (Virginia settentrionale)	us-east-1	Sì
Stati Uniti orientali (Ohio)	us-east-2	Sì

Nome Regione	Identità della regione	Supporto in Amazon EMR
Stati Uniti occidentali (California settentrionale)	us-west-1	Sì
Stati Uniti occidentali (Oregon)	us-west-2	Sì
Asia Pacifico (Mumbai)	ap-south-1	Sì
Asia Pacifico (Osaka-Locale)	ap-northeast-3	Sì
Asia Pacifico (Seul)	ap-northeast-2	Sì
Asia Pacifico (Singapore)	ap-southeast-1	Sì
Asia Pacifico (Sydney)	ap-southeast-2	Sì
Asia Pacifico (Tokyo)	ap-northeast-1	Sì
Canada (Centrale)	ca-central-1	Sì
Europe (Frankfurt)	eu-central-1	Sì
Europa (Irlanda)	eu-west-1	Sì
Europa (Londra)	eu-west-2	Sì
Europe (Paris)	eu-west-3	Sì
Sud America (São Paulo)	sa-east-1	Sì

Personalizzazione dei ruoli IAM

È possibile personalizzare il ruolo del servizio IAM e le autorizzazioni per limitare i privilegi in base ai requisiti di sicurezza. Per personalizzare le autorizzazioni, si consiglia di creare nuovi ruoli e policy. Iniziare con le autorizzazioni nelle policy gestite per i ruoli predefiniti (ad esempio, `AmazonElasticMapReduceforEC2Role` e `AmazonElasticMapReduceRole`). Quindi, copiare e incollare il contenuto in nuove istruzione della policy, modificare le autorizzazioni come appropriato e collegare le policy di autorizzazione modificate ai ruoli creati. È necessario disporre delle

autorizzazioni IAM appropriate per lavorare con i ruoli e le policy. Per ulteriori informazioni, consulta Consentire a utenti e gruppi di creare e modificare i ruoli.

Se crei un ruolo EMR personalizzato per EC2, segui il flusso di lavoro di base che crea automaticamente un profilo di istanza con lo stesso nome. Amazon EC2 consente di creare profili di istanza e ruoli con nomi diversi, ma Amazon EMR non supporta questa configurazione e genera l'errore "invalid instance profile (profilo di istanza non valido)" quando crei il cluster.

> **Important**
>
> Le policy in linea non vengono aggiornate automaticamente quando i requisiti di servizio cambiano. Se si creano e si collegano policy in linea, occorre tenere presente che potrebbero verificarsi aggiornamenti del servizio responsabili di errori di autorizzazione imprevisti. Per ulteriori informazioni, consulta la sezione relativa a Policy gestite e policy inline nella Guida per l'utente IAM e Specifica dei ruoli IAM personalizzati durante la creazione di un cluster.

Per ulteriori informazioni sull'utilizzo dei ruoli IAM, consulta i seguenti argomenti nella Guida per l'utente IAM:

- Creazione di un ruolo per delegare le autorizzazioni a un servizio AWS
- Modifica di un ruolo
- Eliminazione di un ruolo

Specifica dei ruoli IAM personalizzati durante la creazione di un cluster

Quando crei un cluster, devi specificare il ruolo di servizio per Amazon EMR e il ruolo per il profilo dell'istanza Amazon EC2. L'utente che crea i cluster ha bisogno delle autorizzazione per recuperare e assegnare i ruoli ad Amazon EMR e alle istanze EC2. In caso contrario, viene visualizzato un errore Account utente non autorizzato a effettuare la chiamata EC2. Per ulteriori informazioni, consulta Consentire a utenti e gruppi di creare e modificare i ruoli.

Utilizzo della console per specificare i ruoli personalizzati

Quando si crea un cluster, è possibile specificare un ruolo di servizio personalizzato per Amazon EMR, un ruolo personalizzato per il profilo dell'istanza EC2 e un ruolo Auto Scaling personalizzato utilizzando Advanced options (Opzioni avanzate). Quando si utilizzano le Quick options (Opzioni rapide), il ruolo di servizio e il ruolo predefinito per il profilo di istanza EC2 sono specificati. Per ulteriori informazioni, consulta Ruoli di servizio IAM utilizzati da Amazon EMR.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Specifica di ruoli IAM personalizzati con la nuova console

Quando crei un cluster con la nuova console, devi specificare un ruolo di servizio personalizzato per Amazon EMR e un ruolo personalizzato per il profilo dell'istanza EC2. Per ulteriori informazioni, consulta Ruoli di servizio IAM utilizzati da Amazon EMR.

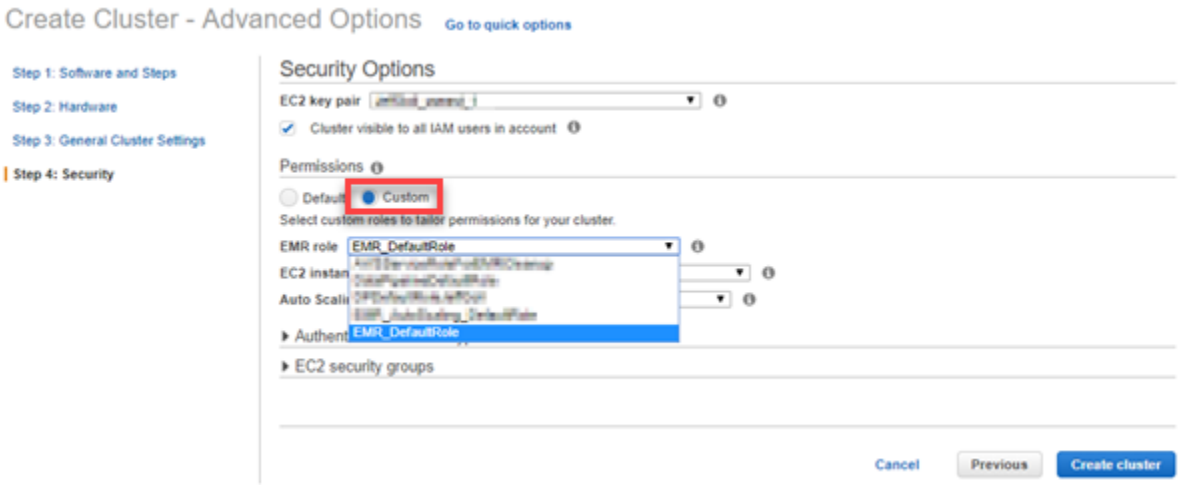
1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Security configuration and permissions (Configurazione e autorizzazioni di sicurezza), cerca i campi IAM role for instance profile (Ruolo IAM per il profilo dell'istanza) e Service role for Amazon EMR (Ruolo di servizio per i campi Amazon EMR). Per ogni tipo di ruolo, selezionare un ruolo dall'elenco. Vengono elencati solo i ruoli all'interno dell'account che dispongono della policy di affidabilità appropriata per quel tipo di ruolo.
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica di ruoli IAM personalizzati con la vecchia console

Quando crei un cluster con la vecchia console, puoi specificare un ruolo di servizio personalizzato per Amazon EMR, un ruolo personalizzato per il profilo dell'istanza EC2 e un ruolo Auto Scaling personalizzato utilizzando Advanced options (Opzioni avanzate). Quando si utilizzano le Quick options (Opzioni rapide), il ruolo di servizio e il ruolo predefinito per il profilo di istanza EC2 sono specificati. Per ulteriori informazioni, consulta Ruoli di servizio IAM utilizzati da Amazon EMR.

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Scegli le impostazioni del cluster appropriate per l'applicazione fino a raggiungere Security Options (Opzioni di protezione). In Permissions (Autorizzazioni), sono selezionati i ruoli Default (Predefiniti) per Amazon EMR.
4. Scegli Custom (Personalizzato).
5. Per ogni tipo di ruolo, selezionare un ruolo dall'elenco. Vengono elencati solo i ruoli all'interno dell'account che dispongono della policy di affidabilità appropriata per quel tipo di ruolo.



6. Scegli altre opzioni appropriate per il cluster, quindi scegli Create Cluster (Crea cluster).

Utilizzo della AWS CLI per specificare i ruoli personalizzati

È possibile specificare un ruolo di servizio per Amazon EMR e un ruolo di servizio per le istanze EC2 del cluster esplicitamente utilizzando le opzioni con il comando `create-cluster` da AWS CLI. L'opzione `--service-role` consente di specificare il ruolo di servizio. Utilizzare l'argomento `InstanceProfile` dell'opzione `--ec2-attributes` per specificare il ruolo del profilo dell'istanza EC2.

Il ruolo Auto Scaling viene specificato mediante un'opzione distinta, `--auto-scaling-role`. Per ulteriori informazioni, consulta Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze .

Specifica di ruoli IAM personalizzati mediante la AWS CLI

- Il comando seguente specifica il ruolo del servizio personalizzato, *MyCustomServiceRoleForEMR*, e un ruolo personalizzato per il profilo dell'istanza EC2, *MyCustomServiceRoleForClusterEC2Instances*, quando si avvia un cluster. Questo esempio utilizza il ruolo Amazon EMR predefinito.

  > **Note**
  >
  > I caratteri di continuazione della riga Linux (\) sono inclusi per la leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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
```

È possibile utilizzare queste opzioni per specificare esplicitamente i ruoli predefiniti anziché utilizzare l'opzione `--use-default-roles`. L'opzione `--use-default-roles` specifica il ruolo del servizio e il ruolo per il profilo dell'istanza EC2 definito nel file `config` per l'AWS CLI.

L'esempio seguente mostra i contenuti di un file `config` per la AWS CLI che specifica i ruoli personalizzati per Amazon EMR. Con questo file di configurazione, quando viene specificata l'opzione `--use-default-roles`, il cluster viene creato utilizzando *MyCustomServiceRoleForEMR* e *MyCustomServiceRoleForClusterEC2Instances*. Per impostazione predefinita, il file `config` specifica l'impostazione predefinita `service_role` come `AmazonElasticMapReduceRole` e l'impostazione predefinita `instance_profile` come `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
```

Configurazione di ruoli IAM per le richieste EMRFS ad Amazon S3

Quando un'applicazione in esecuzione su un cluster fa riferimento ai dati utilizzando il formato `s3://mydata`, Amazon EMR utilizza EMRFS per effettuare la richiesta. Per interagire con Amazon S3, EMRFS assume le policy di autorizzazione associate al Profilo dell'istanza Amazon EC2. Lo stesso profilo dell'istanza Amazon EC2 viene utilizzato indipendentemente dall'utente o dal gruppo che esegue l'applicazione o dalla posizione dei dati in Amazon S3.

Se si dispone di un cluster con più utenti che necessitano di diversi livelli di accesso ai dati in Amazon S3 tramite EMRFS, è possibile impostare una configurazione della sicurezza con i ruoli IAM per EMRFS. EMRFS può assumere un ruolo di servizio per le istanze EC2 del cluster differente in base all'utente o al gruppo da cui proviene la richiesta o in base alla posizione di dati in Amazon S3. Ogni ruolo IAM può avere differenti autorizzazioni per l'accesso ai dati in Amazon S3. Per ulteriori informazioni sull'utilizzo dei ruoli di servizio per le istanze EC2, del cluster, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2).

L'utilizzo di ruoli IAM personalizzati per EMRFS è supportato in Amazon EMR versioni 5.10.0 e successive. Se si utilizza una versione precedente o si dispone di requisiti che i ruoli IAM per EMRFS non possono soddisfare, è possibile creare un provider di credenziali personalizzato. Per ulteriori informazioni, consulta Autorizzazione dell'accesso ai dati EMRFS in Amazon S3.

Quando utilizzi una configurazione di sicurezza per specificare ruoli IAM per EMRFS, configuri mappature di ruoli. Ogni mappatura di ruoli specifica un ruolo IAM che corrisponde a identificatori. Questi identificatori determinano la base per l'accesso ad Amazon S3 tramite EMRFS. Gli identificatori possono essere utenti, gruppi o prefissi Amazon S3 che indicano un percorso di dati. Quando EMRFS invia una richiesta ad Amazon S3, se la richiesta corrisponde alla base per l'accesso, EMRFS dispone delle istanze EC2 del cluster che assumono il ruolo IAM corrispondente per la richiesta. Si applicano le autorizzazioni IAM associate a tale ruolo invece delle autorizzazioni IAM collegate al ruolo di servizio per le istanze EC2 del cluster.

Gli utenti e i gruppi in una mappatura di ruoli sono utenti e gruppi Hadoop definiti nel cluster. Gli utenti e i gruppi sono passati a EMRFS nel contesto dell'applicazione che lo utilizza (ad esempio, la rappresentazione di utente YARN). Il prefisso Amazon S3 può essere un identificatore bucket di qualsiasi profondità (ad esempio, `s3://mybucket` o `s3://mybucket/myproject/mydata`). Puoi specificare più identificatori in una singola mappatura di ruoli, ma devono essere tutti dello stesso tipo.

> **Important**
>
> I ruoli IAM per EMRFS forniscono l'isolamento a livello di applicazione tra gli utenti dell'applicazione. Non offrono l'isolamento a livello di host tra gli utenti sull'host. Qualsiasi utente con accesso al cluster può ignorare l'isolamento per assumere uno qualsiasi dei ruoli.

Quando un'applicazione cluster invia una richiesta ad Amazon S3 tramite EMRFS, EMRFS valuta le mappature di ruoli nell'ordine decrescente in cui sono visualizzate nella configurazione di sicurezza. Se una richiesta inviata tramite EMRFS non corrisponde ad alcun identificatore, EMRFS utilizza nuovamente il ruolo di servizio per le istanze EC2 del cluster. Per questo motivo, è consigliabile che le policy associate a questo ruolo limitino le autorizzazioni per Amazon S3. Per ulteriori informazioni, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2).

Configurazione dei ruoli

Prima di impostare una configurazione di sicurezza con ruoli IAM per EMRFS, pianifica e crea i ruoli e le policy di autorizzazione da associare ai ruoli. Per ulteriori informazioni, consulta Funzionamento dei ruoli per le istanze EC2 nella Guida per l'utente IAM. Durante la creazione di policy di autorizzazione, ti consigliamo di iniziare con la policy gestita associata al ruolo Amazon EMR di default per EC2, per poi modificarla in seguito in base alle tue esigenze. Il nome del ruolo predefinito è `EMR_EC2_DefaultRole` e la policy gestita predefinita da modificare è `AmazonElasticMapReduceforEC2Role`. Per ulteriori informazioni, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2).

Aggiornamento delle policy di affidabilità per le autorizzazioni del ruolo Assume

Ogni ruolo utilizzato da EMRFS deve avere una policy di attendibilità che consente al ruolo di Amazon EMR del cluster per EC2 di assumerlo. Analogamente, il ruolo Amazon EMR del cluster per EC2 deve avere una policy di attendibilità che consente ai ruoli EMRFS di assumerlo.

L'esempio di policy di trust riportata di seguito è collegata ai ruoli per EMRFS. L'istruzione consente al ruolo Amazon EMR predefinito per EC2 di assumere quel ruolo. Se ad esempio si dispone di due ruoli fittizi EMRFS, `EMRFSRole_First` ed `EMRFSRole_Second`, questa istruzione di policy viene aggiunta alla policy di trust per ognuno di essi.

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

```
      {
         "Effect":"Allow",
         "Principal":{
            "AWS":"arn:aws:iam::AWSAcctID:role/EMR_EC2_DefaultRole"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

Inoltre, la seguente istruzione per policy di trust di esempio viene aggiunta al ruolo `EMR_EC2_DefaultRole` per consentire ai due ruoli EMRFS fittizi di assumerlo.

```
{
   "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"
      }
   ]
}
```

Aggiornamento della policy di affidabilità di un ruolo IAM

Aprire la console IAM all'indirizzo https://console.aws.amazon.com/iam/.

1. Scegliere Roles (Ruoli), immettere il nome del ruolo in Search (Cerca), quindi selezionare il relativo Role name (Nome ruolo).
2. Scegliere Trust relationships (Relazioni di trust), quindi Edit trust relationship (Modifica relazione di trust).
3. Aggiungi un'istruzione di attendibilità in base a quanto riportato nel Policy Document (Documento di policy) secondo le linee guida in alto, quindi scegli Update Trust Policy (Aggiorna policy di attendibilità).

Indicazione di un ruolo come utente chiave

Se un ruolo autorizza l'accesso a una posizione in Amazon S3 crittografata mediante una AWS KMS key, assicurati che il ruolo sia specificato come utente di chiavi. Ciò concede al ruolo l'autorizzazione a utilizzare la chiave KMS. Per ulteriori informazioni, consulta Policy delle chiavi in AWS KMS nella Guida per gli sviluppatori di AWS Key Management Service.

Impostazione di una configurazione di sicurezza con ruoli IAM per EMRFS

> **Important**
>
> Se nessuno dei ruoli IAM per EMRFS che specifichi è applicabile, EMRFS utilizza il ruolo Amazon EMR per EC2. Valuta la possibilità di personalizzare questo ruolo per limitare le autorizzazioni per Amazon S3 come appropriato per la tua applicazione e di specificare questo ruolo personalizzato anziché `EMR_EC2_DefaultRole` quando crei un cluster. Per ulteriori informazioni, consulta Personalizzazione dei ruoli IAM e Specifica dei ruoli IAM personalizzati durante la creazione di un cluster.

Specifica dei ruoli IAM per le richieste EMRFS ad Amazon S3 mediante la console

1. Creare una configurazione di sicurezza che specifichi mappature di ruoli:
   a. Nella console di Amazon EMR, seleziona Security configurations (Configurazioni di sicurezza) e Create (Crea).
   b. Digitare un nome in Name (Nome) per la configurazione di sicurezza. Questo nome è utilizzato per specificare la configurazione di sicurezza al momento della creazione di un cluster.
   c. Scegli Use IAM roles for EMRFS requests to Amazon S3 (Utilizza i ruoli IAM per le richieste EMRFS ad Amazon S3).
   d. Seleziona un IAM role (Ruolo IAM) da applicare e in Basis for access (Base per accesso) seleziona un tipo di identificatore (Users [Utenti], Groups [Gruppi) o S3 prefixes [Prefissi S3]) dall'elenco e immetti gli identificatori corrispondenti. Se si utilizzano più identificatori, separarli con una virgola e senza inserire spazi. Per ulteriori informazioni su ogni tipo di identificatore, vedi JSON configuration reference qui sotto.
   e. Scegliere Add role (Aggiungi ruolo) per configurare ulteriori mappature di ruoli come descritto nella fase precedente.

f. Configurare altre opzioni per la configurazione di sicurezza come appropriato e scegliere Create (Crea). Per ulteriori informazioni, consulta Creazione di una configurazione di sicurezza.

2. Specificare la configurazione di sicurezza creata precedentemente alla creazione di un cluster. Per ulteriori informazioni, consulta Impostazione di una configurazione di sicurezza per un cluster.

Specifica dei ruoli IAM per le richieste EMRFS ad Amazon S3 mediante la AWS CLI

1. Utilizzare il comando `aws emr create-security-configuration`, specificando un nome per la configurazione di sicurezza, e i dettagli relativi a tale configurazione in formato JSON.

   L'esempio di comando riportato di seguito crea una configurazione di sicurezza con il nome `EMRFS_Roles_Security_Configuration`. Questa è basata su una struttura JSON nel file `MyEmrfsSecConfig.json`, che viene salvato nella stessa directory in cui viene eseguito il comando.

   ```
   aws emr create-security-configuration --name EMRFS_Roles_Security_Configuration --
   security-configuration file://MyEmrFsSecConfig.json.
   ```

   Utilizza le linee guida seguenti per la struttura del file `MyEmrFsSecConfig.json`. È possibile specificare questa struttura insieme alle strutture per altre opzioni della configurazione di sicurezza. Per ulteriori informazioni, consulta Creazione di una configurazione di sicurezza.

   Di seguito è riportato un esempio di frammento JSON per specificare ruoli IAM personalizzati per EMRFS all'interno di una configurazione di sicurezza. Vengono illustrate le mappature dei ruoli per i tre diversi tipi di identificatori, seguite da un riferimento ai parametri.

   ```
   {
     "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" ]
      }]
    }
  }
}
```

Parametro	Descrizione
`"AuthorizationConfiguration":`	Campo obbligatorio.
`"EmrFsConfiguration":`	Campo obbligatorio. Contiene mappature dei ruoli.
`"RoleMappings":`	Campo obbligatorio. Contiene una o più definizioni di mappatura dei ruoli. Le mappature dei ruoli vengono valutate dall'alto verso il basso nell'ordine in cui vengono visualizzate. Se una mappatura dei ruoli viene valutata come true (vera) per una chiamata EMRFS per i dati in Amazon S3, non vengono valutate altre mappature dei ruoli ed EMRFS utilizza il ruolo IAM specifica to per la richiesta. Le mappature dei ruoli sono costituite dai parametri obbligatori seguenti:
`"Role":`	Specifica l'identificatore ARN di un ruolo IAM nel formato `arn:aws:iam::` *`account-id`* `:role/`*`role-name`* . Questo è il ruolo IAM che Amazon EMR assume se la richiesta EMRFS ad Amazon S3 corrisponde a uno degli `Identifiers` specificato.

Parametro	Descrizione
`"IdentifierType":`	Il valore può essere uno dei seguenti: • `"User"` specifica che gli identificatori sono uno o più utenti Hadoop, i quali possono essere utenti di account Linux o entità Kerberos. Quando la richiesta EMRFS viene originata dall'utente o dagli utenti specificati, viene assunto il ruolo IAM. • `"Prefix"` specifica che l'identificatore è un percorso Amazon S3. Il ruolo IAM viene assunto per le chiamate alla posizione o alle posizioni con i prefissi specifica ti. Ad esempio, il prefisso `s3://mybu cket/` corrisponde a `s3://mybu cket/mydir` e `s3://mybucket/ yetanotherdir` . • `"Group"` specifica che gli identificatori sono uno o più Gruppi Hadoop. Il ruolo IAM viene assunto se la richiesta proviene da un utente nel gruppo o nei gruppi specificati.
`"Identifiers":`	Specifica uno o più identificatori del tipo di identificatore appropriato. Separa più identific atori con virgole senza spazi.

2. Utilizzare il comando `aws emr create-cluster` per creare un cluster e specificare la configurazione di sicurezza creata nella fase precedente.

   L'esempio seguente crea un cluster con le principali applicazioni Hadoop di default installate. Il cluster utilizza la configurazione di sicurezza creata in precedenza come `EMRFS_Roles_Security_Configuration` e un ruolo Amazon EMR personalizzato per EC2, `EC2_Role_EMR_Restrict_S3`, specificato utilizzando l'argomento `InstanceProfile` del parametro `--ec2-attributes`.

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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
```

Utilizzo di policy basate su risorse per l'accesso Amazon EMR ad AWS Glue Data Catalog

Se utilizzi AWS Glue in combinazione con Hive, Spark o Presto in Amazon EMR, AWS Glue supporta policy basate su risorse per controllare l'accesso alle risorse del Data Catalog. Queste risorse includono tabelle, database, connessioni e funzioni definite dall'utente. Per ulteriori informazioni, consulta Policy sulle risorse AWS Glue nella Guida per gli sviluppatori di AWS Glue.

Quando utilizzi policy basate su risorse per limitare l'accesso ad AWS Glue da Amazon EMR, l'entità principale specificata nella policy delle autorizzazioni deve essere il ruolo ARN associato al profilo di istanza EC2 specificato al momento della creazione di un cluster. Ad esempio, per una policy basata sulle risorse collegata a un catalogo, è possibile specificare il ruolo ARN per il ruolo di servizio predefinito per le istanze del cluster EC2, *EMR_EC2_DefaultRole*, come `Principal` utilizzando il formato mostrato nell'esempio seguente:

```
arn:aws:iam::acct-id:role/EMR_EC2_DefaultRole
```

L'*acct-id* può essere diverso dall'ID dell'account AWS Glue. Ciò consente l'accesso da cluster EMR in account diversi. Puoi specificare più entità principali, ognuna da un account diverso.

I ruoli IAM possono essere utilizzati con le applicazioni che richiamano direttamente i servizi AWS

Le applicazioni in esecuzione sulle istanze EC2 di un cluster possono utilizzare il profilo dell'istanza EC2 per ottenere credenziali di protezione temporanee quando si richiamano i servizi AWS.

Le versioni di Hadoop disponibili con Amazon EMR versione 2.3.0 e successive sono già state aggiornate per l'utilizzo di ruoli IAM. Se l'applicazione viene eseguita rigorosamente sull'architettura Hadoop e non richiama direttamente alcun servizio in AWS, dovrebbe funzionare con i ruoli IAM senza apportare alcuna modifica.

Se l'applicazione richiama direttamente i servizi in AWS, è necessario aggiornarla per usufruire dei ruoli IAM. Ciò significa che invece di ottenere le credenziali dell'account da `/etc/hadoop/conf/core-site.xml` sulle istanze EC2 nel cluster, l'applicazione utilizza un SDK per accedere alle risorse utilizzando ruoli IAM o richiama i metadati dell'istanza EC2 per ottenere le credenziali temporanee.

Accesso alle risorse AWS con i ruoli IAM utilizzando un SDK

- Nei seguenti argomenti viene descritto come utilizzare alcuni degli SDK AWS per accedere a credenziali temporanee utilizzando i ruoli IAM. Ogni argomento inizia con una versione di un'applicazione che non utilizza ruoli IAM e poi guida attraverso il processo di conversione di quell'applicazione per utilizzare ruoli IAM.
  - Utilizzo dei ruoli IAM per le istanze Amazon EC2 con SDK per Java nella Guida per gli sviluppatori di AWS SDK for Java
  - Utilizzo dei ruoli IAM per le istanze Amazon EC2 con SDK per .NET nella Guida per gli sviluppatori di AWS SDK for .NET
  - Utilizzo dei ruoli IAM per le istanze Amazon EC2 con SDK per PHP nella Guida per gli sviluppatori di AWS SDK for PHP
  - Utilizzo dei ruoli IAM per le istanze Amazon EC2 con SDK for Ruby nella Guida per gli sviluppatori di AWS SDK for Ruby

Per ottenere le credenziali provvisorie dai metadati dell'istanza EC2

- Richiama il seguente URL da un'istanza EC2 in esecuzione con il ruolo IAM specificato, che restituisce le credenziali di sicurezza temporanee associate (AccessKeyId, SecretAccessKey,

SessionToken ed Expiration). L'esempio seguente utilizza il profilo istanza predefinito per Amazon EMR, `EMR_EC2_DefaultRole`.

```
GET http://169.254.169.254/latest/meta-data/iam/security-
credentials/EMR_EC2_DefaultRole
```

Per ulteriori informazioni sulla scrittura delle applicazioni che utilizzano i ruoli IAM, consulta Concessione dell'accesso alle risorse AWS ad applicazioni in esecuzione sulle istanze Amazon EC2.

Per ulteriori informazioni sulle credenziali di sicurezza temporanee, consulta Utilizzo delle credenziali di sicurezza temporanee nella Guida all'utilizzo di credenziali di sicurezza temporanee.

Consentire a utenti e gruppi di creare e modificare i ruoli

Le entità principali (utenti e gruppi) IAM che creano, modificano e specificano i ruoli per un cluster, inclusi i ruoli predefiniti, devono essere autorizzate a eseguire le seguenti operazioni. Per ulteriori informazioni su ciascuna operazione, consulta Actions (Operazioni) nella Guida di riferimento alle API di IAM.

- `iam:CreateRole`
- `iam:PutRolePolicy`
- `iam:CreateInstanceProfile`
- `iam:AddRoleToInstanceProfile`
- `iam:ListRoles`
- `iam:GetPolicy`
- `iam:GetInstanceProfile`
- `iam:GetPolicyVersion`
- `iam:AttachRolePolicy`
- `iam:PassRole`

L'autorizzazione `iam:PassRole` consente la creazione del cluster. Le restanti autorizzazioni consentono la creazione di ruoli predefiniti.

Per ulteriori informazioni sull'assegnazione di autorizzazioni a un utente IAM, consulta Modifica delle autorizzazioni per un utente nella Guida per l'utente IAM.

Esempi di policy basate su identità per Amazon EMR

Per impostazione predefinita, gli utenti e i ruoli non dispongono dell'autorizzazione per creare o modificare risorse Amazon EMR. Inoltre, non sono in grado di eseguire attività utilizzando la AWS Management Console, AWS CLI o un'API AWS. Un amministratore IAM deve creare policy IAM che concedono a utenti e ruoli l'autorizzazione per eseguire operazioni API specifiche sulle risorse specificate di cui hanno bisogno. L'amministratore deve quindi collegare queste policy a utenti o gruppi che richiedono tali autorizzazioni.

Per informazioni su come creare una policy basata su identità IAM utilizzando questi documenti di policy JSON di esempio, consultare Creazione di policy nella scheda JSON nella Guida per l'utente IAM.

Argomenti

Best practice delle policy per Amazon EMR

Le policy basate su identità sono molto efficaci. Determinano se qualcuno può creare, accedere o eliminare risorse Amazon EMR nell'account. Queste operazioni possono comportare costi aggiuntivi per l'account AWS. Quando crei o modifichi policy basate su identità, segui queste linee guida e raccomandazioni:

- Nozioni di base sull'utilizzo di policy gestite da AWS: per iniziare a utilizzare rapidamente Amazon EMR, utilizza policy gestite da AWS per fornire ai dipendenti le autorizzazioni richieste. Queste policy sono già disponibili nell'account e sono gestite e aggiornate da AWS. Per ulteriori informazioni, consulta Nozioni di base sull'utilizzo delle autorizzazioni con policy gestite da AWS nella Guida per l'utente IAM e Policy gestite da Amazon EMR.

- Assegnare il privilegio minimo: quando si creano policy personalizzate, concedere solo le autorizzazioni richieste per eseguire un'attività. Inizia con un set di autorizzazioni minimo e concedi autorizzazioni aggiuntive quando necessario. Questo è più sicuro che iniziare con autorizzazioni che siano troppo permissive e cercare di limitarle in un secondo momento. Per ulteriori informazioni, consulta Assegnare il privilegio minimo nella Guida per l'utente IAM.
- Abilitare MFA per operazioni sensibili: per una maggiore sicurezza, richiedere agli utenti di utilizzare l'autenticazione a più fattori (MFA) per accedere a risorse sensibili o operazioni API. Per ulteriori informazioni, consulta Utilizzo dell'autenticazione a più fattori (MFA) in AWS nella Guida per l'utente IAM.
- Utilizzare le condizioni della policy per ulteriore sicurezza – Per quanto possibile, definire le condizioni in cui le policy basate su identità consentono l'accesso a una risorsa. Ad esempio, è possibile scrivere condizioni per specificare un intervallo di indirizzi IP consentiti dai quali deve provenire una richiesta. È anche possibile scrivere condizioni per consentire solo le richieste all'interno di un intervallo di date o ore specificato oppure per richiedere l'utilizzo di SSL o MFA. Per ulteriori informazioni, consulta Elementi delle policy JSON di IAM: Condizioni nella Guida per l'utente IAM.

Consentire agli utenti di visualizzare le loro autorizzazioni

Questo esempio mostra in che modo è possibile creare una policy che consente agli utenti di visualizzare le policy inline e gestite che sono collegate alla relativa identità utente. La policy include le autorizzazioni per completare questa azione sulla console o a livello di programmazione utilizzando la AWS CLI o l'API 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":""
      }
   ]
}
```

Policy gestite da Amazon EMR

Il modo più semplice di concedere accesso completo o in sola lettura alle operazioni di Amazon EMR necessarie è utilizzare le policy gestite da IAM per Amazon EMR. Le policy gestite offrono il vantaggio di essere aggiornate automaticamente nel momento in cui i requisiti di autorizzazione cambiano. Se utilizzi policy inline, è possibile che si verifichino degli errori di autorizzazione in caso di modifiche al servizio.

Amazon EMR renderà obsolete le policy gestite esistenti (policy v1), a favore di nuove policy gestite (policy v2). Le nuove policy gestite sono state definite in modo da essere allineate con le best practice AWS. Dopo che le policy gestite v1 renderà obsolete, non sarà possibile allegare queste policy a nuovi ruoli o utenti IAM. I ruoli e gli utenti esistenti che utilizzano policy obsolete possono continuare a utilizzarli. Le policy gestite v2 limitano l'accesso utilizzando i tag. Consentono solo operazioni Amazon EMR specifiche e richiedono risorse cluster contrassegnate con una chiave specifica per EMR. Si consiglia di esaminare attentamente la documentazione prima di utilizzare le nuove policy v2.

Le policy v1 verranno contrassegnate come obsolete con un'icona di avviso accanto a esse nell'elenco Policies (Policy) della console IAM. Le policy obsolete avranno le seguenti caratteristiche:

- Continuano a funzionare per tutti gli utenti, gruppi e ruoli attualmente collegati. Nessuna interruzione.
- Non possono essere collegate a nuovi utenti, gruppi o ruoli. Se una delle policy viene scollegata da un'entità corrente, non è possibile collegarla nuovamente.
- Dopo aver scollegato una policy v1 da tutte le entità correnti, la policy non sarà più visibile e non potrà essere utilizzata.

Nella tabella seguente sono riepilogate le modifiche tra le policy correnti (v1) e le policy v2.

Modifiche alle policy gestite da EMR

Tipo di policy	Nomi delle policy	Scopo della policy	Modifiche alla policy v2
Policy IAM gestite per un accesso completo a EMR da parte di utente, ruolo o gruppo collegato	Policy V1 (da essere resa obsoleta) : **AmazonElasticMapReduceFullAccess** Nome della policy V2 (con ambito): **AmazonEMRFullAccessPolicy_v2**	Consente agli utenti autorizzazioni complete per le operazioni EMR. Include le autorizzazioni iam:PassRole per le risorse.	La policy aggiunge il prerequisito secondo cui gli utenti devono aggiungere tag utente alle risorse prima di poter utilizzare questa policy. Per informazioni, consultare Assegnazione di tag alle risorse per l'utilizzo delle policy gestite. L'operazione iam:PassRole richiede che la condizione iam:PassedToService sia impostata sul servizio specificato. L'accesso ad Amazon EC2, Amazon S3 e ad altri servizi non è consentito per impostazione

Tipo di policy	Nomi delle policy	Scopo della policy	Modifiche alla policy v2
			predefinita. Consulta Policy IAM gestita per l'accesso completo (policy predefinita gestita v2).
Policy IAM gestita per un accesso di sola lettura parte di utente, ruolo o gruppo collegato	Policy V1 (da essere resa obsoleta) : **AmazonElasticMapReduceReadOnlyAccess** Nome della policy V2 (con ambito): **AmazonEMRReadOnlyAccessPolicy_v2**	Consente agli utenti autorizzazioni di sola lettura per le operazioni di Amazon EMR.	Le autorizzazioni si riferiscono esclusivamente alle operazioni elasticmapreduce di sola lettura specifiche. Per impostazione predefinita, l'accesso ad Amazon S3 non è consentito. Consulta Policy IAM gestita per l'accesso di sola lettura (policy predefinita gestita v2).

Tipo di policy	Nomi delle policy	Scopo della policy	Modifiche alla policy v2
Ruolo di servizio EMR predefinito e policy gestita collegata	Nome del ruolo: EMR_DefaultRole Policy V1 (da essere resa obsoleta): AmazonElasticMapReduceRole (ruolo di servizio EMR) Nome della policy v2 (con ambito): **AmazonEMRServicePolicy_v2**	Permette ad Amazon EMR di chiamare altri servizi AWS per conto dell'utente quando effettua il provisioning delle risorse ed esegue operazioni a livello di servizio. Questo ruolo è obbligatorio per tutti i cluster.	Il ruolo di servizio v2 e la policy predefinita v2 sostituiscono il ruolo e la policy obsoleti. La policy aggiunge il prerequisito secondo cui gli utenti devono aggiungere tag utente alle risorse prima di poter utilizzare questa policy. Per informazioni, consultare Assegnazione di tag alle risorse per l'utilizzo delle policy gestite. Per informazioni, consultare Ruolo di servizio per Amazon EMR (ruolo EMR).

Tipo di policy	Nomi delle policy	Scopo della policy	Modifiche alla policy v2
Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2)	Policy V1 (da essere resa obsoleta): EMR_EC2_DefaultRole (profilo dell'istanza) Nome della policy obsoleta: AmazonElasticMapReduceforEC2Role	Consente alle applicazioni in esecuzione su un cluster EMR di accedere ad altre risorse AWS, ad esempio Amazon S3. Ad esempio, se esegui i processi Apache Spark che elaborano i dati da Amazon S3, la policy deve consentire l'accesso a tali risorse.	Sia il ruolo predefinito che la policy predefinita diventeranno presto obsoleti. Non sono previsti ruoli o policy AWS gestiti predefiniti sostitutivi. È necessario fornire una policy basata su risorse o su identità. Ciò significa che, per impostazione predefinita, le applicazioni in esecuzione su un cluster EMR non hanno accesso ad Amazon S3 o ad altre risorse a meno che non vengano aggiunte manualmente alla policy. Per informazioni, consultare Ruolo predefinito e policy gestita.

Tipo di policy	Nomi delle policy	Scopo della policy	Modifiche alla policy v2
Altre policy del ruolo di servizio EC2	Nomi delle policy correnti: AmazonElasticMapReduceforAutoScalingRole, AmazonElasticMapReduceEditorsRole, AmazonEMRCleanupPolicy	Fornisce le autorizzazioni necessarie a EMR per accedere ad altre risorse AWS ed eseguire operazioni se si utilizza la scalabilità automatica, i notebook o per eliminare le risorse EC2.	Nessuna modifica per v2.

Protezione di iam:PassRole

Le policy gestite predefinite con autorizzazioni complete di Amazon EMR incorporano configurazioni di sicurezza `iam:PassRole`, tra cui:

- Autorizzazioni `iam:PassRole` solo per specifici ruoli Amazon EMR predefiniti.
- Condizioni `iam:PassedToService` che consentono di utilizzare la policy solo con servizi AWS specificati, quali `elasticmapreduce.amazonaws.com` e `ec2.amazonaws.com`.

È possibile visualizzare la versione JSON delle policy AmazonEMRFullAccessPolicy_v2 e AmazonEMRServicePolicy_v2 nella console IAM. Ti consigliamo di creare i nuovi cluster con le policy gestite v2.

Per creare policy personalizzate, ti consigliamo di iniziare con le policy gestite e di modificarle in base alle tue esigenze.

Per informazioni su come collegare policy a utenti (entità principali), consulta Utilizzo di policy gestite mediante la AWS Management Console nella Guida per l'utente IAM.

Assegnazione di tag alle risorse per l'utilizzo delle policy gestite

AmazonEMRServicePolicy_v2 e AmazonEMRFullAccessPolicy_v2 dipendono dall'accesso limitato alle risorse che Amazon EMR fornisce o utilizza. L'ambito inferiore si ottiene limitando l'accesso solo

a quelle risorse a cui è associato un tag utente predefinito. Quando si utilizza una di queste due policy, è necessario passare il tag utente predefinito `for-use-with-amazon-emr-managed-policies = true` durante il provisioning del cluster. Amazon EMR propaga automaticamente tale tag. Inoltre, è necessario aggiungere un tag utente alle risorse elencate nella sezione seguente. Se utilizzi la console Amazon EMR per avviare il cluster, consulta la sezione Considerazioni sull'utilizzo della console Amazon EMR per avviare cluster con policy gestite v2.

Per utilizzare le policy gestite, passa il tag utente `for-use-with-amazon-emr-managed-policies = true` durante il provisioning di un cluster con la CLI, l'SDK o un altro metodo.

Quando si passa il tag, Amazon EMR propaga il tag all'ENI della sottorete privata, all'istanza EC2 e ai volumi EBS creati. Amazon EMR assegna automaticamente tag anche ai gruppi di sicurezza creati. Tuttavia, se desideri che Amazon EMR venga avviato con un determinato gruppo di sicurezza, devi taggarlo. Per le risorse che non sono state create da Amazon EMR, è necessario aggiungere i tag a tali risorse. Ad esempio, dovrai taggare le sottoreti Amazon EC2, i gruppi di sicurezza EC2 (se non creati da Amazon EMR) e i VPC (se desideri che Amazon EMR crei gruppi di sicurezza). Per avviare cluster con policy gestite v2 nei VPC, è necessario taggare tali VPC con il tag utente predefinito. Per informazioni, consultare Considerazioni sull'utilizzo della console Amazon EMR per avviare cluster con policy gestite v2.

Assegnazione di tag propagata specificata dall'utente

Amazon EMR contrassegna le risorse create utilizzando i tag Amazon EMR specificati durante la creazione di un cluster. Amazon EMR applica tag alle risorse create durante il ciclo di vita del cluster.

Amazon EMR propaga i tag utente per le seguenti risorse:

- ENI della sottorete privata (interfacce di rete elastiche di accesso ai servizi)
- Istanze EC2
- Volumi EBS
- Modello di avvio di EC2

Gruppi di sicurezza con tag automatici

Amazon EMR tagga i gruppi di sicurezza EC2 creati con il tag richiesto per le policy gestite v2 per Amazon EMR, `for-use-with-amazon-emr-managed-policies`, indipendentemente dai tag specificati nel comando di creazione del cluster. Per un gruppo di sicurezza creato prima dell'introduzione delle policy gestite v2, Amazon EMR non tagga automaticamente il gruppo di

sicurezza. Se si desidera utilizzare policy gestite v2 con i gruppi di sicurezza predefiniti già esistenti nell'account, è necessario taggare manualmente i gruppi di sicurezza con `for-use-with-amazon-emr-managed-policies = true`.

Risorse cluster con tag manuali

Occorre taggare manualmente alcune risorse del cluster in modo che possano essere accessibili dai ruoli predefiniti di Amazon EMR.

- Occorre taggare manualmente i gruppi di sicurezza EC2 e le sottoreti EC2 con il tag `for-use-with-amazon-emr-managed-policies` della policy gestita da Amazon EMR.
- Occorre taggare manualmente un VPC se vuoi che Amazon EMR crei gruppi di sicurezza predefiniti. EMR tenterà di creare un gruppo di sicurezza con il tag specifico se il gruppo di sicurezza predefinito non esiste già.

Amazon EMR tagga automaticamente le seguenti risorse:

- Gruppi di sicurezza EC2 creati da EMR

È necessario aggiungere i tag manualmente alle risorse seguenti:

- Sottorete EC2
- Gruppo di sicurezza EC2

È necessario taggare manualmente le risorse seguenti:

- VPC: solo quando si desidera che Amazon EMR crei gruppi di sicurezza

Considerazioni sull'utilizzo della console Amazon EMR per avviare cluster con policy gestite v2

Puoi eseguire il provisioning di cluster con policy gestite v2 utilizzando la console Amazon EMR. Di seguito sono riportate alcune considerazioni quando utilizzi la console per avviare cluster Amazon EMR.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR. La nuova console non dispone ancora della funzionalità di applicazione di tag automatica e non mostra nemmeno quali risorse (VPC/

sottoreti) devono essere taggate. Per maggiori informazioni sulle differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

- Non è necessario passare il tag predefinito. Amazon EMR aggiunge automaticamente il tag e lo propaga ai componenti appropriati.
- Per i componenti che devono essere taggati manualmente, la vecchia console Amazon EMR tenta di applicare automaticamente i tag se disponi delle autorizzazioni necessarie per taggare le risorse. Se non disponi delle autorizzazioni per taggare le risorse o vuoi utilizzare la nuova console, chiedi all'amministratore di taggare tali risorse.
- Non è possibile avviare cluster con policy gestite v2 a meno che non siano soddisfatti tutti i prerequisiti.
- La vecchia console Amazon EMR mostra quali risorse (VPC/sottorete) devono essere taggate.

Policy IAM gestita per l'accesso completo (policy predefinita gestita v2)

Le policy predefinite gestite da EMR con ambito v2 concedono privilegi di accesso specifici agli utenti. Richiedono un tag di risorsa Amazon EMR predefinito e chiavi di condizione `iam:PassRole` per le risorse utilizzate da Amazon EMR, quali `Subnet` e `SecurityGroup` utilizzate per avviare il cluster.

Per concedere le operazioni necessarie per Amazon EMR, collega la policy `AmazonEMRFullAccessPolicy_v2` gestita. Questa policy gestita predefinita aggiornata sostituisce la policy gestita `AmazonElasticMapReduceFullAccess`.

`AmazonEMRFullAccessPolicy_v2` dipende dall'accesso ridotto alle risorse che Amazon EMR fornisce o utilizza. Quando si utilizza questa policy, è necessario passare il tag utente `for-use-with-amazon-emr-managed-policies = true` durante il provisioning del cluster. Amazon EMR propaga automaticamente tale tag. Inoltre, potrebbe essere necessario aggiungere manualmente un tag utente a tipi specifici di risorse, ad esempio gruppi di sicurezza EC2 che non sono stati creati da Amazon EMR. Per ulteriori informazioni, consulta Assegnazione di tag alle risorse per l'utilizzo delle policy gestite.

La policy `AmazonEMRFullAccessPolicy_v2` protegge le risorse eseguendo le operazioni elencate di seguito:

- Richiede il tag delle risorse con il tag `for-use-with-amazon-emr-managed-policies` predefinito delle policy gestite da Amazon EMR per la creazione del cluster e l'accesso ad Amazon EMR.

- Limita l'operazione `iam:PassRole` a ruoli predefiniti specifici e l'accesso `iam:PassedToService` a servizi specifici.
- Non fornisce più l'accesso ad Amazon EC2, Amazon S3 e ad altri servizi per impostazione predefinita.

Il contenuto di questa policy è mostrato di seguito.

> **Note**
>
> Puoi anche utilizzare il link alla console AmazonEMRFullAccessPolicy_v2 per visualizzare la policy.

```
{
    "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": "*"
        }
    ]
}
```

Policy IAM gestite per l'accesso completo (presto obsolete)

Le policy gestite `AmazonElasticMapReduceFullAccess` e `AmazonEMRFullAccessPolicy_v2` AWS Identity and Access Management (IAM) garantiscono tutte le operazioni necessarie per Amazon EMR e altri servizi.

> **Important**
>
> La policy gestita `AmazonElasticMapReduceFullAccess` diventerà presto obsoleta, pertanto ti consigliamo di non utilizzarla con Amazon EMR. Utilizza invece `AmazonEMRFullAccessPolicy_v2`. Quando il servizio IAM imposterà come obsoleta la

policy v1, non sarà possibile collegarla a un ruolo. Tuttavia, puoi collegare un ruolo esistente a un cluster anche se tale ruolo utilizza la policy obsoleta.

Le policy gestite predefinite con autorizzazioni complete di Amazon EMR incorporano configurazioni di sicurezza `iam:PassRole`, tra cui:

- Autorizzazioni `iam:PassRole` solo per specifici ruoli Amazon EMR predefiniti.
- Condizioni `iam:PassedToService` che consentono di utilizzare la policy solo con servizi AWS specificati, quali `elasticmapreduce.amazonaws.com` e `ec2.amazonaws.com`.

È possibile visualizzare la versione JSON delle policy AmazonEMRFullAccessPolicy_v2 e AmazonEMRServicePolicy_v2 nella console IAM. Ti consigliamo di creare i nuovi cluster con le policy gestite v2.

Puoi visualizzare il contenuto della policy v1 obsoleta nella AWS Management Console all'indirizzo `AmazonElasticMapReduceFullAccess`. L'operazione `ec2:TerminateInstances` nella policy concede all'utente o al ruolo l'autorizzazione a terminare qualsiasi istanza Amazon EC2 associata all'account IAM. Sono incluse le istanze che non fanno parte di un cluster EMR.

Policy IAM gestita per l'accesso di sola lettura (policy predefinita gestita v2).

Per concedere privilegi di sola lettura ad Amazon EMR, collega la policy gestita AmazonEMRReadOnlyAccessPolicy_v2. Questa policy gestita predefinita sostituisce la policy gestita `AmazonElasticMapReduceReadOnlyAccess`. Il contenuto di questa dichiarazione di policy è riportato nel frammento seguente. Rispetto alla policy `AmazonElasticMapReduceReadOnlyAccess`, la policy `AmazonEMRReadOnlyAccessPolicy_v2` non utilizza caratteri jolly per l'elemento `elasticmapreduce`. Al contrario, gli ambiti delle policy v2 predefiniti specificano le operazioni `elasticmapreduce` consentite.

> **Note**
>
> È possibile utilizzare anche il collegamento AWS Management Console `AmazonEMRReadOnlyAccessPolicy_v2` per visualizzare la policy.

```
{
```

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

Policy IAM gestite per l'accesso di sola lettura (presto obsolete)

La policy gestita `AmazonElasticMapReduceReadOnlyAccess` diventerà presto obsoleta. Non è possibile allegare questa policy durante l'avvio di nuovi

cluster. `AmazonElasticMapReduceReadOnlyAccess` è stato sostituito con `AmazonEMRReadOnlyAccessPolicy_v2` come policy gestita predefinita di Amazon EMR. Il contenuto di questa dichiarazione di policy è riportato nel frammento seguente. I caratteri jolly per l'elemento `elasticmapreduce` specificano che solo le operazioni che iniziano con le stringhe specificate sono consentite. Tieni presente che poiché questa policy non nega esplicitamente le operazioni, un'altra dichiarazione di policy può essere utilizzata per concedere l'accesso alle operazioni specificate.

> **Note**
>
> Puoi utilizzare anche la AWS Management Console per visualizzare la policy.

```
{
  "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": "*"
    }
  ]
}
```

Policy gestite da AWS per Amazon EMR

Per aggiungere le autorizzazioni a utenti, gruppi e ruoli, è più semplice utilizzare policy gestite da AWS piuttosto che scrivere autonomamente le policy. La creazione di policy gestite dai clienti IAM che forniscono al tuo team solo le autorizzazioni di cui ha bisogno richiede tempo e competenza. Per iniziare rapidamente, utilizza le nostre policy gestite da AWS. Queste policy coprono i casi d'uso più

comuni e sono disponibili nel tuo account AWS. Per ulteriori informazioni sulle policy gestite da AWS, consulta Policy gestite da AWS nella Guida per l'utente IAM.

I servizi AWS mantengono e aggiornano le policy gestite da AWS. Non è possibile modificare le autorizzazioni nelle policy gestite da AWS. I servizi occasionalmente aggiungono altre autorizzazioni a una policy gestita da AWS per supportare nuove funzionalità. Questo tipo di aggiornamento interessa tutte le identità (utenti, gruppi e ruoli) a cui è collegata la policy. È più probabile che i servizi aggiornino una policy gestita da AWS quando viene avviata una nuova funzionalità o quando diventano disponibili nuove operazioni. I servizi non rimuovono le autorizzazioni da una policy gestita da AWS, pertanto gli aggiornamenti delle policy non interrompono le autorizzazioni esistenti.

Inoltre, AWS supporta policy gestite per le funzioni di processi che coprono più servizi. Ad esempio, la policy gestita da AWS ReadOnlyAccess fornisce accesso in sola lettura a tutti i servizi e le risorse AWS. Quando un servizio avvia una nuova funzionalità, AWS aggiunge autorizzazioni di sola lettura per nuove operazioni e risorse. Per l'elenco e la descrizione delle policy di funzione dei processi, consulta la sezione Policy gestite da AWS per funzioni di processi nella Guida per l'utente IAM.

Aggiornamenti di Amazon EMR sulle policy gestite da AWS

Visualizza i dettagli sugli aggiornamenti alle policy gestite da AWS per Amazon EMR da quando questo servizio ha iniziato a tenere traccia delle modifiche. Per gli avvisi automatici sulle modifiche apportate a questa pagina, sottoscrivere il feed RSS nella pagina della cronologia dei documenti di Amazon EMR.

Modifica	Descrizione	Data
AmazonEMRFullAccessPolicy_v2 e **AmazonEMRReadOnlyAccessPolicy_v2**: aggiornamento a una policy esistente	Aggiunto `elasticmapreduce:ListSupportedInstanceTypes` .	13 luglio 2023

Modifica	Descrizione	Data
AmazonEMRFullAccessPolicy_v2 e **AmazonEMRReadOnlyAccessPolicy_v2**: aggiornamento a una policy esistente	Aggiunto `elasticmapreduce:DescribeReleaseLabel` e `elasticmapreduce:GetAutoTerminationPolicy`.	21 aprile 2022
AmazonEMRFullAccessPolicy_v2: aggiornamento a una policy esistente	Aggiunta di `ec2:DescribeImages` per Utilizzo di un'AMI personalizzata.	15 febbraio 2022
Policy gestite da Amazon EMR	Aggiornato per chiarire l'uso di tag utente predefiniti. Aggiunta della sezione sull'utilizzo del console AWS per avviare cluster con policy gestite v2.	29 settembre 2021
AmazonEMRFullAccessPolicy_v2: aggiornamento a una policy esistente	Modifica delle operazioni `PassRoleForAutoScaling` e `PassRoleForEC2` per utilizzare l'operatore di condizione `StringLike` in modo che corrispondano a `"iam:PassedToService":"application-autoscaling.amazonaws.com*"` e `"iam:PassedToService":"ec2.amazonaws.com*"` rispettivamente.	20 maggio 2021

Modifica	Descrizione	Data
AmazonEMRFullAccessPolicy_v2: aggiornamento a una policy esistente	Rimozione dell'operazione `s3:ListBuckets` non valida e sostituzione con l'operazione `s3:ListAllMyBuckets`. Aggiornamento della creazione del ruolo collegato al servizio (SLR) in modo esplicito fino all'unico SLR di Amazon EMR con principi di servizio espliciti. Gli SLR che possono essere creati sono uguali a quelli precedenti a questa modifica.	23 marzo 2021
AmazonEMRFullAccessPolicy_v2: nuova policy	Amazon EMR ha aggiunto nuove autorizzazioni per l'ambito dell'accesso alle risorse e per aggiungere il prerequisito secondo cui gli utenti devono aggiungere un tag utente predefinito alle risorse prima di poter utilizzare le policy gestite da Amazon EMR. L'operazione `iam:PassRole` richiede che la condizione `iam:PassedToService` sia impostata sul servizio specificato. L'accesso ad Amazon EC2, Amazon S3 e ad altri servizi non è consentito per impostazione predefinita.	11 marzo 2021

Modifica	Descrizione	Data
AmazonEMRServicePolicy_v2 : nuova policy	Aggiunge il prerequisito secondo cui gli utenti devono aggiungere tag utente alle risorse prima di poter utilizzare questa policy.	11 marzo 2021
AmazonEMRReadOnlyAccessPolicy_v2 : nuova policy	Le autorizzazioni si riferiscono no esclusivamente alle operazioni elasticmapreduce di sola lettura specificate. Per impostazione predefinita, l'accesso ad Amazon S3 non è consentito.	11 marzo 2021
Amazon EMR ha iniziato a monitorare le modifiche	Amazon EMR ha iniziato a monitorare le modifiche per le sue policy gestite da AWS.	11 marzo 2021

Policy IAM per l'accesso basato su tag ai cluster e a EMR Notebooks

Puoi utilizzare le condizioni nella policy basata su identità per controllare l'accesso ai cluster e ai EMR Notebooks basati su tag.

Per ulteriori informazioni sull'aggiunta di tag ai cluster, consulta Tagging dei cluster EMR.

I seguenti esempi illustrano differenti scenari e modi per utilizzare gli operatori di condizione con chiavi di condizione Amazon EMR. Queste dichiarazioni di policy IAM sono concepite unicamente per scopi dimostrativi e non devono essere utilizzate negli ambienti di produzione. Esistono vari modi di combinare dichiarazioni di policy per concedere o negare autorizzazioni in base alle tue esigenze. Per ulteriori informazioni sulla pianificazione e sul test di policy IAM, consulta la Guida per l'utente IAM.

> **Important**
>
> Negare esplicitamente l'autorizzazione per operazioni di assegnazione di tag è una possibilità da tenere in debita considerazione. Ciò impedisce agli utenti di concedersi personalmente

autorizzazioni tramite tag di una risorsa che non avevi intenzione di accordare. Se non neghi le operazioni di assegnazione di tag per una risorsa, un utente può modificare i tag e aggirare l'intenzione delle policy basate su tag.

Dichiarazioni di policy basate su identità di esempio per cluster

Gli esempi di seguito dimostrano le policy di autorizzazioni basate su identità che vengono utilizzate per controllare le operazioni che sono consentite con cluster EMR.

Important

L'operazione `ModifyInstanceGroup` in Amazon EMR non richiede che si specifichi un ID cluster. Per questo motivo, negare questa azione basata su tag del cluster richiede un'ulteriore considerazione. Per ulteriori informazioni, consulta Negazione dell'operazione ModifyInstanceGroup.

Argomenti

- Autorizzazione di operazioni solo su cluster con specifici valori di tag
- Richiesta dell'assegnazione di tag a un cluster appena creato
- Autorizzazione di operazioni su cluster con uno specifico tag indipendentemente dal valore di tag

Autorizzazione di operazioni solo su cluster con specifici valori di tag

Gli esempi che seguono mostrano una policy che autorizza un utente a eseguire operazioni in base al tag di cluster *`department`* con il valore *`dev`* e un altro utente a contrassegnare cluster con quello stesso tag. L'esempio di policy finale mostra come negare privilegi per taggare cluster EMR con qualsiasi tag tranne quel tag.

Nell'esempio di policy seguente, l'operatore di condizione `StringEquals` cerca di far corrispondere *`dev`* con il valore del tag *`department`*. Se il tag *`department`* non è stato aggiunto al cluster, oppure non contiene il valore *`dev`*, la policy non viene applicata e le operazioni non sono consentite da questa policy. Se nessun'altra dichiarazione di policy consente le operazioni, l'utente può utilizzare solo i cluster che hanno questo tag con tale valore.

```
{
  "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"
        }
      }
    }
  ]
}
```

Puoi anche specificare più valori di tag utilizzando un operatore di condizione. Ad esempio, per consentire tutte le operazioni su cluster in cui il tag *`department`* contiene il valore *`dev`* o *`test`*, puoi sostituire il blocco di condizione nell'esempio precedente con quanto segue.

```
            "Condition": {
              "StringEquals": {
                "elasticmapreduce:ResourceTag/department":["dev", "test"]
              }
            }
```

Richiesta dell'assegnazione di tag a un cluster appena creato

Come nell'esempio precedente, l'esempio seguente di policy cerca lo stesso tag corrispondente: il valore *`dev`* per il tag *`department`*. In questo esempio, però, la chiave di condizione `RequestTag`

specifica che la policy si applica durante la creazione del tag. Quindi è necessario creare un cluster con un tag che corrisponda al valore specificato.

Per creare un cluster con un tag, devi anche disporre dell'autorizzazione per l'azione `elasticmapredue:AddTags`. Per questa dichiarazione, la chiave di condizione `elasticmapreduce:ResourceTag` garantisce che IAM conceda l'accesso solo alle risorse dei tag con il valore *`dev`* sul tag *`department`*. L'elemento `Resource` viene utilizzato per limitare questa autorizzazione alle risorse del cluster.

Per le risorse `PassRole`, è necessario fornire l'ID o l'alias dell'account AWS, il nome del ruolo di servizio nella dichiarazione `PassRoleForEMR` e il nome del profilo dell'istanza nell'istruzione `PassRoleForEC2`. Per ulteriori informazioni sul formato di ARN IAM, consulta ARN IAM nella Guida per l'utente IAM.

Per ulteriori informazioni sulla corrispondenza delle chiavi di tag, consulta `aws:RequestTag/tag-key` nella Guida per l'utente 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*"
          }
        }
      }
    ]
}
```

Autorizzazione di operazioni su cluster con uno specifico tag indipendentemente dal valore di tag

Puoi anche consentire operazioni solo su cluster con un determinato tag, indipendentemente dal valore di tag. A questo proposito, puoi utilizzare l'operatore `Null`. Per ulteriori informazioni, consulta Operatore di condizione per verificare la presenza di chiavi di condizione nella Guida per l'utente IAM. Ad esempio, per consentire operazioni solo su cluster EMR con il tag *`department`*, indipendentemente dal valore che lo stesso contiene, puoi sostituire i blocchi di condizione nell'esempio precedente con quello illustrato di seguito. L'operatore `Null` cerca il tag *`department`* su un cluster EMR. Se il tag esiste, l'istruzione `Null` restituisce il valore false, corrispondente alla condizione specificata nella dichiarazione di policy, e le operazioni appropriate sono consentite.

```
"Condition": {
  "Null": {
    "elasticmapreduce:ResourceTag/department":"false"
```

```
  }
}
```

La dichiarazione di policy seguente consente a un utente di creare un cluster EMR solo se il cluster avrà un tag *department*, indipendentemente dal valore dello stesso. Per la risorsa `PassRole`, è necessario fornire l'ID account o alias AWS e il nome del ruolo di servizio. Per ulteriori informazioni sul formato di ARN IAM, consulta ARN IAM nella Guida per l'utente di IAM.

Per ulteriori informazioni su come specificare l'operatore di condizione null ("falso"), consulta Operatore di condizione per verificare la presenza di chiavi di condizione nella Guida per l'utente 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*"
                }
            }
        }
    ]
}
```

Dichiarazioni di policy basate su identità di esempio per EMR Notebooks

Le dichiarazioni di policy IAM di esempio in questa sezione illustrano scenari comuni per l'utilizzo di chiavi per limitare le operazioni consentite mediante EMR Notebooks. Finché nessun'altra policy associata al principale (utente) consente le operazioni, le chiavi di contesto della condizione limitano le operazioni consentite come indicato.

Example : consente l'accesso solo ai EMR Notebooks creati da un utente in base all'assegnazione di tag

L'esempio seguente di istruzione di policy, quando collegata a un ruolo o a un utente, consente all'utente di utilizzare solo i notebook che ha creato. Questa istruzione di policy usa il tag predefinito applicato durante la creazione di un notebook.

In questo esempio, l'operatore di condizione `StringEquals` cerca di mettere in corrispondenza una variabile che rappresenta l'ID dell'utente attuale (`{aws:userId}`) con il valore del tag `creatorUserID`. Se il tag `creatorUserID` non è stato aggiunto al notebook, oppure non contiene il valore dell'ID dell'utente corrente, la policy non viene applicata e le operazioni non sono consentite

da questa policy. Se nessun'altra istruzione di policy consente le operazioni, l'utente può utilizzare solo i notebook che hanno questo tag con tale valore.

```
{
    "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 -Richiedere l'assegnazione di tag al notebook durante la creazione

In questo esempio viene utilizzata la chiave di contesto `RequestTag`. L'operazione `CreateEditor` è consentita solo se l'utente non modifica o elimina il tag `creatorUserID` aggiunto per impostazione predefinita. La variabile ${aws:userId}, specifica l'ID dell'utente attualmente attivo, che è il valore predefinito del tag.

L'istruzione della policy può essere utilizzata per garantire che gli utenti non rimuovano il tag `createUserId` o ne modifichino il valore.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
```

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

Questo esempio richiede che l'utente crei il cluster con un tag con la stringa di chiave `dept` e un valore impostato per uno dei seguenti: `datascience`, `analytics`, `operations`.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:CreateEditor"
            ],
            "Effect": "Allow",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:RequestTag/dept": [
                        "datascience",
                        "analytics",
                        "operations"
                    ]
                }
            }
        }
    ]
}
```

Example -Limitare la creazione di notebook ai cluster con tag e richiedere i tag del notebook

Questo esempio consente la creazione di notebook solo se il notebook viene creato con un tag che abbia la stringa di chiave `owner` impostata su uno dei valori specificati. Inoltre, è possibile creare il notebook solo se il cluster dispone di un tag con la stringa di chiave `department` impostata su uno dei valori specificati.

```
{
    "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 -Limitare la possibilità di avviare un notebook basato su tag

Questo esempio limita la possibilità di avviare solo i notebook che dispongono di un tag con la stringa di chiave `owner` impostata su uno dei valori specificati. Poiché l'elemento `Resource` è utilizzato per specificare solo `editor`, la condizione non è valida per il cluster e non è necessario aggiungere tag al cluster.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:editor/*",
```

```
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/owner": [
                        "owner1",
                        "owner2"
                    ]
                }
            }
        }
    ]
}
```

Questo esempio è simile a uno precedente. Tuttavia, il limite si applica solo ai cluster a cui sono applicati tag, non ai notebook.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "elasticmapreduce:StartEditor"
            ],
            "Effect": "Allow",
            "Resource": "arn:aws:elasticmapreduce:*:123456789012:cluster/*",
            "Condition": {
                "StringEquals": {
                    "elasticmapreduce:ResourceTag/department": [
                        "dep1",
                        "dep3"
                    ]
                }
            }
        }
    ]
}
```

In questo esempio viene utilizzato un altro set di tag di notebook e cluster. Consente di avviare un notebook solo se:

- Il notebook dispone di un tag con la stringa di chiave `owner` impostata su uno dei valori specificati

  -e-

- Il cluster dispone di un tag con la stringa di chiave `department` impostata su uno dei valori specificati

```
{
    "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 -Limitare la possibilità di aprire l'editor del notebook basato su tag

Questo esempio consente di aprire l'editor del notebook solo se:

- Il notebook dispone di un tag con la stringa di chiave `owner` impostata su uno dei valori specificati.

  -e-

- Il cluster dispone di un tag con la stringa di chiave `department` impostata su uno dei valori specificati.

```
{
    "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"
                    ]
                }
            }
        }
    ]
}
```

Negazione dell'operazione ModifyInstanceGroup

L'operazione ModifyInstanceGroups in Amazon EMR non richiede che si fornisca un ID cluster con l'operazione. È invece possibile specificare solo un ID del gruppo di istanze. Per questo motivo, una policy di negazione apparentemente semplice per questa operazione basata sull'ID cluster o su un tag cluster potrebbe non avere l'effetto desiderato. Esaminiamo l'esempio di policy seguente.

```
{
 "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"
        }
    ]

}
```

Se un utente con questa policy associata esegue un'operazione `ModifyInstanceGroup` e specifica solo l'ID del gruppo di istanze, la policy non viene applicata. Poiché l'operazione è consentita su tutte le altre risorse, avrà esito positivo.

Una soluzione a questo problema consiste nell'allegare un'istruzione della policy all'identità che utilizza un elemento NotResource per negare qualsiasi operazione `ModifyInstanceGroup` rilasciata senza un ID cluster. La seguente policy di esempio aggiunge tale dichiarazione di negazione in modo che qualsiasi richiesta `ModifyInstanceGroups` abbia esito negativo a meno che non venga specificato un ID cluster. Poiché un'identità deve specificare un ID cluster con l'operazione, le istruzioni negate basate sull'ID cluster sono efficaci.

```
{
```

```
 "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/*"
        }
    ]

}
```

Un problema simile si verifica quando si desidera negare l'operazione `ModifyInstanceGroups` in base al valore associato a un tag cluster. La soluzione è simile. Oltre a un'istruzione di negazione che specifica il valore del tag, è possibile aggiungere un'istruzione della policy che neghi l'operazione `ModifyInstanceGroup` se il tag specificato non è presente, indipendentemente dal valore.

Nell'esempio seguente viene illustrata una policy che, se collegata a un'identità, nega all'identità l'operazione `ModifyInstanceGroups` per qualsiasi cluster con il tag `department` impostato su `dev`. Questa istruzione è efficace solo a causa dell'istruzione di negazione che utilizza la condizione `StringNotLike` per negare l'operazione a meno che il tag `department` non sia 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": "*"
        }
    ],
}
```

Risoluzione dei problemi relativi all'identità e all'accesso di Amazon EMR

Utilizza le informazioni seguenti per eseguire la diagnosi e risolvere i problemi comuni che possono verificarsi durante l'utilizzo di Amazon EMR e IAM.

Argomenti

- Non dispongo dell'autorizzazione per eseguire un'operazione in Amazon EMR
- Non sono autorizzato a eseguire iam:PassRole

- Voglio consentire a persone esterne al mio account AWS di accedere alle mie risorse Amazon EMR

Non dispongo dell'autorizzazione per eseguire un'operazione in Amazon EMR

Se la AWS Management Console indica che non sei autorizzato a eseguire un'operazione, devi contattare l'amministratore per ricevere assistenza. L'amministratore è la persona che ti ha fornito il nome utente e la password.

Il seguente esempio di errore si verifica quando l'utente `mateojackson` prova a utilizzare la console per visualizzare i dettagli relativi a una risorsa *`my-example-widget`* fittizia, ma non dispone di autorizzazioni `EMR:`*`GetWidget`* fittizie.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 EMR:GetWidget on resource: my-example-widget
```

In questo caso, Mateo richiede al suo amministratore di aggiornare le policy per poter accedere alla risorsa *`my-example-widget`* utilizzando l'operazione `EMR:`*`GetWidget`*.

Non sono autorizzato a eseguire iam:PassRole

Se ricevi un errore che indica che non disponi dell'autorizzazione per eseguire l'operazione `iam:PassRole`, per poter passare un ruolo ad Amazon EMR dovrai aggiornare le policy.

Alcuni Servizi AWS consentono di trasmettere un ruolo esistente a tale servizio, invece di creare un nuovo ruolo di servizio o un ruolo collegato ai servizi. Per eseguire questa operazione, è necessario disporre delle autorizzazioni per trasmettere il ruolo al servizio.

Il seguente esempio di errore si verifica quando un utente IAM denominato `marymajor` cerca di utilizzare la console per eseguire un'operazione in Amazon EMR. Tuttavia, l'azione richiede che il servizio disponga delle autorizzazioni concesse da un ruolo di servizio. Mary non dispone delle autorizzazioni per passare il ruolo al servizio.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

In questo caso, le policy di Mary devono essere aggiornate per poter eseguire l'operazione `iam:PassRole`.

Per ulteriore assistenza con l'accesso, contatta l'amministratore AWS. L'amministratore è la persona che ti ha fornito le credenziali di accesso.

Voglio consentire a persone esterne al mio account AWS di accedere alle mie risorse Amazon EMR

È possibile creare un ruolo con il quale utenti in altri account o persone esterne all'organizzazione possono accedere alle tue risorse. È possibile specificare chi è attendibile per l'assunzione del ruolo. Per servizi che supportano policy basate su risorse o liste di controllo accessi (ACL), utilizza tali policy per concedere alle persone l'accesso alle tue risorse.

Per ulteriori informazioni, consulta gli argomenti seguenti:

- Per sapere se Amazon EMR supporta queste funzionalità, consulta Funzionamento di Amazon EMR con IAM.
- Per informazioni su come garantire l'accesso alle risorse negli Account AWS che possiedi, consulta Fornire l'accesso a un utente IAM in un altro Account AWS in tuo possesso nella Guida per l'utente IAM.
- Per informazioni su come fornire l'accesso alle risorse ad Account AWS di terze parti, consulta Fornire l'accesso agli Account AWS di proprietà di terze parti nella Guida per l'utente IAM.
- Per informazioni su come fornire l'accesso tramite la federazione delle identità, consulta Fornire l'accesso a utenti autenticati esternamente (Federazione delle identità) nella Guida per l'utente IAM.
- Per informazioni sulle differenze tra l'utilizzo di ruoli e policy basate su risorse per l'accesso multi-account, consulta Differenza tra i ruoli IAM e le policy basate su risorse nella Guida per l'utente IAM.

Autenticazione nei nodi cluster Amazon EMR

I client SSH possono utilizzare una coppia di chiavi Amazon EC2 per l'autenticazione in istanze cluster. In alternativa, con Amazon EMR rilasci 5.10.0 e successivi, puoi configurare Kerberos per autenticare utenti e connessioni SSH sul nodo primario. E con i rilasci 5.12.0 e successivi di Amazon EMR, puoi autenticarti con LDAP.

Argomenti

- Utilizzo di una coppia di chiavi EC2 per le credenziali SSH
- Utilizzo di Kerberos per l'autenticazione con Amazon EMR
- Utilizzo di server Active Directory o LDAP per l'autenticazione con Amazon EMR

Utilizzo di una coppia di chiavi EC2 per le credenziali SSH

I nodi cluster Amazon EMR vengono eseguiti su istanze Amazon EC2. Puoi connetterti a nodi cluster nello stesso modo in cui puoi connetterti alle istanze Amazon EC2. Puoi utilizzare Amazon EC2 per creare una coppia di chiavi oppure per importare una coppia di chiavi. Quando crei un cluster, puoi specificare la coppia di chiavi Amazon EC2 che verrà utilizzata per le connessioni SSH a tutte le istanze cluster. Puoi anche creare un cluster senza una coppia di chiavi. Questa operazione viene di solito effettuata con cluster transitori che si avviano, eseguono fasi e quindi terminano in automatico.

Il client SSH che utilizzi per connetterti al cluster deve utilizzare il file della chiave privata associato a tale coppia di chiavi. Questo è un file .pem per client SSH che utilizzano Linux, Unix e macOS. Devi impostare le autorizzazioni in modo che solo il proprietario della chiave disponga dell'autorizzazione per accedere al file. Si tratta di un file .ppk per i client SSH che utilizzano Windows ed è in genere creato a partire dal file .pem.

- Per ulteriori informazioni sulla creazione di una coppia di chiavi Amazon EC2, consulta Coppia di chiavi Amazon EC2 nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Per istruzioni sull'utilizzo di PuTTYgen per creare un file .ppk a partire da un file .pem, consulta Conversione della chiave privata mediante PuTTYgen nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Per ulteriori informazioni sull'impostazione delle autorizzazioni del file .pem e su come connettersi a un nodo primario del cluster EMR utilizzando metodi differenti, tra cui `ssh` da Linux o macOS, PuTTY da Windows o la AWS CLI da qualsiasi sistema operativo supportato, consulta Connessione al nodo primario tramite SSH.

Utilizzo di Kerberos per l'autenticazione con Amazon EMR

Amazon EMR rilascio 5.10.0 e successivi supporta Kerberos. Kerberos è un protocollo di autenticazione di rete che utilizza la crittografia con chiave segreta per fornire un'autenticazione efficace affinché le password o altre credenziali non siano inviate in rete in un formato non crittografato.

In Kerberos, i servizi e gli utenti che devono effettuare l'autenticazione sono denominati principali. I principali si trovano in un realm Kerberos. Nel realm, un server Kerberos, denominato Key Distribution Center (KDC), fornisce ai principali i mezzi per eseguire l'autenticazione. A questo proposito, il KDC emette dei ticket per l'autenticazione. Il KDC gestisce un database dei principali nel realm, le relative password e altre informazioni amministrative su ogni principale. Un KDC può anche

accettare credenziali di autenticazione da principali in altri realm; questa condizione è nota come trust tra realm. Inoltre, un cluster EMR può utilizzare un server KDC esterno per autenticare i principali.

La definizione di un trust tra realm o l'uso di un server KDC esterno è di uso comune nell'autenticazione degli utenti da un dominio di Active Directory. Questo consente agli utenti di accedere a un cluster EMR con il proprio account di dominio se si connettono a un cluster o lavorano con applicazioni Big Data utilizzando SSH.

Quando si utilizza l'autenticazione Kerberos, Amazon EMR configura Kerberos per le applicazioni, i componenti e i sottosistemi che installa sul cluster di modo che si autentichino a vicenda.

> **Important**
>
> Amazon EMR non supporta AWS Directory Service for Microsoft Active Directory in un trust fra realm o come server KDC esterno.

Prima di configurare Kerberos utilizzando Amazon EMR, ti consigliamo di acquisire familiarità con i concetti di Kerberos, i servizi eseguiti su un KDC e gli strumenti di amministrazione dei servizi Kerberos. Per ulteriori informazioni, consulta la Documentazione di MIT Kerberos, pubblicata dal Kerberos Consortium.

Argomenti

Applicazioni supportate

In un cluster EMR, i principali Kerberos sono i sottosistemi e i servizi di applicazioni di Big Data eseguiti su tutti i nodi di cluster. Amazon EMR può configurare le applicazioni e i componenti elencati di seguito per utilizzare Kerberos. Ogni applicazione ha un principale utente Kerberos a cui è associato.

Amazon EMR non supporta relazioni di fiducia tra realm con AWS Directory Service for Microsoft Active Directory.

Amazon EMR configura solo le caratteristiche di autenticazione Kerberos open source per le applicazioni e i componenti elencati di seguito. Qualsiasi altra applicazione installata non utilizza Kerberos e ciò può comportare un'incapacità a comunicare con i componenti che utilizzano Kerberos e causare errori nelle applicazioni. Per le applicazioni e i componenti che non utilizzano Kerberos, l'autenticazione non è abilitata. Le applicazioni e i componenti supportati possono variare a seconda dei rilasci di Amazon EMR.

L'interfaccia utente di Livy è l'unica interfaccia utente Web ospitata sul cluster che è Kerberizzato.

- Hadoop MapReduce
- Hbase
- HCatalog
- HDFS
- Hive
  - Non abilitare Hive con l'autenticazione LDAP. Questa operazione può causare problemi di comunicazione con l'applicazione YARN che utilizza Kerberos.
- Hue
  - L'autenticazione utente Hue non è impostata automaticamente e può essere configurata utilizzando l'API di configurazione.
  - Il server Hue utilizza Kerberos. Il front-end Hue (interfaccia utente) non è configurato per l'autenticazione. L'autenticazione LDAP può essere configurata per l'interfaccia utente Hue.
- Livy
  - La rappresentazione di Livy con cluster che utilizzano Kerberos è supportata in Amazon EMR 5.22.0 e rilasci successivi.
- Oozie
- Phoenix
- Presto
  - Presto supporta l'autenticazione Kerberos nei rilasci di Amazon EMR 6.9.0 e successivi.
  - Per utilizzare l'autenticazione Kerberos per Presto, devi abilitare la crittografia in transito.
- Spark
- Tez

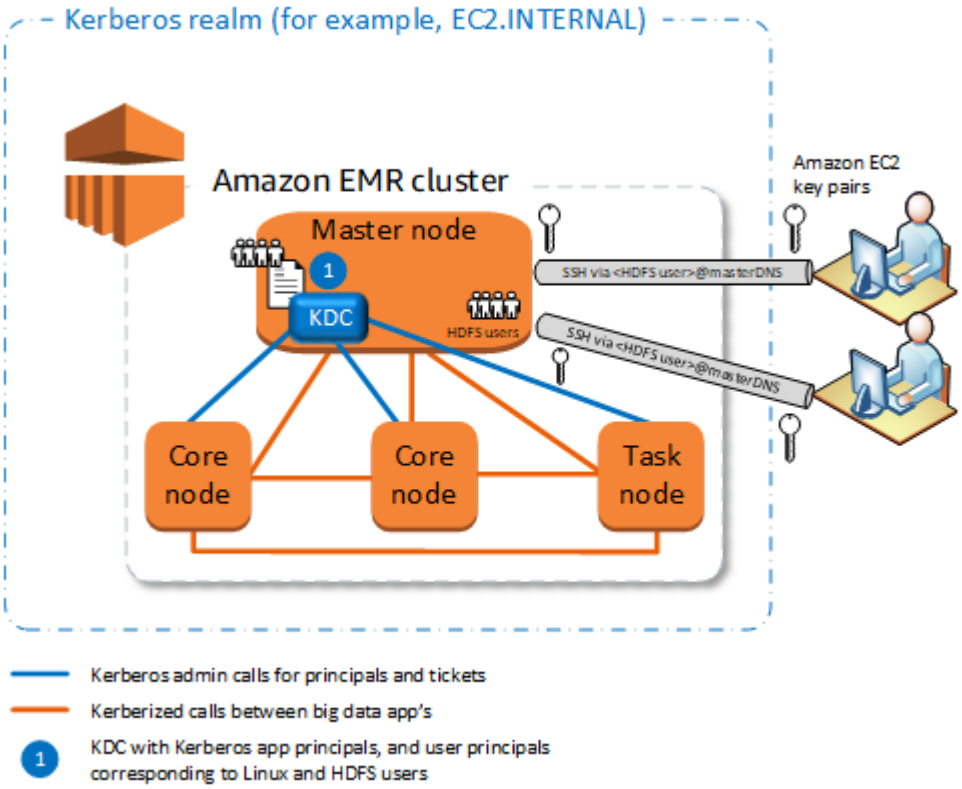
- Trino
  - Trino supporta l'autenticazione Kerberos in Amazon EMR rilasci 6.11.0 e successivi.
  - Per utilizzare l'autenticazione Kerberos per Trino, devi abilitare la crittografia in transito.
- YARN
- Zeppelin
  - Zeppelin è configurato solo per l'utilizzo di Kerberos con l'interprete Spark. Non è configurato per altri interpreti.
  - La rappresentazione dell'utente non è supportata per gli interpreti Zeppelin che utilizzano Kerberos diversi da Spark.
- Zookeeper
  - Il client Zookeeper non è supportato.

Opzioni di architettura di Kerberos

Utilizzando Kerberos con Amazon EMR, è possibile scegliere tra le architetture elencate in questa sezione. Indipendentemente dall'architettura scelta, Kerberos può essere configurato con la stessa procedura. Si crea una configurazione di sicurezza, si specifica la configurazione di sicurezza e le opzioni compatibili di Kerberos specifiche per il cluster quando si crea il cluster stesso, infine si creano directory HDFS per utenti Linux sul cluster corrispondenti ai principali per l'utente nel server KDC. Per una spiegazione delle opzioni di configurazione ed esempi di configurazione per ogni architettura, consulta Configurazione di Kerberos su Amazon EMR.

KDC dedicato del cluster (KDC su nodo primario)

Questa configurazione è disponibile con Amazon EMR 5.10.0 e rilasci successivi.



Vantaggi

- Amazon EMR ha la piena proprietà del server KDC.
- Il server KDC nel cluster EMR è indipendente dalle implementazioni KDC centralizzate, ad esempio Microsoft Active Directory o AWS Managed Microsoft AD.
- L'impatto sulle prestazioni è minimo perché il server KDC gestisce l'autenticazione solo per i nodi locali all'interno del cluster.
- Altri cluster con Kerberos possono facoltativamente fare riferimento al server KDC come KDC esterno. Per ulteriori informazioni, consulta KDC esterno - Nodo primario su un altro cluster.

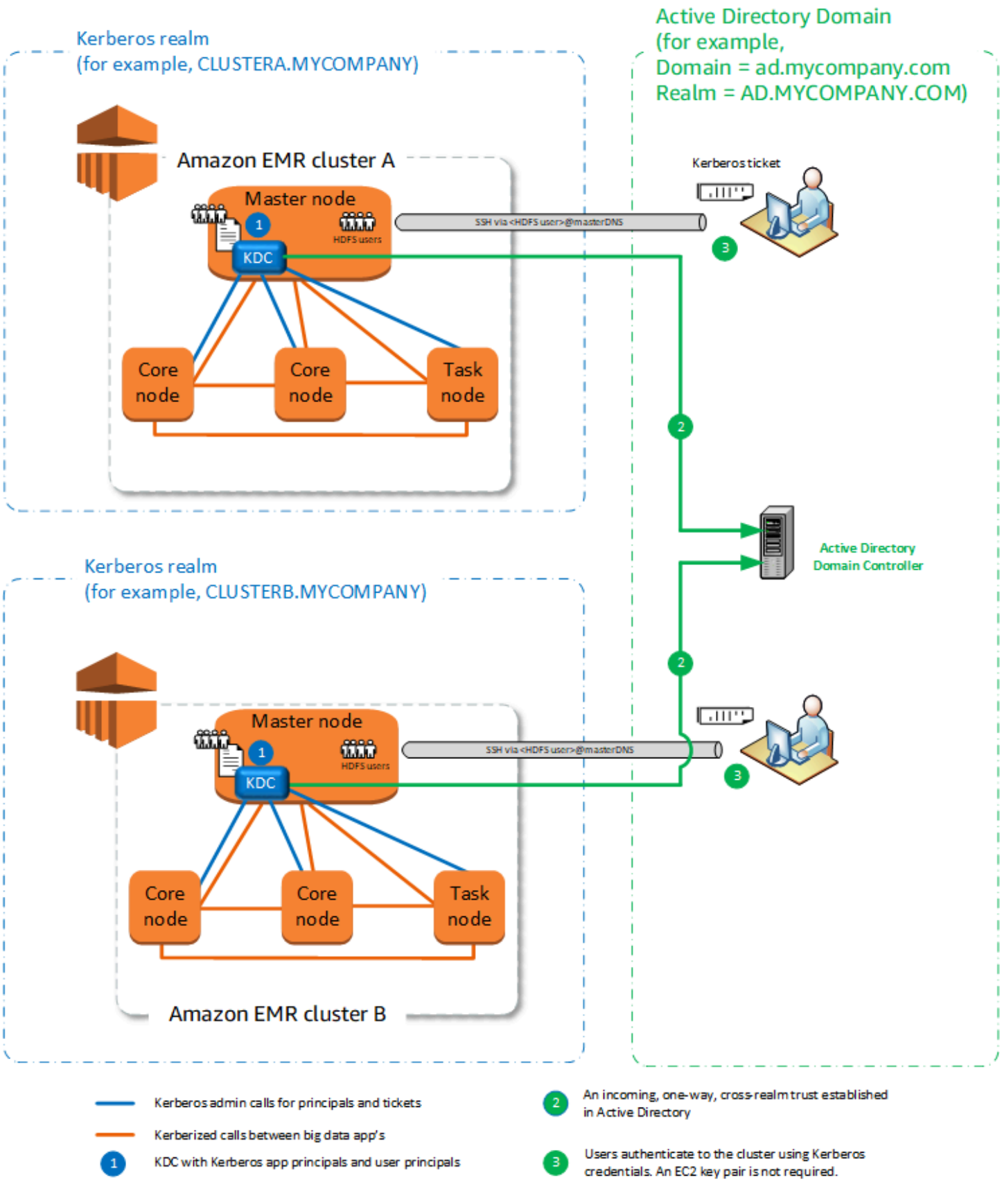
Considerazioni e limitazioni

- I cluster con Kerberos non possono eseguire l'autenticazione tra di loro, quindi le applicazioni non possono interagire. Se le applicazioni del cluster devono interagire, è necessario stabilire un trust tra realm tra i cluster o impostare un cluster come server KDC esterno per altri cluster. Se viene definito un trust tra realm, i server KDC devono avere realm Kerberos diversi.
- È necessario creare utenti Linux sull'istanza EC2 del nodo primario corrispondenti ai principali per l'utente KDC, nonché directory HDFS per ogni utente.

- I principali per l'utente devono utilizzare un file di chiave privata EC2 e credenziali `kinit` per connettersi al cluster tramite SSH.

Fiducia tra realm

In questa configurazione, i principali (di solito gli utenti) di un altro realm Kerberos eseguono l'autenticazione per i componenti applicativi su un cluster EMR con Kerberos, che ha il proprio server KDC. Il server KDC sul nodo primario stabilisce una relazione di trust con un altro KDC utilizzando un principale inter-realm esistente in entrambi i KDC. Il nome e la password del principale corrispondono in modo preciso in ciascun KDC. I trust tra realm sono più comuni nelle implementazioni con Active Directory, come illustrato nel seguente diagramma. Sono supportati anche trust tra realm con un server KDC MIT esterno o un KDC su un altro cluster Amazon EMR.


Kerberos realm
(for example, CLUSTERA.MYCOMPANY)
Amazon EMR cluster A
Master node
1
KDC
HDFS users
SSH via <HDFS user>@masterDNS
Core node
Core node
Task node
Active Directory Domain
(for example,
Domain = ad.mycompany.com
Realm = AD.MYCOMPANY.COM)
Kerberos ticket
3
2
Active Directory
Domain Controller
Kerberos realm
(for example, CLUSTERB.MYCOMPANY)
Master node
1
KDC
HDFS users
SSH via <HDFS user>@masterDNS
Core node
Core node
Task node
Amazon EMR cluster B
2
3
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.

Vantaggi

- Il cluster EMR su cui è installato il server KDC ne mantiene la piena proprietà.
- Con Active Directory, Amazon EMR crea automaticamente utenti Linux corrispondenti alle entità principali per l'utente dal server KDC. È comunque necessario creare directory HDFS per ogni utente. Inoltre, i principali per l'utente nel dominio Active Directory possono accedere ai cluster con Kerberos utilizzando credenziali `kinit`, senza il file di chiave privata EC2. In questo modo viene meno la necessità di condividere il file di chiave privata tra gli utenti del cluster.
- Poiché ogni KDC del cluster gestisce l'autenticazione per i nodi del cluster, sono ridotti al minimo gli effetti della latenza di rete e il sovraccarico di elaborazione per un numero elevato di nodi tra i cluster.

Considerazioni e limitazioni

- Se si sta creando un trust con un realm Active Directory, è necessario fornire un nome utente e una password di Active Directory con le autorizzazioni per l'aggiunta di principali al dominio in cui si crea il cluster.
- Non è possibile stabilire trust tra realm Kerberos con lo stesso nome.
- I trust tra realm devono essere stabiliti in modo esplicito. Ad esempio, se il Cluster A e il Cluster B stabiliscono entrambi un trust tra realm con un server KDC, non hanno intrinsecamente una reciproca relazione di trust e le loro applicazioni non possono autenticarsi l'una con l'altra per interagire.
- I server KDC devono essere gestiti in modo indipendente e coordinati in modo che le credenziali dei principali per l'utente corrispondano in modo preciso.

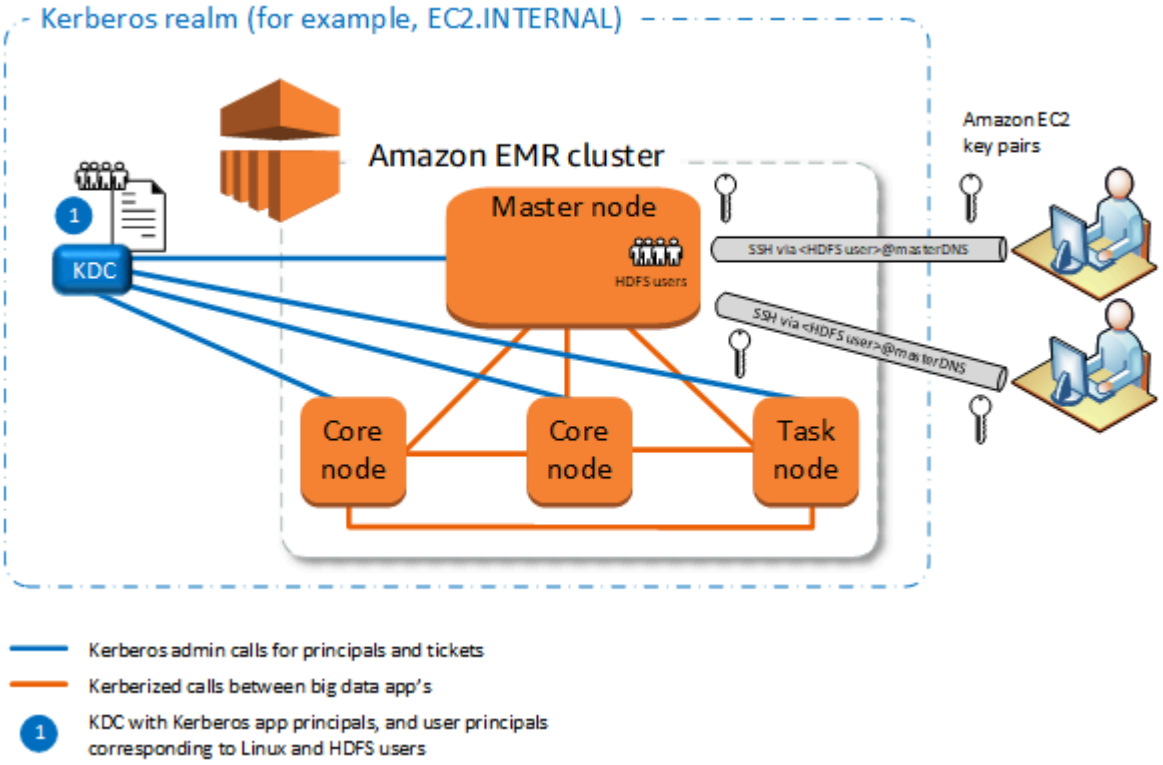
KDC esterno

Le configurazioni con un server KDC esterno sono supportate con Amazon EMR 5.20.0 e versioni successive.

- KDC esterno-KDC MIT
- KDC esterno - Nodo primario su un altro cluster
- KDC esterno: KDC del cluster su un cluster diverso con la relazione di fiducia tra realm Active Directory

KDC esterno-KDC MIT

Questa configurazione consente a uno o più cluster EMR di utilizzare principali definiti e gestiti in un server KDC MIT.



Vantaggi

- La gestione dei principali è consolidata in un unico KDC.
- Più cluster possono utilizzare lo stesso KDC nello stesso realm di Kerberos. Per ulteriori informazioni, consulta Requisiti per l'utilizzo di più cluster con lo stesso KDC.
- La gestione del KDC non comporta rallentamenti nelle prestazioni per un nodo primario su un cluster con Kerberos.

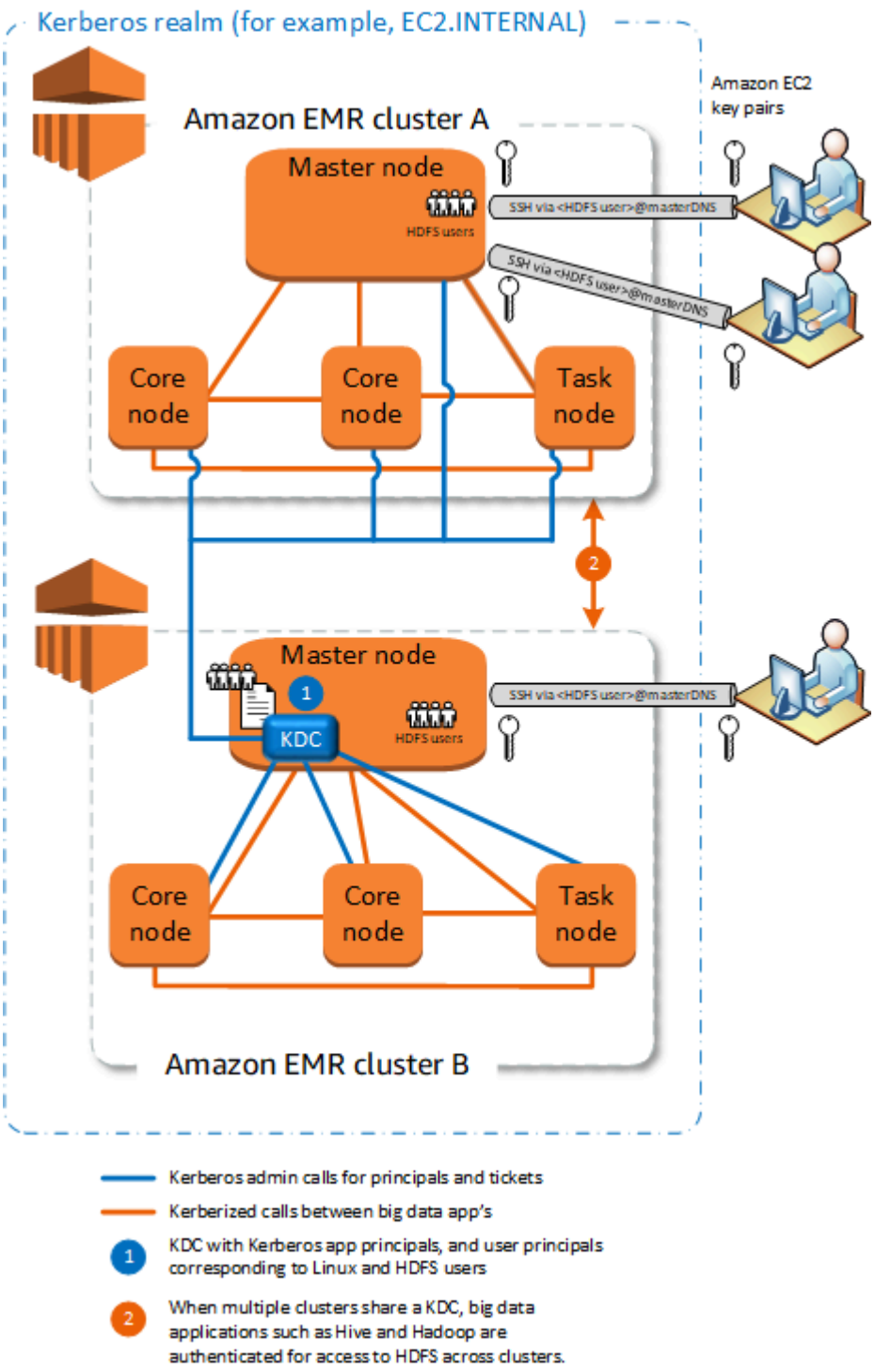
Considerazioni e limitazioni

- È necessario creare, sull'istanza EC2 di ogni nodo primario del cluster con Kerberos, utenti Linux corrispondenti ai principali per l'utente KDC, nonché directory HDFS per ogni utente.
- I principali per l'utente devono utilizzare un file di chiave privata EC2 e credenziali `kinit` per connettersi ai cluster con Kerberos tramite SSH.
- Ogni nodo nei cluster EMR con Kerberos deve avere un instradamento di rete al server KDC.

- L'autenticazione di ogni nodo nei cluster con Kerberos comporta un peso per il server KDC esterno, perciò la configurazione del KDC influisce sulle prestazioni del cluster. Quando si configura l'hardware del server KDC, è opportuno considerare il numero massimo di nodi Amazon EMR da supportare contemporaneamente.
- Le prestazioni del cluster dipendono dalla latenza di rete tra nodi nei cluster con Kerberos e il server KDC.
- La risoluzione dei problemi può essere più difficile a causa delle interdipendenze.

KDC esterno - Nodo primario su un altro cluster

Questa configurazione è quasi identica all'implementazione KDC MIT esterno qui sopra, l'unica differenza è che il KDC è attivo nel nodo primario di un cluster EMR. Per ulteriori informazioni, consulta KDC dedicato del cluster (KDC su nodo primario) e Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory.



Vantaggi

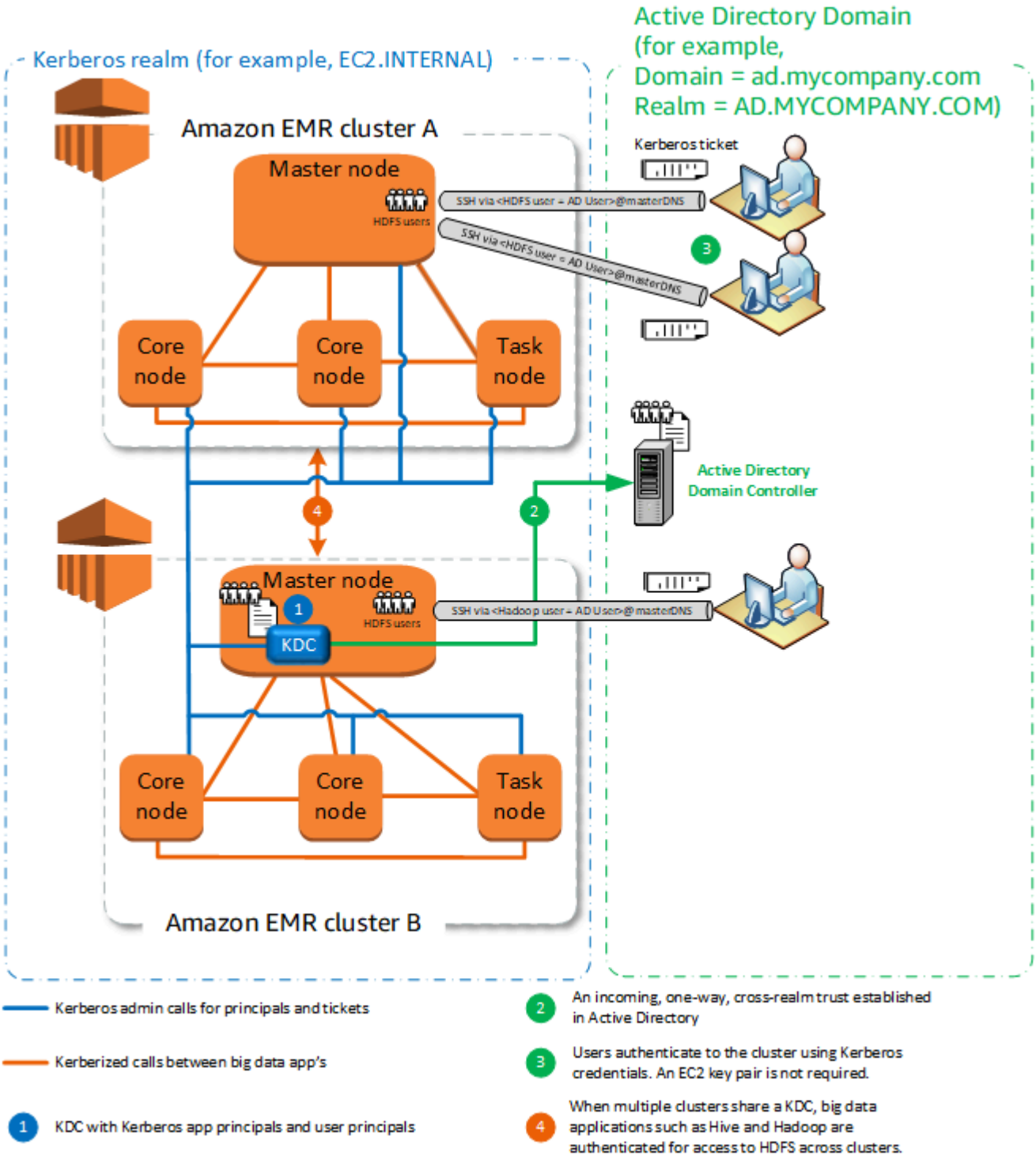
- La gestione dei principali è consolidata in un unico KDC.
- Più cluster possono utilizzare lo stesso KDC nello stesso realm di Kerberos. Per ulteriori informazioni, consulta Requisiti per l'utilizzo di più cluster con lo stesso KDC.

Considerazioni e limitazioni

- È necessario creare, sull'istanza EC2 di ogni nodo primario del cluster con Kerberos, utenti Linux corrispondenti ai principali per l'utente KDC, nonché directory HDFS per ogni utente.
- I principali per l'utente devono utilizzare un file di chiave privata EC2 e credenziali `kinit` per connettersi ai cluster con Kerberos tramite SSH.
- Ogni nodo in ciascun cluster EMR deve avere un instradamento di rete al server KDC.
- L'autenticazione di ogni nodo Amazon EMR nei cluster con Kerberos comporta un peso per il server KDC esterno, perciò la configurazione del KDC influisce sulle prestazioni del cluster. Quando si configura l'hardware del server KDC, è opportuno considerare il numero massimo di nodi Amazon EMR da supportare contemporaneamente.
- Le prestazioni del cluster dipendono dalla latenza di rete tra nodi nei cluster e il server KDC.
- La risoluzione dei problemi può essere più difficile a causa delle interdipendenze.

KDC esterno: KDC del cluster su un cluster diverso con la relazione di fiducia tra realm Active Directory

In questa configurazione, si crea prima un cluster con un server KDC dedicato che dispone di un trust tra realm monodirezionale con Active Directory. Per un tutorial dettagliato, consulta Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory. È quindi possibile avviare ulteriori cluster facendo riferimento al KDC del cluster con il trust come KDC esterno. Per vedere un esempio, consulta KDC esterno del cluster con trust tra realm Active Directory. In questo modo ogni cluster Amazon EMR che utilizza il KDC esterno può autenticare le entità principali definite e gestite in un dominio Microsoft Active Directory.



Vantaggi

- La gestione dei principali è consolidata nel dominio Active Directory.

- Amazon EMR entra nel realm di Active Directory; si elimina così la necessità di creare utenti Linux corrispondenti agli utenti di Active Directory. È comunque necessario creare directory HDFS per ogni utente.
- Più cluster possono utilizzare lo stesso KDC nello stesso realm di Kerberos. Per ulteriori informazioni, consulta Requisiti per l'utilizzo di più cluster con lo stesso KDC.
- I principali per l'utente nel dominio Active Directory possono accedere ai cluster con Kerberos utilizzando credenziali `kinit`, senza il file di chiave privata EC2. In questo modo viene meno la necessità di condividere il file di chiave privata tra gli utenti del cluster.
- Spetta a un solo un nodo primario Amazon EMR l'onere della manutenzione del KDC e solo quel cluster deve essere creato con credenziali di Active Directory per il trust tra realm tra il KDC e Active Directory.

Considerazioni e limitazioni

- Ogni nodo in ogni cluster EMR deve avere un instradamento di rete al server KDC e al controller di dominio di Active Directory.
- L'autenticazione di ogni nodo Amazon EMR comporta un peso per il server KDC esterno, perciò la configurazione del KDC influisce sulle prestazioni del cluster. Quando si configura l'hardware del server KDC, è opportuno considerare il numero massimo di nodi Amazon EMR da supportare contemporaneamente.
- Le prestazioni del cluster dipendono dalla latenza di rete tra nodi nei cluster e il server KDC.
- La risoluzione dei problemi può essere più difficile a causa delle interdipendenze.

Requisiti per l'utilizzo di più cluster con lo stesso KDC

Più cluster possono utilizzare lo stesso KDC nello stesso realm di Kerberos. Tuttavia, se i cluster vengono eseguiti contemporaneamente, potrebbero non riuscire se utilizzano nomi ServicePrincipal Kerberos in conflitto.

Se disponi di più cluster simultanei con lo stesso KDC esterno, assicurati che i cluster utilizzino realm Kerberos diversi. Se i cluster devono utilizzare lo stesso realm Kerberos, assicurati che si trovino in sottoreti diverse e che i relativi intervalli CIDR non si sovrappongano.

Configurazione di Kerberos su Amazon EMR

Questa sezione fornisce dettagli ed esempi di configurazione per configurare Kerberos con architetture comuni. Indipendentemente dall'architettura scelta, i passaggi di base delle

configurazione sono gli stessi e vengono effettuati in tre fasi. Se si utilizza un KDC esterno o si imposta un trust tra realm, è necessario accertarsi che ogni nodo in un cluster disponga di un instradamento di rete al KDC esterno, compresa la configurazione dei gruppi di sicurezza applicabili per consentire il traffico Kerberos in entrata e in uscita.

Fase 1: creazione di una configurazione di sicurezza con le proprietà di Kerberos

La configurazione di sicurezza specifica i dettagli del server KDC Kerberos e consente il riutilizzo della configurazione di Kerberos a ogni creazione di un cluster. Puoi creare una configurazione di sicurezza tramite la console di Amazon EMR, la AWS CLI o l'API EMR. La configurazione di sicurezza può inoltre contenere altre opzioni di sicurezza, ad esempio la crittografia. Per ulteriori informazioni sulla creazione di configurazioni di sicurezza e la specifica di una configurazione di sicurezza al momento della creazione di un cluster, consulta Utilizzo delle configurazioni di sicurezza per impostare la sicurezza del cluster. Per informazioni sulle proprietà di Kerberos in una configurazione di sicurezza, consulta Impostazioni Kerberos per configurazioni di sicurezza.

Fase 2: creazione di un cluster e specifica di attributi Kerberos appositi per il cluster

Quando crei un cluster, specifichi una configurazione di sicurezza Kerberos insieme alle opzioni di Kerberos specifiche del cluster stesso. Quando utilizzi la console di Amazon EMR, sono disponibili solo le opzioni di Kerberos compatibili con la configurazione di sicurezza specificata. Quando utilizzi la AWS CLI o l'API di Amazon EMR, assicurati di specificare opzioni di Kerberos compatibili con la configurazione di sicurezza specificata. Ad esempio, se quando crei un cluster con la CLI specifichi una password principale per un trust tra più realm ma la configurazione di sicurezza specificata non include parametri per questa situazione, si verifica un errore. Per ulteriori informazioni, consulta Impostazioni Kerberos per i cluster.

Fase 3: configurazione del nodo primario del cluster

In base ai requisiti dell'architettura e dell'implementazione, può essere necessaria una configurazione aggiuntiva sul cluster. È possibile farlo dopo la sua creazione o utilizzando operazioni di bootstrap o fasi durante il processo di creazione.

Per ogni utente autenticato Kerberos che si connette al cluster tramite SSH, è necessario accertarsi che vengano creati account Linux corrispondenti agli utenti di Kerberos. Se le entità principali dell'utente vengono fornite da un controller di dominio Active Directory, sia come KDC esterno che attraverso un trust tra realm, Amazon EMR crea automaticamente account Linux. Se Active Directory non viene utilizzato, è necessario creare principali per ciascun utente corrispondenti al rispettivo utente Linux. Per ulteriori informazioni, consulta Configurazione di un cluster per gli utenti HDFS con autenticazione in Kerberos e connessioni SSH.

Ogni utente deve inoltre disporre di una directory HDFS di proprietà, che è necessario creare. Inoltre, SSH deve essere configurato con GSSAPI abilitato per consentire le connessioni dagli utenti autenticati con Kerberos. GSSAPI deve essere abilitato sul nodo primario e l'applicazione client SSH deve essere configurata per l'utilizzo di GSSAPI. Per ulteriori informazioni, consulta Configurazione di un cluster per gli utenti HDFS con autenticazione in Kerberos e connessioni SSH.

Configurazione di sicurezza e impostazioni del cluster per Kerberos su Amazon EMR

Quando crei un cluster che utilizza Kerberos, specifichi la configurazione di sicurezza con attributi Kerberos specifici del cluster. Non puoi specificare un set senza l'altro, altrimenti si verifica un errore.

Questo argomento fornisce una panoramica dei parametri di configurazione disponibili per Kerberos al momento della creazione di una configurazione di sicurezza e di un cluster. Inoltre, sono riportati esempi di CLI per la creazione di configurazioni di sicurezza e cluster compatibili per architetture comuni.

Impostazioni Kerberos per configurazioni di sicurezza

Puoi creare una configurazione di sicurezza che specifica attributi Kerberos mediante la console di Amazon EMR, la AWS CLI o l'API di EMR. La configurazione di sicurezza può inoltre contenere altre opzioni di sicurezza, ad esempio la crittografia. Per ulteriori informazioni, consulta Creazione di una configurazione di sicurezza.

Utilizza i seguenti riferimenti per comprendere le impostazioni di configurazione di sicurezza disponibili per l'architettura Kerberos scelta. Vengono visualizzate le impostazioni della console di Amazon EMR. Per le corrispondenti opzioni con la CLI, consulta Configurazione delle impostazioni di Kerberos mediante la AWS CLI o Esempi di configurazione.

Parametro		Descrizione
Kerberos		Specifica che Kerberos è abilitato per i cluster che utilizzano questa configurazione di protezione. Se un cluster utilizza questa configurazione di protezione, deve avere anche le impostazioni Kerberos specificate o, in caso contrario, genererà un errore.
Provider	KDC dedicato del cluster	Specifica che Amazon EMR crea un KDC sul nodo primario di qualsiasi cluster che utilizza questa configurazione di sicurezza. Quando crei il cluster,

Parametro		Descrizione
		puoi specificare il nome del realm e la password di amministratore KDC. Se necessario, puoi fare riferimento a questo KDC da altri cluster. Crea tali cluster utilizzando una configura zione di sicurezza diversa, specifica un KDC esterno e utilizza il nome del realm e la password di amministr atore KDC specificati per il KDC dedicato al cluster.
	KDC esterno	Disponibili solo in Amazon EMR versione 5.20.0 e successive. Specifica che i cluster che utilizzan o questa configurazione di sicurezza autenticano le entità Kerberos principali utilizzando un server KDC esterno al cluster. Non viene creato un KDC nel cluster. Quando crei il cluster, specifichi il nome del realm e la password di amministratore KDC per il KDC esterno.
Durata del ticket		Facoltativo. Specifica il periodo di validità di un ticket Kerberos emesso dal KDC nei cluster che utilizzano questa configurazione di sicurezza. La durata dei ticket è limitata per motivi di sicurezza . Le applicazioni e i servizi del cluster rinnovano automaticamente i ticket dopo la loro scadenza. Gli utenti che si connettono al cluster tramite SSH utilizzan do le credenziali Kerberos devono eseguire `kinit` dalla linea di comando del nodo primario per rinnovare un ticket dopo la relativa scadenza.

Parametro		Descrizione
Fiducia tra realm		Specifica una relazione di fiducia tra realm tra un KDC dedicato al cluster in cluster che utilizzano questa configurazione di sicurezza e un KDC in un altro realm Kerberos. Le entità principali (in genere gli utenti) di un altro realm vengono autenticate nei cluster che utilizzano questa configurazione. È necessaria una configura zione aggiuntiva nell'altro realm Kerberos. Per ulteriori informazioni, consulta Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory.
Proprietà di fiducia tra realm	Realm (Dominio)	Specifica il nome di realm Kerberos dell'altro realm della relazione di fiducia. Per convenzione, i nomi del realm Kerberos sono uguali al nome di dominio, ma utilizzano solo lettere maiuscole.
	Dominio	Specifica il nome di dominio dell'altro realm della relazione di fiducia.
	Server amministratore	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server di amministrazione nell'altro realm della relazione di fiducia. Generalmente, il server amministratore e il server KDC vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma comunicano su porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 749, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :749`).

Parametro		Descrizione
	Server KDC	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server KDC nell'altro realm della relazione di fiducia. Generalmente, il server KDC e il server amministratore vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma utilizzano porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 88, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :88`).
KDC esterno		Specifica che i cluster KDC esterni vengono utilizzati dal cluster.
Proprietà del KDC esterno	Server amministratore	Specifica il nome di dominio completo (FQDN) o l'indirizzo IP del server amministratore esterno. Generalmente, il server amministratore e il server KDC vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma comunicano su porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 749, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :749`).

Parametro			Descrizione
	Server KDC		Specifica il nome di dominio completo (FQDN) del server KDC esterno. Generalmente, il server KDC e il server amministratore vengono eseguiti sullo stesso computer con lo stesso nome di dominio completo (FQDN), ma utilizzano porte diverse. Se non viene specificata alcuna porta, si utilizza la porta 88, ossia l'impostazione predefinita di Kerberos. Facoltativamente, puoi indicare la porta (ad esempio, `domain.example.com :88`).
	Integrazione con Active Directory		Specifica che l'autenticazione dell'entità Kerberos principale è integrata con un dominio Microsoft Active Directory.
	Proprietà di integrazione con Active Directory	Realm di Active Directory	Specifica il nome del realm Kerberos del dominio Active Directory. Per convenzione, i nomi del realm Kerberos sono uguali al nome di dominio, ma utilizzano solo lettere maiuscole.
		Dominio Active Directory	Specifica il nome del dominio di Active Directory.
		Server Active Directory	Specifica il nome di dominio completo (FQDN) del controller di dominio Microsoft Active Directory.

Impostazioni Kerberos per i cluster

Puoi specificare le impostazioni di Kerberos al momento della creazione di un cluster utilizzando la console Amazon EMR, la AWS CLI oppure l'API EMR.

Utilizza i seguenti riferimenti per comprendere le impostazioni di configurazione di cluster disponibili per l'architettura Kerberos scelta. Vengono visualizzate le impostazioni della console di Amazon EMR. Per le corrispondenti opzioni con la CLI, consulta Esempi di configurazione.

Parametro	Descrizione
Realm (Dominio)	Il nome di realm Kerberos per il cluster. La convenzione Kerberos consiste a impostare un nome identico al nome di dominio, ma in maiuscolo. Ad esempio, per il dominio `ec2.internal` , utilizza `EC2.INTERNAL` come nome di realm.
Password amministratore KDC	La password utilizzata nel cluster per `kadmin` o `kadmin.local` . Queste sono interfacce a riga di comando per il sistema di amministr azione Kerberos V5, che gestisce principali Kerberos, policy sulle password e keytab per il cluster.
Password del principale del trust tra realm (facoltativa)	Richiesta quando si stabilisce un trust tra realm. La password del principale inter-rea lm, che deve essere identica in tutti i realm. Utilizzare una password complessa.
Utente di aggiunta al dominio Active Directory (facoltativo)	Richiesto durante l'utilizzo di Active Directory in un trust tra realm. Questo è il nome di accesso utente per un account Active Directory con l'autorizzazione per aggiungere computer al dominio. Amazon EMR utilizza questa identità per aggiungere il cluster al dominio. Per ulteriori informazioni, consulta the section called “Fase 3: aggiunta di account al dominio per il cluster EMR”.
Password di aggiunta al dominio Active Directory (facoltativa)	La password per l'utente di aggiunta al dominio Active Directory. Per ulteriori informazioni,

Parametro	Descrizione
	consulta the section called “Fase 3: aggiunta di account al dominio per il cluster EMR”.

Esempi di configurazione

I seguenti esempi illustrano le configurazioni di sicurezza e del cluster per scenari comuni. Per brevità sono illustrati i comandi AWS CLI.

KDC locale

I seguenti comandi creano un cluster con un KDC dedicato in esecuzione sul nodo primario. È necessaria una configurazione aggiuntiva sul cluster. Per ulteriori informazioni, consulta Configurazione di un cluster per gli utenti HDFS con autenticazione in Kerberos e connessioni SSH.

Crea una configurazione di sicurezza

```
aws emr create-security-configuration --name LocalKDCSecurityConfig \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ClusterDedicatedKdc",\
"ClusterDedicatedKdcConfiguration": {"TicketLifetimeInHours": 24 }}}}'
```

Crea cluster

```
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 dedicato del cluster con trust tra realm Active Directory

I seguenti comandi creano un cluster con un KDC dedicato in esecuzione sul nodo primario e un trust tra realm con un dominio di Active Directory. È necessaria una configurazione aggiuntiva sul cluster e in Active Directory. Per ulteriori informazioni, consulta Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory.

Crea una configurazione di sicurezza

```
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"}}}}}'
```

Crea cluster

```
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 esterno su un altro cluster

I seguenti comandi creano un cluster che fa riferimento a un KDC dedicato sul nodo primario di un altro cluster per autenticare i principali. È necessaria una configurazione aggiuntiva sul cluster. Per ulteriori informazioni, consulta Configurazione di un cluster per gli utenti HDFS con autenticazione in Kerberos e connessioni SSH.

Crea una configurazione di sicurezza

```
aws emr create-security-configuration --name ExtKDCOnDifferentCluster \
--security-configuration '{"AuthenticationConfiguration": \
{"KerberosConfiguration": {"Provider": "ExternalKdc", \
"ExternalKdcConfiguration": {"KdcServerType": "Single", \
"AdminServer": "MasterDNSOfKDCMaster:749", \
"KdcServer": "MasterDNSOfKDCMaster:88"}}}}'
```

Crea cluster

```
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 esterno del cluster con trust tra realm Active Directory

I seguenti comandi creano un cluster senza KDC. Il cluster fa riferimento a un KDC dedicato in esecuzione sul nodo primario di un altro cluster per autenticare i principali. Il server KDC ha un trust tra realm con controller di dominio Active Directory. È necessaria una configurazione aggiuntiva sul nodo primario con il KDC. Per ulteriori informazioni, consulta Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory.

Crea una configurazione di sicurezza

```
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"}}}}}'
```

Crea cluster

```
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
```

Configurazione di un cluster per gli utenti HDFS con autenticazione in Kerberos e connessioni SSH

Amazon EMR crea client utente autenticati in Kerberos per le applicazioni eseguite sul cluster, ad esempio, l'utente `hadoop`, l'utente `spark` e altri utenti. Puoi anche aggiungere utenti autenticati durante i processi del cluster mediante Kerberos. Gli utenti autenticati possono quindi connettersi al cluster con le relative credenziali Kerberos e lavorare con le applicazioni. Per consentire l'autenticazione dell'utente nel cluster, sono necessarie le seguenti configurazioni:

- Nel cluster deve esistere un account Linux corrispondente all'entità principale di Kerberos nel KDC. Amazon EMR esegue questa operazione automaticamente nelle architetture che si integrano con Active Directory.
- È necessario creare una directory utente HDFS nel nodo primario per ogni utente e assegnare all'utente le autorizzazioni per la directory.
- È necessario configurare il servizio SSH in modo che GSSAPI sia abilitato per il nodo primario. Inoltre, gli utenti devono disporre di un client SSH con GSSAPI abilitato.

Aggiunta di utenti Linux ed entità principali Kerberos al nodo primario

Se non utilizzi Active Directory, dovrai creare gli account Linux sul nodo primario del cluster e aggiungere i principali per tali utenti Linux al KDC. Questo include un principale nel KDC per il nodo primario. Oltre ai principali dell'utente, il KDC in esecuzione sul nodo primario necessita di un principale per l'host locale.

Se l'architettura include l'integrazione con Active Directory, i principali e gli utenti Linux sul server locale KDC, se applicabili, vengono creati automaticamente. Puoi ignorare questa fase. Per ulteriori informazioni, consulta Fiducia tra realm e KDC esterno: KDC del cluster su un cluster diverso con la relazione di fiducia tra realm Active Directory.

> **Important**
>
> Il KDC, insieme al database dei principali, viene perso quando il nodo primario termina perché quest'ultimo utilizza l'archiviazione temporanea. Se si creano utenti per connessioni SSH, è consigliabile stabilire un trust tra realm con un KDC esterno configurato per la disponibilità elevata. In alternativa, se si creano utenti per connessioni SSH utilizzando account Linux, è possibile automatizzare il processo di creazione dell'account utilizzando operazioni di bootstrap e script in modo che possa essere ripetuto quando si crea un nuovo cluster.

Inviare una fase al cluster dopo averla creata oppure quando viene creato il cluster è il modo più semplice per aggiungere utenti e principali di KDC. In alternativa, è possibile connettersi al nodo primario utilizzando una coppia di chiavi EC2 come utente `hadoop` predefinito per l'esecuzione dei comandi. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH.

L'esempio seguente invia uno script Bash `configureCluster.sh` a un cluster già esistente, facendo riferimento al relativo ID di cluster. Lo script viene salvato in 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"]
```

L'esempio seguente mostra il contenuto dello script `configureCluster.sh`. Lo script gestisce anche la creazione di directory utente HDFS e l'abilitazione di GSSAPI per SSH, argomenti che saranno trattati nelle sezioni di seguito.

```
#!/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
```

Aggiunta di directory HDFS utente

Per consentire agli utenti di accedere al cluster ed eseguire processi Hadoop, è necessario aggiungere directory utente HDFS per i relativi account Linux e concedere a ciascun utente la proprietà della sua directory.

Inviare una fase al cluster dopo averla creata oppure quando viene creato il cluster è il modo più semplice per creare directory HDFS. In alternativa, è possibile connettersi al nodo primario utilizzando una coppia di chiavi EC2 come utente `hadoop` predefinito per l'esecuzione dei comandi. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH.

L'esempio seguente invia uno script Bash `AddHDFSUsers.sh` a un cluster già esistente, facendo riferimento al relativo ID di cluster. Lo script viene salvato in 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"]
```

L'esempio seguente mostra il contenuto dello 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
```

Abilitazione di GSSAPI per SSH

Perché gli utenti autenticati in Kerberos possano connettersi al nodo primario tramite SSH, per il servizio SSH deve essere abilitato il metodo di autenticazione GSSAPI. A questo scopo, esegui i

seguenti comandi dalla riga di comando del nodo primario oppure utilizza una fase per eseguirla come script. Dopo aver riconfigurato SSH, devi riavviare il servizio.

```
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
```

Utilizzo di SSH per la connessione a cluster con Kerberos

Questa sezione mostra la procedura per la connessione di un utente autenticato su Kerberos al nodo primario di un cluster EMR.

Ogni computer usato per una connessione SSH deve avere installati un client SSH e applicazioni client Kerberos. Con ogni probabilità nei computer Linux sono inclusi per impostazione predefinita. Ad esempio, OpenSSH è installato nella maggior parte dei sistemi operativi Linux, Unix e macOS. È possibile verificare la presenza di un client SSH digitando ssh nella riga di comando. Se il computer non riconosce il comando, installa un client SSH per la connessione al nodo primario. Il progetto OpenSSH fornisce un'implementazione gratuita della suite completa di strumenti SSH. Per ulteriori informazioni, consulta il sito Web OpenSSH. Gli utenti di Windows possono utilizzare applicazioni come PuTTY come client SSH.

Per ulteriori informazioni sulle connessioni SSH, vedi Connessione a un cluster.

SSH usa GSSAPI per autenticare i client Kerberos ed è necessario abilitare l'autenticazione GSSAPI per il servizio SSH nel nodo primario del cluster. Per ulteriori informazioni, consulta Abilitazione di GSSAPI per SSH. Anche i client SSH devono utilizzare GSSAPI.

Nei seguenti esempi, per *`MasterPublicDNS`*, utilizza il valore visualizzato in Master public DNS (DNS pubblico master) nella scheda Summary (Riepilogo) del riquadro dei dettagli del cluster, ad esempio, *`ec2-11-222-33-44.compute-1.amazonaws.com`*.

Prerequisito per krb5.conf (non Active Directory)

Quando si utilizza una configurazione senza l'integrazione di Active Directory, oltre al client SSH e alle applicazioni client Kerberos, ogni computer client deve disporre di una copia del file `/etc/krb5.conf` corrispondente al file `/etc/krb5.conf` sul nodo primario del cluster.

Per copiare il file krb5.conf

1. Utilizza SSH per connetterti al nodo primario tramite una coppia di chiavi EC2 e l'utente hadoop predefinito, ad esempio `hadoop@MasterPublicDNS`. Per istruzioni dettagliate, consulta Connessione a un cluster.
2. Dal nodo primario, copia i contenuti del file `/etc/krb5.conf`. Per ulteriori informazioni, consulta Connessione a un cluster.
3. Su ogni computer client utilizzato per la connessione al cluster, creare un file `/etc/krb5.conf` identico basato sulla copia creata nella fase precedente.

Utilizzo di Kinit e SSH

Ogni volta che si connette da un computer client con credenziali di Kerberos, l'utente deve prima rinnovare i ticket Kerberos per il proprio utente sul computer client. Inoltre, il client SSH deve essere configurato per utilizzare l'autenticazione GSSAPI.

Per utilizzare SSH per la connessione a un cluster EMR con Kerberos

1. Utilizzare `kinit` rinnovare i ticket Kerberos come nell'esempio seguente

   ```
   kinit user1
   ```

2. Utilizzare un client `ssh` insieme al principale creato nel KDC dedicato al cluster o al nome utente di Active Directory. Assicurarsi che l'autenticazione GSSAPI sia abilitata come illustrato negli esempi seguenti.

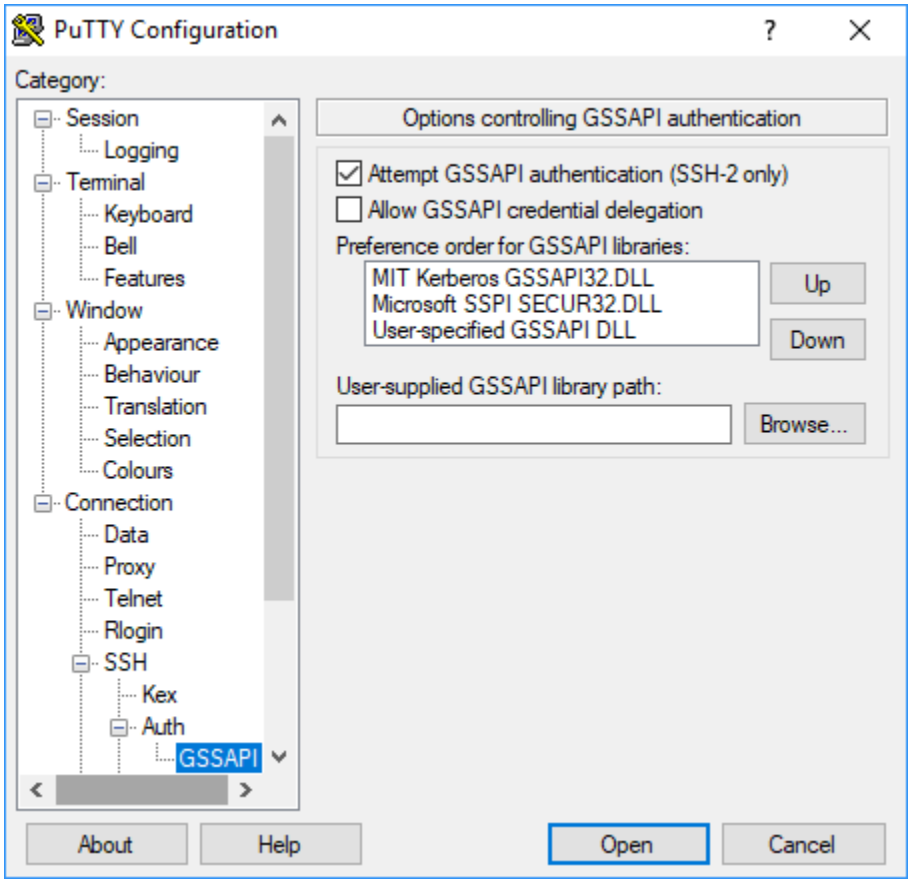
   Esempio: utenti Linux

   L'opzione `-K` specifica l'autenticazione GSSAPI.

   ```
   ssh -K user1@MasterPublicDNS
   ```

   Esempio: utenti Windows (PuTTY)

   Assicurarsi che l'opzione di autenticazione GSSAPI per la sessione sia abilitata come illustrato:



Tutorial: Configurazione di un KDC dedicato al cluster

Questo argomento illustra come creare un cluster con un key distribution center (KDC) dedicato, aggiungendo manualmente account Linux a tutti i nodi cluster, aggiungendo entità principali Kerberos al KDC sul nodo primario e verificando che un client Kerberos sia installato nei computer client.

Per maggiori informazioni sul supporto di Amazon EMR per Kerberos e KDC, nonché sui collegamenti alla documentazione di MIT Kerberos, consulta Utilizzo di Kerberos per l'autenticazione con Amazon EMR.

Fase 1: Creazione del cluster che utilizza Kerberos

1. Creare una configurazione di sicurezza che attiva Kerberos. L'esempio seguente illustra l'utilizzo di un comando `create-security-configuration` con l'AWS CLI che specifica la configurazione di sicurezza come struttura JSON inline. È anche possibile fare riferimento a un file salvato in locale.

```
aws emr create-security-configuration --name MyKerberosConfig \
```

```
--security-configuration '{"AuthenticationConfiguration": {"KerberosConfiguration":
{"Provider": "ClusterDedicatedKdc", "ClusterDedicatedKdcConfiguration":
 {"TicketLifetimeInHours": 24}}}}'
```

2. Creare un cluster che fa riferimento alla configurazione di sicurezza, stabilisce attributi Kerberos per il cluster e aggiunge account Linux utilizzando un'operazione di bootstrap. L'esempio seguente illustra l'utilizzo di un comando `create-cluster` con l'AWS CLI. Il comando fa riferimento alla configurazione di sicurezza creata precedentemente, `MyKerberosConfig`, nonché a uno script semplice, `createlinuxusers.sh`, come un'operazione di bootstrap, che deve essere creato e caricato in Amazon S3 prima di creare il cluster.

```
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
```

   Il codice seguente mostra il contenuto dello script `createlinuxusers.sh`, che aggiunge user1, user2 e user3 a ogni nodo nel cluster. Nella fase successiva, si aggiungeranno questi utenti come principali KDC.

```
#!/bin/bash
sudo adduser user1
sudo adduser user2
sudo adduser user3
```

Fase 2: Aggiunta di entità principali al KDC, creazione di directory utente HDFS e configurazione di SSH

Il KDC in esecuzione sul nodo primario necessita di un principale aggiunto per l'host locale e per ciascun utente creato nel cluster. È anche possibile creare directory HDFS per ogni utente che deve connettersi al cluster ed eseguire processi Hadoop. Analogamente, configurare il servizio SSH per

attivare l'autenticazione GSSAPI, necessaria per Kerberos. Dopo aver attivato GSSAPI, riavviare il servizio SSH.

Il modo più semplice di eseguire queste operazioni è di inviare una fase al cluster. L'esempio seguente invia uno script Bash `configurekdc.sh` al cluster creato nella fase precedente, facendo riferimento al relativo ID di cluster. Lo script viene salvato in Amazon S3. In alternativa, puoi connetterti al nodo primario utilizzando una coppia di chiavi EC2 per eseguire i comandi o inviare la fase durante la creazione del cluster.

```
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"]
```

Il codice seguente mostra il contenuto dello 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
```

Gli utenti aggiunti ora dovrebbero essere in grado di connettersi al cluster tramite SSH. Per ulteriori informazioni, consulta Utilizzo di SSH per la connessione a cluster con Kerberos.

Tutorial: Configurazione di un trust tra realm con un controller di dominio Active Directory

Quando configuri un trust tra realm, autorizzi i principali (di solito utenti) di un altro realm Kerberos a eseguire l'autenticazione a componenti dell'applicazione sul cluster EMR. Il key distribution center (KDC) dedicato al cluster stabilisce una relazione di fiducia con un altro KDC utilizzando un'entità principale tra realm esistente in entrambi i KDC. Il nome e la password del principale corrispondono in modo preciso.

Un trust tra realm richiede che i KDC possano accedere l'uno all'altro sulla rete e risolvere i nomi di dominio mutualmente. Di seguito sono riportate le fasi per stabilire una relazione di trust tra realm con un controller di dominio Microsoft AD eseguito come istanza EC2, nonché un esempio di configurazione di rete che fornisce la connettività e la risoluzione dei nomi di dominio necessarie. È accettabile qualsiasi configurazione di rete che consenta il traffico di rete tra server KDC.

Facoltativamente, dopo aver stabilito un trust tra realm con Active Directory utilizzando un KDC su un cluster, è possibile creare un altro cluster utilizzando un'altra configurazione di sicurezza per fare riferimento al KDC del primo cluster come KDC esterno. Per un esempio di configurazione di sicurezza e di impostazione del cluster, consultare KDC esterno del cluster con trust tra realm Active Directory.

Per maggiori informazioni sul supporto di Amazon EMR per Kerberos e KDC, nonché sui collegamenti alla documentazione di MIT Kerberos, consulta Utilizzo di Kerberos per l'autenticazione con Amazon EMR.

> **Important**
>
> Amazon EMR non supporta relazioni di fiducia tra realm con AWS Directory Service for Microsoft Active Directory.

Fase 1: Configurazione di VPC e sottorete

Fase 1: Configurazione di VPC e sottorete

Le fasi seguenti descrivono la creazione di un VPC e di una sottorete di modo che il KDC dedicato al cluster possa accedere al controller di dominio Active Directory e risolvere il relativo nome di dominio. In queste fasi, la risoluzione dei nomi di dominio viene fornita facendo riferimento al controller di dominio Active Directory come server dei nomi di dominio nel set di opzioni DHCP. Per ulteriori informazioni, consulta Fase 5: Utilizzo di un set di opzioni DHCP per specificare il controller di dominio Active Directory come server DNS di VPC.

Il KDC e il controller di dominio Active Directory devono essere in grado di risolvere uno il nome di dominio dell'altro. Ciò consente ad Amazon EMR di aggiungere computer al dominio e di configurare automaticamente gli account Linux e i parametri SSH corrispondenti sulle istanze del cluster.

Se Amazon EMR non è in grado di risolvere il nome di dominio, è possibile fare riferimento al trust utilizzando l'indirizzo IP del controller di dominio Active Directory. Tuttavia, è necessario aggiungere manualmente gli account Linux, aggiungere i principali corrispondenti al KDC dedicato al cluster e configurare SSH.

Per configurare VPC e sottorete

1. Creazione di un Amazon VPC con una singola sottorete pubblica Per ulteriori informazioni, consulta Fase 1: Creazione del VPC nella Guida alle nozioni di base su Amazon VPC.

   > **Important**
   >
   > Quando si utilizza un controller di dominio Microsoft Active Directory, scegliere un blocco CIDR per il cluster EMR di modo che la lunghezza di ogni indirizzo IPv4 sia inferiore

a nove caratteri (ad esempio, 10.0.0.0/16). Questo accade perché i nomi DNS dei computer cluster sono utilizzati quando si aggiungono i computer alla directory Active Directory. AWS assegna nomi host DNS in funzione dell'indirizzo IPv4, in modo che gli indirizzi IP più lunghi possano generare nomi DNS con più di 15 caratteri. Active Directory ha un limite di 15 caratteri per la registrazione di nomi di computer aggiunti e tronca i nomi più lunghi, condizione che può causare errori imprevedibili.

2. Rimuovere il set di opzioni DHCP di default assegnato al VPC. Per ulteriori informazioni, consulta Modifica di un VPC per non utilizzare opzioni DHCP. Successivamente, aggiungere un nuovo set che specifica il controller di dominio Active Directory come server DNS.
3. Confermare che il supporto DNS è attivato per il VPC, ovvero, che gli hostname DNS e la risoluzione DNS sono attivati (impostazione predefinita). Per ulteriori informazioni, consulta Aggiornamento del supporto DNS per il VPC.
4. Confermare che il VPC dispone di un gateway Internet associato (impostazione predefinita). Per ulteriori informazioni, consulta l'argomento relativo alla creazione e all'associazione di un gateway Internet.

   Note

   Un gateway Internet è utilizzato in questo esempio in quanto si sta creando un nuovo controller di dominio per il VPC. È tuttavia possibile che un gateway Internet non sia necessario per l'applicazione. Il solo requisito è che il KDC dedicato al cluster possa accedere al controller di dominio Active Directory.

5. Creare una tabella di routing personalizzata, aggiungere un instradamento che porta al gateway Internet, quindi associarlo alla sottorete. Per ulteriori informazioni, consulta Creazione di una tabella di routing personalizzata.
6. Quando si avvia l'istanza EC2 per il controller di dominio, questa deve avere un indirizzo IPv4 pubblico statico affinché sia possibile eseguire la connessione alla stessa tramite RDP. Il modo più semplice di procedere è di configurare la sottorete per l'assegnazione automatica degli indirizzi IPv4 pubblici, che non è l'impostazione predefinita quando si crea una sottorete. Per ulteriori informazioni, consulta Modifica dell'attributo di indirizzamento IPv4 pubblico della sottorete. Eventualmente, è possibile assegnare l'indirizzo all'avvio dell'istanza. Per ulteriori informazioni, consulta Assegnazione di un indirizzo IPv4 pubblico durante l'avvio dell'istanza.
7. Al termine, annotare l'ID del VPC e della sottorete. Questi ID saranno utilizzati in seguito quando si avvia il controller di dominio Active Directory e il cluster.

Fase 2: Avvio e installazione del controller di dominio Active Directory

1. Avviare un'istanza EC2 basata sull'AMI di base di Microsoft Windows Server 2016. Consigliamo un tipo di istanza m4.xlarge o superiore. Per ulteriori informazioni, consulta Avvio di un'istanza Marketplace AWS nella Guida per l'utente di Amazon EC2 per le istanze Windows.
2. Prendere nota dell'ID del gruppo di sicurezza associato all'istanza EC2. Servirà per Fase 6: avvio di un cluster EMR che utilizza Kerberos. In questo caso utilizziamo *sg-012xrlmdomain345*. In alternativa, è possibile specificare gruppi di sicurezza diversi per il cluster EMR e per questa istanza che consente il traffico tra di essi. Per ulteriori informazioni, consulta Gruppi di sicurezza Amazon EC2 per le istanze Linux nella Guida per l'utente di Amazon EC2 per le istanze Linux.
3. Connettersi all'istanza EC2 utilizzando RDP. Per ulteriori informazioni, consulta Connessione a un'istanza Windows nella Guida per l'utente di Amazon EC2 per le istanze Windows.
4. Avvia Server Manager per installare e configurare il ruolo dei servizi di dominio Active Directory sul server. Alzare il livello del server a controller di dominio e assegnare un nome di dominio (l'esempio utilizzato qui è *ad.domain.com*). Annotare il nome di dominio in quanto sarà necessario in un secondo momento quando si crea la configurazione di sicurezza EMR e il cluster. Per informazioni sulla configurazione di Active Directory, segui le istruzioni contenute in Come configurare Active Directory (AD) in Windows Server 2016.

   L'istanza viene riavviata al termine della procedura.

Fase 3: aggiunta di account al dominio per il cluster EMR

Esegui una connessione RDP al controller di dominio Active Directory per creare account in utenti e computer Active Directory per ogni utente del cluster. Per ulteriori informazioni, consulta Create a User Account in Active Directory Users and Computers sul sito Microsoft Learn. Annotare User logon name (Nome di accesso utente) di ogni utente. Questi nomi saranno necessari in seguito per la configurazione del cluster.

Inoltre, crea un account con privilegi sufficienti per aggiungere computer al dominio. Questo account viene specificato quando si crea un cluster. Amazon EMR lo utilizza per aggiungere istanze di cluster al dominio. Questo account e la relativa password sono specificati in Fase 6: avvio di un cluster EMR che utilizza Kerberos. Per delegare privilegi di aggiunta di computer all'account, è consigliabile creare un gruppo con privilegi di aggiunta e quindi assegnare l'utente al gruppo. Per istruzioni, consulta Delega dei privilegi di aggiunta di directory nella Guida per l'amministrazione di AWS Directory Service.

Fase 4: Configurazione di un trust in entrata sul controller di dominio Active Directory

I comandi di esempio seguenti creano un trust in Active Directory, ovvero un trust tra realm unidirezionale, in entrata e non transitivo con il KDC dedicato al cluster. L'esempio che utilizziamo per il realm del cluster è *EC2.INTERNAL*. Sostituisci *KDC-FQDN* con il nome del DNS pubblico elencato per il nodo primario Amazon EMR che ospita il KDC. Il parametro `passwordt` specifica la cross-realm principal password (password del principale inter-realm), definita insieme al realm del cluster quando si crea un cluster. Il nome di realm è derivato dal nome di dominio di default in `us-east-1` per il cluster. `Domain` è il dominio di Active Directory in cui si crea il trust, che è minuscolo per convenzione. L'esempio utilizza *ad.domain.com*.

Aprire il prompt dei comandi di Windows con privilegi di amministratore e digitare i comandi seguenti per creare la relazione di trust sul controller di dominio 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
```

Fase 5: Utilizzo di un set di opzioni DHCP per specificare il controller di dominio Active Directory come server DNS di VPC

Ora che il controller di dominio Active Directory è configurato, è necessario configurare il VPC per utilizzarlo come server dei nomi di dominio per la risoluzione dei nomi in VPC. A questo proposito, associare un set di opzioni DHCP. Specifica il Domain name (Nome di dominio) come nome di dominio del cluster, ad esempio `ec2.internal` se il cluster si trova in us-east-1 o *region*`.compute.internal` per le altre Regioni. In Domain name servers (Server nomi di dominio), è necessario specificare l'indirizzo IP del controller di dominio Active Directory (che deve essere accessibile dal cluster) come prima voce, seguito da AmazonProvidedDNS (ad esempio, ***xx.xx.xx.xx***, AmazonProvidedDNS). Per ulteriori informazioni, consulta Modifica dei set di opzioni DHCP.

Fase 6: avvio di un cluster EMR che utilizza Kerberos

1. In Amazon EMR, crea una configurazione di sicurezza che specifica il controller di dominio Active Directory creato nelle fasi precedenti. Un esempio di comando è riportato di seguito. Sostituire il dominio, *ad.domain.com*, con il nome del dominio specificato in Fase 2: Avvio e installazione del controller di dominio 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. Creare il cluster con gli attributi seguenti:

   - Utilizzare l'opzione `--security-configuration` per specificare la configurazione di sicurezza creata. Nell'esempio utilizziamo *`MyKerberosConfig`* .
   - Utilizzare la proprietà `SubnetId` di `--ec2-attributes option` per specificare la sottorete creata in Fase 1: Configurazione di VPC e sottorete. Nell'esempio utilizziamo *`step1-subnet`*.
   - Utilizza `AdditionalMasterSecurityGroups` e `AdditionalSlaveSecurityGroups` dell'opzione `--ec2-attributes` per specificare che il gruppo di sicurezza associato al controller di dominio AD da Fase 2: Avvio e installazione del controller di dominio Active Directory è associato al nodo primario del cluster, nonché ai nodi principali e attività. Nell'esempio utilizziamo *`sg-012xrlmdomain345`* .

   Utilizzare `--kerberos-attributes` per specificare i seguenti attributi Kerberos specifici del cluster:

   - Il realm per il cluster specificato nella configurazione del controller di dominio Active Directory.
   - La password del principale del trust tra realm specificata come `passwordt` in Fase 4: Configurazione di un trust in entrata sul controller di dominio Active Directory.
   - Una `KdcAdminPassword`, che può essere utilizzata per amministrare il KDC dedicato al cluster.
   - La password e il nome di accesso utente dell'account Active Directory con privilegi di aggiunta di computer creati in Fase 3: aggiunta di account al dominio per il cluster EMR.

   L'esempio seguente avvia un cluster che utilizza 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
```

Fase 7: creazione di utenti HDFS e impostazione di autorizzazioni sul cluster per account Active Directory

Quando si configura una relazione di trust con Active Directory, Amazon EMR crea utenti Linux sul cluster per ogni account Active Directory. Ad esempio, il nome di accesso utente `LiJuan` in Active Directory dispone di un account Linux `lijuan`. I nomi utente Active Directory possono contenere lettere maiuscole, ma Linux non supporta la combinazione di maiuscole e minuscole di Active Directory.

Per consentire agli utenti di accedere al cluster ed eseguire processi Hadoop, è necessario aggiungere directory utente HDFS per i relativi account Linux e concedere a ciascun utente la proprietà della sua directory. A questo proposito, consigliamo di eseguire uno script salvato in Amazon S3 come fase di cluster. In alternativa, è possibile eseguire i comandi nello script illustrato di seguito dalla riga di comando sul nodo primario. Utilizza la coppia di chiavi EC2 specificata durante la creazione del cluster per eseguire la connessione al nodo primario tramite SSH come utente Hadoop. Per ulteriori informazioni, consulta Utilizzo di una coppia di chiavi EC2 per le credenziali SSH.

Eseguire questo comando per aggiungere una fase al cluster che esegue uno 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"]
```

Il contenuto del file *AddHDFSUsers.sh* è illustrato di seguito.

```
#!/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
```

Gruppi Active Directory mappati a gruppi Hadoop

Amazon EMR utilizza System Security Services Daemon (SSD) per mappare gruppi Active Directory a gruppi Hadoop. Per confermare le mappature dei gruppi, dopo aver eseguito l'accesso al nodo primario come descritto in Utilizzo di SSH per la connessione a cluster con Kerberos, puoi utilizzare il comando `hdfs groups` per confermare che i gruppi Active Directory a cui l'account Active Directory appartiene siano stati mappati a gruppi Hadoop per l'utente Hadoop corrispondente sul cluster. È inoltre possibile verificare le mappature di gruppi di altri utenti specificando uno o più nomi utente con il comando, ad esempio `hdfs groups` *`lijuan`*. Per ulteriori informazioni, consulta groups nella Apache HDFS Commands Guide.

Utilizzo di server Active Directory o LDAP per l'autenticazione con Amazon EMR

Con i rilasci 6.12.0 e successivi di Amazon EMR, puoi utilizzare il protocollo LDAP over SSL (LDAPS) per avviare un cluster che si integra in modo nativo con il tuo server di identità aziendale. LDAP (Lightweight Directory Access Protocol) è un protocollo applicativo aperto e indipendente dal fornitore che accede e mantiene i dati. LDAP è comunemente usato per l'autenticazione degli utenti su server di identità aziendali ospitati su applicazioni come Active Directory (AD) e OpenLDAP. Con questa integrazione nativa, puoi utilizzare il tuo server LDAP per autenticare gli utenti su Amazon EMR.

I punti salienti dell'integrazione LDAP di Amazon EMR includono:

- Amazon EMR configura le applicazioni supportate per l'autenticazione con LDAP per tuo conto.
- Amazon EMR configura e mantiene la sicurezza per le applicazioni supportate con il protocollo Kerberos. Non è necessario immettere comandi o script.
- Puoi ottenere il controllo granulare degli accessi (FGAC) attraverso l'autorizzazione di Apache Ranger per il database e le tabelle di Hive Metastore. Per ulteriori informazioni, consulta Integrazione di Amazon EMR con Apache Ranger.
- Quando richiedi le credenziali LDAP per accedere a un cluster, ottieni un controllo granulare degli accessi (FGAC) su chi può accedere ai cluster EMR tramite SSH.

Le pagine seguenti forniscono una panoramica concettuale, i prerequisiti e le fasi per avviare un cluster EMR con l'integrazione LDAP di Amazon EMR.

Argomenti

Panoramica di LDAP con Amazon EMR

Lightweight Directory Access Protocol (LDAP) è un protocollo software utilizzato dagli amministratori di rete per gestire e controllare l'accesso ai dati autenticando gli utenti all'interno di una rete aziendale. Il protocollo LDAP memorizza le informazioni in una struttura di directory gerarchica ad albero. Per ulteriori informazioni, consulta Basic LDAP Concepts su LDAP.com.

All'interno della rete aziendale, molte applicazioni potrebbero utilizzare il protocollo LDAP per autenticare gli utenti. Con l'integrazione LDAP di Amazon EMR, i cluster EMR possono utilizzare in modo nativo lo stesso protocollo LDAP con una configurazione di sicurezza aggiuntiva.

Esistono due implementazioni principali del protocollo LDAP supportate da Amazon EMR: Active Directory e OpenLDAP. Sebbene siano possibili altre implementazioni, la maggior parte si adatta agli stessi protocolli di autenticazione di Active Directory o OpenLDAP.

Active Directory (AD)

Active Directory (AD) è un servizio di directory di Microsoft per reti di dominio Windows. AD è incluso nella maggior parte dei sistemi operativi Windows Server e può comunicare con i client tramite i protocolli LDAP e LDAPS. Per l'autenticazione, Amazon EMR tenta di associare un utente alla tua istanza AD utilizzando lo User Principal Name (UPN) come nome distinto e password. L'UPN utilizza il formato standard `username@domain_name`.

OpenLDAP

OpenLDAP è un'implementazione gratuita e open-source del protocollo LDAP. Per l'autenticazione, Amazon EMR tenta di associare un utente alla tua istanza OpenLDAP con il nome di dominio completo (FQDN) come nome distinto e password. L'FQDN utilizza il formato standard `username_attribute=username,LDAP_user_search_base`. In genere, il valore di `username_attribute` è `uid` e il valore `LDAP_user_search_base` contiene gli attributi della struttura ad albero che conduce all'utente. Ad esempio, `ou=People,dc=example,dc=com`.

Altre implementazioni gratuite e open-source del protocollo LDAP in genere seguono un FQDN simile a OpenLDAP per i nomi distinti dei loro utenti.

Componenti LDAP per Amazon EMR

Puoi utilizzare il tuo server LDAP per autenticarti con Amazon EMR e qualsiasi applicazione che l'utente utilizza direttamente sul cluster EMR tramite i seguenti componenti.

Agente segreto

L'agente segreto è un processo in cluster che autentica tutte le richieste degli utenti. L'agente segreto crea l'associazione utente al server LDAP per conto delle applicazioni supportate sul cluster EMR. L'agente segreto viene eseguito come utente `emrsecretagent` e scrive log nella directory `/emr/secretagent/log`. Questi log forniscono dettagli sullo stato della richiesta di autenticazione di ciascun utente e sugli eventuali errori che potrebbero emergere durante l'autenticazione dell'utente.

Daemon dei servizi di sicurezza del sistema (SSSD)

SSSD è un daemon che viene eseguito su ogni nodo di un cluster EMR abilitato per LDAP. SSSD crea e gestisce un utente UNIX per sincronizzare l'identità aziendale remota su ciascun nodo. Le applicazioni basate su YARN come Hive e Spark richiedono l'esistenza di un utente UNIX locale su ogni nodo che esegue una query per un utente.

Supporto e considerazioni sulle applicazioni con LDAP per Amazon EMR

Applicazioni supportate con LDAP per Amazon EMR

> **Important**
>
> Le applicazioni elencate in questa pagina sono le uniche applicazioni supportate da Amazon EMR per LDAP. Per garantire la sicurezza dei cluster, è possibile includere solo applicazioni compatibili con LDAP quando si crea un cluster EMR con il protocollo LDAP abilitato. Se tenti di installare altre applicazioni non supportate, Amazon EMR rifiuta la richiesta di un nuovo cluster.

I rilasci 6.12 e successivi di Amazon EMR supportano l'integrazione LDAP con le seguenti applicazioni:

- Apache Livy
- Apache Hive tramite HiveServer2 (HS2)
- Trino
- Presto
- Hue

Le seguenti applicazioni possono essere installate in un cluster EMR e possono essere configurate per soddisfare le esigenze di sicurezza:

- Apache Spark
- Apache Hadoop

Funzionalità supportate con LDAP per Amazon EMR

Puoi utilizzare le seguenti funzionalità di Amazon EMR con l'integrazione LDAP:

> **Note**
>
> Per proteggere le credenziali LDAP, devi utilizzare la crittografia in transito per proteggere il flusso di dati all'interno e all'esterno del cluster. Per maggiori informazioni sulla crittografia in transito, consulta Crittografia dei dati a riposo e in transito.

- Crittografia dei dati in transito (obbligatoria) e a riposo
- Gruppi di istanze, parchi istanze e istanze Spot
- Riconfigurazione delle applicazioni in un cluster in esecuzione
- Server-Side Encryption (SSE) EMRFS

Caratteristiche non supportate

Considera le seguenti limitazioni quando utilizzi l'integrazione LDAP di Amazon EMR:

- Amazon EMR disabilita i passaggi per i cluster con LDAP abilitato.
- Amazon EMR non supporta i ruoli di runtime e le integrazioni AWS Lake Formation per i cluster con LDAP abilitato.
- Amazon EMR non supporta LDAP con StartTLS.
- Amazon EMR non supporta la modalità ad alta disponibilità (cluster con più nodi primari) per i cluster con LDAP abilitato.
- Non puoi ruotare le credenziali o i certificati di associazione per i cluster con LDAP abilitato. Se uno di questi campi è stato ruotato, ti consigliamo di avviare un nuovo cluster con le credenziali o i certificati di associazione aggiornati.
- È necessario utilizzare basi di ricerca esatte con LDAP. La base di ricerca LDAP per utenti e gruppi non supporta i filtri di ricerca LDAP.

Configurazione e avvio di un cluster EMR con LDAP

Questa sezione spiega come configurare Amazon EMR per l'uso con l'autenticazione LDAP.

Argomenti

Aggiunta di autorizzazioni AWS Secrets Manager al ruolo dell'istanza Amazon EMR

Amazon EMR utilizza un ruolo di servizio IAM per eseguire operazioni per tuo conto per il provisioning e la gestione dei cluster. Il ruolo di servizio per le istanze EC2 del cluster, detto anche

profilo dell'istanza EC2 per Amazon EMR, è un tipo speciale di ruolo di servizio che Amazon EMR assegna a ogni istanza EC2 in un cluster all'avvio.

Per definire le autorizzazioni per fare in modo che un cluster EMR interagisca con i dati Amazon S3 e altri servizi AWS, definisci un profilo di istanza Amazon EC2 personalizzato da utilizzare al posto di `EMR_EC2_DefaultRole` quando avvii il cluster. Per ulteriori informazioni, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2) e Personalizzazione dei ruoli IAM.

Aggiungi le seguenti istruzioni al profilo dell'istanza EC2 predefinito per consentire ad Amazon EMR di contrassegnare le sessioni e accedere all'AWS Secrets Manager che memorizza i certificati 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**
>
> Le richieste dei cluster daranno esito negativo se dimentichi il carattere jolly * alla fine del nome segreto quando imposti le autorizzazioni di Secrets Manager. Il carattere jolly rappresenta le versioni segrete.
> Dovrai inoltre limitare l'ambito della policy di AWS Secrets Manager ai soli certificati di cui il cluster ha bisogno per il provisioning delle istanze.

Creazione della configurazione di sicurezza di Amazon EMR per l'integrazione LDAP

Prima di avviare un cluster EMR con integrazione LDAP, segui i passaggi riportati Creazione di una configurazione di sicurezza per creare una configurazione di sicurezza Amazon EMR per il cluster. Completa le seguenti configurazioni nel blocco `LDAPConfiguration` in `AuthenticationConfiguration` o nei campi corrispondenti nella sezione Configurazioni di sicurezza della console Amazon EMR:

`EnableLDAPAuthentication`

Opzione console: Protocollo di autenticazione: LDAP

Per utilizzare l'integrazione LDAP, imposta questa opzione su `true` o selezionala come protocollo di autenticazione quando crei un cluster nella console. Per impostazione predefinita, `EnableLDAPAuthentication` è `true` quando crei una configurazione di sicurezza nella console Amazon EMR.

`LDAPServerURL`

Opzione console: posizione del server LDAP

La posizione del server LDAP, incluso il prefisso: `ldaps://`*`location_of_server`*.

`BindCertificateARN`

Opzione console: certificato SSL LDAP

L'ARN AWS Secrets Manager che contiene il certificato per firmare il certificato SSL utilizzato dal server LDAP. Se il server LDAP è firmato da un'autorità di certificazione (CA) pubblica, puoi fornire un ARN AWS Secrets Manager con un file vuoto. Per ulteriori informazioni su come archiviare il certificato in Secrets Manager, consulta Archiviazione dei certificati TLS in AWS Secrets Manager.

`BindCredentialsARN`

Opzione console: credenziali di associazione del server LDAP

Un ARN AWS Secrets Manager che contiene le credenziali di associazione dell'utente amministratore LDAP. Le credenziali sono archiviate come oggetto JSON. Esiste una sola coppia chiave-valore in questo segreto; la chiave nella coppia è il nome utente e il valore è la password. Ad esempio, `{"uid=admin,cn=People,dc=example,dc=com": "AdminPassword1"}`. Questo campo è facoltativo, a meno che non abiliti l'accesso SSH per il cluster EMR. In molte

configurazioni, le istanze di Active Directory richiedono credenziali di associazione per consentire a SSSD di sincronizzare gli utenti.

`LDAPAccessFilter`

Opzione console: filtro di accesso LDAP

Specifica il sottoinsieme di oggetti all'interno del server LDAP che possono effettuare l'autenticazione. Ad esempio, se si desidera concedere l'accesso a tutti gli utenti con la classe di oggetti `posixAccount` nel server LDAP, il filtro di accesso viene definito come `(objectClass=posixAccount)`.

`LDAPUserSearchBase`

Opzione console: base di ricerca utenti LDAP

La base di ricerca a cui appartengono gli utenti all'interno del server LDAP. Ad esempio, `cn=People,dc=example,dc=com`.

`LDAPGroupSearchBase`

Opzione console: base di ricerca per gruppi LDAP

La base di ricerca a cui appartengono i gruppi all'interno del server LDAP. Ad esempio, `cn=Groups,dc=example,dc=com`.

`EnableSSHLogin`

Opzione console: accesso SSH

Specifica se consentire o meno l'autenticazione tramite password con credenziali LDAP. Ti consigliamo di abilitare questa opzione. Le coppie di chiavi rappresentano un percorso più sicuro per consentire l'accesso ai cluster EMR. Questo campo è facoltativo e, per impostazione predefinita, corrisponde a `false`.

`LDAPServerType`

Opzione console: tipo di server LDAP

Specifica il tipo di server LDAP a cui si connette Amazon EMR. Le opzioni supportate sono Active Directory e OpenLDAP. Altri tipi di server LDAP potrebbero funzionare, ma Amazon EMR non supporta ufficialmente altri tipi di server. Per ulteriori informazioni, consulta Componenti LDAP per Amazon EMR.

`ActiveDirectoryConfigurations`

Un blocco secondario necessario per le configurazioni di sicurezza che utilizzano il tipo di server Active Directory.

`ADDomain`

Opzione console: dominio Active Directory

Il nome di dominio utilizzato per creare lo User Principal Name (UPN) per l'autenticazione degli utenti con configurazioni di sicurezza che utilizzano il tipo di server Active Directory.

Considerazioni sulle configurazioni di sicurezza con LDAP e Amazon EMR

- Per creare una configurazione di sicurezza con l'integrazione LDAP di Amazon EMR, devi utilizzare la crittografia in transito. Per informazioni sulla crittografia in transito, consulta Crittografia dei dati a riposo e in transito.
- Non è possibile definire la configurazione Kerberos nella stessa configurazione di sicurezza. Amazon EMR fornisce un KDC dedicato all'automazione e gestisce la password di amministratore per questo KDC. Gli utenti non possono accedere alla password di amministratore.
- Non è possibile definire ruoli di runtime IAM e AWS Lake Formation nella stessa configurazione di sicurezza.
- Nel valore dell'`LDAPServerURL` deve essere compreso il protocollo `ldaps://`.
- Il `LDAPAccessFilter` non può essere vuoto.

Utilizzo di LDAP con l'integrazione di Apache Ranger per Amazon EMR

Con l'integrazione LDAP per Amazon EMR, puoi integrare ulteriormente con Apache Ranger. Quando trasferisci gli utenti LDAP in Ranger, puoi associare questi utenti a un server di policy Apache Ranger per eseguire l'integrazione con Amazon EMR e altre applicazioni. Per fare ciò, definisci il campo `RangerConfiguration` all'interno di `AuthorizationConfiguration` nella configurazione di sicurezza che usi con il cluster LDAP. Per ulteriori informazioni su come impostare la configurazione di sicurezza, consulta Creazione di una configurazione di sicurezza EMR.

Quando usi LDAP con Amazon EMR, non devi necessariamente fornire una `KerberosConfiguration` con l'integrazione Amazon EMR per Apache Ranger.

Avvio di un cluster EMR che si autentica con LDAP

Utilizza i seguenti passaggi per avviare un cluster EMR con LDAP o Active Directory.

1. Configurazione dell'ambiente:

   - Assicurati che i nodi del tuo cluster EMR possano comunicare con Amazon S3 e AWS Secrets Manager. Per ulteriori informazioni su come modificare il ruolo del profilo dell'istanza EC2 per comunicare con questi servizi, consulta Aggiunta di autorizzazioni AWS Secrets Manager al ruolo dell'istanza Amazon EMR.
   - Se prevedi di eseguire il tuo cluster EMR in una sottorete privata, dovresti utilizzare uno degli endpoint AWS PrivateLink e Amazon VPC o utilizzare la traduzione degli indirizzi di rete (NAT) per configurare il VPC per comunicare con S3 e Secrets Manager. Per ulteriori informazioni, consulta Endpoint AWS PrivateLink e VPC e Istanze NAT nella Guida alle operazioni di base di Amazon VPC.
   - Assicurati che ci sia connettività di rete tra il tuo cluster Amazon EMR e il server LDAP. I cluster EMR devono accedere al server LDAP tramite la rete. I nodi primario, principale e attività per il cluster comunicano con il server LDAP per la sincronizzazione dei dati degli utenti. Se il tuo server LDAP viene eseguito su Amazon EC2, aggiorna il gruppo di sicurezza EC2 per accettare il traffico dal cluster EMR. Per ulteriori informazioni, consulta Aggiunta di autorizzazioni AWS Secrets Manager al ruolo dell'istanza Amazon EMR.
2. Crea una configurazione di sicurezza di Amazon EMR per l'integrazione LDAP. Per ulteriori informazioni, consulta Creazione della configurazione di sicurezza di Amazon EMR per l'integrazione LDAP.
3. Ora che hai eseguito la configurazione, segui i passaggi indicati in Avvio di un cluster Amazon EMR per avviare il cluster con le seguenti impostazioni:

   - Seleziona il rilascio 6.12 o successivo di Amazon EMR. Consigliamo di utilizzare l'ultimo rilascio di Amazon EMR.
   - Specifica o seleziona solo le applicazioni per il cluster che supportano LDAP. Per un elenco di applicazioni supportate da LDAP con Amazon EMR, consulta Supporto e considerazioni sulle applicazioni con LDAP per Amazon EMR.
   - Applica la configurazione di sicurezza che hai creato nel passaggio precedente.

Esempi di utilizzo di LDAP con Amazon EMR

Dopo aver effettuato il provisioning di un cluster EMR che utilizza l'integrazione LDAP, puoi fornire le credenziali LDAP a qualsiasi applicazione supportata tramite il meccanismo di autenticazione integrato di nome utente e password. In questa pagina sono riportati alcuni esempi.

Utilizzo dell'autenticazione LDAP con Apache Hive

Example - Apache Hive

Il seguente comando di esempio avvia una sessione Apache Hive tramite HiveServer2 e Beeline:

```
beeline -u "jdbc:hive2://$HOSTNAME:10000/default;ssl=true;sslTrustStore=
$TRUSTSTORE_PATH;trustStorePassword=$TRUSTSTORE_PASS"  -n LDAP_USERNAME -
p LDAP_PASSWORD
```

Utilizzo dell'autenticazione LDAP con Apache Livy

Example - Apache Livy

Il comando di esempio seguente avvia una sessione Livy tramite cURL. Sostituisci *ENCODED-KEYPAIR* con una stringa codificata in Base64 per `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
```

Utilizzo dell'autenticazione LDAP con Presto

Example - Presto

Il seguente comando di esempio avvia una sessione Presto tramite la CLI Presto:

```
presto-cli --user "LDAP_USERNAME" --password --catalog hive
```

Dopo aver eseguito questo comando, quando richiesto, inserisci la password LDAP.

Utilizzo dell'autenticazione LDAP con Trino

Example - Trino

Il seguente comando di esempio avvia una sessione Trino tramite la CLI di Trino:

```
trino-cli --user "LDAP_USERNAME" --password --catalog hive
```

Dopo aver eseguito questo comando, quando richiesto, inserisci la password LDAP.

Utilizzo dell'autenticazione LDAP con Hue

Puoi accedere all'interfaccia utente di Hue tramite un tunnel SSH creato sul cluster oppure puoi impostare un server proxy per trasmettere pubblicamente la connessione a Hue. Poiché Hue non funziona in modalità HTTPS per impostazione predefinita, ti consigliamo di utilizzare un livello di crittografia aggiuntivo per garantire che la comunicazione tra i client e l'interfaccia utente di Hue sia crittografata con HTTPS. In questo modo si riduce la possibilità di esporre accidentalmente le credenziali utente in formato di testo normale.

Per utilizzare l'interfaccia utente di Hue, apri l'interfaccia utente di Hue nel browser e inserisci nome utente e password LDAP per accedere. Se le credenziali sono corrette, Hue effettua l'accesso e utilizza la tua identità per eseguire l'autenticazione con tutte le applicazioni supportate.

Utilizzo di SSH per l'autenticazione tramite password e ticket Kerberos per altre applicazioni

> **Important**
>
> Non è consigliabile utilizzare l'autenticazione tramite password su SSH in un cluster EMR.

È possibile utilizzare le credenziali LDAP per accedere a un cluster EMR tramite SSH. A tale scopo, imposta la configurazione di `EnableSSHLogin` su `true` nella configurazione di sicurezza di Amazon EMR utilizzata per avviare il cluster. Quindi, usa il seguente comando per accedere al cluster tramite SSH una volta avviato:

```
ssh username@EMR_PRIMARY_DNS_NAME
```

Dopo aver eseguito questo comando, quando richiesto, inserisci la password LDAP.

Amazon EMR include uno script su cluster che consente agli utenti di generare un file keytab Kerberos e un ticket da utilizzare con le applicazioni supportate che non accettano direttamente le credenziali LDAP. Alcune di queste applicazioni includono `spark-submit`, Spark SQL e PySpark.

Esegui `ldap-kinit` e segui le istruzioni. Se l'autenticazione ha esito positivo, il file keytab Kerberos viene visualizzato nella directory home con un ticket Kerberos valido. Utilizza il ticket Kerberos per eseguire le applicazioni come faresti in qualsiasi ambiente con Kerberos.

Integrazione di Amazon EMR con AWS Lake Formation

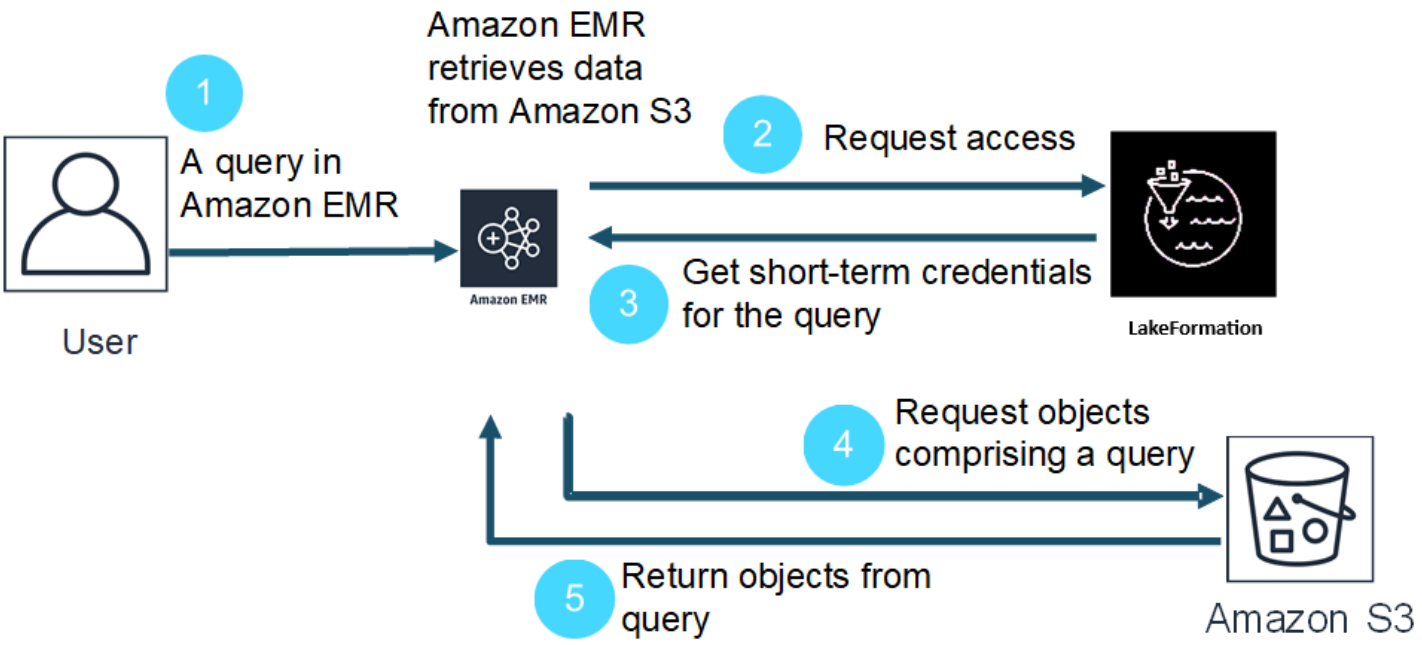
AWS Lake Formation è un servizio gestito che ti aiuta a rilevare, catalogare, pulire e proteggere i dati in un data lake Amazon Simple Storage Service (S3). Lake Formation fornisce l'accesso granulare a livello di colonna a database e tabelle in AWS Glue Data Catalog. Per ulteriori informazioni, consulta Che cos'è AWS Lake Formation?

Con Amazon EMR versione 6.7.0 e successive, puoi applicare il controllo degli accessi basato su Lake Formation ai processi Spark, Hive e Presto inviati ai cluster Amazon EMR. Per l'integrazione con Lake Formation, devi creare un cluster EMR con un ruolo di runtime. Un ruolo di runtime è un ruolo AWS Identity and Access Management (IAM) che puoi associare ai processi o alle query di Amazon EMR. Amazon EMR utilizza quindi questo ruolo per accedere alle risorse AWS. Per ulteriori informazioni, consulta Ruoli di runtime per le fasi di Amazon EMR.

Funzionamento di Amazon EMR con Lake Formation

Dopo aver integrato Amazon EMR con Lake Formation, puoi eseguire query sui cluster Amazon EMR con l'API `Step` o con SageMaker Studio. Quindi, Lake Formation fornisce l'accesso ai dati tramite credenziali temporanee per Amazon EMR. Questo processo è denominato distribuzione di credenziali. Per ulteriori informazioni, consulta Che cos'è AWS Lake Formation?

Di seguito è riportata una panoramica generale sul modo in cui Amazon EMR ottiene l'accesso ai dati protetti dalle policy di sicurezza Lake Formation.



1. Un utente invia una query Amazon EMR per i dati in Lake Formation.
2. Amazon EMR richiede le credenziali temporanee da Lake Formation per consentire all'utente di accedere ai dati.
3. Lake Formation restituisce le credenziali temporanee.
4. Amazon EMR invia la richiesta di query per recuperare dati da Amazon S3.
5. Amazon EMR riceve i dati da Amazon S3, li filtra e restituisce i risultati in base alle autorizzazioni utente definite in Lake Formation.

Per ulteriori informazioni sull'aggiunta di utenti e gruppi ai policy di Lake Formation, consulta Concessione delle autorizzazioni Data Catalog.

Prerequisiti

Prima di integrare Amazon EMR e Lake Formation, è necessario soddisfare i seguenti requisiti:

- Attiva l'autorizzazione dei ruoli di runtime sul cluster Amazon EMR.
- Utilizza il Catalogo dati AWS Glue come archivio di metadati.
- Definisci e gestisci le autorizzazioni in Lake Formation per accedere a database, tabelle e colonne in AWS Glue Data Catalog. Per ulteriori informazioni, consulta Che cos'è AWS Lake Formation?

Argomenti

Abilitazione di Amazon EMR con Lake Formation

Questa sezione illustra come creare una configurazione di sicurezza e impostare Lake Formation per l'utilizzo con Amazon EMR. Descrive inoltre come avviare un cluster con la configurazione di sicurezza creata per Lake Formation.

Passaggio 1: Impostazione di un ruolo di runtime per il cluster EMR

Per utilizzare un ruolo di runtime per il cluster EMR, devi prima creare una configurazione di sicurezza. Con una configurazione di sicurezza puoi applicare opzioni di sicurezza, autorizzazione e autenticazione coerenti in tutti i tuoi cluster.

1. Crea un file denominato `lf-runtime-roles-sec-cfg.json` con la seguente configurazione di sicurezza.

```
{
    "AuthorizationConfiguration":{
        "IAMConfiguration":{
            "EnableApplicationScopedIAMRole":true,
            "ApplicationScopedIAMRoleConfiguration":{
                "PropagateSourceIdentity":true
            }
        },
        "LakeFormationConfiguration":{
            "AuthorizedSessionTagValue":"Amazon EMR"
        },
  "EncryptionConfiguration": {
            "EnableInTransitEncryption": true,
            "InTransitEncryptionConfiguration": {
            "TLSCertificateConfiguration": {<Certificate-configuration>}
        }
    }
}
```

2. Quindi, per assicurarti che il tag di sessione possa autorizzare Lake Formation, imposta la proprietà `LakeFormationConfiguration/AuthorizedSessionTagValue` su `Amazon EMR`.
3. Utilizza il seguente comando per creare una configurazione di sicurezza Amazon EMR.

```
aws emr create-security-configuration \
--name 'iamconfig-with-iam-lf' \
--security-configuration file://lf-runtime-roles-sec-cfg.json
```

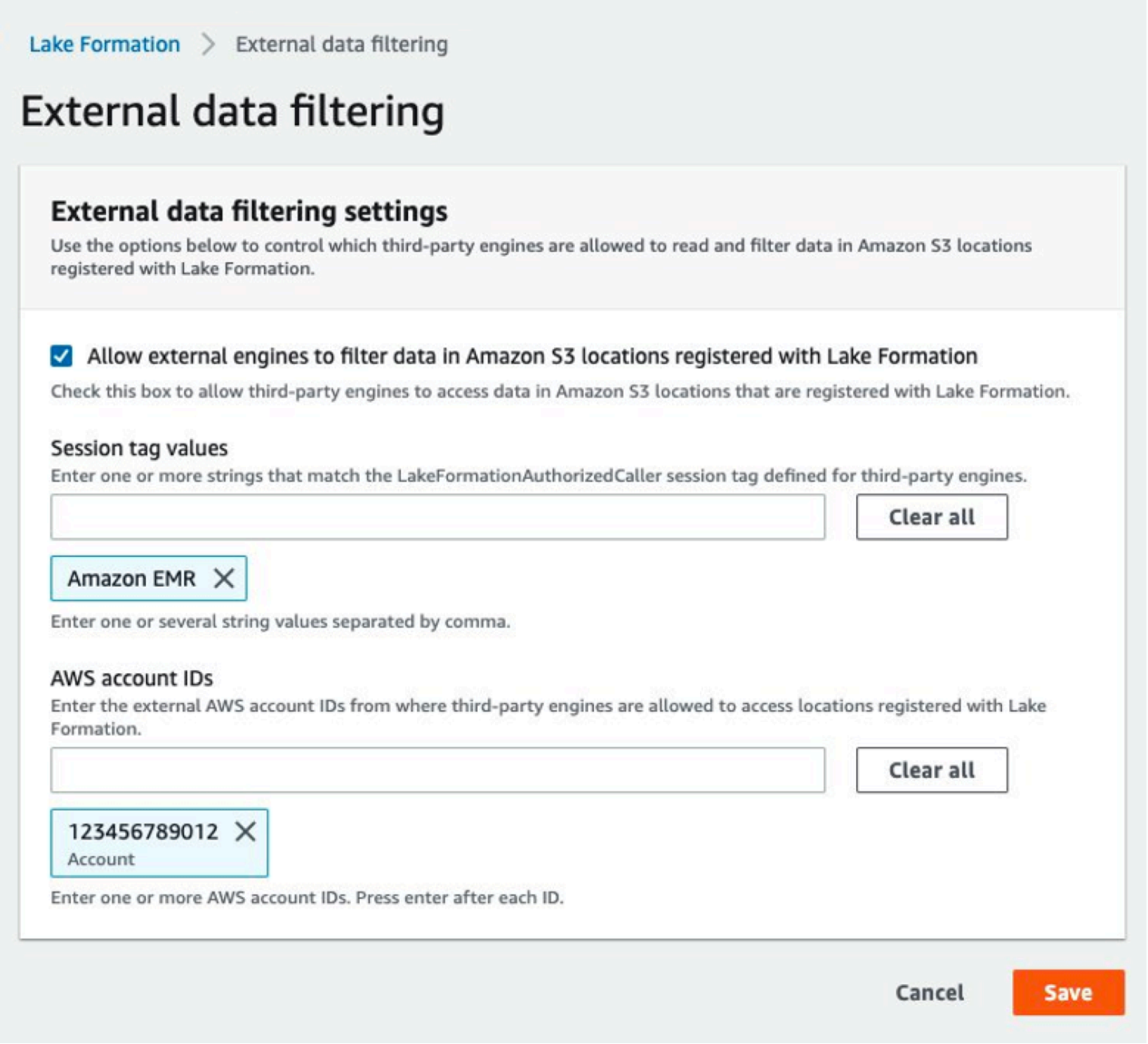
   In alternativa, puoi utilizzare la console Amazon EMR per creare una configurazione di sicurezza con impostazioni personalizzate.

Fase 2: avvio di un cluster Amazon EMR

Ora puoi avviare un cluster EMR con la configurazione di sicurezza creata nel passaggio precedente. Per ulteriori informazioni sulle configurazioni di sicurezza, consulta le sezioni Utilizzo delle configurazioni di sicurezza per impostare la sicurezza del cluster e Ruoli di runtime per le fasi di Amazon EMR.

Passaggio 4: configurazione del controllo degli accessi basato su Lake Formation con i ruoli di runtime di Amazon EMR

Per applicare le autorizzazioni a livello di tabella e colonna con Lake Formation, l'amministratore del data lake per Lake Formation deve impostare `Amazon EMR` come valore per la configurazione del tag di sessione, `AuthorizedSessionTagValue`. Lake Formation utilizza questo tag di sessione per autorizzare i chiamanti e fornire l'accesso al data lake. Puoi impostare questo tag di sessione nella sezione External data filtering (Filtraggio dati esterni) della console Lake Formation. Sostituisci *123456789012* con l'ID del tuo Account AWS.



Per continuare con la configurazione del controllo degli accessi basato su Lake Formation con i ruoli di runtime di Amazon EMR, devi configurare le concessioni di AWS Glue e Lake Formation per i ruoli di runtime di Amazon EMR. Per permettere ai ruoli di runtime IAM di interagire con Lake Formation, concedi loro l'accesso con `lakeformation:GetDataAccess` e `glue:Get*`.

Le autorizzazioni di Lake Formation controllano l'accesso alle risorse di AWS Glue Data Catalog, alle posizioni Amazon S3 e ai dati sottostanti in tali sedi. Le autorizzazioni IAM controllano l'accesso alle API e alle risorse di Lake Formation e AWS Glue. Anche se disponi dell'autorizzazione di Lake Formation per accedere a una tabella nel catalogo dati (SELECT), l'operazione fallisce se non disponi dell'autorizzazione IAM per l'API `glue:Get*`. Per maggiori dettagli sul controllo degli accessi in Lake Formation, consulta la sezione Lake Formation access control overview (Panoramica del controllo degli accessi a Lake Formation).

1. Crea il file `emr-runtime-roles-lake-formation-policy.json` con i seguenti contenuti.

```
{
    "Version": "2012-10-17",
    "Statement": {
        "Sid": "LakeFormationManagedAccess",
        "Effect": "Allow",
        "Action": [
            "lakeformation:GetDataAccess",
            "glue:Get*",
            "glue:Create*",
            "glue:Update*"
        ],
        "Resource": "*"
    }
}
```

2. Crea la policy IAM correlata.

```
aws iam create-policy \
--policy-name emr-runtime-roles-lake-formation-policy \
--policy-document file://emr-runtime-roles-lake-formation-policy.json
```

3. Per assegnare questa policy ai ruoli di runtime IAM, segui i passaggi in Managing AWS Lake Formation permissions.

Ora puoi utilizzare i ruoli di runtime e Lake Formation per applicare le autorizzazioni a livello di tabella e colonna. Puoi anche utilizzare un'identità di origine per controllare le operazioni e monitorare le operazioni con AWS CloudTrail. Per un esempio completo e dettagliato, consulta la sezione Introducing runtime roles for Amazon EMR steps (Introduzione dei ruoli di runtime per le fasi di Amazon EMR).

Apache Hudi e Lake Formation

La versione di rilascio 6.9.0 e successive di Amazon EMR includono un supporto limitato per il controllo degli accessi basato su Lake Formation con Apache Hudi durante la lettura dei dati tramite Spark SQL. Amazon EMR supporta query SELECT con Spark SQL ed è limitato al controllo degli accessi a livello di colonna. Con questa funzione, ora puoi eseguire:

- Query snapshot su tabelle copy-on-write (copia su scrittura) per eseguire query sullo snapshot più recente della tabella in un determinato commit o momento di compattazione.
- Query ottimizzate per la lettura su tabelle merge-on-read per eseguire query sui dati compatti più recenti, che potrebbero non includere gli aggiornamenti più recenti nei file di log non ancora compattati.

La seguente matrice di supporto elenca alcune funzioni principali di Apache Hudi con Lake Formation:

	Copia su scrittura	unisci in lettura
Query snapshot - Spark SQL	Y	N
Lettura di query ottimizzate - Spark SQL	Non applicabile	Y
Query incrementali	N	N
Query temporali	N	N
Query su origini dei dati Spark	N	N
Scrittura di origini dei dati Spark	N	N
DML/DDL	N	N

	Copia su scrittura	unisci in lettura
Tabella dei metadati	N	N

Esecuzione di query per tabelle Hudi

Questa sezione mostra come eseguire le query supportate sopra descritte su un cluster abilitato per Lake Formation. La tabella deve essere una tabella di catalogo registrata.

Questa sezione mostra come eseguire le query supportate su un cluster Lake Formation, come indicato in precedenza

1. Per avviare la shell (interprete di comandi) Spark, utilizza i comandi seguenti.

```
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. Per eseguire query per l'ultima snapshot di tabelle copy-on-write, utilizza i comandi seguenti.

```
select * from <my_hudi_cow_table>
```

```
spark.read.table("<my_hudi_cow_table>")
```

3. Per eseguire query sui dati compatti più recenti delle tabelle MOR, puoi utilizzare la tabella ottimizzata per la lettura che presenta il suffisso _ro:

```
SELECT * from <my_hudi_mor_table>_ro
```

```
spark.read.table("<my_hudi_mor_table>_ro")
```

> **Note**
>
> Le prestazioni delle letture possono essere più lente nei cluster Lake Formation a causa di ottimizzazioni non supportate. Queste funzioni includono l'elenco dei file basato sui metadati

Hudi e i dati ignorati. È preferibile testare le prestazioni dell'applicazione per accertarsi che soddisfi lo SLA.

Considerazioni

Tieni presente le seguenti considerazioni quando utilizzi Amazon EMR con AWS Lake Formation.

- Gli utenti con accesso a una tabella possono accedere a tutte le sue proprietà. Se disponi di un controllo degli accessi basato su Lake Formation su una tabella, controlla la tabella per assicurarti che le proprietà non contengano dati o informazioni sensibili.
- I cluster Amazon EMR con Lake Formation non supportano il fallback di Spark su HDFS quando Spark raccoglie le statistiche delle tabelle. Questo di solito aiuta a ottimizzare le prestazioni delle query.
- Le operazioni che supportano i controlli degli accessi basati su Lake Formation con tabelle Apache Spark e Apache Hive non gestite (da Amazon EMR 6.10.0 e versioni successive) includono `insert into` e `insert overwrite`.
- Le operazioni che supportano i controlli degli accessi basati su Lake Formation con Apache Spark e Apache Hive includono `select`, `describe`, `show database`, `show table`, `show column` e `show partition`.
- Amazon EMR non è in grado di controllare l'accesso alle seguenti operazioni basate su Lake Formation:
  - Scrive su tabelle regolate
  - Il filtro dati Lake Formation
  - Le istruzioni DDL come la tabella CREATE o ALTER
- Esistono differenze di prestazioni tra la stessa query con e senza il controllo degli accessi basato su Lake Formation.

Integrazione di Amazon EMR con Apache Ranger

A partire da Amazon EMR 5.32.0, è possibile avviare un cluster che si integra nativamente con Apache Ranger. Apache Ranger è un framework open source che consente di abilitare, monitorare e gestire la sicurezza completa dei dati attraverso la piattaforma Hadoop. Per ulteriori informazioni, consulta Apache Ranger. L'integrazione nativa consente di utilizzare Apache Ranger per imporre un controllo granulare di accesso ai dati su Amazon EMR.

Questa sezione fornisce una panoramica delle nozioni di base dell'integrazione di Amazon EMR con Apache Ranger. Include inoltre i prerequisiti e le fasi necessari per avviare un cluster Amazon EMR integrato con Apache Ranger.

L'integrazione nativa di Amazon EMR con Apache Ranger offre i seguenti vantaggi principali:

- Controllo dell'accesso granulare ai database e alle tabelle del metastore Hive, che consente di definire policy di filtraggio dei dati a livello di database, tabella e colonna per le applicazioni Apache Spark e Apache Hive. Il filtraggio a livello di riga e il mascheramento dei dati sono supportati con le applicazioni Hive.
- La possibilità di utilizzare le policy Hive esistenti direttamente con Amazon EMR per le applicazioni Hive.
- Controllo dell'accesso ai dati di Amazon S3 a livello di prefisso e oggetto, che consente di definire policy di filtraggio dei dati per l'accesso ai dati S3 utilizzando il file system EMR.
- La possibilità di utilizzare CloudWatch Logs per il controllo centralizzato.
- Amazon EMR installa e gestisce i plug-in Apache Ranger per conto tuo.

Apache Ranger

Apache Ranger è un framework che consente di abilitare, monitorare e gestire la sicurezza completa dei dati attraverso la piattaforma Hadoop.

Apache Ranger è dotata delle seguenti caratteristiche:

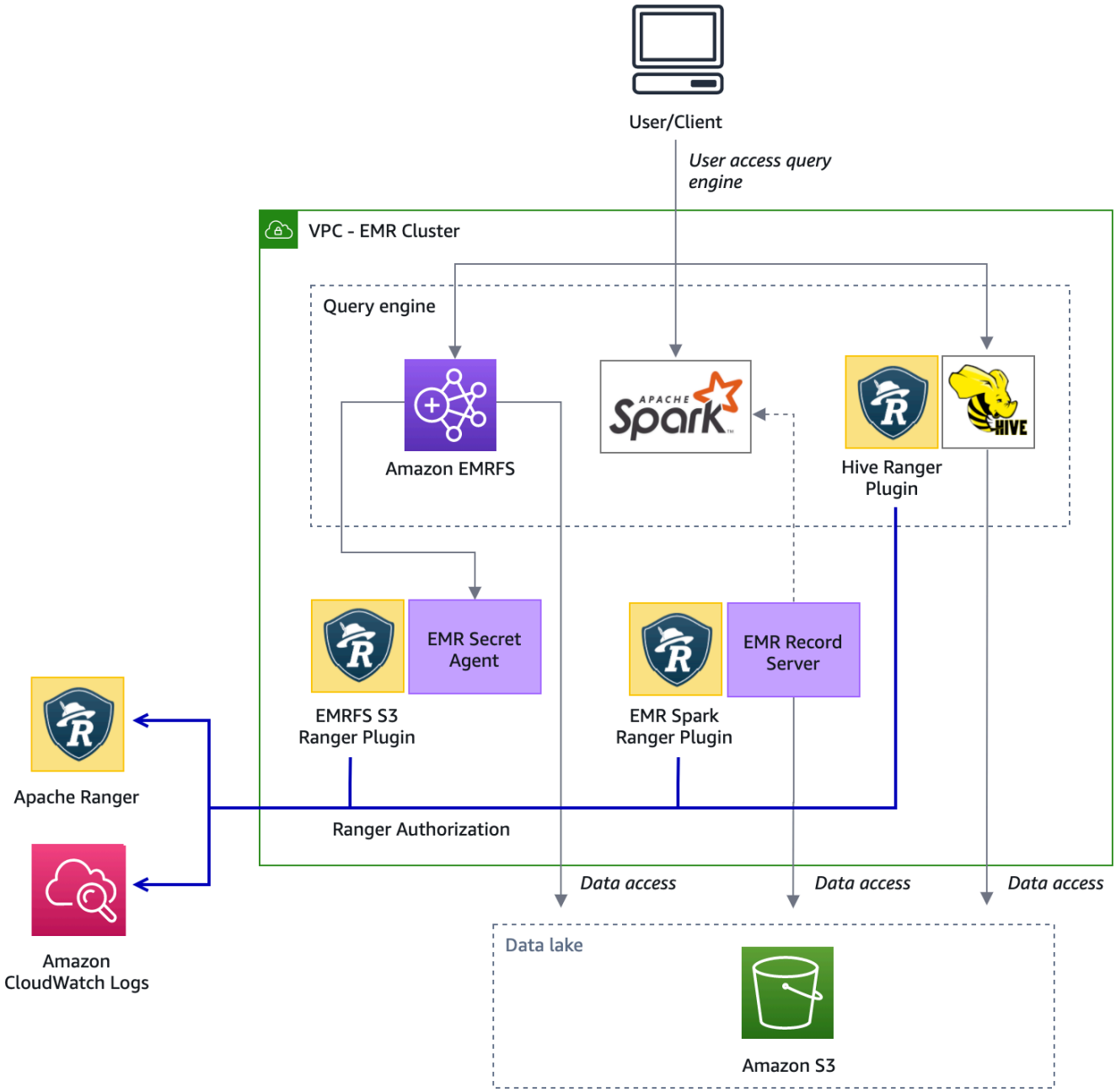
- Amministrazione centralizzata della sicurezza per gestire tutte le attività relative alla sicurezza in un'interfaccia utente centrale o utilizzando le API REST.
- Autorizzazione granulare per eseguire un'operazione specifica o un'operazione con un componente o uno strumento Hadoop, gestito tramite uno strumento di amministrazione centrale.
- Un metodo di autorizzazione standardizzato per tutti i componenti Hadoop.
- Supporto avanzato per vari metodi di autorizzazione.
- Verifica centralizzata dell'accesso degli utenti e delle operazioni amministrative (relative alla sicurezza) all'interno di tutti i componenti di Hadoop.

Apache Ranger utilizza due componenti chiave per l'autorizzazione:

- Server di amministrazione delle policy di Apache Ranger: questo server consente di definire le policy di autorizzazione per le applicazioni Hadoop. Durante l'integrazione con Amazon EMR, puoi definire e applicare policy per Apache Spark e Hive per accedere al Hive Metastore e ai dati di Amazon S3 File System EMR (EMRFS). Per l'integrazione con Amazon EMR, puoi configurare un server di amministrazione delle policy di Apache Ranger esistente.
- Plug-in Apache Ranger: questo plug-in convalida l'accesso di un utente rispetto alle policy di autorizzazione definite nel server di amministrazione delle policy di Apache Ranger. Amazon EMR installa e configura automaticamente il plug-in Apache Ranger per ogni applicazione Hadoop selezionata nella configurazione di Apache Ranger.

Argomenti

Architettura dell'integrazione di Amazon EMR con Apache Ranger



Componenti di Amazon EMR

Amazon EMR consente il controllo granulare degli accessi con Apache Ranger attraverso i seguenti componenti. Fai riferimento al diagramma dell'architettura per una rappresentazione grafica dei componenti Amazon EMR con i plug-in Apache Ranger.

Secret agent (Agente segreto): l'agente segreto archivia in modo sicuro i segreti e distribuisce i segreti ad altri componenti o applicazioni Amazon EMR. I segreti possono includere credenziali utente temporanee, chiavi di crittografia o ticket Kerberos. L'agente segreto viene eseguito su ogni nodo del cluster e intercetta le chiamate al servizio metadati dell'istanza. Per le richieste alle credenziali del ruolo del profilo dell'istanza, l'agente segreto vende le credenziali in base all'utente richiedente e alle risorse richieste dopo aver autorizzato la richiesta con il plug-in EMRFS S3 Ranger. L'agente segreto viene eseguito come utente *`emrsecretagent`* e scrive log nella directory /emr/secretagent/log. Il processo si basa su un set specifico di regole `iptables` per funzionare. È importante assicurarsi che `iptables` non sia disabilitato. Se si personalizza la configurazione `iptables`, le regole della tabella NAT devono essere mantenute e lasciate invariate.

EMR record server (Server di registro EMR): il server di registro riceve le richieste di accesso ai dati da Spark. In seguito autorizza le richieste inoltrando le risorse richieste al plug-in Spark Ranger per Amazon EMR. Il server di registro legge i dati da Amazon S3 e restituisce i dati filtrati a cui l'utente è autorizzato ad accedere in funzione della policy di Ranger. Il server di registro viene eseguito su ogni nodo del cluster come utente emr_record_server e scrive log nella directory /var/log/emr-record-server.

Supporto delle applicazioni e limitazioni

Applicazioni supportate

Attualmente, l'integrazione tra Amazon EMR e Apache Ranger in cui EMR installa i plug-in Ranger supporta le seguenti applicazioni:

- Apache Spark (disponibile con EMR 5.32+ e EMR 6.3+)
- Apache Hive (disponibile con EMR 5.32+ e EMR 6.3+)
- S3 Access tramite EMRFS (disponibile con EMR 5.32+ e EMR 6.3+)

Le seguenti applicazioni possono essere installate in un cluster EMR e possono essere configurate per soddisfare le esigenze di sicurezza:

- Apache Hadoop (disponibile con EMR 5.32+ e EMR 6.3+ compresi YARN e HDFS)
- Apache Livy (disponibile con EMR 5.32+ e EMR 6.3+)
- Apache Zeppelin (disponibile con EMR 5.32+ e EMR 6.3+)
- Apache Hue (disponibile con EMR 5.32+ e EMR 6.3+)

- Ganglia (disponibile con EMR 5.32+ e EMR 6.3+)
- HCatalog (disponibile con EMR 5.32+ e EMR 6.3+)
- Mahout (disponibile con EMR 5.32+ e EMR 6.3+)
- MXNet (disponibile con EMR 5.32+ e EMR 6.3+)
- TensorFlow (disponibile con EMR 5.32+ e EMR 6.3+)
- Tez (disponibile con EMR 5.32+ e EMR 6.3+)
- Trino (disponibile con EMR 6.7+)
- ZooKeeper (disponibile con EMR 5.32+ e EMR 6.3+)

Important

Le applicazioni elencate sopra sono le uniche applicazioni attualmente supportate. Per garantire la sicurezza del cluster, è consentito creare un cluster EMR con solo le applicazioni nell'elenco precedente quando Apache Ranger è abilitato.
Altre applicazioni non sono attualmente supportate. Per garantire la sicurezza del cluster, tentare di installare altre applicazioni causerà il rifiuto del cluster.

Caratteristiche supportate

Le seguenti caratteristiche Amazon EMR possono essere utilizzate con Amazon EMR e Apache Ranger:

- Crittografia dei dati su disco e in transito.
- Autenticazione Kerberos (obbligatoria)
- Gruppi di istanze, parchi istanze e istanze Spot
- Riconfigurazione delle applicazioni in un cluster in esecuzione
- Server-Side Encryption (SSE) EMRFS

Note

Le impostazioni di crittografia di Amazon EMR regolano SSE. Per ulteriori informazioni, consulta Opzioni di crittografia.

Limiti dell'applicazione

Ci sono diverse limitazioni da tenere a mente quando si integra Amazon EMR e Apache Ranger:

- Al momento, non è possibile utilizzare la console per creare una configurazione di sicurezza che specifichi l'opzione di integrazione AWS Ranger nella AWS GovCloud (US) Region. La configurazione della sicurezza può essere eseguita utilizzando la CLI.
- Kerberos deve essere installato nel cluster.
- Le interfacce utente di applicazioni quali YARN Resource Manager, HDFS NameNode e Livy non sono impostate con l'autenticazione per impostazione predefinita.
- Le autorizzazioni predefinite HDFS `umask` sono configurate in modo che gli oggetti creati siano impostati su `world wide readable` per impostazione predefinita.
- Amazon EMR non supporta la modalità ad alta disponibilità (principale multipla) con Apache Ranger.
- Per ulteriori limitazioni, consulta le limitazioni per ogni applicazione.

> **Note**
>
> Le impostazioni di crittografia di Amazon EMR regolano SSE. Per ulteriori informazioni, consulta Opzioni di crittografia.

Limitazioni del plug-in

Ogni plug-in ha limitazioni specifiche. Per le limitazioni del plug-in Apache Hive, consulta Limitazioni del plug-in Apache Hive. Per le limitazioni del plug-in Apache Spark, consulta Limitazioni del plug-in Apache Spark. Per le limitazioni del plug-in EMRFS S3, consulta Limitazioni del plug-in EMRFS S3.

Configurazione di Amazon EMR per Apache Ranger

Prima di installare Apache Ranger, consulta le informazioni contenute in questa sezione per verificare che Amazon EMR sia configurato correttamente.

Argomenti

- Configurazione del server Admin Ranger
- Ruoli IAM per l'integrazione nativa con Apache Ranger

- Creazione di una configurazione di sicurezza EMR
- Archiviazione dei certificati TLS in AWS Secrets Manager
- Avvio di un cluster EMR
- Configurazione dei cluster Amazon EMR abilitati da Zeppelin per Apache Ranger
- Problemi noti

Configurazione del server Admin Ranger

Per l'integrazione di Amazon EMR, i plug-in dell'applicazione Apache Ranger devono comunicare con il server Admin utilizzando TLS/SSL.

Prerequisito: SSL abilitato con il server Admin Ranger

Apache Ranger su Amazon EMR richiede una comunicazione SSL bidirezionale tra i plug-in e il server Admin Ranger. Per garantire che i plug-in comunichino con il server Apache Ranger su SSL, abilita il seguente attributo all'interno di ranger-admin-site.xml sul server Admin Ranger.

```
<property>
    <name>ranger.service.https.attrib.ssl.enabled</name>
    <value>true</value>
</property>
```

Inoltre, sono necessarie le seguenti configurazioni.

```
<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>
```

Certificati TLS

L'integrazione di Apache Ranger con Amazon EMR richiede che il traffico dai nodi di Amazon EMR al server Admin Ranger sia crittografato utilizzando TLS e che i plug-in Ranger si autentichino al server Apache Ranger utilizzando l'autenticazione TLS reciproca bidirezionale. Il servizio Amazon EMR richiede il certificato pubblico del server Admin Ranger (specificato nell'esempio precedente) e il certificato privato.

Certificati del plug-in Apache Ranger

I certificati TLS pubblici del plug-in Apache Ranger devono essere accessibili al server Admin Apache Ranger per convalidare quando i plug-in si connettono. Esistono tre metodi diversi per eseguire questa operazione.

Metodo 1: Configurazione di un truststore nel server Admin Apache Ranger

Compila le seguenti configurazioni in ranger-admin-site.xml per configurare un truststore.

```
<property>
    <name>ranger.truststore.file</name>
    <value><LOCATION TO TRUSTSTORE></value>
</property>

<property>
    <name>ranger.truststore.password</name>
    <value><PASSWORD FOR TRUSTSTORE></value>
</property>
```

Metodo 2: Caricamento del certificato in Java cacerts truststore

Se il tuo server Admin Ranger non specifica un truststore nelle sue opzioni JVM, puoi inserire i certificati pubblici del plug-in nell'archivio cacerts predefinito.

Metodo 3: Creazione di un truststore specificato come parte delle opzioni JVM

In `{RANGER_HOME_DIRECTORY}/ews/ranger-admin-services.sh`, modifica `JAVA_OPTS` per includere "`-Djavax.net.ssl.trustStore=<TRUSTSTORE_LOCATION>`" e "`-Djavax.net.ssl.trustStorePassword=<TRUSTSTORE_PASSWORD>`". Ad esempio, aggiungi la seguente riga dopo il JAVA_OPTS esistente.

```
JAVA_OPTS=" ${JAVA_OPTS} -Djavax.net.ssl.trustStore=${RANGER_HOME}/truststore/
truststore.jck -Djavax.net.ssl.trustStorePassword=changeit"
```

> **Note**
>
> Questa specifica può esporre la password truststore se un utente è in grado di accedere al server Admin Apache Ranger e vedere i processi in esecuzione, ad esempio quando si utilizza il comando `ps`.

Utilizzo di certificati autofirmati

I certificati autofirmati non sono consigliati come certificati. I certificati autofirmati potrebbero non essere revocati o non essere conformi ai requisiti di sicurezza interni.

Installazione della definizione del servizio

Una definizione di servizio viene utilizzata dal server Admin Ranger per descrivere gli attributi delle policy per un'applicazione. Le policy vengono quindi archiviate in un repository di policy per il download dei client.

Per poter configurare le definizioni dei servizi, le chiamate REST devono essere effettuate al server Admin Ranger. Consulta Apache Ranger PublicAPIsv2 per le API richieste nella sezione seguente.

Installazione della definizione del servizio di Apache Spark

Per installare la definizione del servizio di Apache Spark, consulta Plug-in per Apache Spark.

Installazione della definizione del servizio EMRFS

Per installare la definizione del servizio S3 per Amazon EMR, consulta Plug-in EMRFS S3.

Utilizzo della definizione del servizio Hive

Apache Hive può utilizzare la definizione del servizio Ranger esistente fornita con Apache Ranger 2.0 e versioni successive. Per ulteriori informazioni, consulta Plug-in di Apache Hive.

Regole del traffico di rete

Quando Apache Ranger è integrato con il cluster EMR, il cluster deve comunicare con server aggiuntivi e AWS.

Tutti i nodi Amazon EMR, inclusi i nodi principali e attività, devono essere in grado di comunicare con i server Admin Apache Ranger per scaricare le policy. Se il tuo Apache Ranger Admin è in esecuzione su Amazon EC2, devi aggiornare il gruppo di sicurezza per poter prelevare il traffico dal cluster EMR.

Oltre a comunicare con il server Admin Ranger, tutti i nodi devono essere in grado di comunicare con i seguenti servizi AWS:

- Amazon S3
- AWS KMS (se si utilizza EMRFS SSE-KMS)
- Amazon CloudWatch
- AWS STS

Se prevedi di eseguire il cluster EMR all'interno di una sottorete privata, configura il VPC per comunicare con questi servizi utilizzando AWS PrivateLink ed endpoint VPC nella Guida per l'utente di Amazon VPC oppure utilizzando un'istanza NAT (Network Address Translation) nella Guida per l'utente di Amazon VPC.

Ruoli IAM per l'integrazione nativa con Apache Ranger

L'integrazione tra Amazon EMR e Apache Ranger si basa su tre ruoli chiave che è necessario creare prima di avviare il cluster:

- Un profilo dell'istanza Amazon EC2 personalizzato per Amazon EMR
- Un ruolo IAM per Apache Ranger Engines

- Un ruolo IAM per altri servizi AWS

Questa sezione fornisce una panoramica di questi ruoli e delle policy che è necessario includere per ogni ruolo IAM. Per informazioni sulla creazione di questi ruoli, consulta Configurazione del server Admin Ranger.

Profilo dell'istanza EC2

Amazon EMR utilizza un ruolo di servizio IAM per eseguire operazioni per tuo conto per il provisioning e la gestione dei cluster. Il ruolo di servizio per le istanze EC2 del cluster, detto anche profilo dell'istanza EC2 per Amazon EMR, è un tipo speciale di ruolo di servizio assegnato a ogni istanza EC2 in un cluster all'avvio.

Per definire le autorizzazioni per l'interazione del cluster EMR con i dati di Amazon S3 e con il metastore Hive protetto da Apache Ranger e altri servizi AWS, definisci un profilo dell'istanza EC2 personalizzato da utilizzare al posto di `EMR_EC2_DefaultRole` quando avvii il cluster.

Per ulteriori informazioni, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2) e Personalizzazione dei ruoli IAM.

Dovrai aggiungere le seguenti istruzioni al profilo dell'istanza EC2 predefinito per Amazon EMR per poter taggare le sessioni e accedere all'AWS Secrets Manager che memorizza i certificati 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**
>
> Per le autorizzazioni di Secrets Manager, non dimenticare il carattere jolly ("*") alla fine del nome segreto, altrimenti le richieste non avranno esito positivo. Il carattere jolly vale per le versioni segrete.

> **Note**
>
> Limita l'ambito della policy AWS Secrets Manager per solo i certificati necessari per il provisioning.

Ruolo IAM per Apache Ranger

Questo ruolo fornisce ai motori di esecuzione attendibili, ad esempio Apache Hive e Amazon EMR Record Serve, le credenziali per accedere ai dati Amazon S3. Utilizza solo questo ruolo per accedere ai dati di Amazon S3, incluse le chiavi KMS, se utilizzi S3 SSE-KMS.

Questo ruolo deve essere creato con la policy minima riportata nell'esempio seguente.

```
{
  "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

L'asterisco "*" alla fine della risorsa CloudWatch Log deve essere incluso per fornire l'autorizzazione alla scrittura nei flussi di registro.

Note

Se utilizzi la visualizzazione di coerenza EMRFS o la crittografia S3-SSE, aggiungi le autorizzazioni alle tabelle DynamoDB e alle chiavi del servizio di gestione delle chiavi in modo che i motori di esecuzione possano interagire con tali motori.

Il ruolo IAM per Apache Ranger viene assunto dal ruolo del profilo dell'istanza EC2. Utilizza l'esempio seguente per creare una policy di affidabilità che consenta di assumere il ruolo IAM per Apache Ranger dal ruolo del profilo dell'istanza EC2.

```
   {
     "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"]
   }
```

Ruolo IAM per altri servizi AWS

Questo ruolo fornisce agli utenti che non sono motori di esecuzione attendibili le credenziali per interagire con servizi AWS, se necessario. Non utilizzare questo ruolo IAM per consentire l'accesso ai dati di Amazon S3, a meno che non si tratti di dati che devono essere accessibili a tutti gli utenti.

Questo ruolo verrà assunto dal ruolo del profilo dell'istanza EC2. Utilizza l'esempio seguente per creare una policy di affidabilità che consenta di assumere il ruolo IAM per Apache Ranger dal ruolo del profilo dell'istanza EC2.

```
   {
     "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"]
   }
```

Convalida delle autorizzazioni

Consulta Risoluzione dei problemi di Apache Ranger per istruzioni sulla convalida delle autorizzazioni.

Creazione di una configurazione di sicurezza EMR

Creazione di una configurazione di sicurezza di Amazon EMR per Apache Ranger

Prima di lanciare un cluster Amazon EMR integrato con Apache Ranger, crea una configurazione di sicurezza.

Console

Creazione di una configurazione di sicurezza che specifichi l'opzione di integrazione AWS Ranger

1. Nella console di Amazon EMR, seleziona Security configurations (Configurazioni di sicurezza) e in seguito Create (Crea).
2. Digitare un nome in Name (Nome) per la configurazione di sicurezza. Questo nome è utilizzato per specificare la configurazione di sicurezza al momento della creazione di un cluster.
3. In AWS Ranger Integration (Integrazione Ranger), seleziona Enable fine-grained access control managed by Apache Ranger (Abilita controllo granulare degli accessi gestito da Apache Ranger).
4. Seleziona lo IAM role for Apache Ranger (Ruolo IAM per Apache Ranger) da applicare. Per ulteriori informazioni, consulta Ruoli IAM per l'integrazione nativa con Apache Ranger.
5. Seleziona il ruolo IAM per altri servizi AWS da applicare.
6. Configura i plug-in per connetterti al server Admin Ranger immettendo l'ARN di Secrets Manager per il server Admin e l'indirizzo.
7. Seleziona le applicazioni per configurare i plug-in Ranger. Compila l'ARN di Secrets Manager che contiene il certificato TLS privato per il plug-in.

   Se Apache Spark o Apache Hive non vengono configurati e vengono selezionati come applicazione per il cluster, la richiesta ha esito negativo.
8. Configurare altre opzioni per la configurazione di sicurezza come appropriato e scegliere Create (Crea). È necessario abilitare l'autenticazione Kerberos utilizzando il KDC dedicato al cluster o esterno.

> **Note**
>
> Al momento, non è possibile utilizzare la console per creare una configurazione di sicurezza che specifichi l'opzione di integrazione AWS Ranger nella AWS GovCloud (US) Region. La configurazione della sicurezza può essere eseguita utilizzando la CLI.

CLI

Creazione di una configurazione di sicurezza per l'integrazione di Apache Ranger

1. Sostituisci *<ACCOUNT ID>* con il tuo ID account AWS.
2. Sostituisci *<REGION>* con la Regione in cui si trova la risorsa.
3. Specifica un valore per `TicketLifetimeInHours` per determinare il periodo di validità di un ticket Kerberos emesso dal KDC.
4. Specifica l'indirizzo del server Admin Ranger per `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>_"
               }
            }
         }
      }
   }
}
```

I PolicyRespositoryNames sono i nomi di servizio specificati nel server Admin Apache Ranger.

Crea una configurazione di sicurezza Amazon EMR con il seguente comando. Sostituisci security-configuration con un nome a tua scelta. Seleziona questa configurazione per nome quando crei il cluster.

```
aws emr create-security-configuration \
--security-configuration file://./security-configuration.json \
--name security-configuration
```

Configurazione di caratteristiche di sicurezza aggiuntive

Per integrare Amazon EMR con Apache Ranger in modo sicuro, configura le seguenti caratteristiche di sicurezza di EMR:

- Abilita l'autenticazione Kerberos utilizzando il KDC dedicato al cluster o esterno. Per istruzioni, consultare Utilizzo di Kerberos per l'autenticazione con Amazon EMR.
- (Facoltativo) Abilita la crittografia in transito o inattiva. Per ulteriori informazioni, consulta Opzioni di crittografia.

Per ulteriori informazioni, consulta Sicurezza in Amazon EMR.

Archiviazione dei certificati TLS in AWS Secrets Manager

I plug-in Ranger installati su un cluster Amazon EMR e il server Admin Ranger devono comunicare tramite TLS per garantire che i dati delle policy e le altre informazioni inviate non possano essere letti se vengono intercettati. EMR impone inoltre che i plug-in eseguano l'autenticazione al server Admin Ranger fornendo il proprio certificato TLS ed eseguendo l'autenticazione TLS bidirezionale. Questa configurazione richiedeva la creazione di quattro certificati: due coppie di certificati TLS privati e pubblici. Per istruzioni sull'installazione del certificato nel server Admin Ranger, consulta Configurazione del server Admin Ranger. Per completare la configurazione, i plug-in Ranger installati nel cluster EMR necessitano di due certificati: il certificato TLS pubblico del server di amministrazione e il certificato privato che il plug-in utilizzerà per autenticarsi sul server Admin Ranger. Per fornire questi certificati TLS, è necessario che si trovino in AWS Secrets Manager e siano forniti in una configurazione di sicurezza EMR.

> **Note**
>
> Si consiglia, per quanto non sia necessario, di creare una coppia di certificati per ciascuna delle applicazioni per limitare l'impatto in caso di compromissione di uno dei certificati del plug-in.

> **Note**
>
> È necessario monitorare e ruotare i certificati prima della data di scadenza.

Formato del certificato

L'importazione dei certificati nella cartella AWS Secrets Manager è uguale indipendentemente dal fatto che si tratti del certificato del plug-in privato o del certificato del server Admin Ranger pubblico. Prima di importare i certificati TLS, questi devono essere in formato PEM 509x.

Un esempio di certificato pubblico è nel formato:

```
-----BEGIN CERTIFICATE-----
...Certificate Body...
-----END CERTIFICATE-----
```

Un esempio di certificato privato è nel formato:

```
-----BEGIN PRIVATE KEY-----
...Private Certificate Body...
-----END PRIVATE KEY-----
-----BEGIN CERTIFICATE-----
...Trust Certificate Body...
-----END CERTIFICATE-----
```

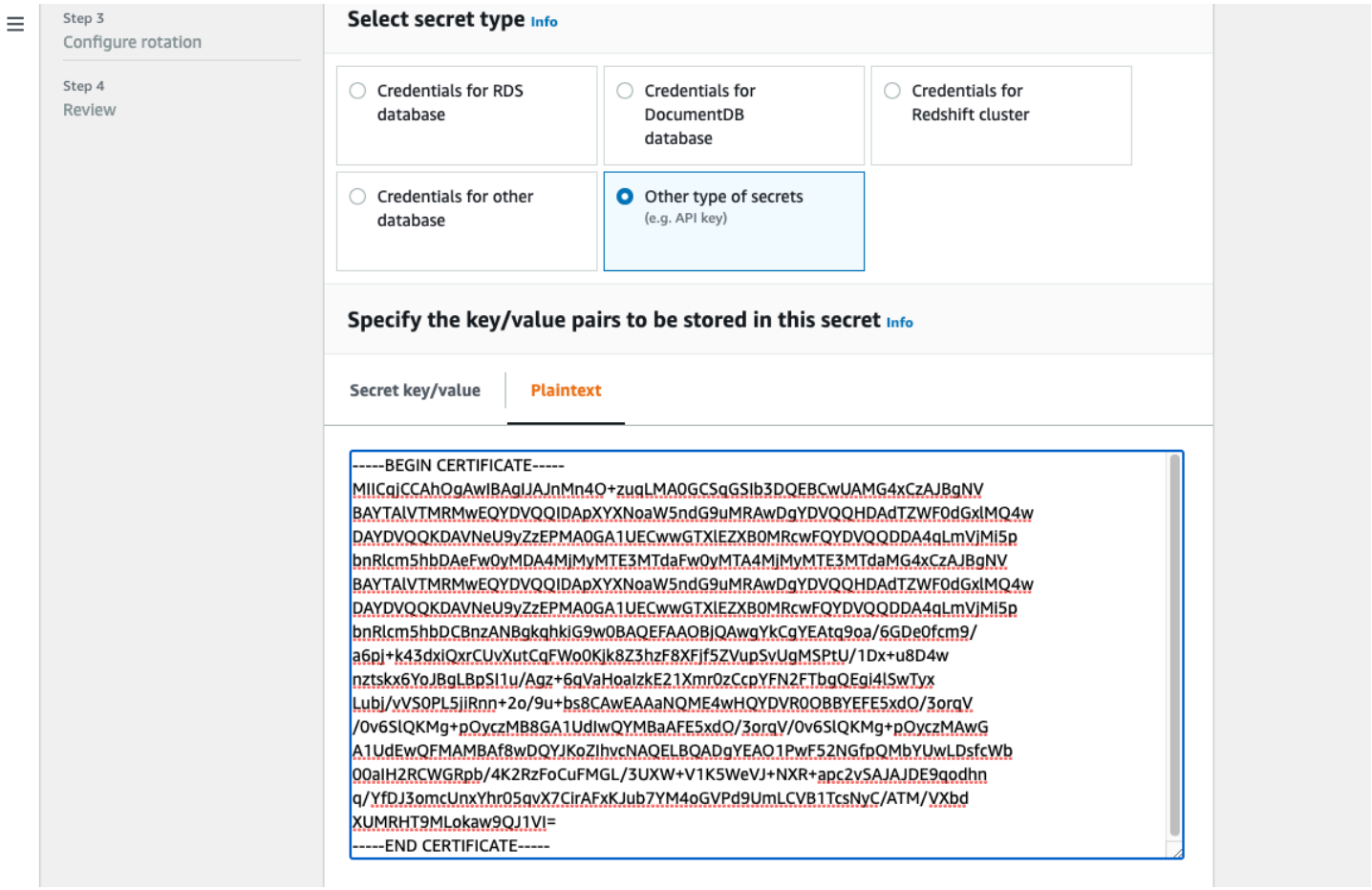
Anche il certificato privato deve contenere un certificato di affidabilità.

Puoi verificare che i certificati siano nel formato corretto eseguendo il seguente comando:

```
openssl x509 -in <PEM FILE> -text
```

Importazione di un certificato in AWS Secrets Manager

Quando crei il tuo segreto nel Secrets Manager, seleziona Other type of secrets (Altro tipo di segreti) in secret type (Tipo di segreto) e incolla il certificato codificato PEM nel campo Plaintext (Testo normale).



Avvio di un cluster EMR

Prima di avviare un cluster Amazon EMR con Apache Ranger, assicurati che ogni componente soddisfi i seguenti requisiti minimi di versione:

- Amazon EMR versione 5.32.0 o successive, o 6.3.0 o successive. Consigliamo di utilizzare la versione del rilascio di Amazon EMR più recente.
- Server Admin Apache Ranger 2.x.

Completa questa procedura:

- Installa Apache Ranger se non lo hai già fatto. Per ulteriori informazioni, consulta Installazione di Apache Ranger 0.5.0.
- Assicurati che ci sia connettività di rete tra il tuo cluster Amazon EMR e il server Admin Apache Ranger. Per informazioni, consultare Configurazione del server Admin Ranger.

- Crea i ruoli IAM necessari. Per informazioni, consultare Ruoli IAM per l'integrazione nativa con Apache Ranger.
- Crea una configurazione di sicurezza EMR per l'installazione di Apache Ranger. Per ulteriori informazioni, consulta Creazione di una configurazione di sicurezza EMR.

Configurazione dei cluster Amazon EMR abilitati da Zeppelin per Apache Ranger

L'argomento illustra come configurare Apache Zeppelin per un cluster Amazon EMR abilitato per Apache Ranger in modo da poter utilizzare Zeppelin come notebook per l'esplorazione interattiva dei dati. Zeppelin è incluso nel rilascio di Amazon EMR versione 5.0.0 e successive. Le versioni precedenti includono Zeppelin come applicazione sandbox. Per ulteriori informazioni, consulta Versioni del rilascio 4.x di Amazon EMR nella Guida ai rilasci di Amazon EMR.

Per impostazione predefinita, Zeppelin è configurato con un log-in e una password di default che non sono sicuri in un ambiente multi-tenant.

Per configurare Zeppelin, completa la procedura riportata di seguito.

1. Modifica del meccanismo di autenticazione.

   Modifica del file `shiro.ini` per implementare il meccanismo di autenticazione preferito. Zeppelin supporta Active Directory, LDAP, PAM e Knox SSO. Consulta Autenticazione Apache Shiro per Apache Zeppelin per ulteriori informazioni.
2. Configurazione di Zeppelin per rappresentare l'utente finale

   Consentire a Zeppelin di rappresentare l'utente finale fa sì che i processi inviati da Zeppelin vengano eseguiti come utente finale. Aggiungi la seguente configurazione a `core-site.xml`:

```
[
  {
    "Classification": "core-site",
    "Properties": {
      "hadoop.proxyuser.zeppelin.hosts": "*",
      "hadoop.proxyuser.zeppelin.groups": "*"
    },
    "Configurations": [
    ]
  }
]
```

Quindi, aggiungi la seguente configurazione a `hadoop-kms-site.xml` in `/etc/hadoop/conf`:

```
[
  {
    "Classification": "hadoop-kms-site",
    "Properties": {
      "hadoop.kms.proxyuser.zeppelin.hosts": "*",
      "hadoop.kms.proxyuser.zeppelin.groups": "*"
    },
    "Configurations": [
    ]
  }
]
```

Puoi anche aggiungere queste configurazioni al tuo cluster Amazon EMR tramite la console seguendo la procedura descritta in Riconfigurazione di un gruppo di istanze nella console.

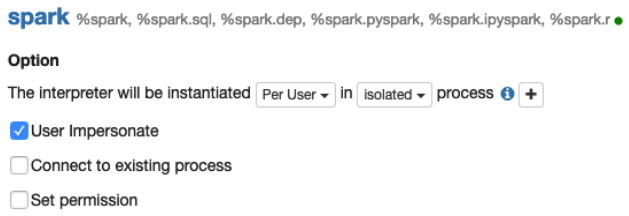
3. Consenti a Zeppelin di eseguire il comando sudo come utente finale

   Crea un file `/etc/sudoers.d/90-zeppelin-user` che contenga quanto segue:

   ```
   zeppelin ALL=(ALL) NOPASSWD:ALL
   ```

4. Modifica le impostazioni degli interpreti per eseguire i processi degli utenti nei propri processi.

   Per tutti gli interpreti, configurali per creare un'istanza degli interpreti "per user" (per utente) nei processi "isolated" (isolati).

5. Modifica **`zeppelin-env.sh`**

   Aggiungi quanto segue a `zeppelin-env.sh` in modo che Zeppelin inizi ad avviare interpreti come utente finale:

   ```
   ZEPPELIN_IMPERSONATE_USER=`echo ${ZEPPELIN_IMPERSONATE_USER} | cut -d @ -f1`
   ```

```
export ZEPPELIN_IMPERSONATE_CMD='sudo -H -u ${ZEPPELIN_IMPERSONATE_USER} bash -c'
```

Aggiungi quanto segue a `zeppelin-env.sh` per modificare le autorizzazioni predefinite del notebook in sola lettura solo per l'autore:

```
export ZEPPELIN_NOTEBOOK_PUBLIC="false"
```

Infine, aggiungi quanto segue a `zeppelin-env.sh` per includere il percorso della classe EMR RecordServer dopo la prima istruzione `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. Riavvia Zeppelin.

   Per riavviare Zeppelin, esegui il comando seguente:

   ```
   sudo systemctl restart zeppelin
   ```

Problemi noti

Problemi noti

C'è un problema noto all'interno del rilascio di Amazon EMR 5.32 in cui le autorizzazioni per `hive-site.xml` sono state modificate in modo che solo gli utenti privilegiati possano leggerlo, in quanto potrebbero esserci credenziali memorizzate al suo interno. Ciò potrebbe impedire a Hue di leggere `hive-site.xml` e fare in modo che le pagine Web vengano ricaricate continuamente. Se si verificano problemi, aggiungi la seguente configurazione per risolvere il problema:

```
[
  {
    "Classification": "hue-ini",
    "Properties": {},
    "Configurations": [
      {
        "Classification": "desktop",
        "Properties": {
```

```
          "server_group":"hive_site_reader"
         },
        "Configurations":[
        ]
      }
    ]
  }
]
```

Sussiste un problema noto per cui il plug-in EMRFS S3 per Apache Ranger attualmente non supporta la caratteristica Security Zone di Apache Ranger. Le restrizioni per il controllo degli accessi definite utilizzando la funzione Security Zone non vengono applicate ai cluster Amazon EMR.

Applicazione UIs

Per impostazione predefinita, l'interfaccia utente di un'applicazione non esegue l'autenticazione. Questo vale per l'interfaccia utente ResourceManager, l'interfaccia utente NodeManager e l'interfaccia utente Livy, tra le altre. Inoltre, qualsiasi utente che abbia la possibilità di accedere alle interfacce utente è in grado di visualizzare le informazioni sui processi di tutti gli altri utenti.

Se questo comportamento non è desiderato, è necessario assicurarsi che un gruppo di sicurezza venga utilizzato per limitare l'accesso alle interfacce utente delle applicazioni da parte degli utenti.

Autorizzazioni HDFS predefinite

Per impostazione predefinita, agli oggetti creati dagli utenti in HDFS vengono concesse autorizzazioni leggibili a livello mondiale. Ciò può tradursi nella leggibilità dei dati da parte degli utenti che non dovrebbero avervi accesso. Per modificare questo comportamento in modo che le autorizzazioni predefinite dei file siano impostate per la lettura e la scrittura solo dall'autore del processo, esegui questa procedura.

Quando si crea il cluster EMR, fornisci la seguente configurazione:

```
[
  {
    "Classification": "hdfs-site",
    "Properties": {
      "dfs.namenode.acls.enabled": "true",
      "fs.permissions.umask-mode": "077",
      "dfs.permissions.superusergroup": "hdfsadmingroup"
    }
```

```
  }
]
```

Inoltre, esegui la seguente operazione di bootstrap:

```
--bootstrap-actions Name='HDFS UMask Setup',Path=s3://elasticmapreduce/hdfs/umask/
umask-main.sh
```

Plug-in di Apache Ranger

I plug-in Apache Ranger convalidano l'accesso di un utente rispetto alle policy di autorizzazione definite nel server amministratore delle policy di Apache Ranger.

Argomenti

- Plug-in di Apache Hive
- Plug-in per Apache Spark
- Plug-in EMRFS S3
- Plugin Trino

Plug-in di Apache Hive

Apache Hive è un motore di esecuzione popolare all'interno dell'ecosistema Hadoop. Amazon EMR fornisce un plug-in Apache Ranger per essere in grado di fornire controlli di accesso granulare per Hive. Il plug-in è compatibile con il server open source Apache Ranger Admin versione 2.0 e successive.

Argomenti

- Funzionalità supportate
- Installazione della configurazione del servizio
- Considerazioni
- Limitazioni

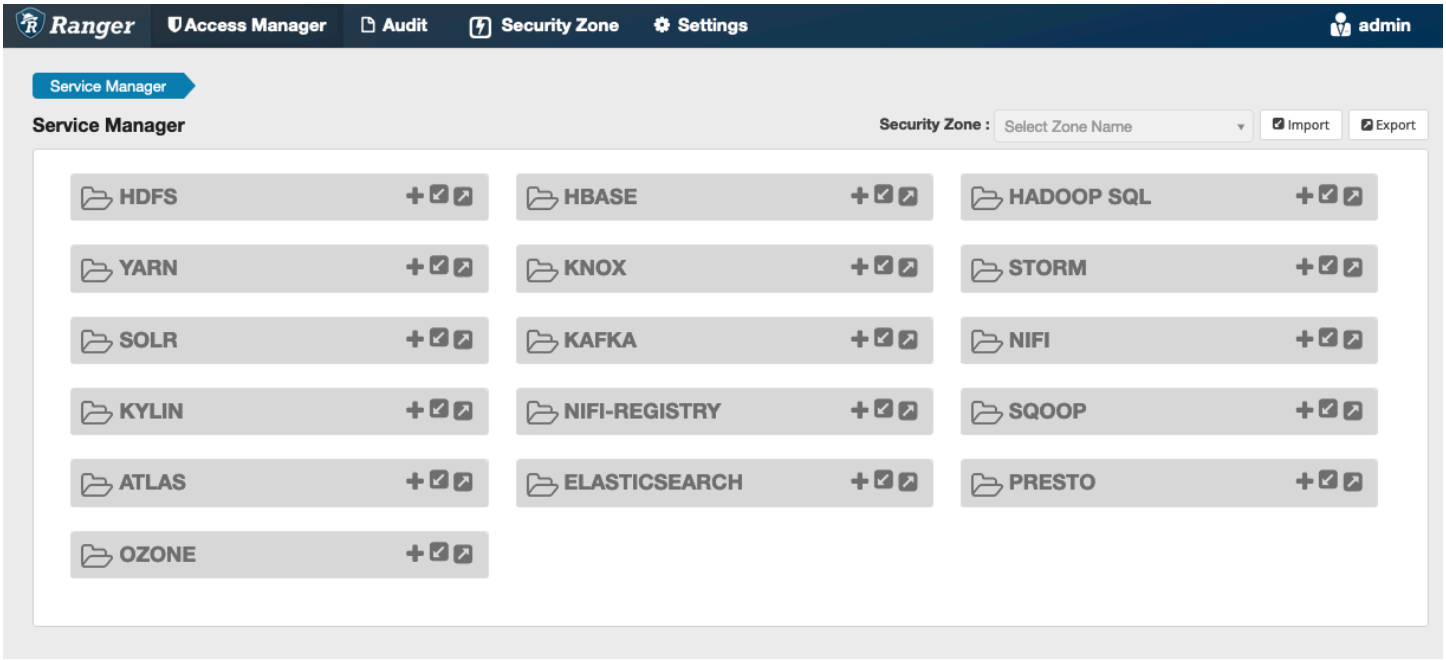
Funzionalità supportate

Il plug-in Apache Ranger per Hive su EMR supporta tutte le funzionalità del plug-in open source, che include database, tabella, controlli di accesso a livello di colonna, filtraggio riga e mascheramento dei

dati. Per una tabella dei comandi Hive e delle autorizzazioni Ranger associate, consulta Comandi Hive per la mappatura delle autorizzazioni Ranger.

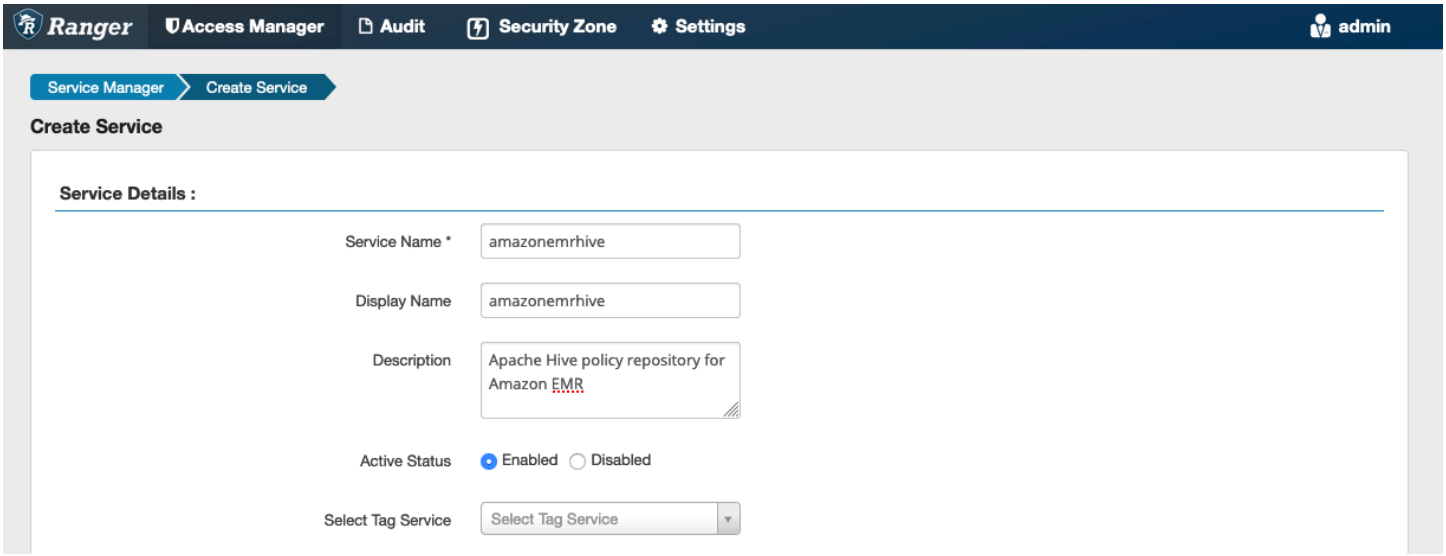
Installazione della configurazione del servizio

Il plugin Apache Hive è compatibile con la definizione del servizio Hive esistente all'interno di Apache Hive Hadoop SQL.



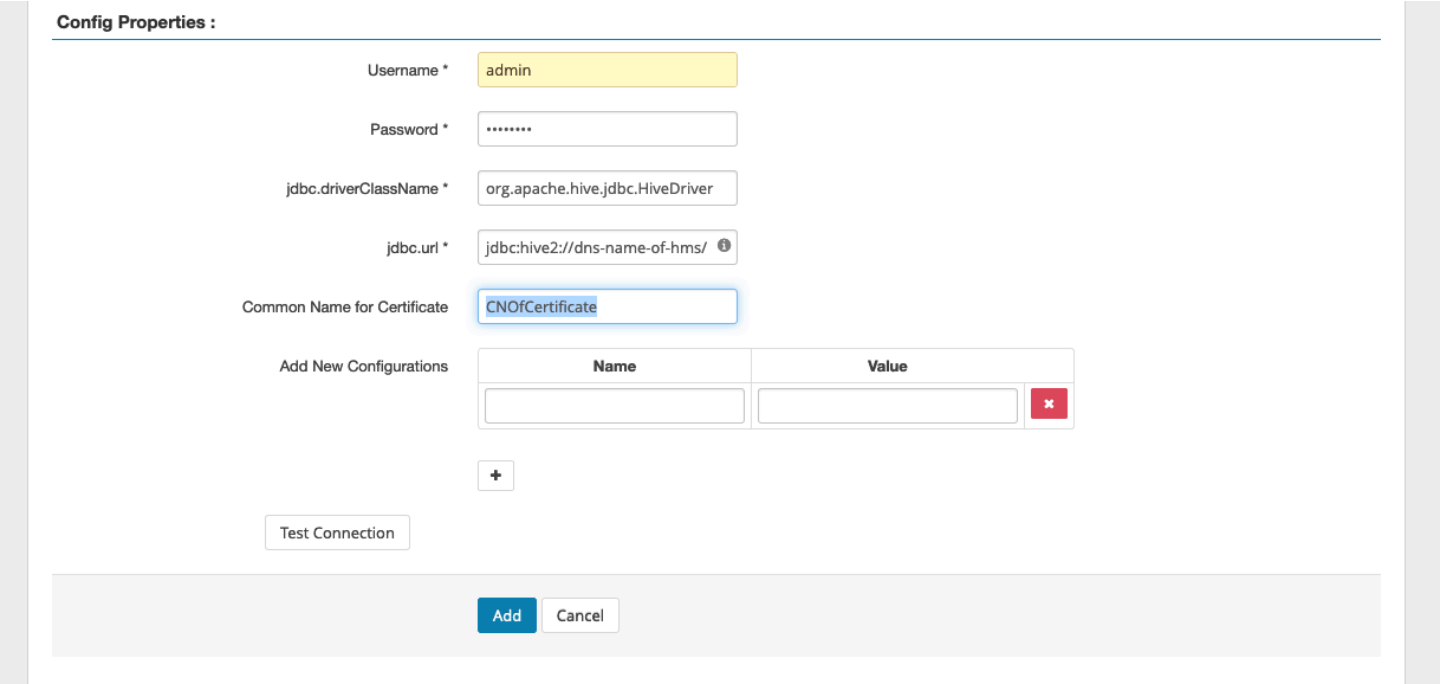
Se non disponi di un'istanza del servizio in Hadoop SQL, come mostrato sopra, puoi crearne una. Fai clic sul pulsante + accanto ad Hadoop SQL.

1. Nome del servizio (se visualizzato): immetti il nome del servizio. Il valore suggerito è **`amazonemrhive`**. Prendi nota di questo nome del servizio: è necessario quando si crea una configurazione di sicurezza EMR.
2. Nome visualizzato: immetti il nome da visualizzare per il servizio. Il valore suggerito è **`amazonemrhive`**.

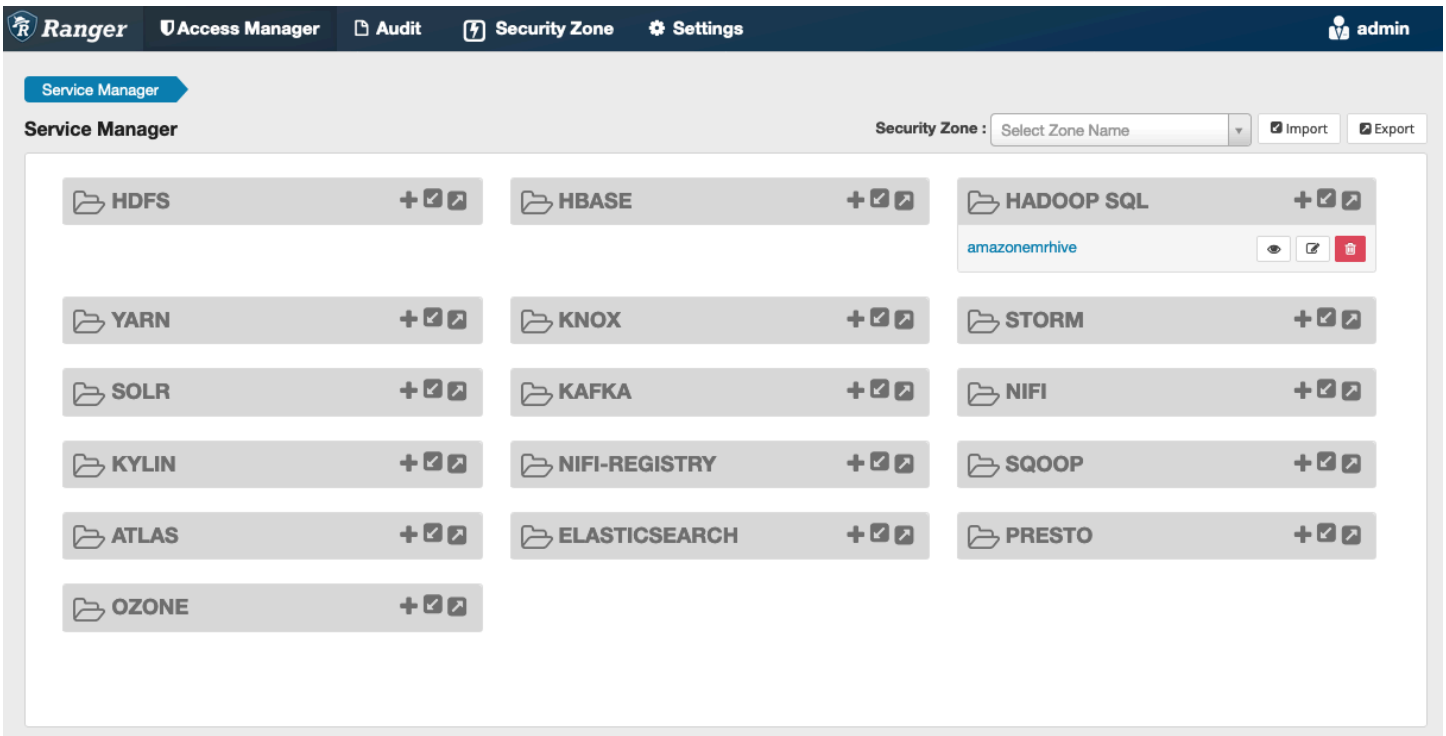


Le proprietà Config di Apache Hive vengono utilizzate per stabilire una connessione al server Admin Apache Ranger con un HIveServer2 per implementare il completamento automatico durante la creazione di policy. Le proprietà riportate di seguito non devono essere accurate se non si dispone di un processo HiveServer2 persistente e possono essere compilate con qualsiasi informazione.

- Nome utente: immetti un nome utente per la connessione JDBC a un'istanza di un'istanza di HiveServer2.
- Password: inserisci la password per il nome utente sopra.
- jdbc.driver.ClassName: immetti il nome della classe JDBC per la connettività Apache Hive. Puoi utilizzare il valore predefinito.
- jdbc.url: immetti la stringa di connessione JDBC da utilizzare per la connessione ad HiveServer2.
- Nome comune per certificato: il campo CN all'interno del certificato utilizzato per connettersi al server Admin da un plug-in client. Questo valore deve corrispondere al campo CN nel certificato TLS creato per il plug-in.



Il pulsante Test Connection (Connessione di prova) verifica se i valori sopra riportati possono essere utilizzati per connettersi correttamente all'istanza HiveServer2. Una volta che il servizio è stato creato correttamente, il Service Manager dovrebbe avere il seguente aspetto:



Considerazioni

Server dei metadati Hive

Il server dei metadati Hive è accessibile solo dai motori attendibili, in particolare Hive e `emr_record_server`, come misura di protezione da accessi non autorizzati. Il server dei metadati Hive è accessibile anche da tutti i nodi del cluster. La porta 9083 richiesta consente a tutti i nodi di accedere al nodo principale.

Autenticazione

Per impostazione predefinita, Apache Hive è configurato per l'autenticazione utilizzando Kerberos, come specificato nella configurazione di sicurezza di EMR. HIveServer2 può essere configurato per autenticare gli utenti utilizzando anche LDAP. Consulta Implementazione dell'autenticazione LDAP per Hive su un cluster Amazon EMR multi-tenant per maggiori informazioni.

Limitazioni

Di seguito sono riportate le attuali limitazioni per il plug-in Apache Hive su Amazon EMR 5.x:

- I ruoli Hive non sono attualmente supportati. Le istruzioni Grant (Concedi) e Revoke (Revoca) non sono supportate.
- La CLI di Hive non è supportata. JDBC/Beeline è l'unico modo autorizzato per connettere Hive.
- La configurazione `hive.server2.builtin.udf.blacklist` deve essere popolata con FDU (funzioni definite dall'utente) ritenute non sicure.

Plug-in per Apache Spark

Amazon EMR ha integrato EMR RecordServer per fornire un controllo di accesso granulare per SparkSQL. RecordServer di EMR è un processo privilegiato in esecuzione su tutti i nodi su un cluster abilitato per Apache Ranger. Quando un driver o un executor Spark esegue un'istruzione SparkSQL, tutte le richieste di metadati e dati passano attraverso il RecordServer. Per ulteriori informazioni su EMR RecordServer, consulta la pagina Componenti di Amazon EMR.

Argomenti

- Funzionalità supportate
- Ridistribuire la definizione del servizio per utilizzare le istruzioni INSERT, ALTER o DDL
- Installazione della definizione del servizio

- Creazione di policy SparkSQL
- Considerazioni
- Limitazioni

Funzionalità supportate

Istruzione SQL/operazione Ranger	STATUS	Rilascio di EMR supportato
SELECT	Supportato	A partire da 5.32
SHOW DATABASES	Supportato	A partire da 5.32
SHOW COLUMNS	Supportato	A partire da 5.32
SHOW TABLES	Supportato	A partire da 5.32
SHOW TABLE PROPERTIES (MOSTRA PROPRIETÀ TABELLA)	Supportato	A partire da 5.32
DESCRIBE TABLE	Supportato	A partire da 5.32
INSERT OVERWRITE	Supportato	A partire da 5.34 e 6.4
INSERT INTO	Supportato	A partire da 5.34 e 6.4
ALTER TABLE	Supportato	A partire da 6.4
CREATE TABLE	Supportato	A partire da 5.35 e 6.7
CREATE DATABASE	Supportato	A partire da 5.35 e 6.7

Istruzione SQL/operazione Ranger	STATUS	Rilascio di EMR supportato
DROP TABLE	Supportato	A partire da 5.35 e 6.7
DROP DATABASE	Supportato	A partire da 5.35 e 6.7
DROP VIEW	Supportato	A partire da 5.35 e 6.7
CREATE VIEW	Non supportato	

Quando si utilizza SparkSQL sono supportate le seguenti caratteristiche:

- Controllo granulare degli accessi sulle tabelle all'interno del metastore Hive; le policy possono essere create a livello di database, tabella e colonna.
- Le policy di Apache Ranger possono includere policy di concessione e di negazione a utenti e gruppi.
- Gli eventi di verifica vengono inviati a CloudWatch Logs.

Ridistribuire la definizione del servizio per utilizzare le istruzioni INSERT, ALTER o DDL

> **Note**
>
> A partire da Amazon EMR 6.4, puoi utilizzare Spark SQL con le istruzioni: INSERT INTO, INSERT OVERWRITE o ALTER TABLE. A partire da Amazon EMR 6.7, puoi utilizzare Spark SQL per creare o rimuovere database e tabelle. Se si dispone di un'installazione esistente sul server Apache Ranger con le definizioni del servizio Apache Spark implementate, utilizzare il codice seguente per ridistribuire le definizioni del servizio.

```
# 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>'
```

Installazione della definizione del servizio

L'installazione della definizione del servizio Apache Spark di EMR richiede l'installazione del server Admin Ranger. Per informazioni, consultare Configurazione del server Admin Ranger.

Segui queste fasi per installare la definizione del servizio Apache Spark:

Fase 1: SSH nel server Admin Apache Ranger.

Ad esempio:

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Fase 2: Download della definizione del servizio e del plug-in del server Admin Apache Ranger

In una directory temporanea, scarica la definizione del servizio. Questa definizione del servizio è supportata dalle versioni Ranger 2.x.

```
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
```

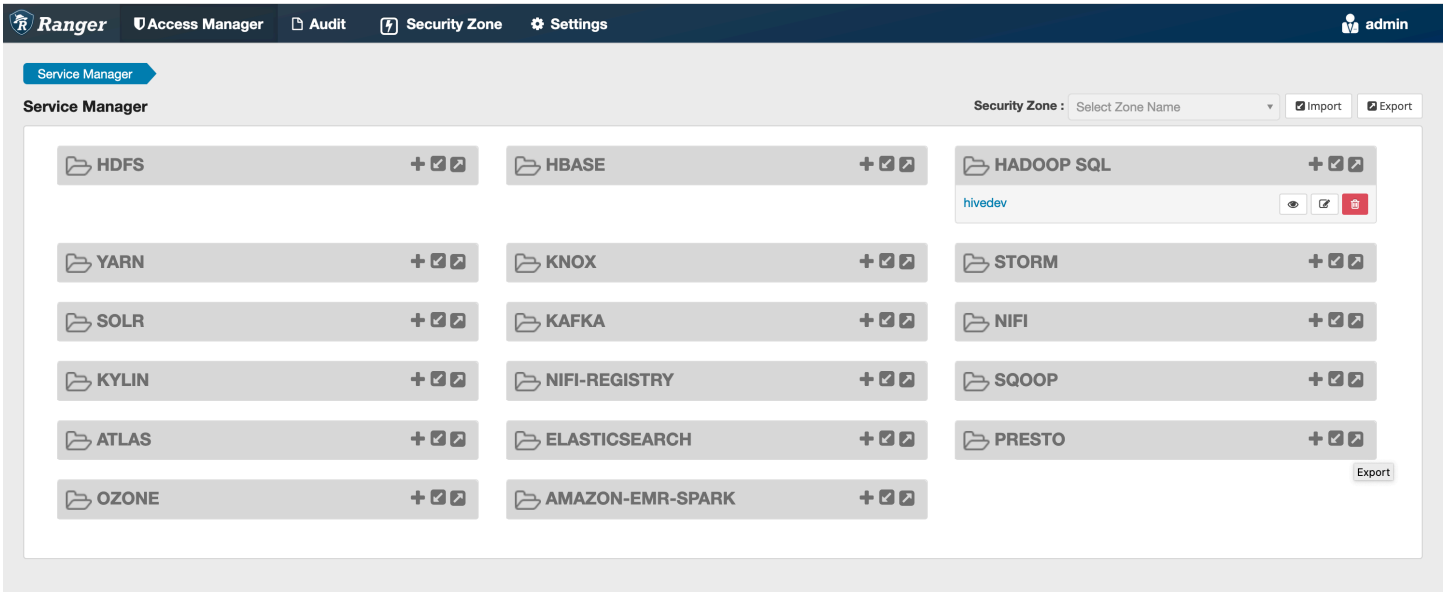
Fase 3: Installazione del plug-in Apache Spark per Amazon EMR

```
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
```

Fase 4: Registrazione della definizione del servizio Apache Spark per 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'
```

Se questo comando viene eseguito correttamente, viene visualizzato un nuovo servizio nell'interfaccia utente del server Admin Ranger denominato "AMAZON-EMR-SPARK", come mostrato nell'immagine seguente (Ranger versione 2.0).

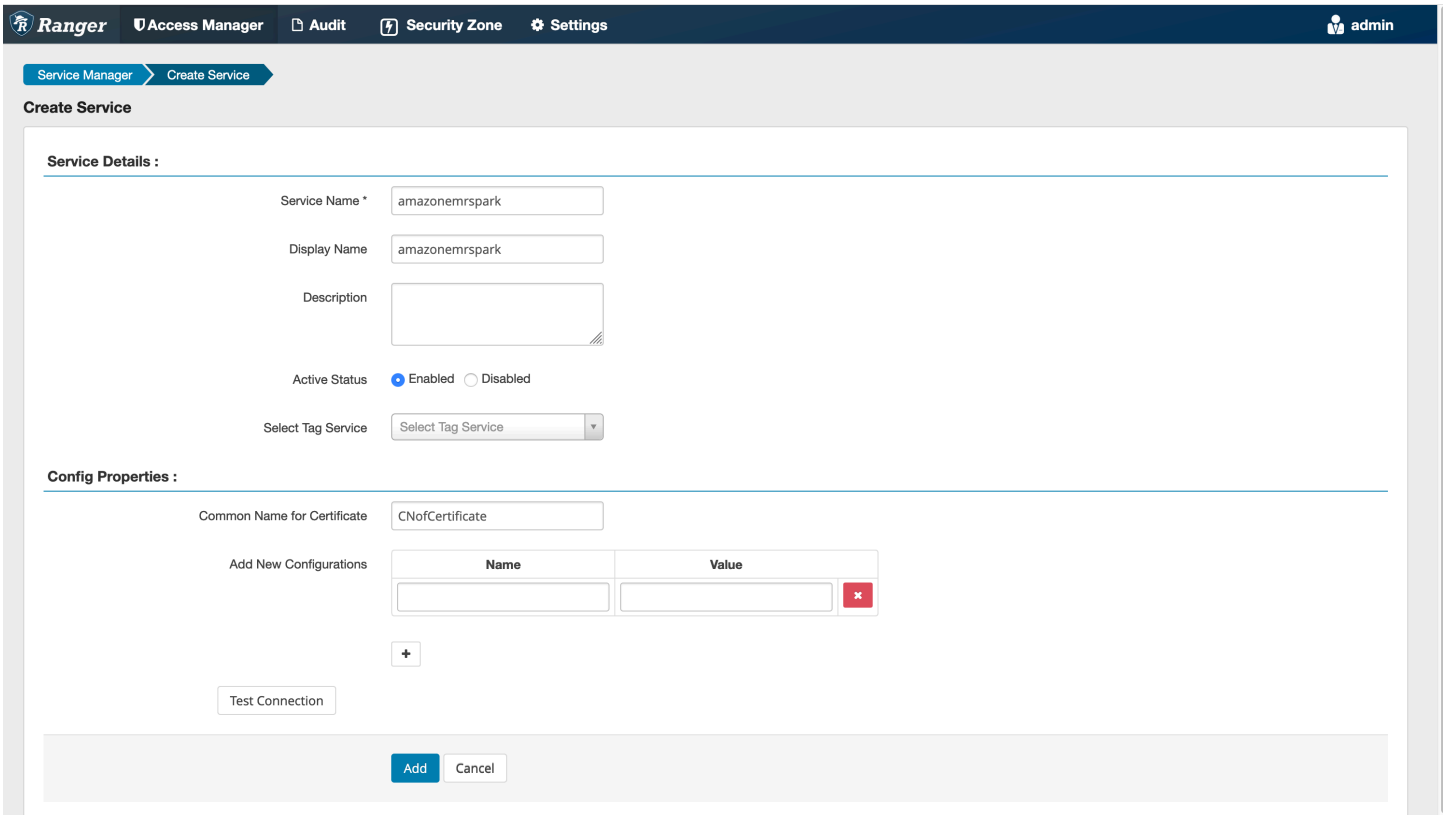


Fase 5: Creazione di un'istanza dell'applicazione AMAZON-EMR-SPARK

Nome del servizio (se visualizzato): il nome del servizio che verrà utilizzato. Il valore suggerito è **`amazonemrspark`**. Prendi nota di questo nome del servizio dal momento che sarà necessario quando crei una configurazione di sicurezza EMR.

Nome visualizzato: il nome da visualizzare per questa istanza. Il valore suggerito è **`amazonemrspark`**.

Nome comune per certificato: il campo CN all'interno del certificato utilizzato per connettersi al server Admin da un plug-in client. Questo valore deve corrispondere al campo CN nel certificato TLS creato per il plug-in.



> **Note**
>
> Il certificato TLS per questo plug-in dovrebbe essere stato registrato nel truststore sul server Admin Ranger. Per ulteriori dettagli, consulta Certificati TLS.

Creazione di policy SparkSQL

Quando si crea una nuova policy, i campi da compilare sono i seguenti:

Nome policy: il nome della policy.

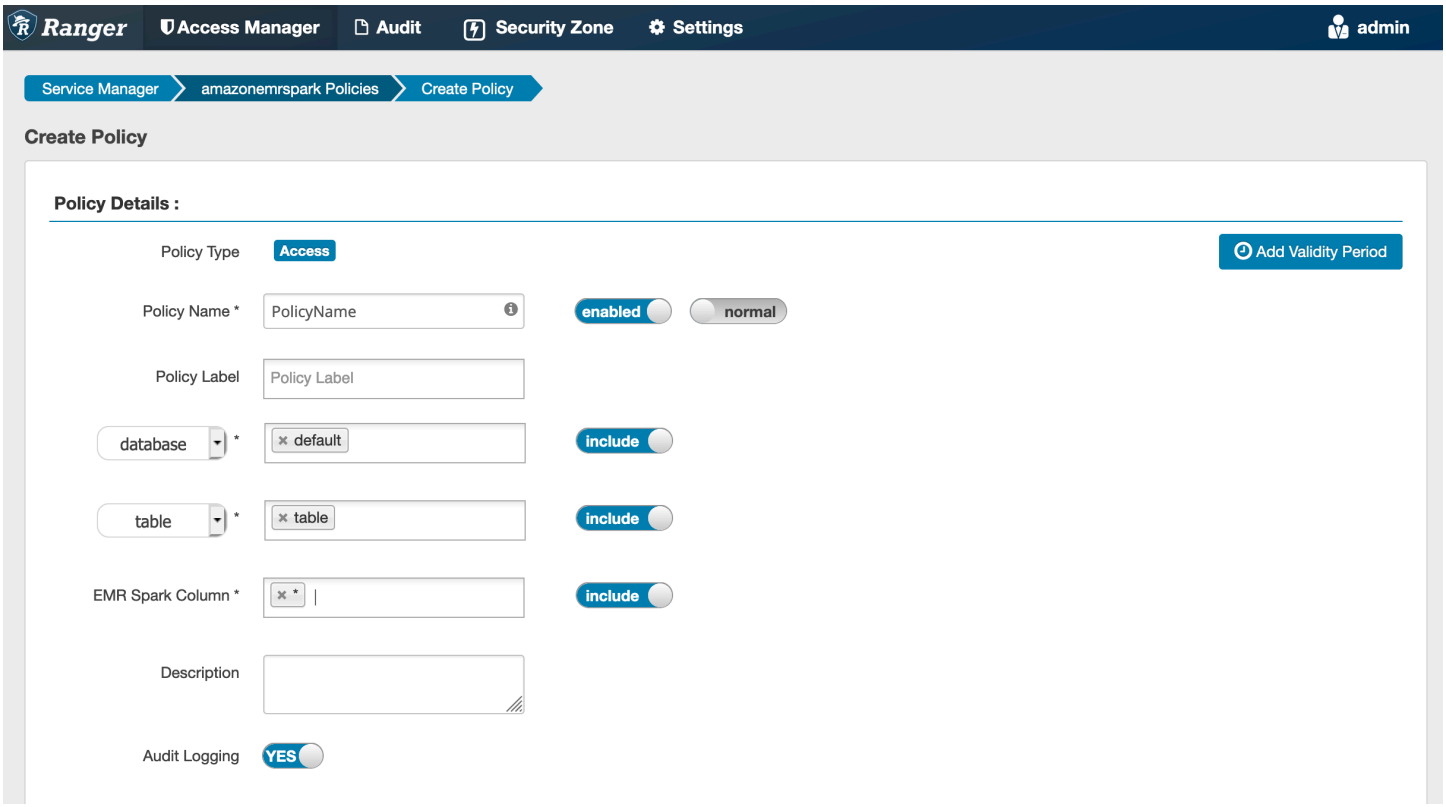
Etichetta policy: un'etichetta che è possibile inserire in questa policy.

Database: il database a cui viene applicata questa policy. Il carattere jolly "*" rappresenta tutti i database.

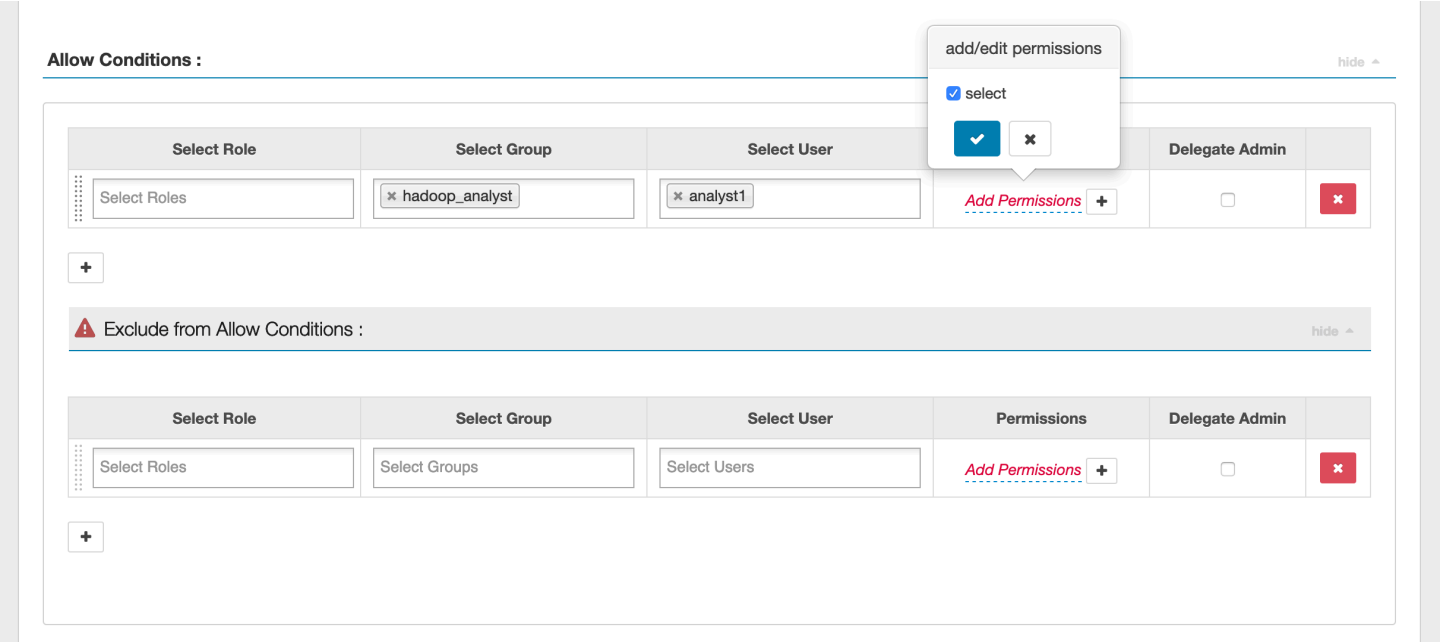
Tabella: le tabelle a cui viene applicata questa policy. Il carattere jolly "*" rappresenta tutte le tabelle.

Colonna Spark EMR: le colonne a cui viene applicata questa policy. Il carattere jolly "*" rappresenta tutte le colonne.

Description (Descrizione): la descrizione di questa policy.



Per specificare gli utenti e i gruppi, immetti gli utenti e i gruppi riportati di seguito per concedere le autorizzazioni. È inoltre possibile specificare delle esclusioni per consentire e negare le condizioni.



Dopo aver specificato le condizioni di autorizzazione e negazione, fai clic su Save (Salva).

Considerazioni

Ogni nodo all'interno del cluster EMR deve essere in grado di connettersi al nodo principale sulla porta 9083.

Limitazioni

Di seguito sono riportate le attuali limitazioni per il plug-in Apache Spark:

- Record Server si connette sempre ad HMS in esecuzione su un cluster Amazon EMR. Configura HMS per la connessione alla modalità remota, se necessario. Non inserire i valori di configurazione all'interno del file di configurazione Hive-site.xml di Apache Spark.
- Le tabelle create utilizzando origini dati Spark su CSV o Avro non sono leggibili utilizzando EMR RecordServer. Utilizza Hive per creare e scrivere dati, per poi leggerli utilizzando Record.
- Le tabelle Delta Lake e Hudi non sono supportate.
- Gli utenti devono avere accesso al database predefinito. Questo è un requisito per Apache Spark.
- Il server Admin Ranger non supporta il completamento automatico.
- Il plug-in SparkSQL per Amazon EMR non supporta i filtri di riga o il mascheramento dei dati.
- Quando si utilizza ALTER TABLE con Spark SQL, una posizione di partizione deve essere la directory figlio di una posizione di tabella. L'inserimento di dati in una partizione in cui la posizione della partizione è diversa da quella della tabella non è supportata.

Plug-in EMRFS S3

Per semplificare la fornitura di controlli degli accessi agli oggetti in S3 in un cluster multi-tenant, il plug-in EMRFS S3 fornisce controlli degli accessi ai dati all'interno di S3 quando vi si accede tramite EMRFS. È possibile consentire l'accesso alle risorse S3 a livello di utente e gruppo.

Per raggiungere questo obiettivo, quando l'applicazione tenta di accedere ai dati all'interno di S3, EMRFS invia una richiesta di credenziali al processo Secret Agent, in cui la richiesta viene autenticata e autorizzata rispetto a un plug-in Apache Ranger. Se la richiesta è autorizzata, l'agente segreto assume il ruolo IAM per Apache Ranger Engines con una policy limitata per generare credenziali che hanno accesso solo alla policy Ranger che ha consentito l'accesso. Le credenziali vengono quindi passate a EMRFS per accedere a S3.

Argomenti

Funzionalità supportate

Il plug-in EMRFS S3 fornisce l'autorizzazione a livello di archiviazione. È possibile creare policy che consentano l'accesso di utenti e gruppi a bucket e prefissi S3. L'autorizzazione è gestita solo rispetto a EMRFS.

Installazione della configurazione del servizio

L'installazione della definizione del servizio Trino richiede la configurazione del server Admin Ranger. Per configurare il server Admin Ranger, consulta Configurazione del server Admin Ranger.

Attenersi alla seguente procedura per installare la definizione del servizio EMRFS.

Fase 1: SSH nel server Admin Apache Ranger.

Ad esempio:

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

Fase 2: Download della definizione del servizio Amazon EMR e del plug-in del server Admin Apache Ranger.

In una directory temporanea, scarica la definizione del servizio Amazon EMR. Questa definizione del servizio è supportata dalle versioni Ranger 2.x.

```
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
```

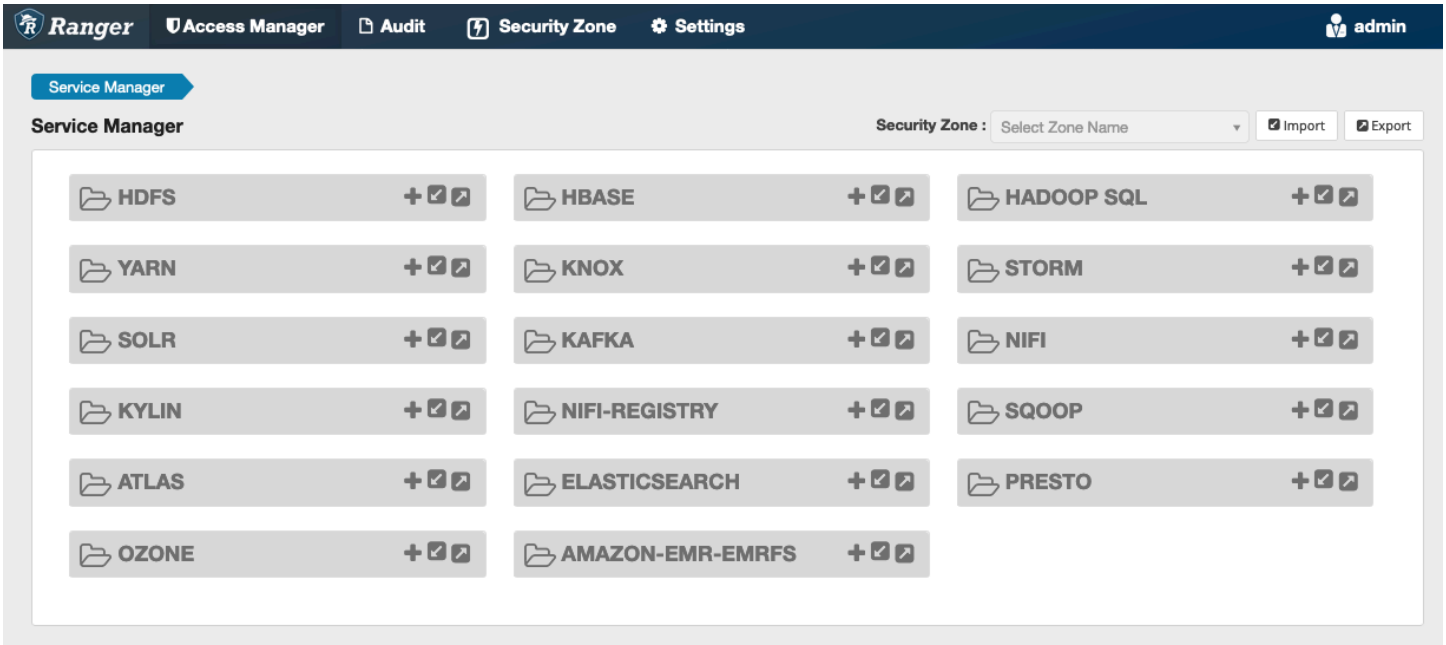
Fase 3: Installazione del plug-in Apache EMRFS S3 per Amazon EMR.

```
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
```

Fase 4: Registrazione della definizione del servizio 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'
```

Se questo comando viene eseguito correttamente, viene visualizzato un nuovo servizio nell'interfaccia utente del server Admin Ranger denominato "AMAZON-EMR-S3", come mostrato nell'immagine seguente (Ranger versione 2.0).



Fase 5: Creazione di un'istanza dell'applicazione AMAZON-EMR-EMRFS.

Crea un'istanza della definizione del servizio.

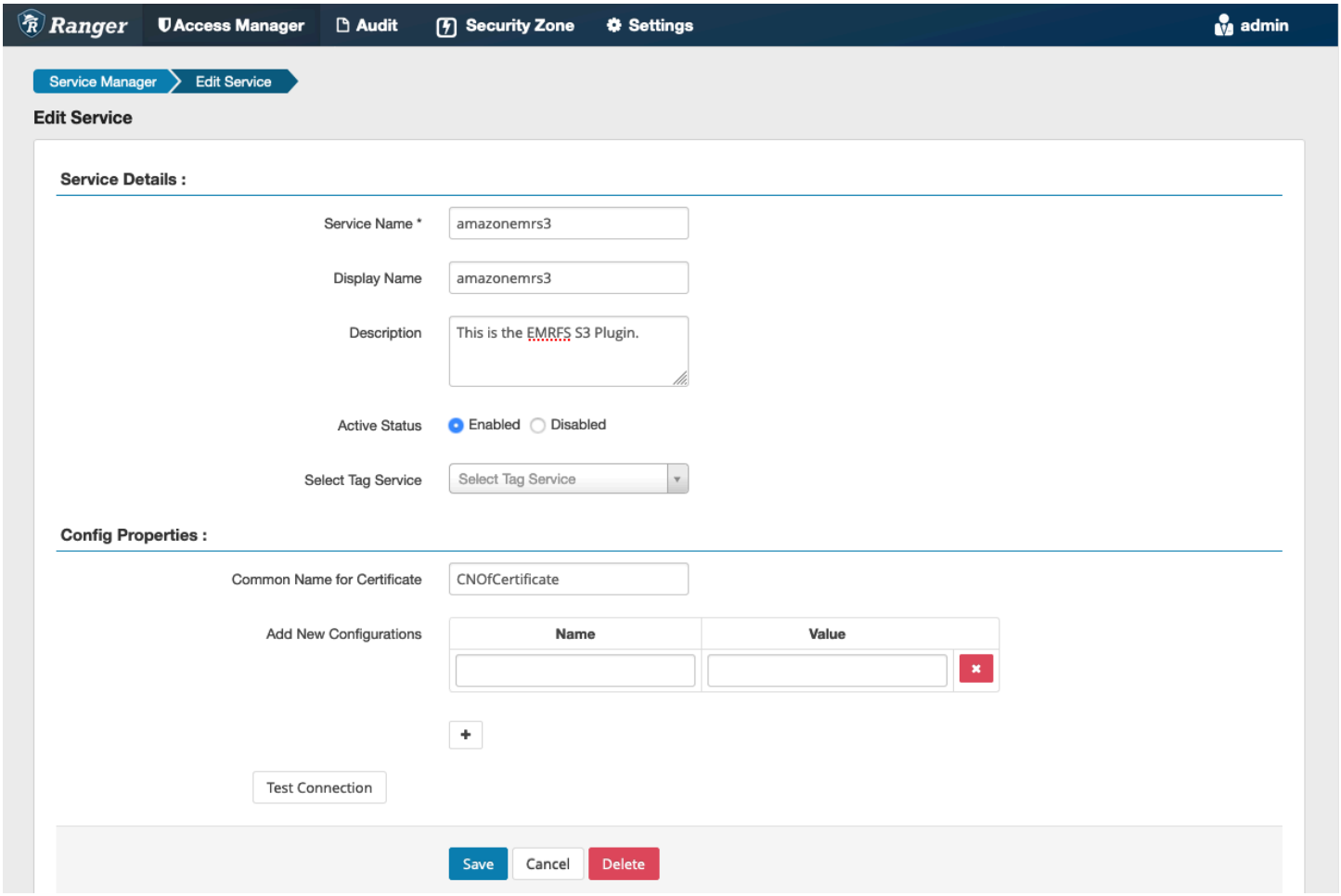
- Fai clic sul pulsante + accanto ad AMAZON-EMR-EMRFS.

Riempi i seguenti campi:

Nome del servizio (se visualizzato): il valore suggerito è**`amazonemrspark`**. Prendi nota di questo nome del servizio dal momento che sarà necessario quando crei una configurazione di sicurezza EMR.

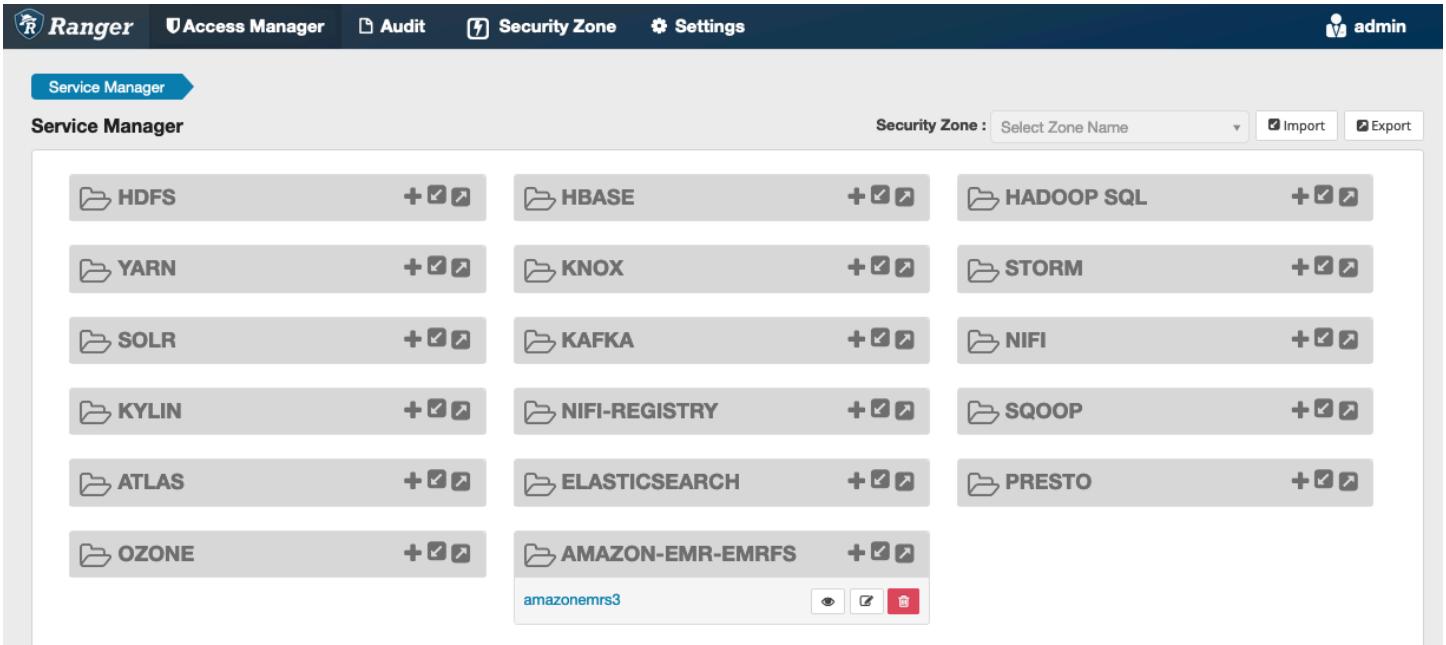
Nome visualizzato: immetti il nome da visualizzare per il servizio. Il valore suggerito è **`amazonemrspark`**.

Nome comune per certificato: il campo CN all'interno del certificato utilizzato per connettersi al server Admin da un plug-in client. Questo valore deve corrispondere al campo CN nel certificato TLS creato per il plug-in.



> **Note**
>
> Il certificato TLS per questo plug-in dovrebbe essere stato registrato nel truststore sul server Admin Ranger. Per ulteriori dettagli, consulta Certificati TLS.

Quando viene creato il servizio, il Service Manager include "AMAZON-EMR-EMRFS", come mostrato nell'immagine seguente.



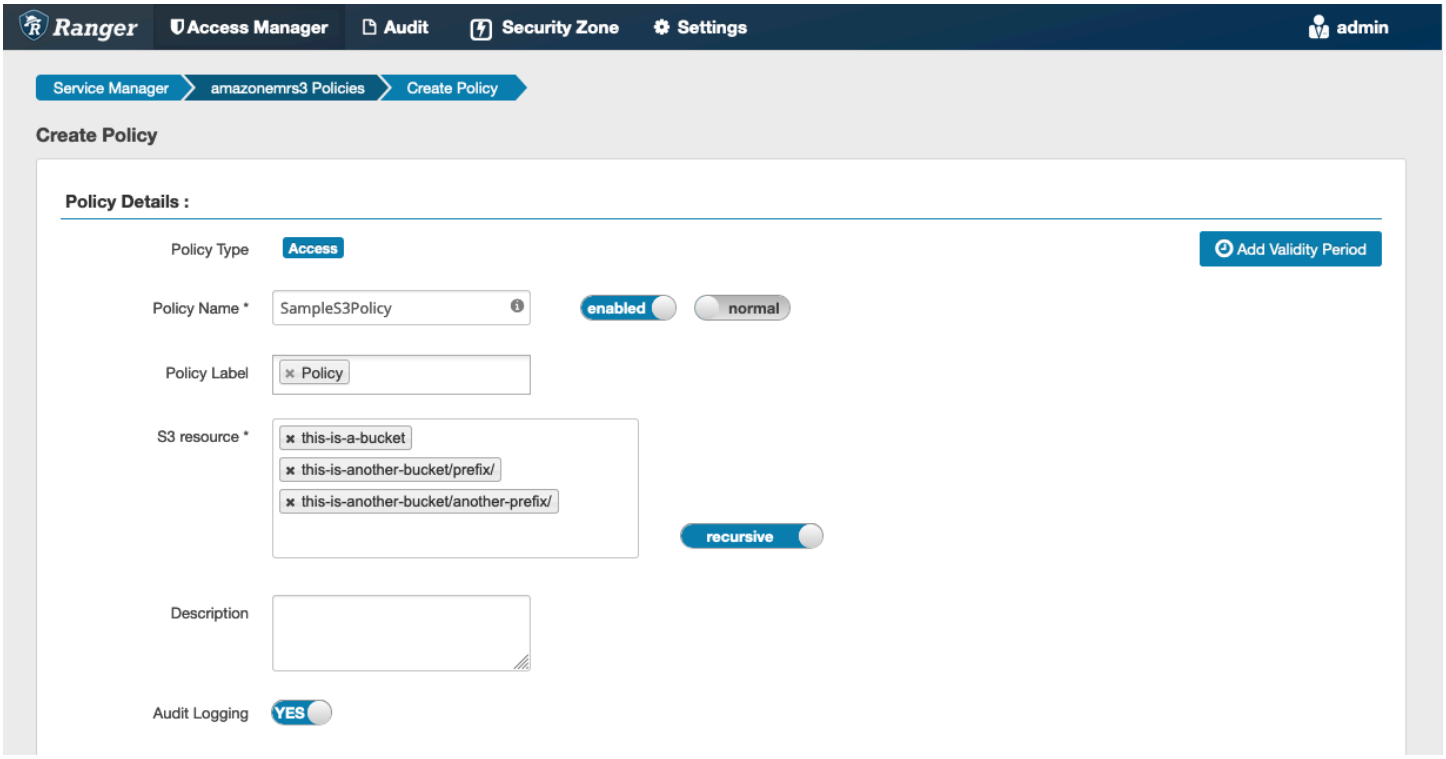
Creazione di policy EMRFS S3

Per creare una nuova policy, nella pagina Create Policy (Crea policy) del Service Manager compila i campi riportati di seguito.

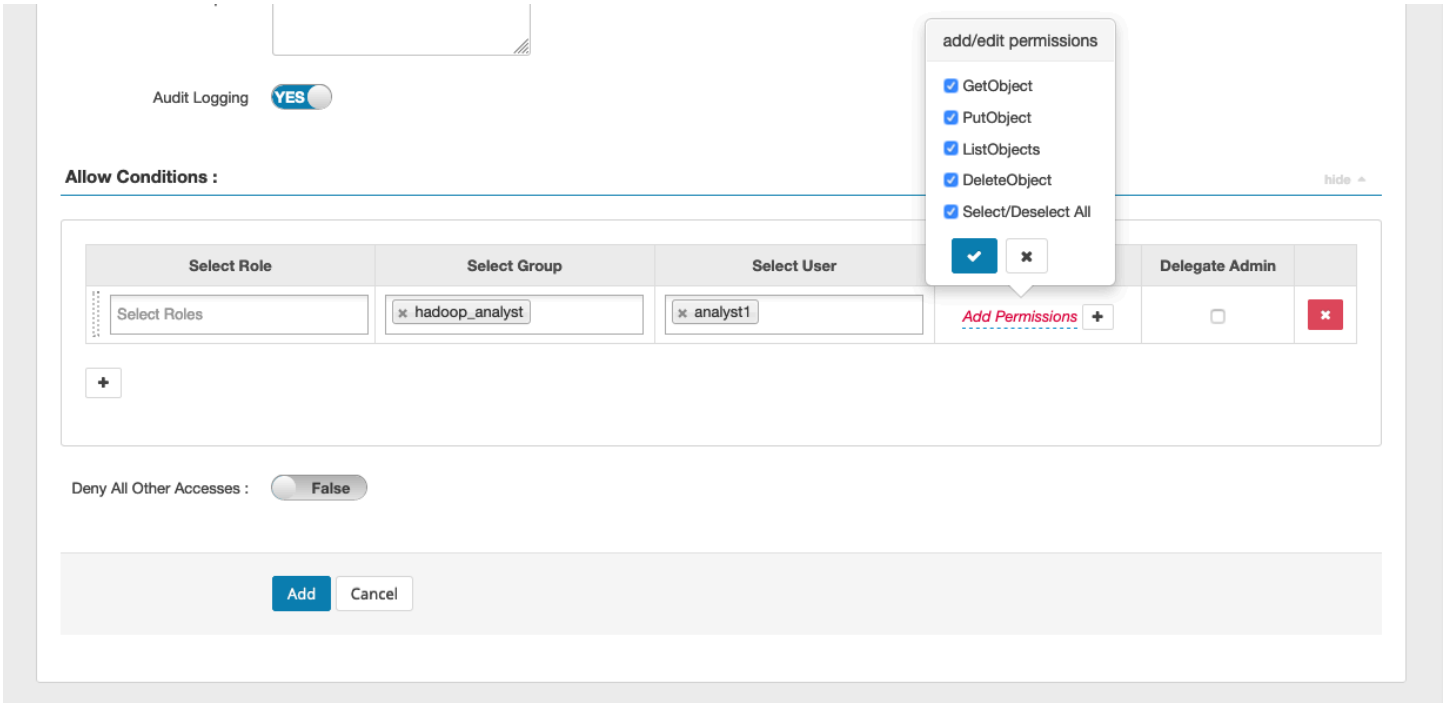
Nome policy: il nome della policy.

Etichetta policy: un'etichetta che è possibile inserire in questa policy.

Risorsa S3: una risorsa che inizia con il bucket e il prefisso facoltativo. Per ulteriori informazioni sulle best practice, consulta Note sull'utilizzo delle policy EMRFS S3. Le risorse nel server Admin Ranger non devono contenere **`s3://`**, **`s3a://`** o **`s3n://`**.



È possibile specificare utenti e gruppi per concedere le autorizzazioni. È inoltre possibile specificare delle esclusioni per consentire e negare le condizioni.



> **Note**
>
> Sono consentite al massimo tre risorse per ogni policy. L'aggiunta di più di tre risorse può causare un errore quando questa policy viene utilizzata in un cluster EMR. L'aggiunta di più di tre policy consente di visualizzare un promemoria relativo al limite di policy.

Note sull'utilizzo delle policy EMRFS S3

Quando si creano policy S3 all'interno di Apache Ranger, occorre tenere a mente alcune considerazioni sull'utilizzo.

Autorizzazioni a più oggetti S3

È possibile utilizzare policy ricorrenti ed espressioni con caratteri jolly per concedere autorizzazioni a più oggetti S3 con prefissi comuni. Le policy ricorrenti forniscono autorizzazioni a tutti gli oggetti con un prefisso comune. Le espressioni con caratteri jolly selezionano più prefissi. Insieme, forniscono le autorizzazioni a tutti gli oggetti con più prefissi comuni, come mostrato nei seguenti esempi.

Example Utilizzo di una policy ricorrente

Supponiamo che necessiti di autorizzazioni per elencare tutti i file parquet in un bucket S3 organizzato come segue.

```
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
```

Innanzitutto, considera i file parquet con il prefisso `s3://sales-reports/americas/year=2000`. Puoi concedere autorizzazioni GetObject a tutti i file in due modi:

Utilizzo di policy non ricorrenti: un'opzione consiste nell'utilizzare due policy non ricorrenti separate, una per la directory e l'altra per i file.

La prima policy concede l'autorizzazione al prefisso `s3://sales-reports/americas/year=2020` (non è presente il `/` finale).

```
- S3 resource = "sales-reports/americas/year=2000"
- permission = "GetObject"
- user = "analyst"
```

La seconda policy utilizza l'espressione jolly per concedere autorizzazioni a tutti i file con prefisso `sales-reports/americas/year=2020/` (nota il `/` finale).

```
- S3 resource = "sales-reports/americas/year=2020/*"
- permission = "GetObject"
- user = "analyst"
```

Utilizzo di una policy ricorrente: un'alternativa più conveniente consiste nell'utilizzare una singola policy ricorrente e concedere l'autorizzazione ricorrente al prefisso.

```
- S3 resource = "sales-reports/americas/year=2020"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Finora, solo i file parquet con il prefisso `s3://sales-reports/americas/year=2000` sono stati inclusi. È ora possibile includere anche i file parquet con un prefisso diverso, `s3://sales-reports/americas/year=2020`, nella stessa policy ricorrente introducendo un'espressione con caratteri jolly come segue.

```
- S3 resource = "sales-reports/americas/year=20?0"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Policy per le autorizzazioni PutObject e DeleteObject

La scrittura di policy per le autorizzazioni `PutObject` e `DeleteObject` per i file su EMRFS richiede particolare attenzione perché, a differenza delle autorizzazioni GetObject, necessita di autorizzazioni ricorrenti aggiuntive concesse al prefisso.

Example Policy per le autorizzazioni PutObject e DeleteObject

Ad esempio, l'eliminazione del file `annual-summary.parquet` richiede non solo un'autorizzazione DeleteObject per il file effettivo,

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "DeleteObject"
- user = "analyst"
```

ma anche una policy che conceda autorizzazioni ricorrenti `GetObject` e `PutObject` per il suo prefisso.

Allo stesso modo, la modifica del file `annual-summary.parquet` richiede non solo un'autorizzazione `PutObject` per il file effettivo,

```
- S3 resource = "sales-reports/americas/year=2020/annual-summary.parquet"
- permission = "PutObject"
- user = "analyst"
```

ma anche una policy che conceda l'autorizzazione ricorrente `GetObject` per il suo prefisso.

```
- S3 resource = "sales-reports/americas/year=2020"
- permission = "GetObject"
- user = "analyst"
- is recursive = "True"
```

Caratteri jolly nelle policy

Sono disponibili due aree in cui è possibile specificare i caratteri jolly. Quando si specifica una risorsa S3, è possibile utilizzare i valori "*" e "?". "*" fornisce la corrispondenza con un percorso S3 e corrisponde a tutto ciò che viene dopo il prefisso. Per esempio, la seguente policy.

```
S3 resource = "sales-reports/americas/*"
```

Questo corrisponde ai seguenti percorsi 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
```

Il carattere jolly "?" corrisponde a un singolo carattere. Ad esempio, per la policy:

```
S3 resource = "sales-reports/americas/year=201?/"
```

Questo corrisponde ai seguenti percorsi S3.

```
sales-reports/americas/year=2019/
sales-reports/americas/year=2018/
sales-reports/americas/year=2017/
```

Caratteri jolly negli utenti

Esistono due caratteri jolly incorporati quando si assegnano utenti per consentire l'accesso agli utenti. Il primo è il carattere jolly "{USER}" che fornisce l'accesso a tutti gli utenti. Il secondo carattere jolly è "{OWNER}" che fornisce l'accesso diretto o l'accesso al proprietario di un particolare oggetto. Tuttavia, il carattere jolly "{USER}" non è attualmente supportato.

Limitazioni

Di seguito sono riportate le attuali limitazioni per il plug-in EMRFS S3:

- Apache Ranger può avere al massimo tre policy.
- L'accesso a S3 deve essere realizzato tramite EMRFS e può essere utilizzato con le applicazioni correlate ad Hadoop. Il seguente elemento non è supportato:

  - Librerie Boto3

  - Kit SDK AWS e CLI AWK

  - Connettore open source S3A
- Le policy di negazione di Apache Ranger non sono supportate.

- Le operazioni su S3 con chiavi con crittografia CSE-KMS non sono attualmente supportate.
- Il supporto tra Regioni non è supportato.
- La caratteristica Security Zone di Apache Ranger non è supportata. Le restrizioni per il controllo degli accessi definite utilizzando la funzione Security Zone non vengono applicate ai cluster Amazon EMR.
- L'utente Hadoop non genera eventi di verifica poiché Hadoop accede sempre al profilo dell'istanza EC2.
- Si consiglia di disabilitare la visualizzazione di coerenza Amazon EMR. S3 ha un'elevata coerenza, pertanto tale visualizzazione non è più necessaria. Per maggiori informazioni, consulta Forte coerenza di Amazon S3.
- Il plug-in EMRFS S3 effettua numerose chiamate STS. Si consiglia di eseguire test di carico su un account di sviluppo e monitorare il volume delle chiamate STS. Si consiglia inoltre di effettuare una richiesta STS per aumentare i limiti del servizio AssumeRole.

Plugin Trino

Trino (precedentemente PrestoSQL) è un motore di query SQL che consente di eseguire query su più origini dati, come HDFS, archiviazione di oggetti, database relazionali e database NoSQL. Grazie a questo plugin è possibile eseguire query sui dati indipendentemente dalla posizione in cui si trovano, senza la necessità di effettuare la migrazione di dati in un percorso centrale. Amazon EMR mette a disposizione un plugin Apache Ranger per fornire un controllo granulare degli accessi a Trino. Il plug-in è compatibile con il server open source Apache Ranger Admin versione 2.0 e successive.

Argomenti

Funzionalità supportate

Il plugin Apache Ranger per Trino su Amazon EMR supporta tutte le funzionalità del motore di query Trino, che è protetto da un controllo granulare degli accessi comprendente controlli degli accessi a livello di database, tabella e colonna, filtraggio di riga e mascheramento dei dati. Le policy di Apache

Ranger possono includere policy di concessione e di negazione a utenti e gruppi. Anche gli eventi di verifica vengono inviati a CloudWatch Logs.

Installazione della configurazione del servizio

L'installazione della definizione del servizio Trino richiede la configurazione del server Admin Ranger. Per configurare il server Admin Ranger, consulta Configurazione del server Admin Ranger.

Attenersi alla seguente procedura per installare la definizione del servizio Trino.

1. SSH nel server Admin Apache Ranger.

```
ssh ec2-user@ip-xxx-xxx-xxx-xxx.ec2.internal
```

2. Disinstalla il plugin del server Presto, se presente. Esegui il comando seguente. Se viene visualizzato l'errore "Service not found" (Servizio non trovato), il plugin del server Presto non è stato installato sul server. Passare alla fase successiva.

```
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'
```

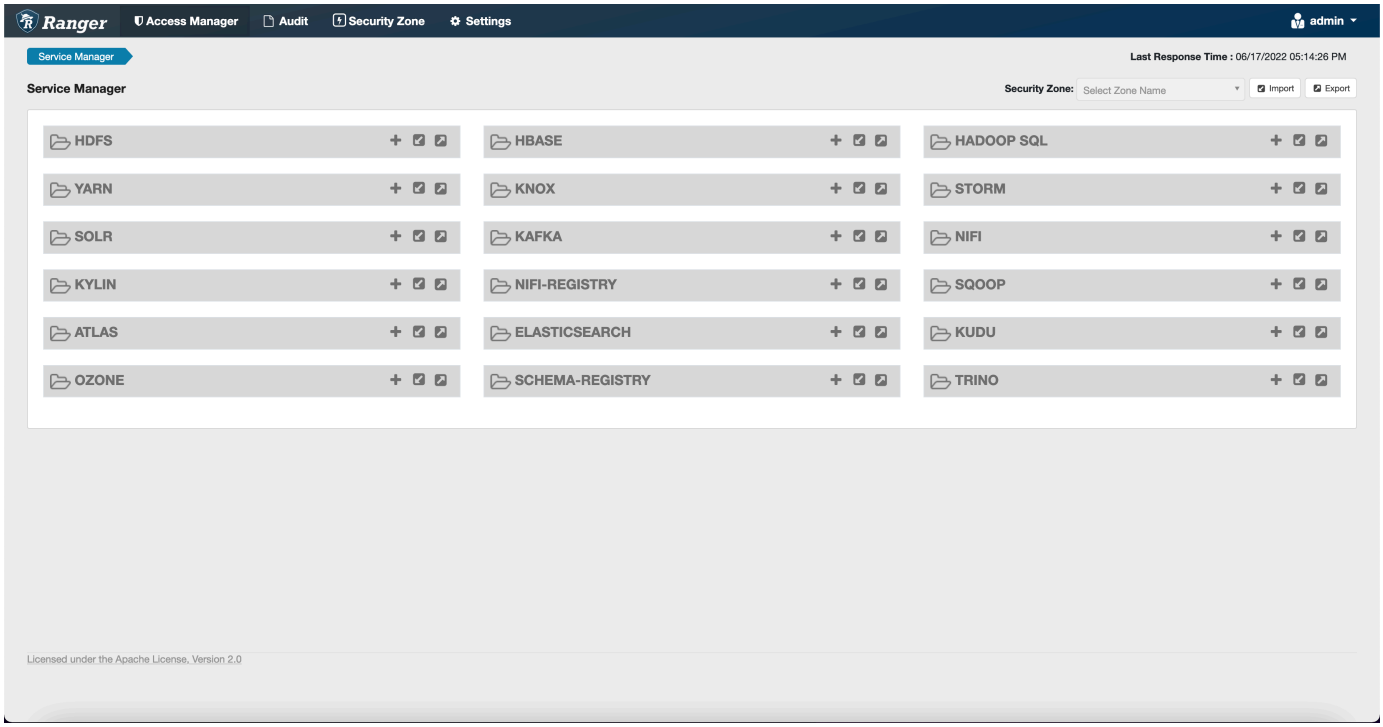
3. Scarica la definizione del servizio e il plugin del server Admin Apache Ranger. In una directory temporanea, scarica la definizione del servizio. Questa definizione del servizio è supportata dalle versioni Ranger 2.x.

```
wget https://s3.amazonaws.com/elasticmapreduce/ranger/service-definitions/
version-2.0/ranger-servicedef-amazon-emr-trino.json
```

4. Registra la definizione del servizio Apache Trino per 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'
```

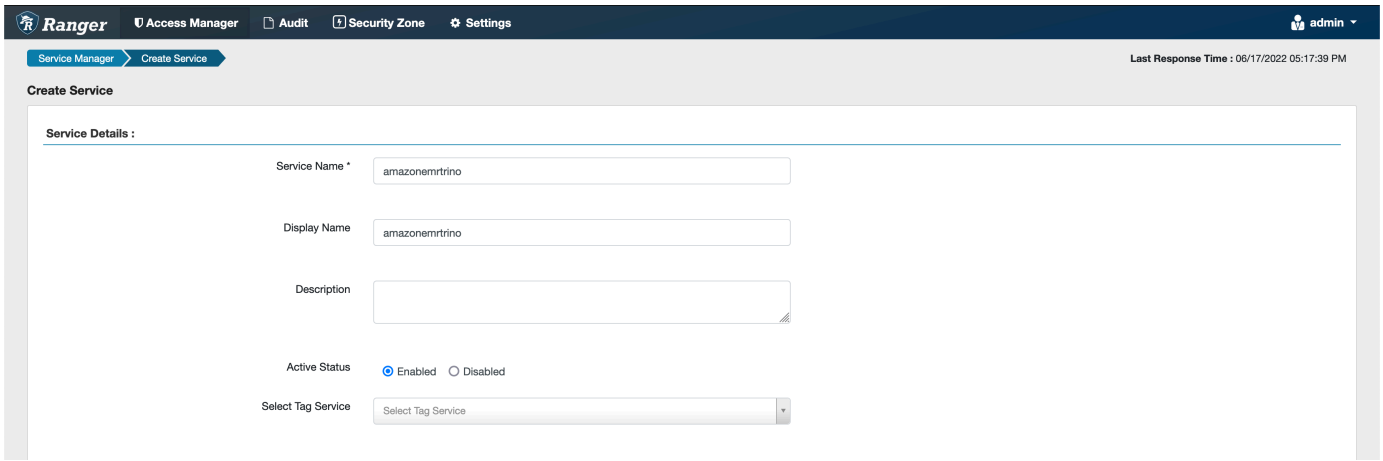
   Se questo comando viene eseguito correttamente, viene visualizzato un nuovo servizio nell'interfaccia utente di Admin Apache Ranger denominato `TRINO`, come mostrato nell'immagine seguente.



5. Crea un'istanza dell'applicazione `TRINO` inserendo le informazioni seguenti.

   Service Name (Nome del servizio): il nome del servizio che verrà utilizzato. Il valore suggerito è `amazonemrtrino`. Prendi nota di questo nome del servizio dal momento che sarà necessario durante la creazione di una configurazione di sicurezza per Amazon EMR.

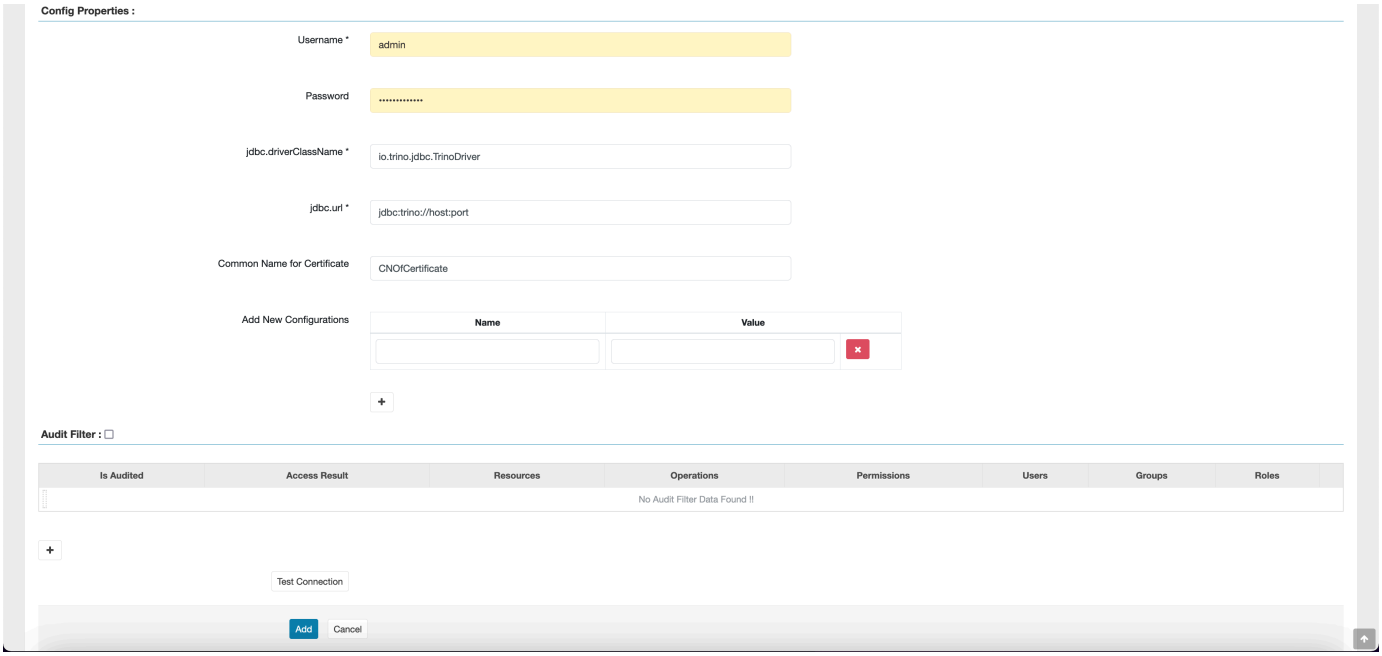
   Nome visualizzato: il nome da visualizzare per questa istanza. Il valore suggerito è `amazonemrtrino`.



   jdbc.driver.ClassName: il nome della classe JDBC per la connettività Trino. Puoi usare il valore predefinito.

jdbc.url: la stringa di connessione JDBC da utilizzare per la connessione a un coordinatore Trino.

Nome comune per certificato: il campo CN all'interno del certificato utilizzato per connettersi al server Admin da un plug-in client. Questo valore deve corrispondere al campo CN nel certificato TLS creato per il plug-in.



Il certificato TLS per questo plugin dovrebbe essere stato registrato nel trust store sul server Admin Ranger. Per ulteriori informazioni, consulta la sezione Certificati TLS.

Creazione di policy Trino

Quando crei una nuova policy, compila i campi seguenti.

Nome policy: il nome della policy.

Etichetta policy: un'etichetta che è possibile inserire in questa policy.

Catalog (Catalogo): il catalogo a cui viene applicata questa policy. Il carattere jolly "*" rappresenta tutti i cataloghi.

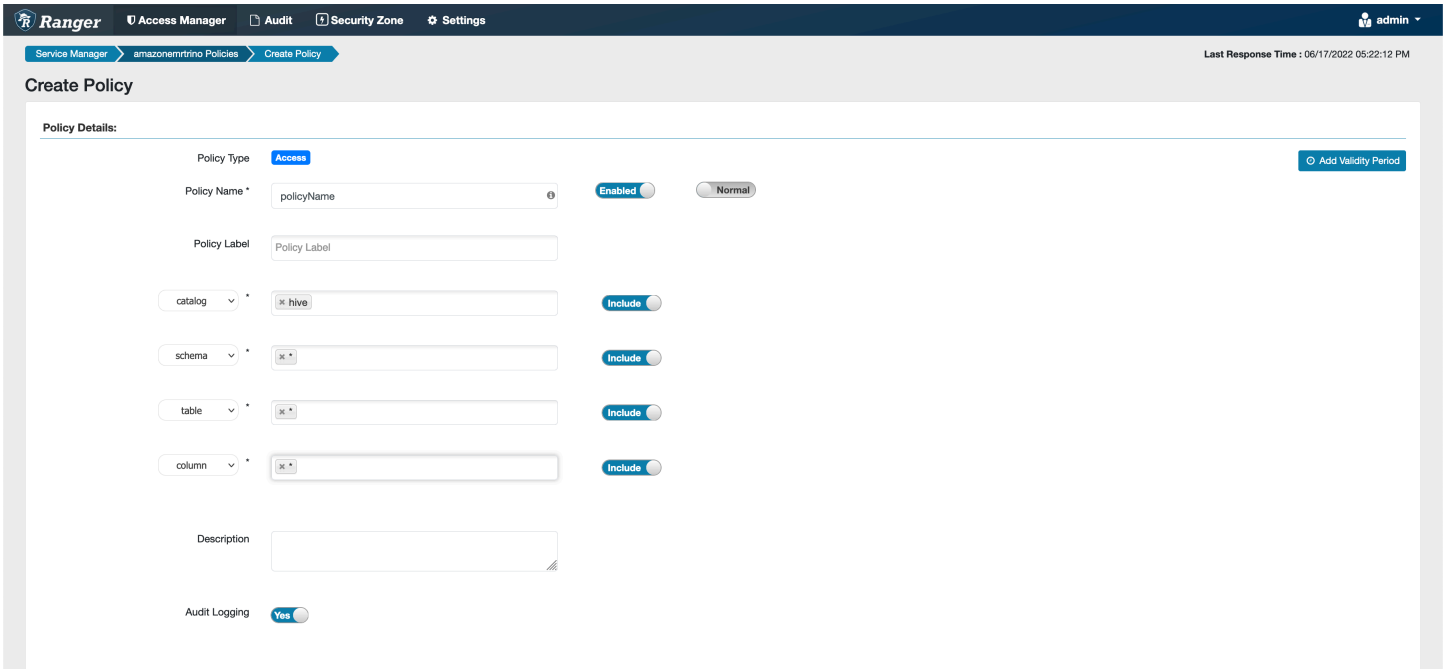
Schema (Schema): gli schemi a cui viene applicata questa policy. Il carattere jolly "*" rappresenta tutti gli schemi.

Tabella: le tabelle a cui viene applicata questa policy. Il carattere jolly "*" rappresenta tutte le tabelle.

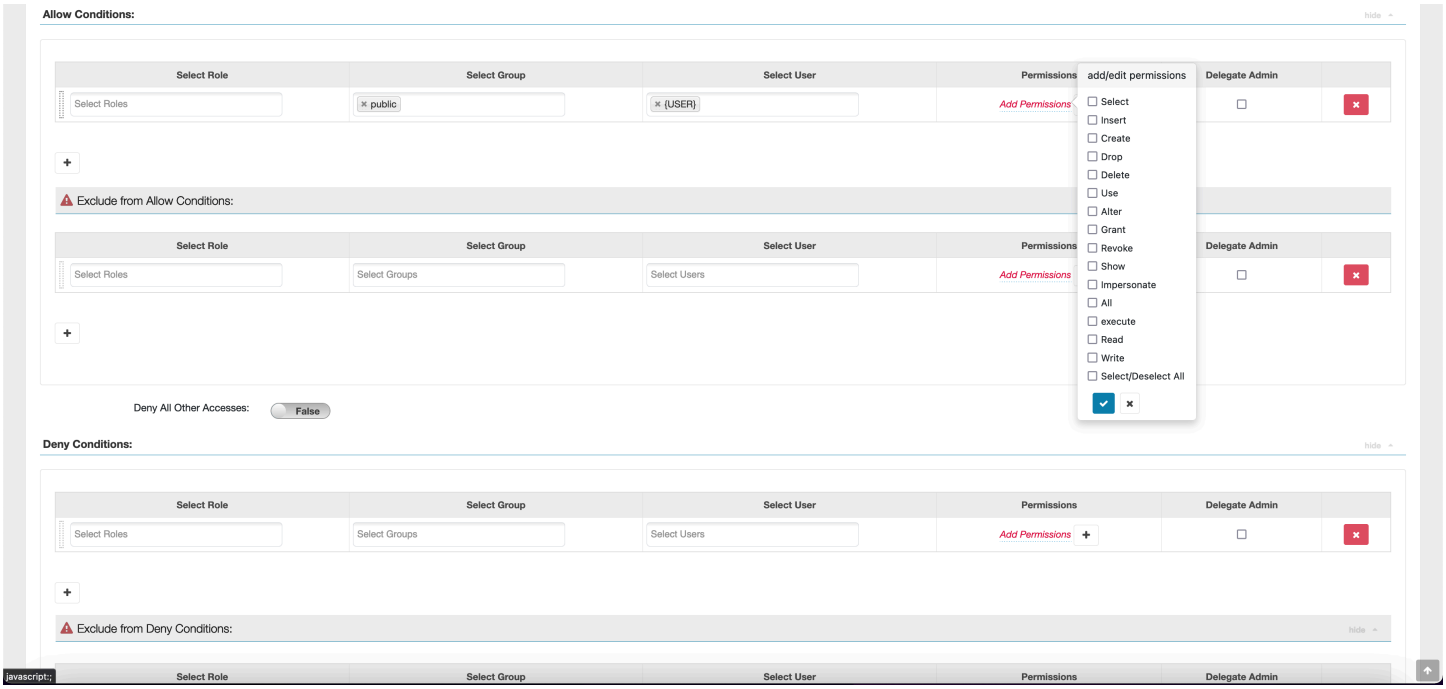
Column (Colonna): le colonne a cui viene applicata questa policy. Il carattere jolly "*" rappresenta tutte le colonne.

Description (Descrizione): la descrizione di questa policy.

Esistono altri tipi di policy, ad esempio Trino User (Utente Trino) (per l'accesso alla rappresentazione di utenti), Trino System/Session Property (Proprietà sistema/sessione Trino) (per modificare le proprietà del sistema o della sessione del motore), Functions/Procedures (Funzioni/procedure) (per consentire chiamate a funzioni o procedure) e URL (per consentire l'accesso in lettura/scrittura al motore nei percorsi dati).



Per concedere autorizzazioni a utenti e gruppi specifici, inserirli. È inoltre possibile specificare delle esclusioni per consentire e negare le condizioni.



Dopo aver specificato le condizioni di autorizzazione e negazione, scegli Save (Salva).

Considerazioni

Quando si creano policy Trino all'interno di Apache Ranger, occorre tenere a mente alcune considerazioni sull'utilizzo.

Server dei metadati Hive

Il server dei metadati Hive è accessibile solo dai motori attendibili, in particolare il motore Trino, come misura di protezione da accessi non autorizzati. Il server dei metadati Hive è accessibile anche da tutti i nodi del cluster. La porta 9083 richiesta consente a tutti i nodi di accedere al nodo principale.

Autenticazione

Per impostazione predefinita, Trino è configurato per l'autenticazione utilizzando Kerberos, come specificato nella configurazione di sicurezza di Amazon EMR.

Crittografia in transito obbligatoria

Il plugin Trino richiede che la crittografia in transito sia abilitata nella configurazione di sicurezza di Amazon EMR. Per abilitare la crittografia, consulta Crittografia in transito.

Limitazioni

Di seguito sono riportate le attuali limitazioni per il plugin Trino:

- Il server Admin Ranger non supporta il completamento automatico.

Risoluzione dei problemi di Apache Ranger

Seguono alcuni dei problemi comunemente diagnosticati relativi all'utilizzo di Apache Ranger.

Raccomandazioni

- Test utilizzando un cluster a nodo principale singolo: il provisioning dei cluster a nodo master singolo è più rapido rispetto a un cluster a più nodi e può ridurre il tempo per ogni iterazione di test.
- Imposta la modalità di sviluppo sul cluster. Quando avvii il cluster EMR, imposta il parametro `--additional-info"` su:

  ```
  '{"clusterType":"development"}'
  ```

  Questo parametro può essere impostato solo tramite la CLI AWS o l'SDK AWS e non è disponibile tramite la console Amazon EMR. Quando questo contrassegno è impostato e il master non riesce a eseguire il provisioning, il servizio Amazon EMR mantiene il cluster attivo per un po' di tempo prima di disattivarlo. Questo arco temporale è molto utile per sondare i vari file di log prima che il cluster venga terminato.

Insuccesso del provisioning del cluster EMR

Esistono diversi motivi alla base del mancato avvio di un cluster Amazon EMR. Di seguito sono riportati alcuni modi per diagnosticare il problema.

Verifica dei log di provisioning EMR

Amazon EMR utilizza Puppet per installare e configurare le applicazioni in un cluster. Esaminando i registri verranno forniti dettagli sulla presenza o meno di errori durante la fase di provisioning di un cluster. I log sono accessibili sul cluster o su S3 se sono configurati per essere inviati a S3.

I log vengono archiviati in `/var/log/provision-node/apps-phase/0/{UUID}/puppet.log` sul disco e `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/provision-node/apps-phase/0/{UUID}/puppet.log.gz.`.

Messaggi di errore comuni

Messaggio di errore	Causa
Puppet (err): Systemd start for emr-record-server failed! journalctl log for emr-record-server:	Impossibile avviare EMR Record Server. Consulta "Log di EMR Record Server" di seguito.
Puppet (err): Systemd start for emr-record-server failed! journalctl log for emrsecretagent:	Impossibile avviare EMR Secret Agent. Consulta "Controllo dei log Secret Agent" di seguito.
/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	Il certificato TLS privato in Secrets Manager per il certificato del plug-in Apache Ranger non è nel formato corretto o non è un certificato privato. Consulta Certificati TLS per i formati dei certificati.
/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-XXXXXXXXXXXX is not authorized to perform: secretsmanager:GetSecretValue on resource: arn:aws:secretsmanager:us-east-1:XXXXXXXXXX:secret:AdminServer-XXXXX	Il ruolo del profilo dell'istanza EC2 non dispone delle autorizzazioni corrette per recuperare i certificati TLS da Secret Agent.

Controlla i log di Secret Agent

I log di Secret Agent si trovano in `/emr/secretagent/log/` su un nodo EMR o nella directory `s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/daemons/secretagent/` in S3.

Messaggi di errore comuni

Messaggio di errore	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)	L'eccezione precedente indica che il ruolo del profilo dell'istanza EC2 EMR non dispone delle autorizzazioni per assumere il ruolo RangerPluginDataAccessRole. Per informazioni, consultare Ruoli IAM per l'integrazione nativa con Apache Ranger.
ERROR qtp54617902-149: Web App Exception Occurred javax.ws.rs.NotAllowedException: HTTP 405 Method Not Allowed	Puoi ignorare questi errori senza correre rischi.

Controllo dei log di EMR Record Server (per SparkSQL)

I log di EMR Record Server sono disponibili in /var/log/emr-record-server/ su un nodo EMR oppure possono essere consultati nella directory s3://<LOG LOCATION>/<CLUSTER ID>/node/<EC2 INSTANCE ID>/daemons/emr-record-server/ in S3.

Messaggi di errore comuni

Messaggio di errore	Causa

Messaggio di errore	Causa
InstanceMetadataServiceResourceFetcher:105 - [] Fail to retrieve token com.amazonaws.SdkClientException: Failed to connect to service endpoint	EMR Secret Agent non è riuscito a presentarsi o sta riscontrando un problema. Verifica la presenza di errori nei log di Secret Agent e nello script puppet per determinare se si sono verificati errori di provisioning.

Le query producono un errore imprevisto

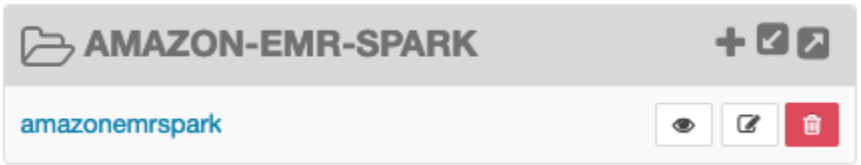
Controlla i log dei plug-in di Apache Ranger (Apache Hive, EMR Record Server, EMR Secret Agent, ecc.)

Questa sezione è comune a tutte le applicazioni che si integrano con il plug-in Ranger, come Apache Hive, EMR Record Server ed EMR Secret Agent.

Messaggi di errore comuni

Messaggio di errore	Causa
ERROR PolicyRefresher:272 - [] PolicyRefresher(serviceName=policy-repository): failed to find service. Will clean up local cache of policies (-1)	Questo messaggio di errore indica che il nome del servizio fornito nella configurazione di sicurezza EMR non corrisponde a un repository delle policy di servizio nel server Admin Ranger.

Se all'interno del server Admin Ranger il servizio AMAZON-EMR-SPARK è simile al seguente, è necessario inserire **`amazonemrspark`** come nome del servizio.



Controllo del traffico di rete con gruppi di sicurezza

I gruppi di sicurezza agiscono da firewall virtuale per le istanze EC2 nel cluster allo scopo di controllare il traffico in entrata e quello in uscita. Ogni gruppo di sicurezza ha un set di regole che controlla il traffico in entrata verso le istanze e un set di regole per controllare il traffico in uscita. Per ulteriori informazioni, consulta Gruppi di sicurezza Amazon EC2 per le istanze Linux nella Guida per l'utente di Amazon EC2 per le istanze Linux.

È possibile utilizzare due classi di gruppi di sicurezza con Amazon EMR: gruppi di sicurezza gestiti da Amazon EMR e gruppi di sicurezza aggiuntivi.

A ogni cluster sono associati gruppi di sicurezza gestiti. È possibile utilizzare i gruppi di sicurezza gestiti predefiniti creati da Amazon EMR oppure specificare gruppi di sicurezza gestiti personalizzati. In entrambi i casi, Amazon EMR aggiunge automaticamente regole per i gruppi di sicurezza gestiti che un cluster deve utilizzare per comunicare tra istanze del cluster e servizi AWS.

I gruppi di sicurezza aggiuntivi sono facoltativi. È possibile specificarli in aggiunta ai gruppi di sicurezza gestiti per definire l'accesso alle istanze del cluster. Altri gruppi di sicurezza contengono solo regole definite. Amazon EMR non li modifica.

Le regole che Amazon EMR crea nei gruppi di sicurezza gestiti consentono al cluster di comunicare tra i componenti interni. Per consentire agli utenti e alle applicazioni di accedere a un cluster dall'esterno del cluster, è possibile modificare le regole per i gruppi di sicurezza gestiti oppure è possibile creare ulteriori gruppi di sicurezza con regole aggiuntive o entrambe le cose.

> **Important**
>
> Modificare le regole per i gruppi di sicurezza gestiti può avere conseguenze impreviste. Potresti inavvertitamente bloccare il traffico necessario per il corretto funzionamento dei cluster e causare errori in quanto i nodi non sono raggiungibili. Pianifica attentamente e verifica le configurazioni per i gruppi di sicurezza prima dell'implementazione.

Puoi, quindi, specificare i gruppi di sicurezza solo al momento della creazione di un cluster. Essi non possono essere aggiunti a un cluster o alle istanze del cluster mentre il cluster è in esecuzione, ma è possibile modificare, aggiungere o rimuovere regole dai gruppi di sicurezza esistenti. Le regole diventano effettive non appena vengono salvate.

Per impostazione predefinita, i gruppi di sicurezza sono restrittivi Se non si aggiunge una regola che consente il traffico, il traffico viene rifiutato. Se è presente più di una regola che si applica allo stesso traffico e alla stessa origine, viene applicata la regola più permissiva. Ad esempio, se si dispone di una regola che consente SSH dall'indirizzo IP 192.0.2.12/32 e un'altra regola che consente l'accesso a tutto il traffico TCP dall'intervallo 192.0.2.0/24, ha precedenza la regola che consente tutto il traffico TCP dall'intervallo che include 192.0.2.12. In questo caso, il client all'intervallo 192.0.2.12 può avere più accesso di quanto si desideri.

> **Important**
>
> Presta attenzione quando modifichi le regole per i gruppi di sicurezza per aprire le porte. Assicurati di aggiungere regole che permettano solo il traffico da client affidabili e autenticati per i protocolli e le porte necessari per eseguire i tuoi carichi di lavoro.

È possibile configurare il blocco degli accessi pubblici per Amazon EMR in ogni Regione utilizzata per impedire la creazione del cluster quando una regola consente l'accesso pubblico su una porta non aggiunta a un elenco di eccezioni. Per gli account AWS creati dopo luglio 2019, il blocco degli accessi pubblici per Amazon EMR è attivato per impostazione predefinita. Per gli account AWS che hanno creato un cluster prima di luglio 2019, il blocco degli accessi pubblici per Amazon EMR è disattivato per impostazione predefinita. Per ulteriori informazioni, consulta Utilizzo del blocco dell'accesso pubblico di Amazon EMR.

Argomenti

- Utilizzo dei gruppi di sicurezza gestiti da Amazon EMR
- Utilizzo di gruppi di sicurezza aggiuntivi
- Specifica dei gruppi di sicurezza gestiti da Amazon EMR e aggiuntivi
- Specifica dei gruppi di sicurezza EC2 per EMR Notebooks
- Utilizzo del blocco dell'accesso pubblico di Amazon EMR

> **Note**
>
> Amazon EMR mira a utilizzare alternative inclusive per termini di settore potenzialmente offensivi o non inclusivi come "master" e "slave". Siamo passati a una nuova terminologia per promuovere un'esperienza più inclusiva e facilitare la comprensione dei componenti del servizio.

Ora descriviamo i "nodi" come istanze e descriviamo i tipi di istanze Amazon EMR come primarioi, principale e attività. Durante la transizione, potresti ancora trovare riferimenti precedenti a termini obsoleti, come quelli relativi ai gruppi di sicurezza per Amazon EMR.

Utilizzo dei gruppi di sicurezza gestiti da Amazon EMR

> **Note**
>
> Amazon EMR mira a utilizzare alternative inclusive per termini di settore potenzialmente offensivi o non inclusivi come "master" e "slave". Siamo passati a una nuova terminologia per promuovere un'esperienza più inclusiva e facilitare la comprensione dei componenti del servizio.
> Ora descriviamo i "nodi" come istanze e descriviamo i tipi di istanze Amazon EMR come primarioi, principale e attività. Durante la transizione, potresti ancora trovare riferimenti precedenti a termini obsoleti, come quelli relativi ai gruppi di sicurezza per Amazon EMR.

I diversi gruppi di sicurezza gestiti sono associati all'istanza primaria e alle istanze principali e attività in un cluster. Quando si crea un cluster in una sottorete privata, è necessario un ulteriore gruppo di sicurezza gestito per l'accesso ai servizi. Per ulteriori informazioni sul ruolo dei gruppi di sicurezza gestiti per quanto riguarda la configurazione di rete, consulta Opzioni di Amazon VPC.

Quando vengono specificati i gruppi di sicurezza gestiti per un cluster, è necessario utilizzare lo stesso tipo di gruppo di sicurezza, predefinito o personalizzato, per tutti i gruppi di sicurezza gestiti. Ad esempio, non è possibile specificare un gruppo di sicurezza personalizzato per l'istanza primaria, quindi non specificare un gruppo di sicurezza personalizzato per le istanze principali e le istanze attività.

Se utilizzi i gruppi di sicurezza gestiti predefiniti, non è necessario specificarli quando si crea un cluster. Amazon EMR utilizza automaticamente le impostazioni predefinite. Inoltre, se le impostazioni predefinite non esistono ancora nel VPC del cluster, Amazon EMR le crea. Amazon EMR le crea anche se vengono specificate in modo esplicito e non esistono ancora.

Puoi modificare le regole nei gruppi di sicurezza gestiti dopo la creazione dei cluster. Quando crei un nuovo cluster, Amazon EMR controlla le regole nei gruppi di sicurezza gestiti specificati e quindi crea tutte le regole in entrata mancanti necessarie per il nuovo cluster, oltre alle regole che possono essere state aggiunte in precedenza. Salvo diversamente specificato, ciascuna regola per gruppi di

sicurezza gestiti da Amazon EMR predefiniti viene anche aggiunta a gruppi di sicurezza gestiti da Amazon EMR personalizzati da te specificati.

I gruppi di sicurezza gestiti predefiniti sono i seguenti:

- ElasticMapReduce-primary

  Per le regole in questo gruppo di sicurezza, consulta Gruppo di sicurezza gestito da Amazon EMR per l'istanza primaria (sottoreti pubbliche).
- ElasticMapReduce-core

  Per le regole in questo gruppo di sicurezza, consulta Gruppo di sicurezza gestito da Amazon EMR per le istanze principali e attività (sottoreti pubbliche).
- ElasticMapReduce-Primary-Private

  Per le regole in questo gruppo di sicurezza, consulta Gruppo di sicurezza gestito da Amazon EMR per l'istanza primaria (sottoreti private).
- ElasticMapReduce-Core-Private

  Per le regole in questo gruppo di sicurezza, consulta Gruppo di sicurezza gestito da Amazon EMR per le istanze principali e attività (sottoreti private).
- ElasticMapReduce-ServiceAccess

  Per le regole in questo gruppo di sicurezza, consulta Gruppo di sicurezza gestito da Amazon EMR per l'accesso ai servizi (sottoreti private).

Gruppo di sicurezza gestito da Amazon EMR per l'istanza primaria (sottoreti pubbliche)

Il gruppo di sicurezza gestito predefinito per l'istanza primaria nelle sottoreti pubbliche ha come Nome gruppo ElasticMapReduce-primary. Include le seguenti regole: Se specifichi un gruppo di sicurezza gestito personalizzato, Amazon EMR aggiungerà le stesse regole al gruppo di sicurezza personalizzato.

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Regole in entrata				

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Tutte le regole ICMP-IPv4	Tutti	N/D	L'ID del gruppo di sicurezza gestito dell'istanza primaria. In altre parole, lo stesso gruppo di sicurezza in cui appare la regola.	Tali regole riflessive consentono il traffico in entrata da qualsiasi istanza associata al gruppo di sicurezza specificato. Utilizzando `ElasticMapReduce-primary` predefinito per più cluster consente ai nodi principali e di task di tali cluster di comunicare tra loro su ICMP o su qualsiasi porta TCP o UDP. Specifica i gruppi di sicurezza gestiti personalizzati per limitare l'accesso tra cluster.
Tutte le regole TCP	TCP	Tutti		
Tutte le regole UDP	UDP	Tutti		
Tutte le regole ICMP-IPV4	Tutti	N/D	L'ID del gruppo di sicurezza gestito specificato per nodi principali e di task.	Tali regole consentono il traffico ICMP in entrata e il traffico su qualsiasi porta TCP o UDP da qualsiasi istanza principale e di attività associata al gruppo di sicurezza specificato, anche se le istanze si trovano in cluster diversi.
Tutte le regole TCP	TCP	Tutti		
Tutte le regole UDP	UDP	Tutti		
Personalizza	TCP	8443	Vari intervalli di indirizzi IP Amazon	Tali regole consentono al cluster manager di comunicare con il nodo primario.

Concessione dell'accesso SSH alle origini attendibili per il gruppo di sicurezza primario con la vecchia console

Per modificare i gruppi di sicurezza, devi disporre dell'autorizzazione per gestire i gruppi di sicurezza per il VPC in cui si trova il cluster. Per ulteriori informazioni, consulta Modifica delle autorizzazioni per

un utente e la Policy di esempio che consente di gestire i gruppi di sicurezza EC2 nella Guida per l'utente IAM.

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Cluster. Scegli il Nome del cluster che desideri modificare.
3. Scegli il link Gruppi di sicurezza per master in Sicurezza e accesso.
4. Scegli ElasticMapReduce-master dall'elenco.
5. Scegliere la scheda Inbound rules (Regole in entrata), quindi scegliere Edit inbound rules (Modifica le regole in entrata).
6. Verifica la presenza di una regola in entrata che consenta l'accesso pubblico con le seguenti impostazioni. Se esiste, scegli Elimina per rimuoverla.

   - Tipo

     SSH
   - Porta

     22
   - Origine

     Personalizzata 0.0.0.0/0

   > **Warning**
   >
   > Prima di dicembre 2020, il gruppo di sicurezza ElasticMapReduce-master aveva una regola preconfigurata per consentire il traffico in entrata sulla porta 22 da tutte le origini. Questa regola è stata creata per semplificare le connessioni SSH iniziali al nodo primario. Consigliamo vivamente di rimuovere questa regola in entrata e limitare il traffico alle origini affidabili.

7. Scorri fino alla fine dell'elenco di regole e seleziona Aggiungi regola.
8. Per Tipo, seleziona SSH.

   Selezionando SSH si inserisce in automatico TCP per Protocollo e 22 per Intervallo porta.

9. Per origine, seleziona Il mio IP per aggiungere in automatico il tuo indirizzo IP come indirizzo di origine. Puoi anche aggiungere un intervallo di indirizzi IP affidabili del client Custom (Personalizzato), o creare regole aggiuntive per altri client. In molti ambienti di rete, gli indirizzi IP vengono allocati in modo dinamico, perciò, nel futuro, potrebbe essere necessario aggiornare gli indirizzi IP per client affidabili.
10. Seleziona Salva.
11. Facoltativamente, scegli ElasticMapReduce-slave dall'elenco e ripeti le fasi descritte in precedenza per consentire l'accesso SSH dei client ai nodi principali e nodi attività.

Gruppo di sicurezza gestito da Amazon EMR per le istanze principali e attività (sottoreti pubbliche)

Il gruppo di sicurezza gestito predefinito per le istanze principali e attività nelle sottoreti pubbliche ha come Nome gruppo ElasticMapReduce-core. Il gruppo di sicurezza gestito predefinito ha le regole seguenti e Amazon EMR aggiunge le stesse regole se si specifica un gruppo di sicurezza gestito personalizzato.

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Regole in entrata				
Tutte le regole ICMP-IPV4	Tutti	N/D	L'ID del gruppo di sicurezza gestito specificato per le istanze principali e di attività. In altre parole, lo stesso gruppo di sicurezza in cui appare la regola.	Tali regole riflessive consentono il traffico in entrata da qualsiasi istanza associata al gruppo di sicurezza specificato. Utilizzando `ElasticMapReduce-core` predefinito per più cluster consente alle istanze principali e di attività di tali cluster di comunicare tra loro su ICMP o su qualsiasi porta TCP o UDP. Specifica i gruppi di sicurezza gestiti personalizzati per limitare l'accesso tra cluster.
Tutte le regole TCP	TCP	Tutti		
Tutte le regole UDP	UDP	Tutti		

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Tutte le regole ICMP-IPV4	Tutti	N/D	L'ID del gruppo di sicurezza gestito dell'istanza primaria.	Tali regole consentono tutto il traffico ICMP in entrata e il traffico su qualsiasi porta TCP o UDP da tutte le istanze primarie associate al gruppo di sicurezza specificato, anche se le istanze si trovano in cluster diversi.
Tutte le regole TCP	TCP	Tutti		
Tutte le regole UDP	UDP	Tutti		

Gruppo di sicurezza gestito da Amazon EMR per l'istanza primaria (sottoreti private)

Il gruppo di sicurezza gestito predefinito per l'istanza primaria nelle sottoreti private ha come Nome gruppo ElasticMapReduce-Primary-Private. Il gruppo di sicurezza gestito predefinito ha le regole seguenti e Amazon EMR aggiunge le stesse regole se si specifica un gruppo di sicurezza gestito personalizzato.

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Regole in entrata				
Tutte le regole ICMP-IPv4	Tutti	N/D	L'ID del gruppo di sicurezza gestito dell'istanza primaria. In altre parole, lo stesso gruppo di sicurezza	Tali regole riflessive consentono il traffico in entrata da qualsiasi istanza associata al gruppo di sicurezza specificato e raggiungibile dall'interno di una sottorete privata. Utilizzando `ElasticMapReduce-Primary-Private` predefinito per più cluster consente ai nodi principali e di task di tali cluster di

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Tutte le regole TCP	TCP	Tutti	in cui appare la regola.	comunicare tra loro su ICMP o su qualsiasi porta TCP o UDP. Specifica i gruppi di sicurezza gestiti personalizzati per limitare l'accesso tra cluster.
Tutte le regole UDP	UDP	Tutti		
Tutte le regole ICMP-IPV4	Tutti	N/D	L'ID gruppo del gruppo di sicurezza gestito specificato per i nodi principali e di task.	Tali regole consentono il traffico ICMP in entrata e il traffico su qualsiasi porta TCP o UDP da qualsiasi istanza principale e di attività associata al gruppo di sicurezza specificato e raggiungibile dalla sottorete privata, anche se le istanze si trovano in cluster diversi.
Tutte le regole TCP	TCP	Tutti		
Tutte le regole UDP	UDP	Tutti		
HTTPS (8443)	TCP	8443	L'ID del gruppo di sicurezza gestito per l'accesso ai servizi in una sottorete privata.	Questa regola consente al cluster manager di comunicare con il nodo primario.
Regole in uscita				
Tutto il traffico	Tutti	Tutti	0.0.0.0/0	Fornisce l'accesso in uscita a Internet.

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
TCP personalizzato	TCP	9443	L'ID del gruppo di sicurezza gestito per l'accesso ai servizi in una sottorete privata.	Tieni presente che la regola in uscita predefinita "Tutto il traffico" sopra descritta è un requisito minimo per Amazon EMR 5.30.0 e versioni successive. **Note** Amazon EMR non aggiunge questa regola quando si utilizza un gruppo di sicurezza gestito personalizzato.
TCP personalizzato	TCP	80 (http) o 443 (https)	L'ID del gruppo di sicurezza gestito per l'accesso ai servizi in una sottorete privata.	Se la regola in uscita predefinita "Tutto il traffico" sopra descritta viene rimossa, tieni presente che si tratta di un requisito minimo per Amazon EMR 5.30.0 e versioni successive e per la connessione ad Amazon S3 tramite https. **Note** Amazon EMR non aggiunge questa regola quando si utilizza un gruppo di sicurezza gestito personalizzato.

Gruppo di sicurezza gestito da Amazon EMR per le istanze principali e attività (sottoreti private)

Il gruppo di sicurezza gestito predefinito per le istanze principali e attività nelle sottoreti private ha come Nome gruppo ElasticMapReduce-Core-Private. Il gruppo di sicurezza gestito predefinito ha le regole seguenti e Amazon EMR aggiunge le stesse regole se si specifica un gruppo di sicurezza gestito personalizzato.

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Regole in entrata				
Tutte le regole ICMP-IPV4	Tutti	N/D	L'ID del gruppo di sicurezza gestito specificato per le istanze principali e di attività. In altre parole, lo stesso gruppo di sicurezza in cui appare la regola.	Tali regole riflessive consentono il traffico in entrata da qualsiasi istanza associata al gruppo di sicurezza specificato. Utilizzando `ElasticMapReduce-core` predefinito per più cluster consente alle istanze principali e di attività di tali cluster di comunicare tra loro su ICMP o su qualsiasi porta TCP o UDP. Specifica i gruppi di sicurezza gestiti personalizzati per limitare l'accesso tra cluster.
Tutte le regole TCP	TCP	Tutti		
Tutte le regole UDP	UDP	Tutti		
Tutte le regole ICMP-IPV4	Tutti	N/D	L'ID del gruppo di sicurezza gestito dell'istanza primaria.	Tali regole consentono tutto il traffico ICMP in entrata e il traffico su qualsiasi porta TCP o UDP da tutte le istanze primarie associate al gruppo di sicurezza specificato, anche se le istanze si trovano in cluster diversi.
Tutte le regole TCP	TCP	Tutti		
Tutte le regole UDP	UDP	Tutti		
HTTPS (8443)	TCP	8443	L'ID del gruppo di sicurezza gestito per l'accesso ai servizi in una sottorete privata.	Questa regola consente al cluster manager di comunicare con i nodi principali e di task.

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Regole in uscita				
Tutto il traffico	Tutti	Tutti	0.0.0.0/0	Consulta Modifica di regole in uscita qui di seguito.
TCP personalizzato	TCP	80 (http) o 443 (https)	L'ID del gruppo di sicurezza gestito per l'accesso ai servizi in una sottorete privata.	Se la regola in uscita predefinita "Tutto il traffico" sopra descritta viene rimossa, tieni presente che si tratta di un requisito minimo per Amazon EMR 5.30.0 e versioni successive per la connessione ad Amazon S3 tramite https. **Note** Amazon EMR non aggiunge questa regola quando si utilizza un gruppo di sicurezza gestito personalizzato.

Modifica di regole in uscita

Per impostazione predefinita, Amazon EMR crea questo gruppo di sicurezza con regole in uscita che consentono tutto il traffico in uscita su protocolli e porte. L'autorizzazione di tutto il traffico in uscita è selezionata perché diverse applicazioni Amazon EMR e cliente eseguibili su cluster Amazon EMR potrebbero richiedere regole di uscita diverse. Amazon EMR non è in grado di prevedere queste impostazioni specifiche durante la creazione di gruppi di sicurezza predefiniti. È possibile definire l'ambito in uscita nei gruppi di sicurezza in modo da includere solo le regole che si adattano ai casi d'uso e alle policy di sicurezza. Questo gruppo di sicurezza richiede almeno le seguenti regole in uscita, ma alcune applicazioni potrebbero richiedere un'uscita aggiuntiva.

Tipo	Protocollo	Intervallo porte	Destinazione	Informazioni
Tutte le regole TCP	TCP	Tutti	pl-*xxxxxxxx*	Elenco dei prefissi Amazon S3 gestiti `com.amazonaws.` *MyRegion*`.s3`.
All Traffic	Tutti	Tutti	sg-*xxxxxxxxx xxxxxxxx*	L'ID del gruppo di sicurezza `ElasticMapReduce-Core-Private` .
All Traffic	Tutti	Tutti	sg-*xxxxxxxxx xxxxxxxx*	L'ID del gruppo di sicurezza `ElasticMapReduce-Primary-Private` .
TCP personalizzato	TCP	9443	sg-*xxxxxxxxx xxxxxxxx*	L'ID del gruppo di sicurezza `ElasticMapReduce-ServiceAccess` .

Gruppo di sicurezza gestito da Amazon EMR per l'accesso ai servizi (sottoreti private)

Il gruppo di sicurezza gestito predefinito per l'accesso ai servizi nelle sottoreti private ha come Group Name (Nome gruppo) ElasticMapReduce-ServiceAccess. Dispone di regole in entrata e di regole in uscita che consentono il traffico su HTTPS (porta 8443, porta 9443) per altri i gruppi di sicurezza gestiti nelle sottoreti private. Tali regole consentono al cluster manager di comunicare con il nodo primario e con i nodi principale e attività. Le stesse regole sono necessarie se si utilizzano gruppi di sicurezza personalizzati.

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
Regole in ingresso richieste per i cluster Amazon EMR con Amazon EMR versione 5.30.0 e successive.				
TCP personalizzato	TCP	9443	L'ID del gruppo di sicurezza	

Tipo	Protocollo	Intervallo porte	Origine	Informazioni
			gestito dell'istanza primaria.	Questa regola consente la comunicazione tra il gruppo di sicurezza dell'istanza primaria e il gruppo di sicurezza di accesso al servizio.
Regole in uscita richieste per tutti i cluster Amazon EMR				
TCP personalizzato	TCP	8443	L'ID del gruppo di sicurezza gestito dell'istanza primaria.	Tali regole consentono al cluster manager di comunicare con il nodo primario e con i nodi principale e attività.
TCP personalizzato	TCP	8443	L'ID del gruppo di sicurezza gestito specificato per le istanze principali e di attività.	Tali regole consentono al cluster manager di comunicare con il nodo primario e con i nodi principale e attività.

Utilizzo di gruppi di sicurezza aggiuntivi

Sia se si utilizzano i gruppi di sicurezza gestiti predefiniti o se si specificano gruppi di sicurezza gestiti personalizzati, è possibile utilizzare i gruppi di sicurezza aggiuntivi. I gruppi di sicurezza aggiuntivi offrono la flessibilità necessaria per definire l'accesso tra diversi cluster e dai client, risorse e applicazioni esterne.

Si consideri lo scenario riportato di seguito come un esempio: Disponi di più cluster che devono comunicare reciprocamente ma desideri consentire l'accesso SSH in entrata all'istanza primaria solo a un determinato sottoinsieme di cluster. Per eseguire questa operazione, puoi utilizzare lo stesso set di gruppi di sicurezza gestiti per i cluster. Quindi crei gruppi di sicurezza aggiuntivi che consentono l'accesso SSH in entrata da client affidabili e specifichi i gruppi di sicurezza aggiuntivi per l'istanza primaria per ogni cluster nel sottoinsieme.

È possibile applicare fino a quattro gruppi di sicurezza aggiuntivi per l'istanza primaria , quattro per le istanze principali e attività e quattro per l'accesso ai servizi (in sottoreti private). Se necessario, puoi specificare lo stesso gruppo di sicurezza aggiuntivo per le istanze primarie, principali, attività e per

l'accesso ai servizi. Il numero massimo di gruppi di sicurezza e di regole nell'account è soggetto ai limiti dell'account. Per ulteriori informazioni, consulta Limiti del gruppo di sicurezza nella Guida per l'utente di Amazon VPC.

Specifica dei gruppi di sicurezza gestiti da Amazon EMR e aggiuntivi

È possibile specificare i gruppi di sicurezza utilizzando la AWS Management Console, l'AWS CLI, oppure l'API di Amazon EMR. Se non specifichi i gruppi di sicurezza, Amazon EMR crea i gruppi di sicurezza predefiniti. Specificare i gruppi di sicurezza aggiuntivi è facoltativo. Puoi assegnare i gruppi di sicurezza aggiuntivi per le istanze primarie, principali, attività e per l'accesso ai servizi (solo sottoreti private).

New console

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

Specifica dei gruppi di sicurezza con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. In Networking (Reti), seleziona la freccia accanto ai EC2 security groups (firewall) (Gruppi di sicurezza EC2 [firewall]) per espandere questa sezione. In Primary node (Nodo primario) e Core and task nodes (Nodi core e attività), i gruppi di sicurezza gestiti da Amazon EMR predefiniti sono selezionati per impostazione predefinita. Se utilizzi una sottorete privata, hai la possibilità di selezionare un gruppo di sicurezza anche per Service access (Accesso al servizio).
4. Per modificare il gruppo di sicurezza gestito da Amazon EMR, utilizza il menu a discesa Choose security groups (Scegli gruppi di sicurezza) per selezionare un'opzione diversa dall'elenco di opzioni Amazon EMR-managed security group (Gruppo di sicurezza gestito da Amazon EMR). Hai un gruppo di sicurezza gestito da Amazon EMR sia per Primary node (Nodo primario) sia per Core and task nodes (Nodi core e attività).

5. Per aggiungere gruppi di sicurezza personalizzati, utilizza lo stesso menu a discesa Choose security groups (Scegli gruppi di sicurezza) per selezionare fino a quattro gruppi di sicurezza personalizzati dall'elenco di opzioni Custom security group (Gruppi di sicurezza personalizzati). È possibile avere fino a quattro gruppi di sicurezza personalizzati sia per Primary node (Nodo primario) sia per Core and task nodes (Nodi core e attività).
6. Scegli qualsiasi altra opzione applicabile al cluster.
7. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Specifica dei gruppi di sicurezza con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Seleziona Create cluster (Crea cluster), Go to advanced options (Vai alle opzioni avanzate).
3. Scegli le opzioni del cluster fino a Fase 4: sicurezza.
4. Scegli EC2 Security Groups (Gruppi di sicurezza EC2) per espandere la sezione.

   In EMR managed security groups (Gruppi di sicurezza gestiti EMR), i gruppi di sicurezza gestiti predefiniti sono selezionati per impostazione predefinita. Se un'impostazione di default non esiste nel VPC per Master, Core & Task (Principale e Attività) o per Service Access (Accesso ai servizi) (solo sottorete privata), Create (Crea) viene visualizzato prima del nome del gruppo di sicurezza associato.
5. Se utilizzi i gruppi di sicurezza gestiti personalizzati, selezionali dagli elenchi EMR managed security groups (Gruppi di sicurezza gestiti EMR).

   Se selezioni un gruppo di sicurezza gestito personalizzato, viene inviato un messaggio che segnala di selezionare un gruppo di sicurezza personalizzato per le altre istanze. Per un cluster è possibile usare solo gruppi di sicurezza gestiti predefiniti o solo gruppi di sicurezza gestiti personalizzati.
6. Facoltativamente in Additional security groups (Gruppi di sicurezza aggiuntivi), scegli l'icona a forma di matita, seleziona fino a quattro gruppi di sicurezza dall'elenco e quindi scegli Assign security groups (Assegna gruppi di sicurezza). Ripetere l'operazione per ogni Master, Core & Task (Core e Task) e Service Access (Accesso ai servizi) come desiderato.
7. Scegli Crea cluster.

Specifica dei gruppi di sicurezza con la AWS CLI

Per specificare i gruppi di sicurezza utilizzando la AWS CLI viene usato il comando `create-cluster` con i seguenti parametri dell'opzione `--ec2-attributes`:

Parametro	Descrizione
`EmrManagedPrimarySecurityGroup`	Utilizza questo parametro per specificare un gruppo di sicurezza gestito personalizzato per l'istanza primaria. Se è specificato questo parametro, è necessario specificare anche `EmrManagedCoreSecurityGroup` . Per i cluster in sottoreti private deve essere specificato anche `ServiceAccessSecurityGroup` .
`EmrManagedCoreSecurityGroup`	Utilizza questo parametro per specificare un gruppo di sicurezza gestito personalizzato per le istanze principali e di attività. Se è specificato questo parametro, è necessario specificare anche `EmrManagedPrimarySecurityGroup` . Per i cluster in sottoreti private deve essere specificato anche `ServiceAccessSecurityGroup` .
`ServiceAccessSecurityGroup`	Utilizza questo parametro per specificare un gruppo di sicurezza gestito personalizzato per l'accesso ai servizi, che si applica solo ai cluster nelle sottoreti private. Il gruppo di sicurezza specificato come `ServiceAccessSecurityGroup` non deve essere utilizzato per alcun altro scopo e deve essere riservato anche ad Amazon EMR. Se è specificato questo parametro, è necessario specificare anche `EmrManagedPrimarySecurityGroup` .

Parametro	Descrizione
`AdditionalPrimarySecurityGroups`	Utilizza questo parametro per specificare fino a quattro gruppi di sicurezza aggiuntivi per l'istanza primaria.
`AdditionalCoreSecurityGroups`	Utilizza questo parametro per specificare fino a quattro gruppi di sicurezza aggiuntivi per le istanze principali e di attività.

Example Specifica di gruppi di sicurezza gestiti da Amazon EMR e gruppi di sicurezza aggiuntivi personalizzati

L'esempio seguente specifica gruppi di sicurezza gestiti da Amazon EMR personalizzati per un cluster in una sottorete privata, più gruppi di sicurezza aggiuntivi per l'istanza primaria e un singolo gruppo di sicurezza aggiuntivo per le istanze principale e attività.

> **Note**
>
> I caratteri di continuazione della riga Linux (\) sono inclusi per questioni di leggibilità. Possono essere rimossi o utilizzati nei comandi Linux. Per Windows, rimuoverli o sostituirli con un accento circonflesso (^).

```
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-xxxxxxxxxxxx,\
EmrManagedCoreSecurityGroup=sg-xxxxxxxxxxx,\
AdditionalPrimarySecurityGroups=['sg-xxxxxxxxxxx',\
'sg-xxxxxxxxxxx','sg-xxxxxxxxxx'],\
AdditionalCoreSecurityGroups=sg-xxxxxxxxxxx \
--instance-type m5.xlarge
```

Per ulteriori informazioni, consulta create-cluster nella Guida di riferimento ai comandi della AWS CLI.

Specifica dei gruppi di sicurezza EC2 per EMR Notebooks

Al momento della creazione di un notebook EMR, per controllare il traffico di rete tra il notebook EMR e il cluster Amazon EMR vengono utilizzati due gruppi di sicurezza quando si usa l'editor del notebook. I gruppi di sicurezza predefiniti dispongono di regole minime che consentono solo il traffico di rete tra il servizio EMR Notebooks e i cluster a cui sono collegati i notebook.

Un notebook EMR utilizza Apache Livy per comunicare con il cluster tramite un proxy utilizzando la porta TCP 18888. Quando si creano gruppi di sicurezza personalizzati con regole su misura per l'ambiente, è possibile limitare il traffico di rete in modo che solo un sottoinsieme di notebook sia in grado di eseguire il codice all'interno dell'editor del notebook su determinati cluster. Il cluster utilizza la sicurezza personalizzata in aggiunta ai gruppi di sicurezza predefiniti per il cluster. Per ulteriori informazioni, consulta Controllo del traffico di rete con i gruppi di sicurezza nella Guida alla gestione di Amazon EMR e Specifica dei gruppi di sicurezza EC2 per EMR Notebooks.

Gruppo di sicurezza EC2 predefinito per l'istanza primaria

Il gruppo di sicurezza EC2 predefinito per l'istanza primaria è associato all'istanza primaria in aggiunta ai gruppi di sicurezza del cluster dell'istanza primaria.

Nome gruppo: ElasticMapReduceEditors-Livy

Regole

- In entrata

  Consenti la porta TCP 18888 da qualsiasi risorsa del gruppo di sicurezza predefinito EC2 per EMR Notebooks.
- In uscita

  Nessuno

Gruppi di sicurezza EC2 predefiniti per EMR Notebooks

Il gruppo di sicurezza EC2 predefinito per il notebook EMR è associato all'editor del notebook per qualsiasi notebook EMR a cui è assegnato.

Nome gruppo: ElasticMapReduceEditors-Editor

Regole

- In entrata

  Nessuno

- In uscita

  Consenti la porta TCP 18888 a qualsiasi risorsa del gruppo di sicurezza predefinito EC2 per EMR Notebooks.

Gruppo di sicurezza EC2 personalizzato per EMR Notebooks quando si associano notebook con repository Git

Per collegare un repository Git al notebook, il gruppo di sicurezza per il notebook EMR deve includere una regola in uscita per consentire al notebook di instradare il traffico a Internet. Si consiglia di creare un nuovo gruppo di sicurezza per questo scopo. L'aggiornamento del gruppo di sicurezza predefinito ElasticMapReduceEditors-editor può fornire le stesse regole in uscita ad altri notebook collegati a questo gruppo di sicurezza.

Regole

- In entrata

  Nessuno

- In uscita

  Consenti al notebook di instradare il traffico a Internet tramite il cluster, come illustrato nell'esempio seguente: Ad esempio, viene utilizzato il valore 0.0.0.0/0. È possibile modificare questa regola per specificare gli indirizzi IP dei repository basati su Git.

Tipo	Protocollo	Intervallo porte	Destinazione
Regola TCP personalizzata	TCP	18888	SG-
HTTPS	TCP	443	0.0.0.0/0

Utilizzo del blocco dell'accesso pubblico di Amazon EMR

Il blocco dell'accesso pubblico (BPA) di Amazon EMR impedisce l'avvio di un cluster in una sottorete pubblica se il cluster dispone di una configurazione di sicurezza che consente il traffico in entrata da indirizzi IP pubblici su una porta.

> **Important**
>
> Il blocco dell'accesso pubblico è abilitato per impostazione predefinita. Per aumentare la protezione degli account, ti consigliamo di mantenerlo abilitato.

Informazioni sul blocco dell'accesso pubblico

Puoi utilizzare la configurazione a livello di account del blocco dell'accesso pubblico per gestire a livello centrale l'accesso della rete pubblica ai cluster Amazon EMR.

Quando un utente dal tuo Account AWS avvia un cluster, Amazon EMR controlla le regole delle porte nel gruppo di sicurezza per il cluster e le confronta con le regole del traffico in entrata. Se per il gruppo di sicurezza è impostata una regola in entrata che consente l'apertura delle porte agli indirizzi IP pubblici IPv4 0.0.0.0/0 o IPv6: :/0 e tali porte non sono specificate come eccezioni per il tuo account, Amazon EMR non consente all'utente di creare il cluster.

Se un utente modifica le regole del gruppo di sicurezza per un cluster in esecuzione in una sottorete pubblica in modo da avere una regola di accesso pubblico che viola la configurazione BPA del tuo account, Amazon EMR revoca la nuova regola se dispone dell'autorizzazione per farlo. Se Amazon EMR non dispone dell'autorizzazione per revocare la regola, crea un evento nel pannello di controllo di AWS Health che descrive la violazione. Per concedere l'autorizzazione alla revoca della regola ad Amazon EMR, consulta Configura Amazon EMR per revocare le regole dei gruppi di sicurezza.

Il blocco dell'accesso pubblico è abilitato per impostazione predefinita per tutti i cluster di ogni Regione AWS per il tuo Account AWS. Il BPA si applica all'intero ciclo di vita di un cluster, ma non si applica ai cluster creati in sottoreti private. È possibile configurare eccezioni alla regola BPA; la porta 22 è un'eccezione per impostazione predefinita. Per ulteriori informazioni sull'impostazione delle eccezioni, consulta Configurazione del blocco degli accessi pubblici.

Configurazione del blocco degli accessi pubblici

È possibile aggiornare i gruppi di sicurezza e la configurazione del blocco dell'accesso pubblico negli account in qualsiasi momento.

Puoi attivare e disattivare le impostazioni del blocco dell'accesso pubblico (BPA) con la AWS Management Console, la AWS Command Line Interface (AWS CLI) e l'API Amazon EMR. Le impostazioni si applicano all'account in base alla regione. Per preservare la sicurezza del cluster, si consiglia di mantenere abilitato il blocco dell'accesso pubblico.

New console

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

Configurazione del blocco dell'accesso pubblico con la nuova console

1. Accedi alla AWS Management Console, quindi apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. Nella barra di navigazione in alto, seleziona la Regione che desideri configurare, se non è già selezionata.
3. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Block public access (Blocca accesso pubblico).
4. In Block public access settings (Impostazioni blocco accesso pubblico) completare la procedura seguente.

A…	Esegui questa operazione…
Attivare o disattivare il blocco degli accessi pubblici	Scegli Edit (Modifica), scegli Turn on (Attiva) oppure Turn off (Disattiva) seconda dei casi, quindi scegli Save (Salva).
Modificare le porte nell'elenco delle eccezioni	1. Scegli Edit (Modifica) e cerca la sezione Port range exceptions (Eccezioni dell'intervallo di porte). 2.

A…	Esegui questa operazione…
	Per aggiungere porte all'elenco delle eccezioni, scegliere Add a port range (Aggiungi un intervallo di porte) e immettere una nuova porta o un nuovo intervallo di porte. Ripetere l'operazione per ogni porta o intervallo di porte da aggiungere. 3. Per rimuovere una porta o un intervall o di porte, scegli Remove (Rimuovi) accanto alla voce nell'elenco degli intervalli di porte. 4. Seleziona Salva.

Old console

Configurazione del blocco dell'accesso pubblico con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. Nella barra di navigazione in alto, verifica che sia selezionata la Regione che desideri configurare.
3. Scegliere Block public access (Blocca accesso pubblico).
4. In Block public access settings (Impostazioni blocco accesso pubblico) completare la procedura seguente.

A…	Esegui questa operazione…
Attivare o disattivare il blocco degli accessi pubblici	Scegliere Modifica, scegliere On (Attivato) oppure Off (Disattivato), quindi scegliere il segno di spunta per confermare.

A…	Esegui questa operazione…
Modificare le porte nell'elenco delle eccezioni	1. In Exceptions (Eccezioni) scegliere Modifica. 2. Per aggiungere porte all'elenco delle eccezioni, scegliere Add a port range (Aggiungi un intervallo di porte) e immettere una nuova porta o un nuovo intervallo di porte. Ripetere l'operazione per ogni porta o intervallo di porte da aggiungere. 3. Per rimuovere una porta o un intervall o di porte, scegliere la x accanto alla voce nell'elenco Port ranges (Intervalli di porte). 4. Seleziona Salva modifiche.

AWS CLI

Per configurare il blocco degli accessi pubblici tramite la AWS CLI

- Utilizzare il comando `aws emr put-block-public-access-configuration` per configurare il blocco degli accessi pubblici come mostrato negli esempi seguenti.

A…	Esegui questa operazione…
Attivare il blocco degli accessi pubblici	Impostare `BlockPublicSecurityGroupRules` su `true` come mostrato nell'esempio seguente. Per consentire l'avvio del cluster, nessun gruppo di sicurezza associato a un cluster può avere una regola in entrata che consente l'accesso pubblico. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=true`
Disattivare il blocco degli accessi pubblici	Impostare `BlockPublicSecurityGroupRules` su `false` come mostrato nell'esempio seguente. I gruppi di sicurezza associati a un cluster possono avere regole in entrata che consentono l'accesso pubblico su qualsiasi porta. Questa configurazione non è consigliata. `aws emr put-block-public-access-configuration --block-public-access-configuration BlockPublicSecurityGroupRules=false`

A…	Esegui questa operazione…
Attivare il blocco degli accessi pubblici e specificare le porte come eccezioni	L'esempio seguente attiva il blocco degli accessi pubblici e specifica la porta 22 e le porte 100-101 come eccezioni. Ciò consente la creazione di cluster se per un gruppo di sicurezza associato è specifica ta una regola in entrata che consente l'accesso pubblico sulla porta 22, sulla porta 100 o sulla porta 101. `aws emr put-block-public-access-configuration --block-public-acc ess-configuration '{ "BlockPub licSecurityGroupRules": true, "PermittedPublicSecurityGro upRuleRanges": [ { "MinRange": 22, "MaxRange": 22 }, { "MinRange ": 100, "MaxRange": 101 } ] }'`

Configura Amazon EMR per revocare le regole dei gruppi di sicurezza

Amazon EMR necessita dell'autorizzazione per revocare le regole dei gruppi di sicurezza e conformarsi alla configurazione del blocco dell'accesso pubblico. Puoi utilizzare uno dei seguenti approcci per concedere ad Amazon EMR l'autorizzazione necessaria:

- Consigliato Collega la policy gestita da `AmazonEMRServicePolicy_v2` al ruolo di servizio. Per ulteriori informazioni, consulta Ruolo di servizio per Amazon EMR (ruolo EMR).
- Crea una nuova policy inline che consenta l'operazione `ec2:RevokeSecurityGroupIngress` sui gruppi di sicurezza. Per ulteriori informazioni su come modificare una policy di autorizzazione dei ruoli, consulta Modifica di una policy di autorizzazioni dei ruoli con la console IAM, l'API AWS e la AWS CLI nella Guida per l'utente IAM.

Risoluzione delle violazioni dell'accesso pubblico

Se si verifica una violazione del blocco dell'accesso pubblico, puoi mitigarla con una delle seguenti azioni:

- Se desideri accedere a un'interfaccia Web sul tuo cluster, usa una delle opzioni descritte in Visualizzazione di interfacce Web ospitate su cluster Amazon EMR per accedere all'interfaccia tramite SSH (porta 22).
- Per consentire il traffico verso il cluster da indirizzi IP specifici anziché dall'indirizzo IP pubblico, aggiungi una regola del gruppo di sicurezza. Per ulteriori informazioni, consulta Aggiunta di regole a un gruppo di sicurezza nella Guida alle operazioni di base di Amazon EC2.
- (Non consigliato) Puoi configurare le eccezioni BPA di Amazon EMR per includere la porta o l'intervallo di porte desiderati. Quando specifichi un'eccezione BPA, introduci un rischio con una porta non protetta. Se intendi specificare un'eccezione, devi rimuoverla appena non è più necessaria. Per ulteriori informazioni, consulta Configurazione del blocco degli accessi pubblici.

Identificazione dei cluster associati alle regole dei gruppi di sicurezza

Potrebbe essere necessario identificare tutti i cluster associati a una determinata regola dei gruppi di sicurezza o trovare la regola dei gruppi di sicurezza per un determinato cluster.

- Se conosci il gruppo di sicurezza, puoi identificare i cluster associati, se trovi le interfacce di rete per tale gruppo di sicurezza. Per ulteriori informazioni, consulta Come posso trovare le risorse associate a un gruppo di sicurezza di Amazon EC2? su AWS re:Post. Le istanze Amazon EC2 collegate a queste interfacce di rete verranno contrassegnate con l'ID del cluster a cui appartengono.
- Se desideri trovare i gruppi di sicurezza per un cluster noto, segui i passaggi riportati in Visualizzazione dello stato e dei dettagli di un cluster. Puoi trovare i gruppi di sicurezza per il cluster nel riquadro Rete e sicurezza della console o nel campo `Ec2InstanceAttributes` della AWS CLI.

Convalida della conformità per Amazon EMR

Revisori di terze parti valutano la sicurezza e la conformità di Amazon EMR come parte di più programmi di conformità di AWS. Questi includono SOC, PCI, FedRAMP, HIPAA e altri.

Per un elenco di servizi AWS che rientrano nell'ambito di programmi di conformità specifici, consultare Servizi AWS coperti dal programma di compliance. Per informazioni generali, consultare Programmi per la conformità di AWS.

Puoi scaricare i report di audit di terze parti utilizzando AWS Artifact. Per ulteriori informazioni, consulta l'argomento Download dei rapporti in AWS Artifact.

La tua responsabilità di conformità durante l'utilizzo di Amazon EMR è determinata dalla riservatezza dei dati, dagli obiettivi di conformità dell'azienda e dalle leggi e normative applicabili. Se l'utilizzo di Amazon EMR è soggetto a conformità con standard quali HIPAA, PCI o FedRAMP, AWS fornisce risorse utili:

- Guide Quick Start per la sicurezza e conformità: queste guide all'implementazione illustrano considerazioni relative all'architettura e forniscono fasi per l'implementazione di ambienti di base incentrati sulla sicurezza e sulla conformità su AWS.
- Architecting for HIPAA Security and Compliance Whitepaper: questo whitepaper descrive in che modo le aziende possono utilizzare AWS per creare applicazioni conformi ai requisiti HIPAA.
- Risorse per la conformità di AWS: questa raccolta di workbook e guide potrebbe essere utile al tuo settore e alla tua posizione.
- AWS Config: questo servizio AWS valuta il livello di conformità delle configurazioni delle risorse con pratiche interne, linee guida e regolamenti di settore.
- AWS Security Hub: questo servizio AWS fornisce una visione completa dello stato di sicurezza all'interno di AWS che consente di verificare la conformità con gli standard e le best practice di sicurezza del settore.

Resilienza in Amazon EMR

L'infrastruttura globale di AWS è basata su Regioni e zone di disponibilità AWS. AWS Le Regioni forniscono più zone di disponibilità fisicamente separate e isolate che sono connesse tramite reti altamente ridondanti, a bassa latenza e velocità effettiva elevata. Con le zone di disponibilità, è possibile progettare e gestire le applicazioni e database che eseguono il failover automatico tra zone di disponibilità senza interruzioni. Le zone di disponibilità sono più disponibili, tolleranti ai guasti e scalabili rispetto alle infrastrutture a data center singolo o multiplo.

Per ulteriori informazioni sulle Regioni e le Zone di disponibilità AWS, consulta Infrastruttura globale di AWS.

Oltre all'infrastruttura globale AWS, Amazon EMR offre numerose funzionalità per supportare la resilienza dei dati e le esigenze di backup.

- Integrazione con Amazon S3 tramite EMRFS
- Supporto per più nodi master

Sicurezza dell'infrastruttura in Amazon EMR

In qualità di servizio gestito, Amazon EMR è protetto dalle procedure di sicurezza di rete globale AWS descritte nel whitepaper Amazon Web Services: Panoramica dei processi di sicurezza.

Utilizza le chiamate API pubblicate di AWS per accedere ad Amazon EMR tramite la rete. I client devono supportare Transport Layer Security (TLS) 1.2. I client devono, inoltre, supportare le suite di cifratura con PFS (Perfect Forward Secrecy), ad esempio Ephemeral Diffie-Hellman (DHE) o Elliptic Curve Ephemeral Diffie-Hellman (ECDHE). La maggior parte dei sistemi moderni, come Java 7 e versioni successive, supporta tali modalità.

Inoltre, le richieste devono essere firmate utilizzando un ID chiave di accesso e una chiave di accesso segreta associata a un principale IAM. In alternativa, è possibile utilizzare AWS Security Token Service (AWS STS) per generare le credenziali di sicurezza temporanee per sottoscrivere le richieste.

Argomenti

- Connessione ad Amazon EMR utilizzando un endpoint VPC di interfaccia

Connessione ad Amazon EMR utilizzando un endpoint VPC di interfaccia

Puoi connetterti direttamente ad Amazon EMR utilizzando un endpoint VPC di interfaccia (AWS PrivateLink) nel tuo cloud privato virtuale (VPC, Virtual Private Cloud) invece di connetterti tramite Internet. Quando utilizzi un endpoint VPC di interfaccia, la comunicazione tra il VPC e Amazon EMR avviene interamente all'interno della rete AWS. Ogni endpoint VPC è rappresentato da una o più interfacce di rete elastiche (ENI) con indirizzi IP privati nelle sottoreti del VPC.

L'endpoint VPC di interfaccia connette il tuo VPC direttamente ad Amazon EMR senza alcun Internet gateway, dispositivo NAT, connessione VPN o connessione AWS Direct Connect. Le istanze presenti nel tuo VPC non richiedono indirizzi IP pubblici per comunicare con l'API di Amazon EMR.

Per utilizzare Amazon EMR tramite il VPC, devi connetterti da un'istanza che si trova all'interno del VPC o connettere la rete privata al VPC utilizzando Amazon Virtual Private Network (VPN) o AWS Direct Connect. Per informazioni su Amazon VPN, consulta Connessioni VPN nella Guida per l'utente di Amazon Virtual Private Cloud. Per informazioni su AWS Direct Connect, consulta Creazione di una connessione nella Guida per l'utente di AWS Direct Connect.

Puoi creare un endpoint VPC di interfaccia per connetterti ad Amazon EMR utilizzando la console AWS o i comandi della AWS Command Line Interface (AWS CLI). Per ulteriori informazioni, consulta Creazione di un endpoint di interfaccia.

Se dopo aver creato un endpoint VPC di interfaccia abiliti nomi host DNS privati per l'endpoint, l'endpoint Amazon EMR predefinito restituisce il tuo endpoint VPC. L'endpoint del nome del servizio predefinito per Amazon EMR è nel formato seguente.

```
elasticmapreduce.Region.amazonaws.com
```

Se non abiliti nomi host DNS privati, Amazon VPC fornisce un nome di endpoint DNS che puoi utilizzare nel formato seguente:

```
VPC_Endpoint_ID.elasticmapreduce.Region.vpce.amazonaws.com
```

Per ulteriori informazioni, consulta Endpoint VPC di interfaccia (AWS PrivateLink) nella Guida per l'utente di Amazon VPC.

Amazon EMR supporta l'esecuzione di chiamate a tutte le sue operazioni API all'interno del VPC.

Puoi collegare le policy di endpoint VPC a un endpoint VPC per controllare l'accesso per le entità principali IAM. Puoi inoltre associare i gruppi di sicurezza a un endpoint VPC per controllare l'accesso in ingresso e in uscita in base all'origine e alla destinazione del traffico di rete, ad esempio un intervallo di indirizzi IP. Per ulteriori informazioni, consulta Controllo dell'accesso ai servizi con endpoint VPC.

Creazione di una policy di endpoint VPC per Amazon EMR

Puoi creare una policy per gli endpoint di Amazon VPC per Amazon EMR per specificare quanto segue:

- Il principale che può o non può eseguire azioni
- Le azioni che possono essere eseguite

- Le risorse sui cui si possono eseguire le azioni

Per ulteriori informazioni, consultare Controllo degli accessi ai servizi con endpoint VPC in Guida per l'utente di Amazon VPC.

Example - Policy di endpoint VPC per negare tutti gli accessi da un account AWS specificato

La seguente policy di endpoint VPC nega all'account AWS *123456789012* ogni accesso alle risorse utilizzando l'endpoint.

```
{
    "Statement": [
        {
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": "*"
        },
        {
            "Action": "*",
            "Effect": "Deny",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "123456789012"
                ]
            }
        }
    ]
}
```

Example - Policy di endpoint VPC per consentire l'accesso VPC solo a un'entità principale IAM (utente) specificata

La seguente policy di endpoint VPC consente l'accesso completo solo a un utente *lijuan* nell'account AWS *123456789012*. A tutte le altre entità principali IAM viene negato l'accesso utilizzando l'endpoint.

```
{
    "Statement": [
        {
```

```
            "Action": "*",
            "Effect": "Allow",
            "Resource": "*",
            "Principal": {
                "AWS": [
                    "arn:aws:iam::123456789012:user/lijuan"
                ]
            }
        }]
}
```

Example - Policy di endpoint VPC per consentire operazioni EMR di sola lettura

La seguente policy di endpoint VPC consente solo all'account AWS *123456789012* di eseguire le operazioni Amazon EMR specificate.

Le operazioni specificate forniscono l'accesso di sola lettura equivalente per Amazon EMR. Tutte le altre azioni sul VPC vengono negate per l'account specificato. A tutti gli altri account viene negato l'accesso. Per un elenco delle operazioni Amazon EMR, consulta Operazioni, risorse e chiavi di condizione per 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 - Policy di endpoint VPC per negare l'accesso a un cluster specificato

La seguente policy di endpoint VPC consente l'accesso completo per tutti gli account e le entità principali, ma nega l'accesso per l'account AWS *123456789012* alle operazioni eseguite nel cluster Amazon EMR con ID cluster *j-A1B2CD34EF5G*. Altre operazioni Amazon EMR che non supportano le autorizzazioni a livello di risorsa per i cluster sono comunque consentite. Per l'elenco delle operazioni Amazon EMR e dei tipi di risorse corrispondenti, consulta Operazioni, risorse e chiavi di condizione per 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"
                ]
            }
        }
    ]
}
```

Gestione di cluster

Dopo aver avviato il cluster, puoi monitorarlo e gestirlo. Amazon EMR fornisce vari strumenti che ti consentono di connetterti al cluster e di controllarlo.

Argomenti

Connessione a un cluster

Quando si esegue un cluster Amazon EMR, spesso è sufficiente eseguire un'applicazione per analizzare i dati e raccogliere l'output da un bucket Amazon S3. In altri casi, potresti voler interagire con il nodo primario mentre il cluster è in esecuzione. Ad esempio, potresti volerti connettere al nodo primario per eseguire query interattive, controllare i file di log, eseguire il debug di un problema con il cluster, monitorare le prestazioni utilizzando un'applicazione come Ganglia in esecuzione sul nodo primario e così via. Le sezioni seguenti descrivono le tecniche che è possibile utilizzare per connettersi al nodo primario.

In un cluster EMR, il nodo primario è un'istanza Amazon EC2 che coordina le istanze EC2 in esecuzione come nodi attività e core. Il nodo primario espone un nome DNS pubblico che è possibile utilizzare per connettersi ad esso. Per impostazione predefinita, Amazon EMR crea regole del gruppo di sicurezza per il nodo primario e i nodi core e attività, che determinano la modalità di accesso ai nodi.

> **Note**
>
> È possibile connettersi al nodo primario solo mentre il cluster è in esecuzione. Quando il cluster termina, l'istanza EC2 che funge da nodo primario viene terminata e non è più disponibile. Per connettersi al nodo primario, è necessario anche autenticarsi al cluster. È

possibile utilizzare Kerberos per l'autenticazione, o specificare una coppia di chiavi private Amazon EC2 quando si avvia il cluster. Per ulteriori informazioni sulla configurazione di Kerberos e sulla connessione, vedere Utilizzo di Kerberos per l'autenticazione con Amazon EMR. Quando avvii un cluster dalla console, la coppia di chiavi Amazon EC2 private è specificata nella sezione Security and Access (Sicurezza e accesso) della pagina Create Cluster (Crea cluster).

Per impostazione predefinita, il gruppo di sicurezza ElasticMapReduce-master non consente l'accesso SSH in entrata. Potrebbe essere necessario aggiungere una regola in entrata che consenta l'accesso SSH (porta TCP 22) dalle sorgenti a cui si desidera accedere. Per ulteriori informazioni sulla modifica delle regole del gruppo di sicurezza, consulta Aggiunta di regole a un gruppo di sicurezza nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Important

Non modificare le altre regole del gruppo di sicurezza ElasticMapReduce-master. La modifica di queste regole può interferire con il funzionamento del cluster.

Argomenti

Prima di connetterti: autorizza il traffico in entrata

Prima di effettuare la connessione a un cluster Amazon EMR, devi autorizzare il traffico SSH in entrata (porta 22) da parte di client affidabili come l'indirizzo IP del computer. A tale scopo, modifica le regole del gruppo di sicurezza gestito per i nodi a cui desideri connetterti. Ad esempio, nelle istruzioni seguenti viene illustrato come aggiungere una regola in ingresso per l'accesso SSH al gruppo di sicurezza predefinito ElasticMapReduce-master.

Per ulteriori informazioni sull'utilizzo dei gruppi di sicurezza tramite Amazon EMR, consulta Controllo del traffico di rete con gruppi di sicurezza.

New console

Concessione dell'accesso SSH alle origini attendibili per il gruppo di sicurezza primario con la nuova console

Per modificare i gruppi di sicurezza, devi disporre dell'autorizzazione per gestire i gruppi di sicurezza per il VPC in cui si trova il cluster. Per ulteriori informazioni, consulta Modifica delle autorizzazioni per un utente e la Policy di esempio che consente di gestire i gruppi di sicurezza EC2 nella Guida per l'utente IAM.

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona il cluster da aggiornare. Si apre la pagina dei dettagli del cluster. La scheda Properties (Proprietà) in questa pagina sarà preselezionata.
3. In Networking (Reti) nella scheda Properties (Proprietà), seleziona la freccia accanto a EC2 security groups (firewall) (Gruppi di sicurezza EC2 [firewall]) per espandere questa sezione. In Primary node (Nodo primario), seleziona il collegamento al gruppo di sicurezza. Si apre la console EC2.
4. Seleziona la scheda Inbound rules (Regole in entrata), quindi scegli Edit inbound rules (Modifica regola in entrata).
5. Verifica la presenza di una regola in entrata che consenta l'accesso pubblico con le seguenti impostazioni. Se esiste, scegli Elimina per rimuoverla.

   - Tipo

     SSH
   - Porta

     22
   - Origine

     Personalizzata 0.0.0.0/0

> **Warning**
>
> Prima di dicembre 2020, il gruppo di sicurezza ElasticMapReduce-master aveva una regola preconfigurata per consentire il traffico in entrata sulla porta 22 da tutte le origini. Questa regola è stata creata per semplificare le connessioni SSH iniziali al nodo primario. Consigliamo vivamente di rimuovere questa regola in entrata e limitare il traffico alle origini affidabili.

6. Scorri fino alla fine dell'elenco di regole e seleziona Aggiungi regola.
7. Per Tipo, seleziona SSH. Questa selezione inserisce automaticamente TCP per Protocol (Protocollo) e 22 per Port Range (Intervallo porte).
8. Per origine, seleziona Il mio IP per aggiungere in automatico il tuo indirizzo IP come indirizzo di origine. Puoi anche aggiungere un intervallo di indirizzi IP affidabili del client Custom (Personalizzato), o creare regole aggiuntive per altri client. In molti ambienti di rete, gli indirizzi IP vengono allocati in modo dinamico, perciò, nel futuro, potrebbe essere necessario aggiornare gli indirizzi IP per client affidabili.
9. Seleziona Salva.
10. Facoltativamente, torna al passaggio 3, scegli Core and task nodes (Nodi core e attività) e ripeti i passaggi da 4 a 8. Ciò garantisce l'accesso al client SSH ai nodi core e attività.

Old console

Concessione dell'accesso SSH alle origini attendibili per il gruppo di sicurezza primario con la vecchia console

Per modificare i gruppi di sicurezza, devi disporre dell'autorizzazione per gestire i gruppi di sicurezza per il VPC in cui si trova il cluster. Per ulteriori informazioni, consulta Modifica delle autorizzazioni per un utente e la Policy di esempio che consente di gestire i gruppi di sicurezza EC2 nella Guida per l'utente IAM.

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Cluster. Scegli il Nome del cluster che desideri modificare.
3. Scegli il link Gruppi di sicurezza per master in Sicurezza e accesso.

4. Scegli ElasticMapReduce-master dall'elenco.
5. Scegliere la scheda Inbound rules (Regole in entrata), quindi scegliere Edit inbound rules (Modifica le regole in entrata).
6. Verifica la presenza di una regola in entrata che consenta l'accesso pubblico con le seguenti impostazioni. Se esiste, scegli Elimina per rimuoverla.

   - Tipo

     SSH
   - Porta

     22
   - Origine

     Personalizzata 0.0.0.0/0

   > **Warning**
   >
   > Prima di dicembre 2020, il gruppo di sicurezza ElasticMapReduce-master aveva una regola preconfigurata per consentire il traffico in entrata sulla porta 22 da tutte le origini. Questa regola è stata creata per semplificare le connessioni SSH iniziali al nodo primario. Consigliamo vivamente di rimuovere questa regola in entrata e limitare il traffico alle origini affidabili.

7. Scorri fino alla fine dell'elenco di regole e seleziona Aggiungi regola.
8. Per Tipo, seleziona SSH.

   Selezionando SSH si inserisce in automatico TCP per Protocollo e 22 per Intervallo porta.
9. Per origine, seleziona Il mio IP per aggiungere in automatico il tuo indirizzo IP come indirizzo di origine. Puoi anche aggiungere un intervallo di indirizzi IP affidabili del client Custom (Personalizzato), o creare regole aggiuntive per altri client. In molti ambienti di rete, gli indirizzi IP vengono allocati in modo dinamico, perciò, nel futuro, potrebbe essere necessario aggiornare gli indirizzi IP per client affidabili.
10. Seleziona Salva.
11. Facoltativamente, scegli ElasticMapReduce-slave dall'elenco e ripeti le fasi descritte in precedenza per consentire l'accesso SSH dei client ai nodi principali e nodi attività.

Connessione al nodo primario tramite SSH

Secure Shell (SSH) è un protocollo di rete che può essere utilizzato per creare una connessione sicura a un computer remoto. Dopo aver effettuato il collegamento, il terminale del computer locale si comporta come se fosse in esecuzione sul computer remoto. I comandi emessi localmente vengono eseguiti sul computer remoto e l'output del comando dal computer remoto viene visualizzato nella finestra del terminale.

Quando si utilizza SSH con AWS, ci si connette a un'istanza EC2, che è un server virtuale in esecuzione nel cloud. Quando si lavora con Amazon EMR, l'impiego più comune di SSH è quello di connettersi all'istanza EC2 che sta agendo come nodo primario del cluster.

L'utilizzo di SSH per connettersi al nodo primario consente di monitorare e interagire con il cluster. È possibile eseguire comandi Linux sul nodo primario, eseguire applicazioni come Hive e Pig in modo interattivo, sfogliare le directory, leggere i file di log e così via. Si può anche creare un tunnel nella connessione SSH per visualizzare le interfacce Web ospitate sul nodo primario. Per ulteriori informazioni, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Per connettersi al nodo primario utilizzando SSH, è necessario il nome DNS pubblico del nodo primario. Inoltre, il gruppo di sicurezza associato al nodo primario deve avere una regola in entrata che consenta il traffico SSH (porta TCP 22) da un'origine che include il client da cui viene creata la connessione SSH. Potrebbe essere necessario aggiungere una regola per consentire una connessione SSH dal client. Per ulteriori informazioni sulla modifica delle regole del gruppo di sicurezza, consulta Controllo del traffico di rete con gruppi di sicurezza e Aggiunta di regole a un gruppo di sicurezza nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Recupero del nome DNS pubblico del nodo primario

È possibile recuperare il nome DNS pubblico primario utilizzando la console Amazon EMR e la AWS CLI.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Recupero del nome DNS pubblico del nodo primario con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster del quale vuoi recuperare il nome DNS pubblico.
3. Prendi nota del valore Primary node public DNS (DNS pubblico del nodo primario) che compare nella sezione Summary (Riepilogo) della pagina dei dettagli del cluster.

Old console

Recupero del nome DNS pubblico del nodo primario con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Nella pagina Cluster List (Elenco cluster), selezionare il link per il proprio cluster.
3. Prendi nota del valore Master public DNS (DNS pubblico master) che compare nella sezione Summary (Riepilogo) della pagina Cluster Details (Dettagli cluster).

> **Note**
>
> È inoltre possibile scegliere il collegamento SSH per le istruzioni sulla creazione di una connessione SSH con il nodo primario.

CLI

Recupero del nome DNS pubblico del nodo primario con la AWS CLI

1. Per recuperare l'identificatore del cluster, digitare il seguente comando:

```
aws emr list-clusters
```

L'output elenca i cluster, tra cui gli ID del cluster. Prendere nota dell'ID del cluster per il cluster a cui ci si connette.

```
"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. Per elencare le istanze del cluster, incluso il nome DNS pubblico del cluster, inserisci uno dei seguenti comandi. Sostituisci *j-2AL4XXXXXX5T9* con l'ID cluster restituito dal comando precedente.

```
aws emr list-instances --cluster-id j-2AL4XXXXXX5T9
```

O:

```
aws emr describe-cluster --cluster-id j-2AL4XXXXXX5T9
```

L'output elenca le istanze del cluster, compresi i nomi DNS e gli indirizzi IP. Prendere nota del valore per `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": "###.##.#.###"
```

Per ulteriori informazioni, consulta Comandi di Amazon EMR nella AWS CLI.

Connessione al nodo primario tramite SSH e una chiave privata Amazon EC2 su Linux, Unix e Mac OS X

Per creare una connessione SSH autenticata con un file chiave privato, è necessario specificare la coppia di chiavi Amazon EC2 private all'avvio di un cluster. Per ulteriori informazioni sull'accesso alla coppia di chiavi, consulta Coppia di chiavi Amazon EC2 nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Molto probabilmente il tuo computer Linux include un client SSH per impostazione predefinita. Ad esempio, OpenSSH è installato nella maggior parte dei sistemi operativi Linux, Unix e macOS. È possibile verificare la presenza di un client SSH digitando ssh nella riga di comando. Se il computer non riconosce il comando, installa un client SSH per la connessione al nodo primario. Il progetto OpenSSH fornisce un'implementazione gratuita della suite completa di strumenti SSH. Per ulteriori informazioni, consulta il sito Web OpenSSH.

Le seguenti istruzioni illustrano l'apertura di una connessione SSH al nodo primario Amazon EMR su Linux, Unix e Mac OS X.

Per configurare le autorizzazioni per i file chiave privati della coppia di chiavi

Prima di poter utilizzare la chiave privata della coppia di chiavi Amazon EC2 per creare una connessione SSH, è necessario impostare le autorizzazioni sul file `.pem` in modo che solo il proprietario della chiave abbia il permesso di accedere al file. Questo è necessario per creare una connessione SSH usando il terminale o la AWS CLI.

1. Assicurati di aver consentito il traffico SSH in entrata. Per istruzioni, consultare Prima di connetterti: autorizza il traffico in entrata.
2. Individuare il file `.pem`. Queste istruzioni presuppongono che il file sia denominato `mykeypair.pem` e che venga memorizzato nella directory principale dell'utente corrente.
3. Digitare il seguente comando per impostare le autorizzazioni. Sostituisci *`~/mykeypair.pem`* con il percorso completo e il nome di file della chiave privata della coppia di chiavi. Ad esempio `C:/Users/<username>/.ssh/mykeypair.pem`.

```
chmod 400 ~/mykeypair.pem
```

Se non si impostano le autorizzazioni sul file `.pem`, si riceverà un errore che indica che il file chiave non è protetto e la chiave verrà rifiutata. Per connettersi, è sufficiente impostare le autorizzazioni sul file chiave privata della coppia di chiavi la prima volta che lo si utilizza.

Connessione al nodo primario utilizzando il terminale

1. Apri una finestra del terminale. In Mac OS X, selezionare Applications > Utilities > Terminal (Applicazioni > Utility > Terminale). In altre distribuzioni Linux, la finestra terminal si trova generalmente in Applications > Accessories > Terminal (Applicazioni > Accessori >; Terminale).
2. Per stabilire una connessione al nodo primario, inserisci il seguente comando. Sostituisci *ec2-###-##-##-###.compute-1.amazonaws.com* con il nome DNS pubblico priamario del cluster e sostituisci *~/mykeypair.pem* con il percorso completo e il nome del file `.pem`. Ad esempio `C:/Users/<username>/.ssh/mykeypair.pem`.

   ```
   ssh hadoop@ec2-###-##-##-###.compute-1.amazonaws.com -i ~/mykeypair.pem
   ```

   > **Important**
   >
   > Quando si esegue la connessione al nodo primario di Amazon EMR occorre utilizzare il nome di login `hadoop`, altrimenti potrebbe verificarsi un errore simile a `Server refused our key`.

3. Un avviso indica che non è possibile verificare l'autenticità dell'host a cui ci si connette. Per continuare, digitare `yes`.
4. Una volta terminato il lavoro sul nodo primario, inserisci il seguente comando per chiudere la connessione SSH.

   ```
   exit
   ```

Se riscontri problemi nell'utilizzo di SSH per la connessione al nodo primario, consulta la sezione Troubleshoot connecting to your instance (Risoluzione dei problemi di connessione all'istanza).

Connessione al nodo primario tramite SSH su Windows

Gli utenti Windows possono utilizzare un client SSH come PuTTY per connettersi al nodo primario. Prima di connettersi al nodo primario Amazon EMR, è necessario scaricare e installare PuTTY e PuTTYgen. Questi strumenti possono essere scaricati dalla pagina di download di PuTTY.

PuTTY non supporta a livello nativo il formato di file per la chiave privata della coppia di chiavi (`.pem`) generato da Amazon EC2. Si utilizza PuTTYgen per convertire il file chiave nel formato PuTTY richiesto (`.ppk`). È necessario convertire la chiave privata nel formato `.ppk` prima di tentare una connessione al nodo primario tramite PuTTY.

Per maggiori informazioni sulla conversione della chiave, consulta Conversione della chiave privata utilizzando PuTTYgen nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Connessione al nodo primario usando PuTTY

1. Assicurati di aver consentito il traffico SSH in entrata. Per istruzioni, consultare Prima di connetterti: autorizza il traffico in entrata.
2. Aprire `putty.exe`. È anche possibile avviare PuTTY dall'elenco dei programmi di Windows.
3. Se necessario, nell'elenco Category (Categoria) selezionare Session (Sessione).
4. Per Host Name (or IP address) (Nome host (o Indirizzo IP)), digitare `hadoop@`*`MasterPublicDNS`*. Ad esempio: `hadoop@`*`ec2-###-##-##-###.compute-1.amazonaws.com`*.
5. Nell'elenco Category (Categoria) scegliere Connection > SSH (Connessione > SSH), Auth (Autenticazione).
6. Per Private key file for authentication (File chiave privata per autenticazione), scegliere Browse (Sfoglia) e selezionare il file `.ppk` generato.
7. Seleziona Open (Apri), quindi Yes (Sì) per ignorare l'avviso di sicurezza di PuTTY.

> **Important**
>
> Quando accedi al nodo primario, inserisci `hadoop` nel caso in cui venga richiesto un nome utente.

8. Una volta terminato il lavoro sul nodo primario puoi chiudere la connessione SSH chiudendo PuTTY.

> **Note**
>
> Per evitare il timeout della connessione SSH, scegliere Connection (Connessione) nell'elenco Category (Categoria) e selezionare l'opzione Enable TCP_keepalives (Abilita TCP_keepalives). Se si dispone di una sessione SSH attiva in PuTTY, è possibile modificare le impostazioni aprendo il contesto (clic con il tasto destro del mouse) per la barra del titolo di PuTTY e scegliendo Change Settings (Modifica impostazioni).

Se riscontri problemi nell'utilizzo di SSH per la connessione al nodo primario, consulta la sezione Troubleshoot connecting to your instance (Risoluzione dei problemi di connessione all'istanza).

Connessione al nodo primario tramite la AWS CLI

È possibile creare una connessione SSH con il nodo primario utilizzando la AWS CLI su Windows e su Linux, Unix e Mac OS X. Indipendentemente dalla piattaforma, occorrono il nome DNS pubblico del nodo primario e la chiave privata della coppia di chiavi Amazon EC2. Se si utilizza AWS CLI su Linux, Unix o Mac OS X, è necessario anche impostare le autorizzazioni per il file della chiave privata (`.pem` o `.ppk`) come mostrato in Per configurare le autorizzazioni per i file chiave privati della coppia di chiavi.

Connessione al nodo primario con la AWS CLI

1. Assicurati di aver consentito il traffico SSH in entrata. Per istruzioni, consultare Prima di connetterti: autorizza il traffico in entrata.
2. Per recuperare l'identificatore del cluster, digitare:

   ```
   aws emr list-clusters
   ```

   L'output elenca i cluster, tra cui gli ID del cluster. Prendere nota dell'ID del cluster per il cluster a cui ci si connette.

   ```
   "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. Inserisci il seguente comando per aprire una connessione SSH al nodo primario. Sostituisci nell'esempio seguente *j-2AL4XXXXXX5T9* con l'ID del cluster e sostituisci *~/mykeypair.key* con il percorso completo e il nome del file `.pem` (per Linux, Unix e Mac OS X) o del file `.ppk` (per Windows). Ad esempio `C:\Users\<username>\.ssh\mykeypair.pem`.

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

4. Una volta terminato il lavoro sul nodo primario, chiudi la finestra AWS CLI.

Per ulteriori informazioni, consulta Comandi Amazon EMR nella AWS CLI. Se riscontri problemi nell'utilizzo di SSH per la connessione al nodo primario, consulta la sezione Troubleshoot connecting to your instance (Risoluzione dei problemi di connessione all'istanza).

Porte dei servizi Amazon EMR

> **Note**
>
> Di seguito sono riportate le interfacce e le porte dei servizi per i componenti su Amazon EMR. Questo non è l'elenco completo delle porte di servizio. I servizi non predefiniti, come le porte SSL e diversi tipi di protocolli, non sono elencati.

> **Important**
>
> Presta attenzione quando modifichi le regole per i gruppi di sicurezza per aprire le porte. Assicurati di aggiungere regole che permettano solo il traffico da client affidabili e autenticati per i protocolli e le porte necessari per eseguire i tuoi carichi di lavoro.

Componente	Descrizione del servizio	Servizio in esecuzione per impostazione predefinita	Porta	Chiave di configurazione
Hadoop	HTTP KMS REST API	Sì	9600	hadoop.kms.http.port
HDFS	Interfaccia utente Web di Namenode	Sì	9870	dfs.namenode.http-address
	NameNode RPC	Sì	8020	dfs.namenode.rpc-indirizzo-pc
	Interfaccia utente Web di DataNode	Sì	9864	dfs.datanode.http.address
	Datanode HTTP per il trasferimento di dati	Sì	9866	dfs.datanode.address
	Datanode RPC per il trasferimento dei dati	Sì	9867	dfs.datanode.ipc.address
Hive	HiveServer2 Thrift	Sì	10000	hive.server2.thrift.port
	HiveServer2 HTTP	No	10001	hive.server2.thrift.http.port
	Interfaccia utente Web di HiveServer 2	Sì	10002	hive.server2.webui.port
	Hive Metastore	Sì	9083	hive.metastore.port/metastore.thrift.port

Componente	Descrizione del servizio	Servizio in esecuzione per impostazione predefinita	Porta	Chiave di configurazione
	WebHCat	No	50111	templeton.port
	Servizio di gestione daemon (RPC) LLAP	No	15004	hive.llap.management.rpc.port
	Porta shuffle YARN per shuffle ospitato da LLAP-Daemon	No	15551	hive.llap.daemon.yarn.shuffle.port
	RPC daemon LLAP	No	Dinamico	hive.llap.daemon.rpc.port
	Interfaccia utente Web daemon LLAP	No	15002	hive.llap.daemon.web.port
	Servizio di output daemon LLAP	No	15003	hive.llap.daemon.output.service.port
Oozie		Sì	11000	
Tez	Interfaccia utente Tez	Sì	8080	
YARN	Shuffle	Sì	13562	mapreduce.shuffle.port
	Localizzatore RPC	Sì	8040	yarn.nodemanager.localizer.address

Componente	Descrizione del servizio	Servizio in esecuzione per impostazione predefinita	Porta	Chiave di configurazione
		Sì	8041	
	Indirizzo di NM Webapp	Sì	8042	yarn.nodemanager.webapp.address
	Applicazione web RM	Sì	8088	yarn.resourcemanager.webapp.address
		Sì	8025	
	Pianificatore	Sì	8030	yarn.resourcemanager.scheduler.address
	Interfaccia del gestore delle applicazioni	Sì	8032	yarn.resourcemanager.address
	Interfaccia amministratore di RM	Sì	8033	yarn.resourcemanager.admin.address
	Interfaccia utente Web JobHistory Server	Sì	19888	mapreduce.jobhistory.webapp.address
	Interfaccia utente Web amministratore server JobHistory	Sì	10033	mapreduce.jobhistory.admin.address

Componente	Descrizione del servizio	Servizio in esecuzione per impostazione predefinita	Porta	Chiave di configurazione
	Server JobHistory (RPC)	Sì	10020	mapreduce.jobhistory.address
	Application Timeline Server (RPC)	Sì	10200	yarn.timeline-service.indirizzo
	Interfaccia utente Web HTTP di Application Timeline Server	Sì	8188	yarn.timeline-service.webapp.address
	Interfaccia utente Web HTTPS di Application Timeline Server	No	8190	yarn.timeline-service.webapp.https.address
		Sì	20888	
Zookeeper	Porta client	Sì	2181	
		Sì	37301	
		Sì	8341	

Visualizzazione di interfacce Web ospitate su cluster Amazon EMR

Important

È possibile configurare un gruppo di sicurezza personalizzato per consentire l'accesso in entrata a queste interfacce Web. Tenere presente che qualsiasi porta su cui si consente il traffico in entrata rappresenta una potenziale vulnerabilità per la sicurezza. Esaminare

attentamente i gruppi di sicurezza personalizzati per assicurarsi di ridurre al minimo le vulnerabilità. Per ulteriori informazioni, consulta Controllo del traffico di rete con gruppi di sicurezza.

Hadoop e altre applicazioni installate sul cluster EMR pubblicano interfacce utente come siti Web ospitati sul nodo primario. Per motivi di sicurezza, quando utilizzi i gruppi di sicurezza gestiti da Amazon EMR, questi siti Web sono disponibili solo sul server Web locale del nodo primario. Per questo motivo, devi connetterti al nodo primario per visualizzare le interfacce Web. Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH. Hadoop pubblica anche le interfacce utente come siti Web ospitati sui nodi di task e principali. Questi siti web sono disponibili anche solo sui server web locali dei nodi.

La tabella seguente elenca le interfacce Web che puoi visualizzare sulle istanze di cluster. Queste interfacce di Hadoop sono disponibili su tutti i cluster. Per le interfacce dell'istanza master, sostituisci *master-public-dns-name* con il valore DNS pubblico master elencato nella scheda Riepilogo del cluster nella console Amazon EMR. Per le interfacce dell'istanza di task e principale, sostituire *coretask-public-dns-name* con il valore Public DNS name (Nome DNS pubblico) elencato per l'istanza. Per trovare un'istanza Nome DNS pubblico, nella console Amazon EMR, scegli il cluster dall'elenco, quindi la scheda Hardware, l'ID del gruppo di istanze che contiene l'istanza a cui connetterti, infine annota il Nome DNS pubblico indicato per l'istanza.

Nome dell'interfaccia	URI
Flink History Server (EMR versione 5.33 e successive)	http://*master-public-dns-name* :8082/
Ganglia	http://*master-public-dns-name* /ganglia/
Hadoop HDFS NameNode (versione EMR precedente alla 6.x)	https://*master-public-dns-name* :50470/
Hadoop HDFS NameNode	http://*master-public-dns-name* :50070/
Hadoop HDFS DataNode	http://*coretask-public-dns-name* :50075/
Hadoop HDFS NameNode (EMR versione 6.x)	https://*master-public-dns-name* :9870/

Nome dell'interfaccia	URI
Hadoop HDFS DataNode (versione EMR precedente alla 6.x)	https://*coretask-public-dns-name* :50475/
Hadoop HDFS DataNode (EMR versione 6.x)	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/

Poiché sul nodo primario sono disponibili diverse interfacce specifiche per l'applicazione che non sono disponibili sui nodi core e attività, le istruzioni di questo documento sono specifiche per il nodo primario Amazon EMR. È possibile accedere alle interfacce Web sui nodi core e attività nello stesso modo in cui si accede alle interfacce Web sul nodo primario.

Esistono diversi modi per accedere alle interfacce Web sul nodo primario. Il metodo più semplice e veloce è utilizzare SSH per connettersi al nodo primario e usare il browser basato su testo, Lynx, per visualizzare i siti Web nel proprio client SSH. Tuttavia, Lynx è un browser basato su testo con un'interfaccia utente limitata che non è in grado di visualizzare la grafica. L'esempio seguente mostra come aprire l'interfaccia ResourceManager di Hadoop utilizzando Lynx (gli URL Lynx sono forniti anche quando si accede al nodo primario usando SSH).

```
lynx http://ip-###-##-##-###.us-west-2.compute.internal:8088/
```

Esistono altre due opzioni per accedere alle interfacce Web sul nodo primario che forniscono caratteristiche complete per il browser. Scegli una delle seguenti opzioni:

- Opzione 1 (raccomandata per gli utenti più esperti): utilizzare un client SSH per connettersi al nodo primario, configurare il tunneling SSH con l'inoltro porta locale e utilizzare un browser Internet per aprire le interfacce Web ospitate sul nodo primario. Questo metodo consente di configurare l'accesso all'interfaccia web senza utilizzare un proxy SOCKS.
- Opzione 2 (raccomandata per i nuovi utenti): utilizzare un client SSH per connettersi al nodo primario, configurare il tunneling SSH con inoltro dinamico delle porte e configurare il browser Internet per l'utilizzo di un componente aggiuntivo, ad esempio FoxyProxy per Firefox o SwitchyOmega per Chrome, per gestire le impostazioni del proxy SOCKS. Questo metodo consente di filtrare automaticamente gli URL in base a modelli di testo e di limitare le impostazioni proxy a domini che corrispondono alla forma del nome DNS del nodo primario. Per ulteriori informazioni su come configurare FoxyProxy per Firefox e Google Chrome, vedere Opzione 2, parte 2: configurazione delle impostazioni del proxy per visualizzare i siti Web ospitati nel nodo primario.

> **Note**
>
> Se modifichi la porta in cui un'applicazione viene eseguita tramite la configurazione del cluster, il collegamento ipertestuale alla porta non verrà aggiornato nella console di Amazon EMR. Questo avviene perché la console non dispone di una caratteristica per leggere la configurazione `server.port`.

Con Amazon EMR versione 5.25.0 o successiva, puoi accedere all'interfaccia utente del server della cronologia Spark dalla console senza configurare un proxy Web tramite una connessione SSH. Per ulteriori informazioni, consulta Accesso con un clic a Spark History Server persistente.

Argomenti

- Opzione 1: impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta locale
- Opzione 2, parte 1: impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta dinamico

- Opzione 2, parte 2: configurazione delle impostazioni del proxy per visualizzare i siti Web ospitati nel nodo primario

Opzione 1: impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta locale

Per connetterti al server Web locale sul nodo primario, crea un tunnel SSH tra il computer e il nodo primario. È noto anche come inoltro porta. Se non desideri utilizzare un proxy SOCKS, puoi impostare un tunnel SSH per il nodo primario utilizzando l'inoltro porta locale. Con l'inoltro porta locale, si specificano le porte locali inutilizzate che vengono impiegate per inoltrare il traffico a specifiche porte remote sul server Web locale del nodo primario.

La configurazione di un tunnel SSH utilizzando l'inoltro porta locale richiede il nome DNS pubblico del nodo primario e il file di chiave privata della coppia di chiavi. Per ulteriori informazioni sull'individuazione del nome DNS pubblico master, vedere Recupero del nome DNS pubblico del nodo primario con la vecchia console. Per ulteriori informazioni sull'accesso alla coppia di chiavi, consulta Coppia di chiavi Amazon EC2 nella Guida per l'utente di Amazon EC2 per le istanze Linux. Per ulteriori informazioni sui siti da visualizzare nel nodo primario, consulta la sezione Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta locale con OpenSSH

Per impostare un tunnel SSH utilizzando l'inoltro porta locale nel terminale

1. Assicurati di aver consentito il traffico SSH in entrata. Per istruzioni, consultare Prima di connetterti: autorizza il traffico in entrata.
2. Apri una finestra del terminale. In Mac OS X, selezionare Applications > Utilities > Terminal (Applicazioni > Utility > Terminale). In altre distribuzioni Linux, la finestra terminal si trova generalmente in Applications > Accessories > Terminal (Applicazioni > Accessori >; Terminale).
3. Digita il seguente comando per aprire un tunnel SSH sulla tua macchina locale. Questo comando di esempio accede all'interfaccia Web di ResourceManager inoltrando il traffico sulla porta locale 8157 (una porta locale non utilizzata scelta casualmente) alla porta 8088 sul server Web locale del nodo master.

   Nel comando, sostituisci *~/mykeypair.pem* con il percorso e il nome del file con il tuo file `.pem` e sostituisci *ec2-###-##-##-###.compute-1.amazonaws.com* con il nome DNS pubblico principale del cluster. Per accedere a un'interfaccia Web diversa, sostituisci 8088 con il numero

```
ssh -i ~/mykeypair.pem -N -L 8157:ec2-###-##-##-
###.compute-1.amazonaws.com:8088 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
```

`-L` indica l'uso dell'inoltro porta locale che consente di specificare una porta locale utilizzata per inoltrare i dati alla porta remota identificata sul server Web locale del nodo master.

Dopo l'emissione di questo comando, il terminale rimane aperto e non risponde.

4. Per aprire l'interfaccia Web ResourceManager nel browser, digita: `http://localhost:8157/` nella barra degli indirizzi.
5. Una volta terminato il lavoro con le interfacce Web sul nodo primario, chiudi le finestre del terminale.

Opzione 2, parte 1: impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta dinamico

Per connetterti al server Web locale sul nodo primario, crea un tunnel SSH tra il computer e il nodo primario. È noto anche come inoltro porta. Se crei il tunnel SSH utilizzando l'inoltro porte dinamico, tutto il traffico instradato verso una specifica porta locale inutilizzata viene inoltrato al server Web locale sul nodo primario. In questo modo viene creato un proxy SOCKS. È quindi possibile configurare il browser Internet in modo che utilizzi un componente aggiuntivo, ad esempio FoxyProxy o SwitchyOmega, per gestire le impostazioni del proxy SOCKS.

L'utilizzo di un componente aggiuntivo per la gestione del proxy consente di filtrare automaticamente gli URL in base a modelli di testo e di limitare le impostazioni proxy a domini che corrispondono alla forma del nome DNS pubblico del nodo primario. Il componente aggiuntivo del browser gestisce automaticamente l'attivazione e la disattivazione del proxy quando si passa dalla visualizzazione di siti Web ospitati sul nodo primario a quella di siti Web su Internet.

Prima di iniziare, devi disporre del nome DNS pubblico del nodo primario e del file di chiave privata della coppia di chiavi. Per informazioni sull'individuazione del nome DNS pubblico primario, consulta Recupero del nome DNS pubblico del nodo primario con la vecchia console. Per ulteriori informazioni sull'accesso alla coppia di chiavi, consulta Coppia di chiavi Amazon EC2 nella Guida per l'utente di Amazon EC2 per le istanze Linux. Per ulteriori informazioni sui siti da visualizzare nel nodo primario, consulta la sezione Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta dinamico su OpenSSH

Impostazione di un tunnel SSH utilizzando l'inoltro porta dinamico con OpenSSH

1. Assicurati di aver consentito il traffico SSH in entrata. Per istruzioni, consultare Prima di connetterti: autorizza il traffico in entrata.
2. Apri una finestra del terminale. In Mac OS X, selezionare Applications > Utilities > Terminal (Applicazioni > Utility > Terminale). In altre distribuzioni Linux, la finestra terminal si trova generalmente in Applications > Accessories > Terminal (Applicazioni > Accessori >; Terminale).
3. Digitare il seguente comando per aprire un tunnel SSH sulla macchina locale. Sostituisci *~/mykeypair.pem* con il percorso e il nome file del tuo file `.pem`, sostituisci *8157* con un numero di porta locale non utilizzato, quindi sostituisci *c2-###-##-##-###.compute-1.amazonaws.com* con il nome DNS pubblico primario del cluster.

   ```
   ssh -i ~/mykeypair.pem -N -D 8157 hadoop@ec2-###-##-##-###.compute-1.amazonaws.com
   ```

   Dopo l'emissione di questo comando, il terminale rimane aperto e non risponde.

   > **Note**
   >
   > `-D` indica l'uso dell'inoltro porta dinamico, che consente di specificare una porta locale utilizzata per inoltrare i dati verso tutte le porte remote sul server Web locale del nodo primario. L'inoltro dinamico della porta crea un proxy SOCKS locale ascolto sulla porta specificato nel comando.

4. Dopo che il tunnel è attivo, configurare un proxy SOCKS per il browser. Per ulteriori informazioni, consulta Opzione 2, parte 2: configurazione delle impostazioni del proxy per visualizzare i siti Web ospitati nel nodo primario.
5. Una volta terminato il lavoro con le interfacce Web sul nodo primario, chiudi la finestra del terminale.

Impostazione di un tunnel SSH utilizzando l'inoltro porta dinamico con la AWS CLI

È possibile creare una connessione SSH con il nodo primario utilizzando AWS CLI su Windows e su Linux, Unix e Mac OS X. Se si utilizza AWS CLI su Linux, Unix o Mac OS X, è necessario impostare le autorizzazioni sul file `.pem` come mostrato in Per configurare le autorizzazioni per i file chiave privati della coppia di chiavi. Se si utilizza AWS CLI su Windows, PuTTY deve apparire nella variabile

d'ambiente del percorso, altrimenti si potrebbe ricevere un errore come ad esempio OpenSSH or PuTTY not available (OpenSSH o PuTTY non disponibile).

Per impostare un tunnel SSH utilizzando l'inoltro porta dinamico con la AWS CLI

1. Assicurati di aver consentito il traffico SSH in entrata. Per istruzioni, consultare Prima di connetterti: autorizza il traffico in entrata.
2. Crea una connessione SSH con il nodo primario come illustrato nella Connessione al nodo primario tramite la AWS CLI.
3. Per recuperare l'identificatore del cluster, digitare:

   ```
   aws emr list-clusters
   ```

   L'output elenca i cluster, tra cui gli ID del cluster. Prendere nota dell'ID del cluster per il cluster a cui ci si connette.

   ```
   "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. Per aprire un tunnel SSH con il nodo primario utilizzando l'inoltro alla porta dinamico, digita il seguente comando. Sostituire nell'esempio seguente *j-2AL4XXXXXX5T9* con l'ID del cluster e sostituire *~/mykeypair.key* con la posizione e il nome del file `.pem` (per Linux, Unix e Mac OS X) o del file `.ppk` (per Windows).

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

> **Note**
>
> Il comando socks configura automaticamente l'inoltro dinamico della porta sulla porta locale 8157. Al momento, questa impostazione non può essere modificata.

5. Dopo che il tunnel è attivo, configurare un proxy SOCKS per il browser. Per ulteriori informazioni, consulta Opzione 2, parte 2: configurazione delle impostazioni del proxy per visualizzare i siti Web ospitati nel nodo primario.
6. Una volta terminato il lavoro con le interfacce Web sul nodo primario, chiudi la finestra di AWS CLI.

   Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Configurazione di un tunnel SSH per il nodo primario utilizzando PuTTY

Gli utenti Windows possono usare un client SSH come PuTTY per creare un tunnel SSH al nodo primario. Prima di connettersi al nodo primario Amazon EMR, è necessario scaricare e installare PuTTY e PuTTYgen. Questi strumenti possono essere scaricati dalla pagina di download di PuTTY.

PuTTY non supporta a livello nativo il formato di file per la chiave privata della coppia di chiavi (`.pem`) generato da Amazon EC2. Si utilizza PuTTYgen per convertire il file chiave nel formato PuTTY richiesto (`.ppk`). È necessario convertire la chiave privata nel formato `.ppk` prima di tentare una connessione al nodo primario tramite PuTTY.

Per maggiori informazioni sulla conversione della chiave, consulta Conversione della chiave privata utilizzando PuTTYgen nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Impostazione di un tunnel SSH utilizzando l'inoltro porta dinamico su PuTTY

1. Assicurati di aver consentito il traffico SSH in entrata. Per istruzioni, consultare Prima di connetterti: autorizza il traffico in entrata.
2. Fai doppio clic su `putty.exe` per avviare PuTTY. È anche possibile avviare PuTTY dall'elenco dei programmi di Windows.

> **Note**
>
> Se disponi già di una sessione SSH attiva con il nodo primario, puoi aggiungere un tunnel facendo clic con il pulsante destro del mouse sulla barra del titolo di PuTTY e scegliendo Change Settings (Modifica impostazioni).

3. Se necessario, nell'elenco Category (Categoria) selezionare Session (Sessione).
4. Nel campo Host Name (Nome host) digitare **hadoop@***MasterPublicDNS*. Ad esempio: **hadoop@***ec2-###-##-##-###.compute-1.amazonaws.com*.
5. Nell'elenco Category (Categoria), espandere Connection > SSH (Connessione > SSH), quindi scegliere Auth (Autenticazione).
6. Per Private key file for authentication (File chiave privata per autenticazione), scegliere Browse (Sfoglia) e selezionare il file `.ppk` generato.

> **Note**
>
> PuTTY non supporta a livello nativo il formato di file per la chiave privata della coppia di chiavi (`.pem`) generato da Amazon EC2. Si utilizza PuTTYgen per convertire il file chiave nel formato PuTTY richiesto (`.ppk`). È necessario convertire la chiave privata nel formato `.ppk` prima di tentare una connessione al nodo primario tramite PuTTY.

7. Nell'elenco Category (Categoria), espandere Connection > SSH (Connessione > SSH), quindi scegliere Tunnels (Tunnel).
8. Nel campo Source port (Porta sorgente), digita `8157` (un numero di porta locale inutilizzato), quindi seleziona Add (Aggiungi).
9. Lasciare vuoto il campo Destination (Destinazione).
10. Selezionare le opzioni Dynamic (Dinamico) e Auto (Automatico).
11. Scegliere Open (Apri).
12. Scegliere Yes (Sì) per ignorare l'avviso di sicurezza PuTTY.

> **Important**
>
> Quando accedi al nodo primario, inserisci `hadoop` nel caso in cui venga richiesto un nome utente.

13. Dopo che il tunnel è attivo, configurare un proxy SOCKS per il browser. Per ulteriori informazioni, consulta Opzione 2, parte 2: configurazione delle impostazioni del proxy per visualizzare i siti Web ospitati nel nodo primario.
14. Una volta terminato il lavoro con le interfacce Web sul nodo primario, chiudi la finestra di PuTTY.

Opzione 2, parte 2: configurazione delle impostazioni del proxy per visualizzare i siti Web ospitati nel nodo primario

Se si utilizza un tunnel SSH con inoltro dinamico delle porte, è necessario utilizzare un add-on per la gestione dei proxy SOCKS per controllare le impostazioni proxy nel browser. L'utilizzo di uno strumento per la gestione del proxy SOCKS consente di filtrare automaticamente gli URL in base a modelli di testo e di limitare le impostazioni proxy a domini che corrispondono alla forma del nome DNS pubblico del nodo primario. Il componente aggiuntivo del browser gestisce automaticamente l'attivazione e la disattivazione del proxy quando si passa dalla visualizzazione di siti Web ospitati sul nodo primario a quella di siti Web su Internet. Per gestire le impostazioni del proxy, configurare il browser in modo che utilizzi un componente aggiuntivo, ad esempio FoxyProxy o SwitchyOmega.

Per ulteriori informazioni sulla creazione di un tunnel SSH, consulta Opzione 2, parte 1: impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta dinamico. Per ulteriori informazioni sulle interfacce Web, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Includi le seguenti impostazioni quando configuri il componente aggiuntivo proxy:

- Utilizza localhost come indirizzo host.
- Utilizza lo stesso numero di porta locale che hai selezionato per stabilire il tunnel SSH con il nodo primario in Opzione 2, parte 1: impostazione di un tunnel SSH sul nodo primario utilizzando l'inoltro porta dinamico. Ad esempio, porta *8157*. Questa porta deve anche corrispondere al numero di porta utilizzato in PuTTY o in un qualsiasi altro emulatore di terminale utilizzato per il collegamento.
- Specifica il valore del protocollo SOCKS v5. SOCKS v5 consente di impostare facoltativamente l'autorizzazione utente.
- URL Patterns (Modelli URL)

  I seguenti modelli URL dovrebbero essere nella whitelist e specificati con un tipo di modello jolly:

  - I modelli *ec2*.*compute*.amazonaws.com* e *10*.amazonaws.com* corrispondono al nome DNS pubblico dei cluster nelle regioni US.

- I modelli *ec2*.compute* e *10*.compute* corrispondono al nome DNS pubblico dei cluster in tutte le altre regioni.
- Un modello 10.* fornisce l'accesso ai file di log JobTracker in Hadoop. Modificare questo filtro se entra in conflitto con il piano di accesso della rete.
- I modelli *.ec2.internal* e *.compute.internal* devono corrispondere ai nomi DNS privati (interni) dei cluster rispettivamente nella regione `us-east-1` e in tutte le altre regioni.

Esempio: configurazione di FoxyProxy per Firefox

L'esempio seguente mostra una configurazione FoxyProxy Standard (versione 7.5.1) per Mozilla Firefox.

FoxyProxy fornisce una serie di strumenti di gestione dei proxy. Consente di utilizzare un server proxy per gli URL associati a modelli corrispondenti ai domini utilizzati da istanze Amazon EC2 nel cluster Amazon EMR.

Installazione e configurazione di FoxyProxy utilizzando Mozilla Firefox

1. In Firefox, vai all'indirizzo https://addons.mozilla.org/, cerca FoxyProxy Standard e segui le istruzioni per aggiungere FoxyProxy a Firefox.
2. Mediante un editor di testo, crea un file JSON denominato `foxyproxy-settings.json` dalla configurazione di esempio riportata di seguito:

```
{
  "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. Apri la pagina Manage Your Extensions (Gestisci estensioni) di Firefox (vai a about:addons, quindi seleziona Extensions (Estensioni).
4. Seleziona FoxyProxy Standard (Standard FoxyProxy), quindi seleziona il pulsante che mostra più opzioni (il pulsante con i tre punti di sospensione).
5. Seleziona Options (Opzioni) dal menu a discesa.
6. Seleziona Import Settings (Importa impostazioni) dal menu a sinistra.
7. In Import Settings (Importa impostazioni), scegli Import Settings (Importa impostazioni) sotto Import Settings from FoxyProxy 6.0+ (Importa impostazioni da FoxyProxy 6.0+), quindi apri il percorso del file `foxyproxy-settings.json` che hai creato, seleziona il file e scegli Open (Apri).
8. Scegli OK quando il sistema richiede di sovrascrivere le impostazioni esistenti e salvare la nuova configurazione.

Esempio: Configurazione di SwitchyOmega per Chrome

Nell'esempio seguente viene illustrato come configurare l'estensione SwitchyOmega per Google Chrome. SwitchyOmega consente di configurare, gestire e selezionare più proxy.

Installazione e configurazione di SwitchyOmega utilizzando Google Chrome

1. Vai a https://chrome.google.com/webstore/category/extensions, cerca Proxy SwitchyOmega e aggiungilo a Chrome.
2. Seleziona New profile (Nuovo profilo) e immetti `emr-socks-proxy` come nome del profilo.
3. Seleziona PAC profile (Profilo PAC) e in seguito Create (Crea). I file Proxy Auto-Configuration (PAC) (Configurazione automatica proxy (PAC)) consentono di definire un elenco di autorizzazioni per le richieste del browser che devono essere inoltrate a un server proxy Web.
4. Nel campo PAC Script (Script PAC), sostituisci il contenuto con lo script seguente che definisce quali URL devono essere inoltrati tramite il server proxy Web. Se hai specificato un numero di porta diverso durante la configurazione del tunnel SSH, sostituisci *8157* con il tuo numero di porta.

```
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. In Actions (Operazioni), scegli Apply changes (Applica modifiche) per salvare le impostazioni del proxy.
6. Nella barra degli strumenti di Chrome, seleziona SwitchyOmega e seleziona il profilo `emr-socks-proxy`.

Accesso a un'interfaccia Web nel browser

Per aprire un'interfaccia Web, inserisci il nome DNS pubblico del nodo primario o core seguito dal numero di porta dell'interfaccia scelta nella barra degli indirizzi del browser. L'esempio seguente mostra l'URL che dovrai immettere per connetterti a Spark HistoryServer.

```
http://master-public-dns-name:18080/
```

Per istruzioni su come recuperare il nome DNS pubblico di un nodo, consulta Recupero del nome DNS pubblico del nodo primario. Per un elenco completo degli URL delle interfacce Web, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Invio di lavoro a un cluster

Questa sezione descrive i metodi che puoi utilizzare per l'invio di lavoro a un cluster Amazon EMR. Per inviare un lavoro, puoi aggiungere fasi oppure puoi inviare processi Hadoop al nodo primario in modo interattivo.

Quando invii fasi a un cluster, considera le seguenti regole di comportamento:

- Un ID fase può contenere fino a 256 caratteri.
- Puoi avere fino a 256 fasi PENDING e RUNNING in un cluster.
- Puoi inviare processi al nodo primario in modo interattivo anche se hai 256 fasi attive in esecuzione sul cluster. Puoi inviare un numero illimitato di fasi durante l'intero ciclo di vita di un cluster di lunga durata, ma soltanto 256 fasi possono essere RUNNING (IN ESECUZIONE) o PENDING (IN SOSPESO) in un determinato momento.
- Con Amazon EMR versioni 4.8.0 e successive, a eccezione della versione 5.0.0, puoi annullare le fasi in sospeso. Per ulteriori informazioni, consulta Annullamento delle fasi.
- Con Amazon EMR versioni 5.28.0 e successive, puoi annullare sia le fasi in sospeso che quelle in esecuzione. Puoi inoltre scegliere di eseguire più fasi in parallelo per migliorare l'utilizzo del cluster e risparmiare sui costi. Per ulteriori informazioni, consulta Considerazioni sull'esecuzione di più fasi in parallelo.

> **Note**
>
> Per ottenere prestazioni ottimali, ti consigliamo di archiviare operazioni di bootstrap personalizzate, script e altri file che desideri utilizzare con Amazon EMR in un bucket Amazon S3 che si trova nella stessa Regione AWS del tuo cluster.

Argomenti

- Aggiunta di fasi a un cluster con la console di gestione Amazon EMR
- Aggiunta di fasi a un cluster con la AWS CLI

- Considerazioni sull'esecuzione di più fasi in parallelo
- Visualizzazione delle fasi
- Annullamento delle fasi

Aggiunta di fasi a un cluster con la console di gestione Amazon EMR

Utilizza le procedure seguenti per aggiungere fasi a un cluster utilizzando la AWS Management Console. Per informazioni dettagliate su come inviare fasi per applicazioni Big Data specifiche, consulta le seguenti sezioni della Guida ai rilasci di Amazon EMR:

- Invio di una fase JAR personalizzata
- Invio di una fase di streaming Hadoop
- Invio di una fase Spark
- Invio di una fase Pig
- Esecuzione di un comando o di uno script come fase
- Passaggio di valori in fasi per l'esecuzione di script Hive

Aggiungere fasi durante la creazione del cluster

Dalla AWS Management Console, puoi aggiungere fasi quando crei un cluster.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Aggiunta di fasi durante la creazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.

3. In Steps (Fasi), scegli Add step (Aggiungi fase). Inserisci i valori appropriati nei campi della finestra di dialogo Add step (Aggiungi fase). Per informazioni sulla formattazione degli argomenti delle fasi, consulta Aggiungere argomenti di fase. Le opzioni variano a seconda del tipo di fase. Per aggiungere la fase e chiudere la finestra di dialogo, scegli Aggiungi fase.
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Aggiunta di fasi durante la creazione di un cluster con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/home. Seleziona Create Cluster - Advanced Options (Crea un cluster - Opzioni avanzate).
2. Nella pagina Fase 1: Software e fasi per Fasi (facoltative), selezionare Run multiple steps in parallel to improve cluster utilization and save cost (Esegui più fasi in parallelo per migliorare l'utilizzo del cluster e risparmiare sui costi). Il valore predefinito per il livello di concorrenza è 10. Puoi scegliere tra 2 e 256 fasi che possono essere eseguite in parallelo.

   > **Note**
   >
   > L'esecuzione di più fasi in parallelo è supportata solo con Amazon EMR versione 5.28.0 e successive.

3. Per After last step completes (Dopo il completamento dell'ultimo passaggio), scegliere Cluster enters waiting state (Cluster entra in stato di attesa) o Auto-terminate the cluster (Chiudi automaticamente il cluster).
4. Scegliere Step type (Tipo fase), quindi Add step (Aggiungi fase).
5. Digitare i valori appropriati nei campi della finestra di dialogo Add step (Aggiungi fase). Per informazioni sulla formattazione degli argomenti delle fasi, consulta Aggiungere argomenti di fase. Le opzioni variano a seconda del tipo di fase. Se è stato abilitato Esegui più fasi in parallelo per migliorare l'utilizzo del cluster e risparmiare sui costi, l'unica opzione disponibile per Azione in caso di errore è Continua. Quindi, scegliere Add (Aggiungi).

Aggiungere fasi a un cluster in esecuzione

Con la AWS Management Console, puoi aggiungere fasi a un cluster con l'opzione di terminazione automatica disattivata.

New console

Aggiunta di fasi a un cluster in esecuzione con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Nella scheda Steps (Fasi) della pagina dei dettagli del cluster, scegli Add step (Aggiungi fase). Per clonare una fase esistente, scegli il menu a discesa Actions (Operazioni) e seleziona Clone step (Clona fase).
4. Inserisci i valori appropriati nei campi della finestra di dialogo Add step (Aggiungi fase). Le opzioni variano a seconda del tipo di fase. Per aggiungere la fase e uscire dalla finestra di dialogo, scegli Add step (Aggiungi fase).

Old console

Aggiunta di fasi a un cluster in esecuzione con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/home. Nella pagina Cluster List (Elenco cluster), selezionare il link per il proprio cluster.
2. Nella pagina Cluster Details (Dettagli cluster), scegliere la scheda Steps (Fasi) .
3. Nella scheda Steps (Fasi), scegliere Add step (Aggiungi fase).
4. Digitare i valori appropriati nei campi della finestra di dialogo Add step (Aggiungi fase), quindi selezionare Add (Aggiungi). Le opzioni variano a seconda del tipo di fase.

Modificare il livello di concorrenza delle fasi in un cluster in esecuzione

Con la AWS Management Console, puoi modificare il livello di simultaneità delle fasi in un cluster in esecuzione.

> **Note**
>
> Puoi eseguire più fasi in parallelo solo con Amazon EMR versione 5.28.0 e successive.

New console

Modifica del livello di concorrenza delle fasi in un cluster in esecuzione con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare. Il cluster deve essere in esecuzione per modificare l'attributo di simultaneità.
3. Nella scheda Steps (Fasi) della pagina dei dettagli del cluster, cerca la sezione Attributes (Attributi). Seleziona Edit (Modifica) per modificare la simultaneità. Inserisci un valore compreso tra 1 e 256.

Old console

Modifica del livello di simultaneità delle fasi in un cluster in esecuzione con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/home. Nella pagina Cluster List (Elenco cluster), selezionare il link per il proprio cluster.
2. Nella pagina Cluster Details (Dettagli cluster), scegliere la scheda Steps (Fasi) .
3. Per Concurrency (Concorrenza), scegliere Change (Cambia). Selezionare un nuovo valore per il livello di concorrenza delle fasi, quindi salvare.

Aggiungere argomenti di fase

Quando utilizzi la AWS Management Console per aggiungere una fase al cluster, puoi specificare argomenti per quella fase nel campo Argomenti. È necessario separare gli argomenti con spazi bianchi e circondare gli argomenti di stringa costituiti da caratteri e spazio bianco con virgolette.

Example : Argomenti corretti

I seguenti argomenti di esempio sono formattati correttamente per la AWS Management Console, con virgolette attorno all'argomento finale della stringa.

```
bash -c "aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

Puoi inoltre inserire ogni argomento su una riga separata per la leggibilità come mostrato nell'esempio seguente.

```
bash
-c
"aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh ."
```

Example : Argomenti errati

I seguenti argomenti di esempio sono formattati in modo improprio per la AWS Management Console. Notate che l'argomento finale della stringa, `aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .`, contiene spazi bianchi e non è circondato da virgolette.

```
bash -c aws s3 cp s3://DOC-EXAMPLE-BUCKET/my-script.sh .
```

Aggiunta di fasi a un cluster con la AWS CLI

Le seguenti procedure illustrano come aggiungere fasi a un cluster appena creato e a un cluster in esecuzione con la AWS CLI. In entrambi gli esempi, viene utilizzato il sottocomando `--steps` per aggiungere fasi al cluster.

Per aggiungere fasi durante la creazione del cluster

- Digitare il comando seguente per creare un cluster e aggiungere una fase Apache Pig. Assicurati di sostituire *myKey* con il nome della coppia di chiavi 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**
>
> L'elenco di argomenti cambia in funzione del tipo di fase.

Per impostazione predefinita, il livello di concorrenza delle fasi è 1. Puoi impostare il livello di simultaneità delle fasi utilizzando il parametro `StepConcurrencyLevel` quando crei un cluster.

L'output è un identificatore di cluster simile a quanto segue.

```
{
    "ClusterId": "j-2AXXXXXXGAPLF"
}
```

Per aggiungere una fase a un cluster in esecuzione

- Digitare il comando seguente per aggiungere una fase a un cluster in esecuzione. Sostituisci *j-2AXXXXXXGAPLF* con l'ID del tuo cluster.

```
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"}]'
```

L'output è un identificatore di fase simile a quanto segue.

```
{
    "StepIds": [
 "s-Y9XXXXXXAPMD"
    ]
}
```

Per modificare StepConcurrencyLevel in un cluster in esecuzione

1. In un cluster in esecuzione, puoi modificare il `StepConcurrencyLevel` con l'API `ModifyCluster`. Ad esempio, digita il comando seguente per aumentare `StepConcurrencyLevel` fino a `10`. Sostituisci *`j-2AXXXXXXGAPLF`* con l'ID del tuo cluster.

   ```
   aws emr modify-cluster --cluster-id j-2AXXXXXXGAPLF --step-concurrency-level 10
   ```

2. L'output è simile a quello riportato di seguito.

   ```
   {
   "StepConcurrencyLevel": 10
   }
   ```

Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR nella AWS CLI, consulta la Guida di riferimento ai comandi della AWS CLI.

Considerazioni sull'esecuzione di più fasi in parallelo

- Le fasi in esecuzione in parallelo possono essere completate in qualsiasi ordine, ma le fasi in sospeso nella coda passano allo stato in esecuzione nell'ordine in cui sono state inviate.
- Quando si seleziona un livello di simultaneità delle fasi per il cluster, è necessario considerare se il tipo di istanza del nodo primario soddisfa o meno i requisiti di memoria dei carichi di lavoro degli utenti. Il processo di esecuzione della fase principale viene eseguito sul nodo primario per ogni fase. L'esecuzione di più fasi in parallelo richiede più memoria e utilizzo della CPU del nodo primario rispetto all'esecuzione di una fase alla volta.
- Per ottenere la pianificazione complessa e la gestione delle risorse delle fasi simultanee, è possibile utilizzare caratteristiche di programmazione YARN come `FairScheduler` o `CapacityScheduler`. Ad esempio, è possibile utilizzare `FairScheduler` con un set `queueMaxAppsDefault` per impedire l'esecuzione di più di un certo numero di processi contemporaneamente.
- Il livello di concorrenza delle fasi è soggetto alle configurazioni dei gestori delle risorse. Ad esempio, se YARN è configurato con un solo parallelismo di `5`, allora è possibile avere solo cinque applicazioni YARN in esecuzione in parallelo anche se `StepConcurrencyLevel` è impostato su `10`. Per ulteriori informazioni sulla configurazione dei gestori delle risorse, consulta Configurazione delle applicazioni nella Guida al rilascio di Amazon EMR.

- Non è possibile aggiungere una fase con un `ActionOnFailure` diverso da CONTINUE (CONTINUA) mentre il livello di concorrenza della fase del cluster è maggiore di 1.
- Se il livello di concorrenza della fase di un cluster è maggiore di uno, la caratteristica `ActionOnFailure` della fase non si attiverà.
- Se un cluster ha un livello di concorrenza della fase 1 ma ha più fasi in esecuzione, `TERMINATE_CLUSTER ActionOnFailure` potrebbe attivarsi, ma `CANCEL_AND_WAIT ActionOnFailure` non lo farà. Questo caso limite si verifica quando il livello di concorrenza della fase del cluster era maggiore di uno, ma si è abbassato durante l'esecuzione di più fasi.
- È possibile utilizzare la scalabilità automatica di EMR per aumentare o ridurre in base alle risorse YARN evitando per evitare conflitti tra risorse. Per ulteriori informazioni, consulta Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze nella Guida alla gestione di Amazon EMR.
- Quando si riduce il livello di concorrenza delle fasi, EMR consente di completare tutte le fasi in esecuzione prima di ridurne il numero. Se le risorse sono esaurite perché il cluster esegue troppi passaggi simultanei, è consigliabile annullare manualmente tutte le fasi in esecuzione per liberare risorse.

Visualizzazione delle fasi

Il numero totale di record di fasi che puoi visualizzare (indipendentemente dallo stato) è 1.000. Questo totale include le fasi inviate dall'utente e le fasi di sistema. Quando lo stato delle fasi inviate dall'utente diventa COMPLETED o FAILED, è possibile aggiungere ulteriori fasi inviate dall'utente al cluster fino a che non viene raggiunto il limite di 1.000 fasi. Quando si raggiunge il limite di 1.000 fasi aggiunte a un cluster, l'invio di ulteriori fasi comporta la rimozione dei record delle fasi inviate dall'utente meno recenti. Questi record non vengono rimossi dai file di log, ma non sono più visualizzati nella console e nemmeno quando utilizzi l'AWS CLI o l'API per recuperare informazioni del cluster. I record relativi alle fasi di sistema non vengono mai rimossi.

Le informazioni sulle fasi che puoi visualizzare dipendono dal meccanismo utilizzato per recuperare le informazioni del cluster. La tabella seguente indica le informazioni sulle fasi restituite da ciascuna delle opzioni disponibili.

Opzione	DescribeJobFlow o --describe --jobflow	ListSteps o list-steps
SDK	256 fasi	1.000 fasi
CLI di Amazon EMR	256 fasi	N/A
AWS CLI	N/A	1.000 fasi
API	256 fasi	1.000 fasi

Annullamento delle fasi

Puoi annullare le fasi in sospeso e in esecuzione dalla AWS Management Console, dalla AWS CLI o dall'API Amazon EMR.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Annullamento di fasi con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Nella scheda Steps (Fasi) della pagina dei dettagli del cluster, seleziona la casella di controllo accanto alla fase da annullare. Scegli il menu a discesa Actions (Operazioni) e seleziona Cancel steps (Annulla fasi).
4. Nella finestra di dialogo Cancel the step (Annulla fase), scegli se annullare la fase e attenderne l'uscita oppure annullare la fase e forzarne l'uscita. Quindi scegli Conferma.

5. Lo stato delle fasi nella tabella Steps (Fasi) diventa `CANCELLED`.

Old console

Annullamento di fasi con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Nella pagina Cluster Details (Dettagli del cluster), espandere la sezione Steps (Fasi).
3. Per ogni fase che si desidera annullare, selezionare dall'elenco Steps (Fasi). Quindi scegliere Cancel step (Annulla fase).
4. Nella finestra di dialogo Cancel step (Annulla passaggio) mantenere l'opzione predefinita Cancel the step and wait for it to exit (Annulla la fase e attendi l'uscita). Se si desidera terminare immediatamente la fase senza attendere il completamento di alcun processo, scegliere Cancel the step and force it to exit (Annulla la fase e forza l'uscita).
5. Scegliere Cancel step (Annulla passaggio).

CLI

Annullamento di fasi con la AWS CLI

- Utilizzare il comando `aws emr cancel-steps`, specificando il cluster e le fasi da annullare. L'esempio seguente illustra un comando dell'AWS CLI per annullare due fasi.

```
aws emr cancel-steps --cluster-id j-2QUAXXXXXXXXX \
--step-ids s-3M8DXXXXXXXXX s-3M8DXXXXXXXXX \
--step-cancellation-option SEND_INTERRUPT
```

Con Amazon EMR versione 5.28.0, è possibile scegliere una delle due seguenti opzioni di annullamento per il parametro `StepCancellationOption` durante l'annullamento delle fasi.

- `SEND_INTERRUPT`: questa è l'opzione predefinita. Quando riceve una richiesta di annullamento fase, EMR invia un segnale `SIGTERM` alla fase. Aggiungi un gestore di segnale `SIGTERM` alla logica di fase per rilevare questo segnale e terminare i processi discendenti della fase o attendi che questi vengano completati.

- `TERMINATE_PROCESS`: quando questa opzione è selezionata, EMR invia un segnale `SIGKILL` alla fase e a tutti i relativi processi discendenti, i quali vengono terminati immediatamente.

Considerazioni per l'annullamento delle fasi

- L'annullamento di una fase in esecuzione o in sospeso rimuove tale fase dal numero di fasi attive.
- L'annullamento di una fase in esecuzione non consente l'avvio di una fase in sospeso, presupponendo l'assenza di modifiche a `stepConcurrencyLevel`.
- L'annullamento di una fase in esecuzione non attiva il `ActionOnFailure` della fase.
- Per EMR 5.32.0 e versioni successive, `SEND_INTERRUPT StepCancellationOption` invia un segnale `SIGTERM` al processo figlio della fase. Sarebbe opportuno rilevare questo segnale ed eseguire una pulizia e uno spegnimento corretti. `TERMINATE_PROCESS StepCancellationOption` invia un segnale `SIGKILL` al processo figlio della fase e a tutti i relativi processi discendenti; tuttavia, i processi asincroni non sono interessati.

Visualizzazione e monitoraggio di un cluster

Amazon EMR fornisce vari strumenti utili per raccogliere informazioni sul cluster. Puoi accedere a le informazioni relative al cluster dalla console, dall'interfaccia a riga di comando (CLI) o a livello di codice. Le interfacce Web Hadoop standard e i file di log sono disponibili sul nodo primario. Puoi inoltre utilizzare i servizi di monitoraggio come CloudWatch e Ganglia per tenere traccia delle prestazioni del cluster.

La cronologia delle applicazioni è disponibile anche dalla console utilizzando l'interfaccia utente dell'applicazione "persistente" per Spark History Server a partire da Amazon EMR 5.25.0. Con Amazon EMR 6.x, sono disponibili anche il server della timeline YARN persistente e le interfacce utente Tez. Questi servizi sono ospitati fuori dal cluster, quindi è possibile accedere alla cronologia delle applicazioni per 30 giorni dopo la chiusura del cluster, senza la necessità di una connessione SSH o di un proxy Web. Consulta Visualizzazione della cronologia dell'applicazione

Argomenti

- Visualizzazione dello stato e dei dettagli di un cluster
- Debug migliorato delle fasi
- Visualizzazione della cronologia dell'applicazione
- Visualizzare file di log di

- Visualizzazione di istanze cluster in Amazon EC2
- Eventi e parametri CloudWatch
- Visualizzazione dei parametri dell'applicazione cluster con Ganglia
- Registrazione delle chiamate API di Amazon EMR in AWS CloudTrail

Visualizzazione dello stato e dei dettagli di un cluster

Dopo la creazione di un cluster, puoi monitorarne lo stato e ottenere informazioni dettagliate sull'esecuzione dello stesso e sugli errori che possono verificarsi, anche dopo la terminazione del cluster. Per scopi di riferimento, Amazon EMR conserva i metadati relativi ai cluster terminati per due mesi; dopo questo periodo, i metadati vengono eliminati. Non puoi eliminare cluster dalla cronologia dei cluster, ma mediante AWS Management Console puoi utilizzare il comando Filter (Filtro) e con AWS CLI puoi utilizzare opzioni con il comando `list-clusters` per concentrare l'attenzione sui cluster che ti interessano.

Puoi accedere alla cronologia dell'applicazione archiviata nel cluster per una settimana dal momento in cui viene registrata, indipendentemente dal fatto che il cluster sia in esecuzione o terminato. Inoltre, le interfacce utente delle applicazioni persistenti archiviano la cronologia delle applicazioni fuori cluster per 30 giorni dopo la chiusura di un cluster. Consulta Visualizzazione della cronologia dell'applicazione

Per ulteriori informazioni sugli stati del cluster, ad esempio Waiting (In attesa) e Running (In esecuzione), consulta Comprensione del ciclo di vita del cluster.

Visualizzazione dei dettagli del cluster con la AWS Management Console

L'elenco Cluster nella pagina https://console.aws.amazon.com/emr elenca tutti i cluster nel tuo account e nella tua Regione AWS, inclusi i cluster terminati. L'elenco contiene le informazioni seguenti per ogni cluster: Nome e ID, Stato e Dettagli stato, Data di creazione, Tempo trascorso dal momento dell'esecuzione del cluster e Ore istanze normalizzate accumulate per tutte le istanze EC2 nel cluster. Questo elenco è il punto di partenza per monitorare lo stato dei cluster. È concepito per eseguire il drill-down dei dettagli di ogni cluster per scopi di analisi e di risoluzione dei problemi.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Visualizzazione delle informazioni di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da visualizzare.
3. Utilizza il pannello Summary (Riepilogo) per visualizzare informazioni di base sulla configurazione del cluster, ad esempio lo stato del cluster, le applicazioni open source che Amazon EMR ha installato nel cluster e la versione di Amazon EMR utilizzata per creare il cluster. Consulta ciascuna scheda in Summary (Riepilogo) per visualizzare le informazioni come descritto nella sezione seguente.

Old console

Visualizzazione delle informazioni di un cluster con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Per visualizzare un riepilogo abbreviato delle informazioni sul cluster, seleziona la freccia rivolta verso il basso accanto al collegamento del cluster in Name (Nome). La riga del cluster viene espansa per fornire ulteriori informazioni su cluster, hardware, fasi e operazioni di bootstrap. Utilizzare i collegamenti in questa sezione per eseguire il drill-down delle specifiche. Ad esempio, fare clic su un collegamento in Steps (Fasi) per accedere ai file di log delle fasi, visualizzare il JAR associato alla fase, eseguire il drill-down dei processi e delle attività della fase e accedere ai file di log.

3. Per visualizzare le informazioni dettagliate, scegli il collegamento del cluster in Name (Nome) per aprire la pagina dei dettagli del cluster. Le seguenti informazioni sono disponibili nella pagina dei dettagli del cluster nella vecchia console:

Scheda (vecchia console)	Descrizione (vecchia console)
Proprietà	Utilizza questa scheda per visualizzare il sistema operativo, la terminazione e le configurazioni di sicurezza del cluster, le informazioni sul VPC e sulla sottorete e la posizione di archiviazione dei log in Amazon S3.
Bootstrap actions (Operazioni di bootstrap)	Utilizzare questa scheda per visualizzare lo stato di qualsiasi operazione di bootstrap che il cluster esegue quando avviato. Le operazioni di bootstrap sono utilizzate per le installazioni software personalizzate e la configurazione avanzata. Per ulteriori informazioni, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo.
Monitoraggio	Utilizza questa scheda per visualizzare i parametri chiave del funzionamento del cluster. È possibile visualizzare dati a livello del cluster e a livello dei nodi nonché informazioni sulle operazioni di I/O e sullo storage dei dati.
Istanze	Utilizza questa scheda per visualizzare informazioni sui nodi nel cluster, tra cui ID delle istanze EC2, nomi DNS, volumi EBS e altro ancora.
Fasi	Utilizzare questa scheda per visualizzare lo stato e accedere ai file di log per le fasi

Scheda (vecchia console)	Descrizione (vecchia console)
	inviate. Per ulteriori informazioni sulle fasi, consulta Invio di lavoro a un cluster.
Applicazioni	Utilizzare questa scheda per visualizzare i dettagli dell'applicazione dell'interfaccia utente Tez e del server della timeline YARN persistente fuori cluster. È inoltre possibile visualizzare informazioni relative alle applicazioni installate, alle configurazioni del cluster e ai gruppi di istanze. Le interfacce utente delle applicazioni in cluster sono disponibili mentre il cluster è in esecuzione.
Eventi	Utilizzare questa scheda per visualizzare il log di eventi per il cluster. Per ulteriori informazioni, consulta Monitoraggio di eventi di Amazon EMR con CloudWatch.
Tag	Utilizza questa scheda per visualizzare tutti i tag che hai applicato al cluster.

Visualizzazione dei dettagli del cluster con la AWS CLI

I seguenti esempi illustrano come recuperare informazioni dettagliate sul cluster mediante l'AWS CLI. Per ulteriori informazioni sui comandi disponibili, consulta la Guida di riferimento ai comandi della AWS CLI per Amazon EMR. Puoi utilizzare il comando describe-cluster per visualizzare dettagli a livello del cluster, tra cui stato, configurazione hardware e software, impostazioni VPC, operazioni di bootstrap, gruppi di istanze e così via. Per ulteriori informazioni sugli stati del cluster, consulta Comprensione del ciclo di vita del cluster. L'esempio seguente illustra l'utilizzo del comando `describe-cluster`, seguito da esempi del comando list-clusters.

Example Visualizzazione dello stato del cluster

Per utilizzare il comando `describe-cluster`, è necessario l'ID del cluster. Questo esempio illustra come ottenere un elenco di cluster creati entro un determinato intervallo di tempo, nonché l'utilizzo di uno degli ID di cluster restituiti per visualizzare ulteriori informazioni sullo stato di un singolo cluster.

Il comando seguente descrive il cluster *j-1K48XXXXXXHCB*, che sostituisci con l'ID del cluster.

```
aws emr describe-cluster --cluster-id j-1K48XXXXXXHCB
```

L'output del comando è simile a quanto segue:

```
{
    "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 Visualizzazione dei cluster per data di creazione

Per recuperare i cluster creati entro un determinato intervallo di tempo, utilizza il comando `list-clusters` con i parametri `--created-after` e `--created-before`.

Il comando seguente elenca tutti i cluster creati tra il 9 ottobre 2019 e il 12 ottobre 2019.

```
aws emr list-clusters --created-after 2019-10-09T00:12:00 --created-
before 2019-10-12T00:12:00
```

Example Visualizzazione dei cluster per stato

Per elencare i cluster per stato, utilizza il comando `list-clusters` con il parametro `--cluster-states`. Lo stato dei cluster validi può essere: STARTING, BOOTSTRAPPING, RUNNING, WAITING, TERMINATING, TERMINATED e TERMINATED_WITH_ERRORS.

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

Puoi anche utilizzare i parametri seguenti per elencare tutti i cluster negli stati specificati.

- `--active` filtra i cluster il cui stato è STARTING, BOOTSTRAPPING, RUNNING, WAITING o TERMINATING.
- `--terminated` filtra i cluster il cui stato è TERMINATED.
- Il parametro `--failed` filtra i cluster il cui stato è TERMINATED_WITH_ERRORS.

I seguenti comandi restituiscono lo stesso risultato.

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

```
aws emr list-clusters --terminated
```

Per ulteriori informazioni sugli stati del cluster, consulta Comprensione del ciclo di vita del cluster.

Debug migliorato delle fasi

Se una fase di Amazon EMR ha esito negativo e hai inviato il lavoro utilizzando l'operazione API Step con un'AMI versione 5.x o successiva, in alcuni casi Amazon EMR può identificare e restituire la causa principale dell'errore nella fase insieme al nome del file di log pertinente e una parte della traccia di stack dell'applicazione mediante l'API. Ad esempio, è possibile identificare i seguenti errori:

- Un errore Hadoop comune come una directory di output già esistente, una directory di input inesistente o un'applicazione con memoria insufficiente.
- Errori Java, ad esempio un'applicazione compilata con una versione incompatibile di Java o eseguita con una classe principale introvabile.
- Un problema di accesso agli oggetti archiviati in Amazon S3.

Queste informazioni sono disponibili mediante le operazioni API DescribeStep e ListSteps. Il campo FailureDetails dello StepSummary restituito da tali operazioni. Per accedere alle informazioni FailureDetails, utilizza la CLI AWS, la console oppure l'SDK AWS.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

La nuova console Amazon EMR non offre il debug delle fasi. Tuttavia, è possibile visualizzare i dettagli della terminazione del cluster con la procedura seguente.

Visualizzazione dei dettagli sugli errori con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da visualizzare.
3. Prendi nota del valore Status (Stato) nella sezione Summary (Riepilogo) della pagina dei dettagli del cluster. Se lo stato è Terminated with errors (Terminato con errori), passa con il mouse sul testo per visualizzare i dettagli degli errori del cluster.

Old console

Visualizzazione dei dettagli sugli errori con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegliere Cluster List (Elenco cluster) e selezionare un cluster.
3. Selezionare l'icona a forma di freccia accanto a ogni fase per visualizzare ulteriori dettagli. Se la fase non è andata a buon fine e Amazon EMR è in grado di identificare la causa principale, vengono visualizzati i dettagli dell'errore.

CLI

Visualizzazione dei dettagli sugli errori con la AWS CLI

- Per ottenere i dettagli sugli errori di una fase con la AWS CLI, utilizza il comando `describe-step`.

```
aws emr describe-step --cluster-id j-1K48XXXXXHCB --step-id s-3QM0XXXXXM1W
```

L'output risulterà simile al seguente:

```
{
  "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"
  }
```

```
}
```

Visualizzazione della cronologia dell'applicazione

È possibile visualizzare i dettagli delle applicazioni Spark History Server e Timeline Server di YARN dallla pagina dei dettagli del cluster nella console. La cronologia dell'applicazione di Amazon EMR facilita la risoluzione dei problemi e l'analisi dei processi attivi e della cronologia dei processi.

New console

La sezione Application user interfaces (Interfacce utente delle applicazioni) della scheda Applications (Applicazioni) offre diverse opzioni di visualizzazione, a seconda dello stato del cluster e delle applicazioni installate nel cluster.

- Accesso dall'esterno del cluster alle interfacce utente delle applicazioni persistenti: a partire dalla versione 5.25.0 di Amazon EMR, i collegamenti alle interfacce utente delle applicazioni persistenti sono disponibili per l'interfaccia utente di Spark e per Spark History Service. Con Amazon EMR versione 5.30.1 e successive, anche Tez UI e il timeline server di YARN hanno interfacce utente di applicazioni persistenti. Il server timeline YARN e l'interfaccia utente Tez sono applicazioni open source che forniscono parametri e strumenti visivi per i cluster attivi e terminati. L'interfaccia utente Spark fornisce dettagli sulle fasi e le attività del pianificatore, le dimensioni di RDD e l'utilizzo della memoria, le informazioni ambientali e le informazioni sugli esecutori attivi. Le interfacce utente dell'applicazione persistente vengono eseguite fuori dal cluster, pertanto le informazioni e i registri del cluster sono disponibili per 30 giorni dopo la chiusura di un'applicazione. A differenza delle interfacce utente dell'applicazione nel cluster, le interfacce utente dell'applicazione persistente non richiedono l'impostazione di un proxy Web tramite una connessione SSH.
- Interfacce utente delle applicazioni nel cluster: esistono diverse interfacce utente della cronologia di applicazioni che possono essere eseguite in un cluster. Le interfacce utente nel cluster sono ospitate sul nodo master e richiedono l'impostazione di una connessione SSH al server Web. Le interfacce utente delle applicazioni nel cluster mantengono la cronologia delle applicazioni per una settimana dopo la fine di un'applicazione. Per ulteriori informazioni e istruzioni su come configurare un tunnel SSH, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

A eccezione di Spark History Server, del Timeline Server di YARN e delle applicazioni Hive, la cronologia delle applicazioni nel cluster può essere visualizzata solo mentre il cluster è in esecuzione.

Old console

La scheda Interfacce utente applicazione fornisce diverse opzioni di visualizzazione:

- Accesso fuori cluster alle interfacce utente delle applicazioni persistenti: a partire dalla versione 5.25.0 di Amazon EMR, i collegamenti alle interfacce utente delle applicazioni persistenti sono disponibili per Spark. Con Amazon EMR versione 5.30.1 e successive, anche Tez UI e il timeline server di YARN hanno interfacce utente di applicazioni persistenti. Il server timeline YARN e l'interfaccia utente Tez sono applicazioni open source che forniscono parametri e strumenti visivi per i cluster attivi e terminati. L'interfaccia utente Spark fornisce dettagli sulle fasi e le attività del pianificatore, le dimensioni di RDD e l'utilizzo della memoria, le informazioni ambientali e le informazioni sugli esecutori attivi. Le interfacce utente dell'applicazione persistente vengono eseguite fuori dal cluster, pertanto le informazioni e i registri del cluster sono disponibili per 30 giorni dopo la chiusura di un'applicazione. A differenza delle interfacce utente dell'applicazione nel cluster, le interfacce utente dell'applicazione persistente non richiedono l'impostazione di un proxy Web tramite una connessione SSH.
- Interfacce utente delle applicazioni nel cluster: esistono diverse interfacce utente della cronologia di applicazioni che possono essere eseguite in un cluster. Le interfacce utente nel cluster sono ospitate sul nodo master e richiedono l'impostazione di una connessione SSH al server Web. Le interfacce utente delle applicazioni nel cluster mantengono la cronologia delle applicazioni per una settimana dopo la fine di un'applicazione. Per ulteriori informazioni e istruzioni su come configurare un tunnel SSH, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

  A eccezione di Spark History Server, del Timeline Server di YARN e delle applicazioni Hive, la cronologia delle applicazioni nel cluster può essere visualizzata solo mentre il cluster è in esecuzione.
- Cronologia delle applicazioni di alto livello: con Amazon EMR rilasci da 5.8.0 a 5.36.0 e rilasci della serie 6.x fino a 6.8.0, è possibile visualizzare un riepilogo della cronologia delle applicazioni nella vecchia console EMR, inclusi i parametri chiave delle attività e degli esecutori di fase. A partire dal 23 gennaio 2023, Amazon EMR interromperà la cronologia

delle applicazioni di alto livello per tutte le versioni. Se utilizzi Amazon EMR versione 5.25.0 o successive, ti consigliamo di utilizzare invece l'interfaccia utente persistente dell'applicazione.

Visualizzazione di interfacce utente delle applicazioni persistenti

A partire da Amazon EMR versione 5.25.0, è possibile connettersi ai dettagli dell'applicazione Spark History Server persistente ospitata fuori cluster utilizzando la pagina Summary (Riepilogo) o la scheda Application user interfaces (Interfacce utente applicazioni) nella console. Le interfacce utente delle applicazioni persistenti Timeline Server di YARN e Tez sono disponibili a partire dalla versione Amazon EMR 5.30.1. L'accesso al collegamento con un clic alla cronologia delle applicazioni persistenti offre i seguenti vantaggi:

- È possibile analizzare e risolvere rapidamente i processi attivi e la cronologia dei processi senza configurare un proxy Web tramite una connessione SSH.
- È possibile accedere alla cronologia delle applicazioni e ai file di log pertinenti per cluster attivi e chiusi. I log sono disponibili per 30 giorni dopo la fine dell'applicazione.

Nella scheda Application user interfaces (Interfacce utente delle applicazioni) o nella pagina Summary (Riepilogo) del cluster nella vecchia console Amazon EMR 5.30.1 o 6.x, scegli YARN timeline server, Tez UI (Interfaccia utente di Tez) o il collegamento Spark history server.

L'interfaccia utente dell'applicazione si apre in una nuova scheda del browser. Per ulteriori informazioni, consulta Monitoraggio e strumentazione.

È possibile visualizzare i log dei container YARN tramite i collegamenti sul server cronologia Spark, sul server della timeline YARN e sull'interfaccia utente Tez.

> **Note**
>
> Per accedere ai log del container YARN da Spark History Server e dall'interfaccia utente Tez è necessario abilitare la registrazione ad Amazon S3 per il cluster. Se la registrazione non è abilitata, i collegamenti ai log del container YARN non funzioneranno.

Raccolta di log

Per abilitare l'accesso con un solo clic alle interfacce utente delle applicazioni persistenti, Amazon EMR raccoglie due tipi di log:

- Ilog di eventi dell'applicazione vengono raccolti in un bucket di sistema EMR. I log di eventi vengono crittografati mentre sono inattivi utilizzando la crittografia lato server con chiavi gestite da Amazon S3 (SSE-S3). Se si utilizza una sottorete privata per il cluster, assicurarsi di includere `“arn:aws:s3:::prod.MyRegion.appinfo.src/*”` nell'elenco delle risorse della policy Amazon S3 per la sottorete privata. Per ulteriori informazioni, consulta Policy Amazon S3 minima per sottorete privata.
- I log del container YARN vengono raccolti in un bucket Amazon S3 di proprietà. È necessario abilitare la registrazione per il cluster per accedere ai log del container YARN. Per ulteriori informazioni, consulta Configurazione della registrazione e del debug di cluster.

Se è necessario disabilitare questa caratteristica per motivi di privacy, è possibile arrestare il daemon utilizzando uno script bootstrap quando si crea un cluster, come illustrato nell'esempio seguente.

```
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"]
```

Dopo aver eseguito questo script bootstrap, Amazon EMR non raccoglierà i log di eventi di Spark History Server o del Timeline Server di YARN nel bucket di sistema EMR. Nessuna informazione sulla cronologia delle applicazioni sarà disponibile nella scheda Interfacce utente applicazione e si perderà l'accesso a tutte le interfacce utente dell'applicazione dalla console.

File di log degli eventi Spark di grandi dimensioni

In alcuni casi, i job Spark a esecuzione prolungata, come lo streaming Spark, e i job di grandi dimensioni, come le query SQL Spark, possono generare registri di eventi di grandi dimensioni. Con registri di eventi di grandi dimensioni, puoi utilizzare rapidamente lo spazio su disco sulle istanze di calcolo e su `OutOfMemory` errori riscontrati durante il caricamento delle UI persistenti. Per evitare questi problemi, si consiglia di abilitare la caratteristica di rolling e compattazione del log eventi Spark. Questa caratteristica è disponibile solo nella versione Amazon EMR mr-6.1.0 e successive. Per ulteriori dettagli su rolling e compattazione, consulta Applicazione della compattazione ai file di log degli eventi nella documentazione di Spark.

Per attivare la funzione di rolling e compattazione del log degli eventi di Spark, attiva le seguenti impostazioni di configurazione di Spark.

- `spark.eventLog.rolling.enabled`— Attiva la rolling del log degli eventi in base alle dimensioni. Questa impostazione è disattivata per impostazione predefinita.
- `spark.eventLog.rolling.maxFileSize`— Quando la rolling è attivata, specifica la dimensione massima del file di log degli eventi prima che venga eseguito il rollover. Il valore predefinito è 128 MB.
- `spark.history.fs.eventLog.rolling.maxFilesToRetain`— Specifica il numero massimo di file di log eventi non compatti da mantenere. Per impostazione predefinita, tutti i file di log degli eventi vengono mantenuti. Imposta un numero inferiore per compattare i registri degli eventi più vecchi. Il valore più basso è 1.

Nota che la compattazione tenta di escludere eventi con file di log eventi obsoleti, come i seguenti. Se elimina gli eventi, non li vedrai più nell'interfaccia utente di Spark History Server.

- Eventi per i lavori terminati e relativi eventi relativi alla fase o all'attività.
- Eventi per esecutori licenziati.
- Eventi per le interrogazioni SQL completate e i relativi eventi relativi a job, stage e attività.

Per avviare un cluster con funzionalità di rolling e compattazione abilitate

1. Creare un file `spark-configuration.json` con la seguente configurazione.

```
[
   {
     "Classification": "spark-defaults",
        "Properties": {
           "spark.eventLog.rolling.enabled": true,
           "spark.history.fs.eventLog.rolling.maxFilesToRetain": 1
        }
   }
]
```

2. Crea un cluster con la configurazione Spark Rolling Compaction come segue.

```
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
```

Considerazioni e limitazioni

L'accesso con un solo clic alle interfacce utente dell'applicazione persistente presenta attualmente le seguenti limitazioni.

- Ci sarà un ritardo di almeno due minuti quando i dettagli dell'applicazione vengono visualizzati nell'interfaccia utente di Spark History Server.
- Questa caratteristica è attiva solo quando la directory dei log di eventi per l'applicazione si trova in HDFS. Per impostazione predefinita, Amazon EMR archivia i log di eventi in una directory di HDFS. Se si modifica la directory predefinita in un file system diverso, ad esempio Amazon S3, questa funzionalità non sarà attiva.
- Al momento, questa caratteristica non è disponibile per i cluster EMR con più nodi principali o per i cluster EMR integrati con AWS Lake Formation.
- Per abilitare l'accesso con un solo clic alle interfacce utente dell'applicazione persistente, è necessario disporre dell'autorizzazione per l'operazione `DescribeCluster` per Amazon EMR. Se si nega l'autorizzazione di un principal IAM a questa operazione, occorrono circa cinque minuti per la propagazione della modifica dell'autorizzazione.
- Se si riconfigurano le applicazioni in un cluster in esecuzione, la cronologia delle applicazioni non sarà disponibile tramite l'interfaccia utente dell'applicazione.
- Per ogni Account AWS, il limite predefinito per le interfacce utente delle applicazioni attive è 200.
- Puoi accedere alle interfacce utente delle applicazioni dalla console nelle Regioni Stati Uniti orientali (Virginia settentrionale e Ohio), Stati Uniti occidentali (California settentrionale e Oregon), Canada (Centrale), Europa (Francoforte, Irlanda, Londra, Parigi e Stoccolma), Asia Pacifico (Mumbai, Seoul, Singapore, Sydney e Tokyo), Sud America (San Paolo), Cina (Pechino) gestita da Sinnet e Cina (Ningxia) gestita da NWCD.

Visualizzazione della cronologia di applicazioni di alto livello

> **Note**
>
> Ti consigliamo di utilizzare l'interfaccia persistente dell'applicazione per un'esperienza utente migliore, in grado di conservare la cronologia delle applicazioni fino a un massimo di 30 giorni. La cronologia delle applicazioni di alto livello descritta in questa pagina non è disponibile nella nuova console Amazon EMR (https://console.aws.amazon.com/emr). Per ulteriori informazioni, consulta Visualizzazione di interfacce utente delle applicazioni persistenti.

Con Amazon EMR rilasci da 5.8.0 a 5.36.0 e rilasci della serie 6.x fino a 6.8.0, puoi visualizzare la cronologia di applicazioni di alto livello dalla scheda Application user interfaces (Interfacce utente delle applicazioni) nella vecchia console Amazon EMR. Una Application user interface (Interfaccia utente dell'applicazione) di Amazon EMR conserva il riepilogo della cronologia delle applicazioni per 7 giorni dopo il completamento di un'applicazione.

Considerazioni e limitazioni

Considera i seguenti limiti quando utilizzi la scheda Interfacce utente applicazioni nella precedente console di Amazon EMR.

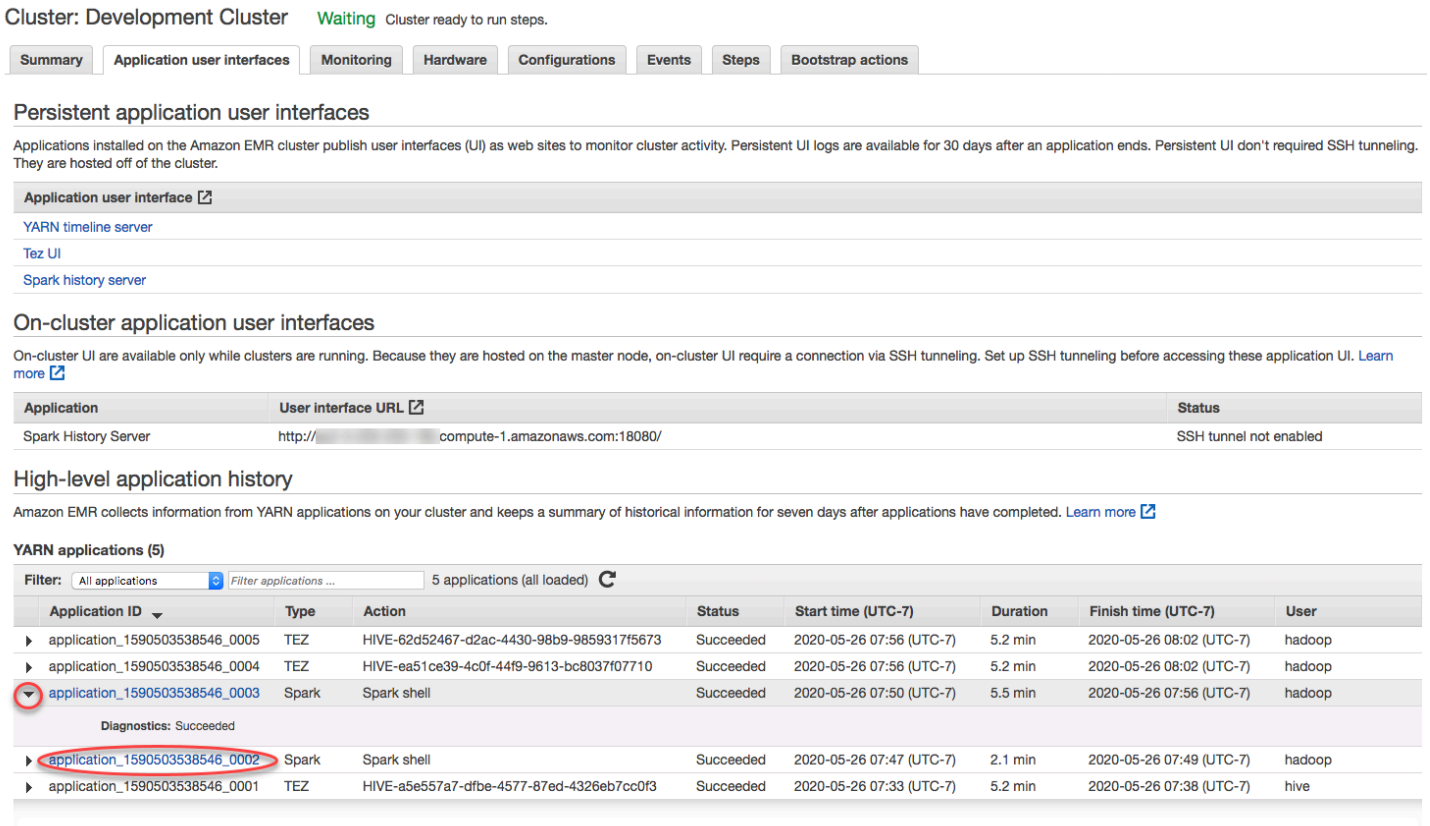
- Puoi accedere alla funzionalità di cronologia delle applicazioni di alto livello solo se utilizzi Amazon EMR versioni da 5.8.0 a 5.36.0 e versioni della serie 6.x fino a 6.8.0. A partire dal 23 gennaio 2023, Amazon EMR interromperà la cronologia delle applicazioni di alto livello per tutte le versioni. Se utilizzi Amazon EMR versione 5.25.0 o successive, ti consigliamo di utilizzare invece l'interfaccia utente persistente dell'applicazione.
- La caratteristica della cronologia delle applicazioni di alto livello non supporta le applicazioni Spark Streaming.
- L'accesso con un clic alle interfacce utente delle applicazioni persistenti non è attualmente disponibile per i cluster Amazon EMR con più nodi principali o per i cluster Amazon EMR integrati con AWS Lake Formation.

Esempio: Visualizzazione di una cronologia delle applicazioni di alto livello

Nella sequenza seguente viene illustrato un drill-down nei dettagli del processo attraverso un'applicazione Spark o YARN utilizzando la scheda Application user interfaces (Interfacce utente delle applicazioni) nella pagina dei dettagli del cluster nella vecchia console.

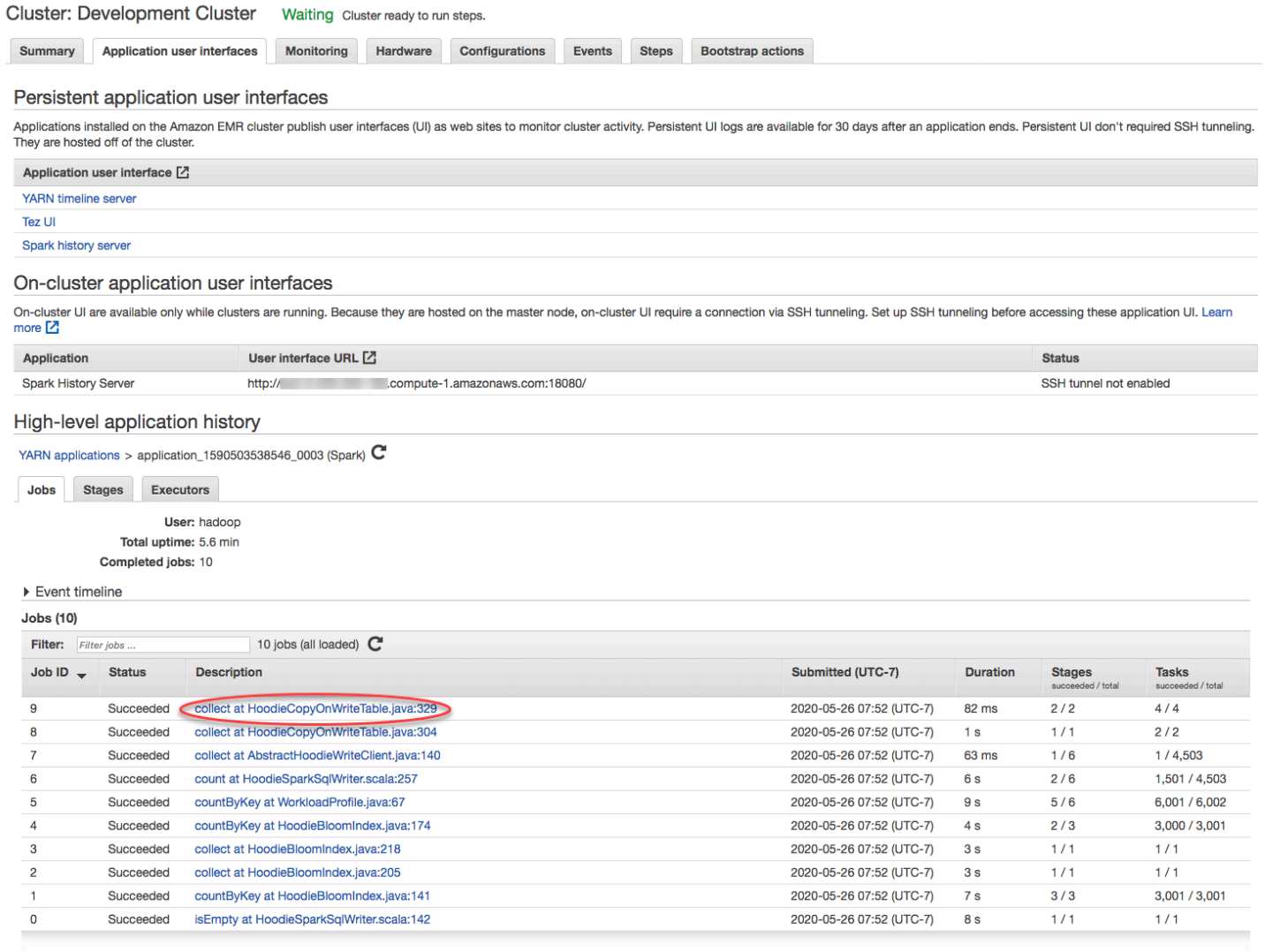
Per visualizzare i dettagli del cluster, seleziona un Name (Nome) del cluster dall'elenco Clusters (Cluster). Per visualizzare informazioni sui log del container YARN, è necessario abilitare la registrazione per il cluster. Per ulteriori informazioni, consulta Configurazione della registrazione e del debug di cluster. Per la cronologia delle applicazioni Spark, le informazioni fornite nella tabella di riepilogo sono solo un sottoinsieme delle informazioni disponibili tramite l'interfaccia utente di Spark History Server.

Nella scheda Application user interfaces (Interfacce utente applicazioni), sotto High-level application history (Cronologia delle applicazioni di alto livello), è possibile espandere una riga per visualizzare il riepilogo diagnostico per un'applicazione Spark o selezionare un Application ID (ID applicazione) per visualizzare i dettagli di un'applicazione diversa.

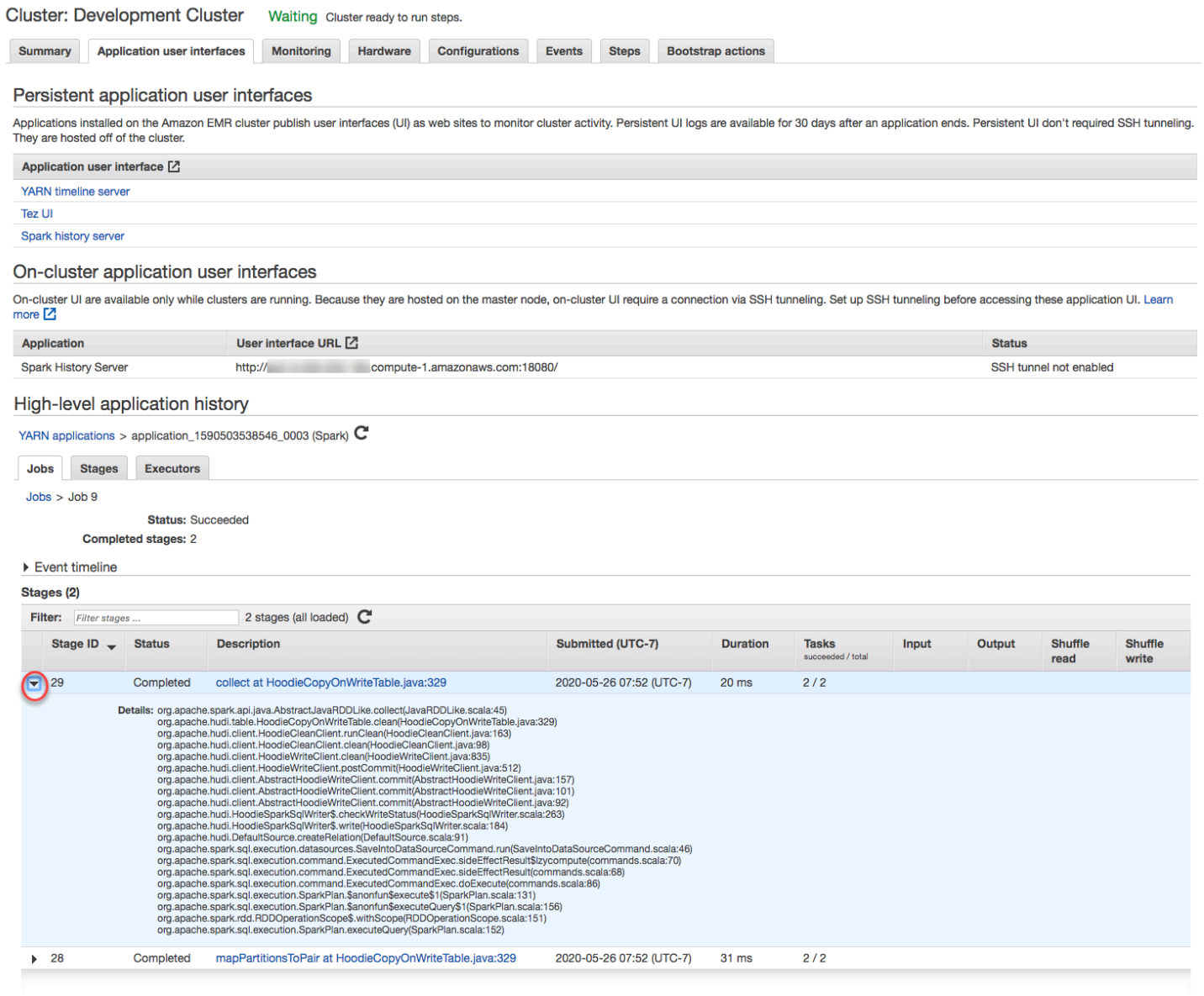


Quando si seleziona un Application ID (ID applicazione), l'interfaccia utente cambia per mostrare i dettagli della YARN application (Applicazione YARN) relativi a tale applicazione. Nella scheda

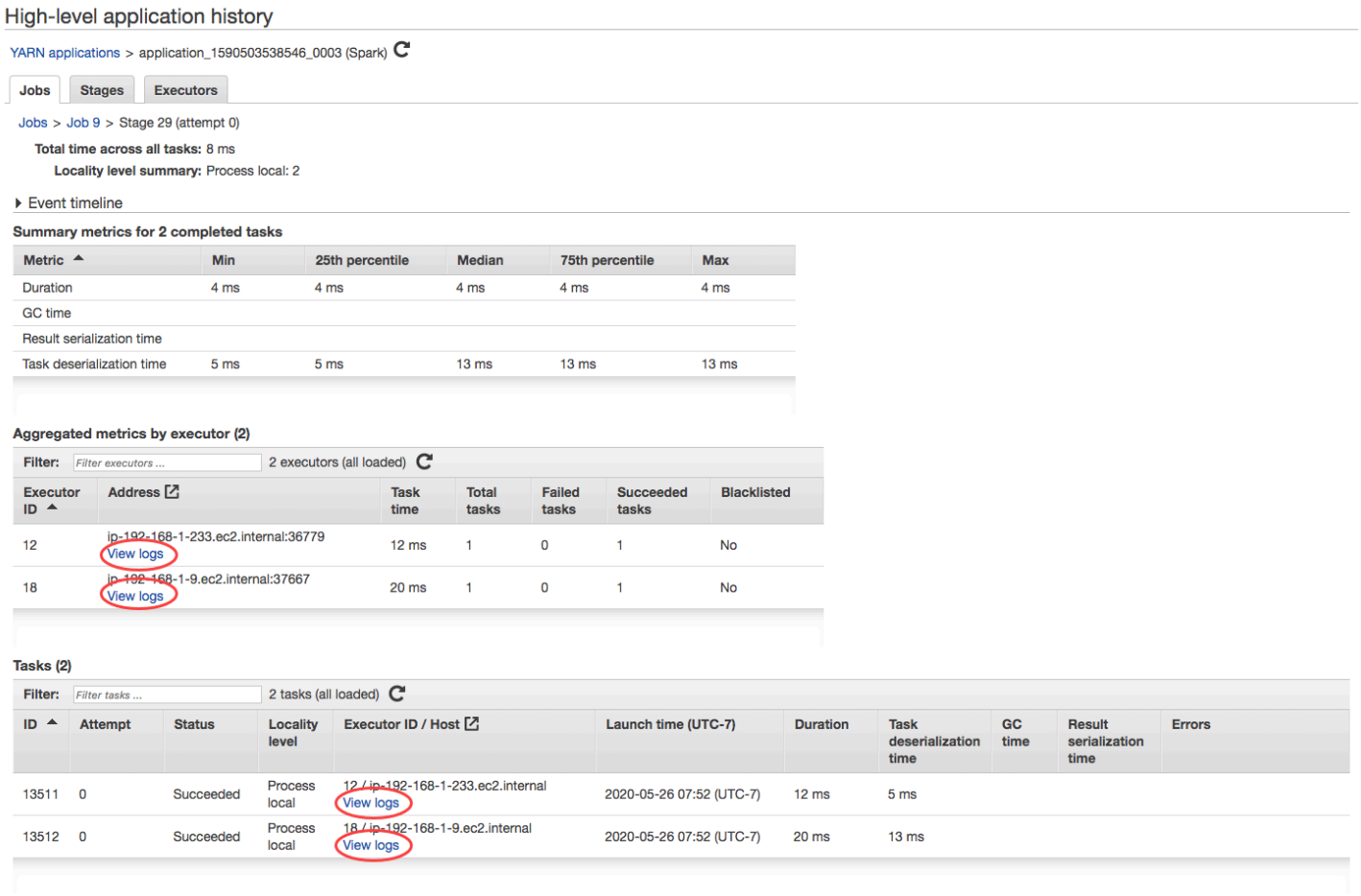
Jobs (Processi) dei dettagli della YARN application (Applicazione YARN), è possibile scegliere la Description (Descrizione) di un processo per visualizzarne i dettagli.



Nella pagina dei dettagli del processo, puoi espandere le informazioni sulle singole fasi del processo e selezionare il collegamento Description (Descrizione) per vedere i dettagli della fase.



Nella pagina dei dettagli della fase è possibile visualizzare i parametri principali delle attività e degli executor della fase. È inoltre possibile visualizzare i log delle attività e degli executor utilizzando i collegamenti View logs (Visualizza log).



Visualizzare file di log di

Amazon EMR e Hadoop producono entrambi file di log che comunicano lo stato sul cluster. Per impostazione predefinita, questi file vengono scritti nel nodo primario nella directory `/mnt/var/log/`. A seconda di come il cluster è stato configurato quando è stato avviato, questi log possono anche essere archiviati in Amazon S3 ed essere visualizzati tramite lo strumento di debug grafico.

Esistono molti tipi di log scritti nel nodo primario. Amazon EMR scrive log di stato della fase, dell'operazione di bootstrap e dell'istanza. Apache Hadoop scrive i log per comunicare l'elaborazione di processi, attività e tentativi di attività. Hadoop registra inoltre i log dei suoi daemon. Per ulteriori informazioni sui log scritti da Hadoop, vai a http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/ClusterSetup.html.

Visualizzazione di file di log sul nodo primario

Nella tabella seguente vengono elencati alcuni dei file di log che si trovano sul nodo primario.

Ubicazione	Descrizione
/emr/instance-controller/log/bootstrap-actions	Log scritti durante l'elaborazione delle operazioni di bootstrap.
/mnt/var/log/hadoop-state-pusher	Log scritti dal processo pusher dello stato di Hadoop.
/emr/instance-controller/log	Log del controller istanze.
/emr/instance-state	Log degli stati istanza. Contengono informazioni su CPU, stato della memoria e thread del garbage collector del nodo.
/emr/service-nanny	Log scritti dal processo nanny del servizio.
/mnt/var/log/*application*	Log specifici di un'applicazione, ad esempio Hadoop, Spark, o Hive.
/mnt/var/log/hadoop/steps/*N*	Log di fase contenenti informazioni sull'elaborazione della fase. Il valore di *N* indica lo stepId assegnato da Amazon EMR. Ad esempio, un cluster ha due fasi: `s-1234ABCDEFGH` e `s-5678IJKLMNOP`. La prima fase si trova in `/mnt/var/log/hadoop/steps/s-1234ABCDEFGH/` e la seconda in `/mnt/var/log/hadoop/steps/s-5678IJKLMNOP/`. I log di fase scritti da Amazon EMR sono i seguenti. • controller: informazioni sull'elaborazione della fase. Se la fase ha esito negativo durante il caricamento, puoi trovare la traccia di stack in questo log. • syslog: descrive l'esecuzione dei processi Hadoop nella fase.

Ubicazione	Descrizione
	• stderr: il canale di errore standard di Hadoop durante l'elaborazione della fase. • stdout: il canale di uscita standard di Hadoop durante l'elaborazione della fase.

Visualizzazione dei file di log sul nodo primario con la AWS CLI.

1. Utilizza SSH per connetterti al nodo primario come descritto in Connessione al nodo primario tramite SSH.
2. Passare alla directory contenente le informazioni sul file di log che si desidera visualizzare. La tabella precedente fornisce un elenco dei tipi di file di log che sono disponibili e dove è possibile trovarli. L'esempio seguente mostra il comando per passare al log della fase con un ID, `s-1234ABCDEFGH`.

   ```
   cd /mnt/var/log/hadoop/steps/s-1234ABCDEFGH/
   ```

3. Utilizzare un visualizzatore file preferito per visualizzare il file di log. L'esempio seguente utilizza il comando `less` Linux per visualizzare i file di log `controller`.

   ```
   less controller
   ```

Visualizzazione dei file di log archiviati in Amazon S3

Per impostazione predefinita, i cluster Amazon EMR avviati utilizzando la console archiviano automaticamente i file di log in Amazon S3. Puoi specificare il tuo percorso dei log, oppure consentire alla console di generare automaticamente un percorso dei log per te. Per cluster avviati utilizzando la CLI o l'API, occorre configurare l'archiviazione dei log Amazon S3 manualmente.

Quando Amazon EMR è configurato per archiviare i file di log in Amazon S3, i file vengono archiviati nel percorso S3 specificato, nella cartella /*cluster-id*/, dove *cluster-id* è l'ID del cluster.

Nella tabella seguente vengono elencati alcuni dei file di log che si trovano su Amazon S3.

Ubicazione	Descrizione
/*cluster-id* /node/	Log dei nodi, inclusi log di operazioni di bootstrap, stato istanza e applicazioni per il nodo. I log per ogni nodo vengono archiviati in una cartella etichettata con l'identificatore dell'istanza EC2 di tale nodo.
/*cluster-id* /node/*instance-id* /*application*	I log creati da ogni applicazione o daemon associato a un'applicazione. Ad esempio, il log del server Hive si trova in */cluster-id/node/instance-id/hive/hive-server.log* .
/*cluster-id* /steps/*step-id*/	Log di fase contenenti informazioni sull'elaborazione della fase. Il valore di */step-id* indica l'ID di fase assegnato da Amazon EMR. Ad esempio, un cluster ha due fasi: `s-1234ABCDEFGH` e `s-5678IJKLMNOP` . La prima fase si trova in `/mnt/var/log/hadoop/steps/s-1234ABCDEFGH/` e la seconda in `/mnt/var/log/hadoop/steps/s-5678IJKLMNOP/` . I log di fase scritti da Amazon EMR sono i seguenti. • controller: informazioni sull'elaborazione della fase. Se la fase ha esito negativo durante il caricamento, puoi trovare la traccia di stack in questo log. • syslog: descrive l'esecuzione dei processi Hadoop nella fase. • stderr: il canale di errore standard di Hadoop durante l'elaborazione della fase.

Ubicazione	Descrizione
	• stdout: il canale di uscita standard di Hadoop durante l'elaborazione della fase.
/*cluster-id* /containers	Log del container applicazioni. I log per ogni applicazione YARN vengono salvati in queste posizioni.
/*cluster-id* /hadoop-mapreduce/	I log che contengono informazioni sui dettagli di configurazione e la cronologia dei processi MapReduce.

Visualizzazione dei file di log archiviati in Amazon S3 con la console Amazon S3

1. Accedi alla AWS Management Console e apri la console Amazon S3 all'indirizzo https://console.aws.amazon.com/s3/.
2. Apri il bucket S3 specificato quando hai configurato il cluster per archiviare i file di log in Amazon S3.
3. Passare al file di log contenente le informazioni da visualizzare. La tabella precedente fornisce un elenco dei tipi di file di log che sono disponibili e dove è possibile trovarli.
4. Scarica l'oggetto file di log per visualizzarlo. Per istruzioni, consulta Download di un oggetto.

Visualizzazione dei file di log nello strumento di debug

Amazon EMR non abilita automaticamente lo strumento di debug. che deve essere configurato quando si avvia il cluster. Tieni presente che la nuova console Amazon EMR non offre lo strumento di debug.

Visualizzazione dei log del cluster con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Dalla pagina Cluster List (Elenco dei cluster), seleziona l'icona dei dettagli accanto al cluster che desideri visualizzare.

In questo modo si apre la pagina Cluster Details (Dettagli del cluster). Nella sezione Steps (Fasi), i collegamenti a destra di ogni fase visualizzano i vari tipi di log disponibili per la fase. Questi log sono generati da Amazon EMR.

3. Per visualizzare un elenco dei processi Hadoop associati a una determinata fase, seleziona il collegamento View Jobs (Visualizza processi) a destra della fase.
4. Per visualizzare un elenco delle attività Hadoop associate a un determinato processo, seleziona il collegamento View Tasks (Visualizza attività) a destra del processo.
5. Per visualizzare un elenco dei tentativi effettuati per completare una determinata attività, seleziona la casella di controllo View Attempts (Visualizza tentativi) a destra dell'attività.
6. Per visualizzare i log generati da un tentativo di attività, seleziona la casella di controllo stderr, stdout e syslog a destra del tentativo di attività.

Lo strumento di debug visualizza i collegamenti ai file di log dopo che questi vengono caricati da Amazon EMR nel bucket su Amazon S3. Poiché i file di log vengono caricati in Amazon S3 ogni 5 minuti, il caricamento dei file di log può richiedere alcuni minuti dopo che la fase è stata terminata.

Amazon EMR aggiorna periodicamente lo stato di processi Hadoop, attività e tentativi di attività nello strumento di debug. Puoi fare clic su Refresh List (Aggiorna elenco) nei riquadri di debug per ottenere lo stato più recente di questi elementi.

Visualizzazione di istanze cluster in Amazon EC2

Per gestire meglio le risorse, Amazon EC2 consente di assegnare metadati alle risorse sotto forma di tag. Ciascun tag Amazon EC2 è formato da una chiave e da un valore. I tag consentono di categorizzare le risorse Amazon EC2 in modi diversi, ad esempio, per scopo, proprietario o ambiente.

È possibile cercare e filtrare le risorse in base ai tag che vengono aggiunti. I tag che assegni alle risorse tramite il tuo account AWS sono disponibili solo per te. Altri account che condividono la risorsa non possono visualizzare i tag.

Amazon EMR etichetta automaticamente ogni istanza EC2 che avvia con coppie chiave-valore. Le chiavi identificano il cluster e il gruppo di istanze a cui appartiene l'istanza. In questo modo è più semplice filtrare le istanze EC2 per mostrare, ad esempio, solo quelle appartenenti a un particolare cluster o per mostrare tutte le istanze attualmente in esecuzione nel gruppo di istanze dell'attività. Ciò è particolarmente utile se si eseguono più cluster contemporaneamente o si gestisce un gran numero di istanze EC2.

Queste sono le coppie predefinite di valori chiave che Amazon EMR assegna:

Chiave	Valore	Definizione di valore
aws:elasticmapreduce:job-flow-id	*job-flow-identifier*	L'ID del cluster per cui viene eseguito il provisioning dell'istanza. Viene visualizzato nel formato `j-XXXXXXXXXXXXX` e può contenere fino a 256 caratteri.
aws:elasticmapreduce:instance-group-role	*group-role*	Il tipo di gruppo di istanze, immesso come uno dei seguenti valori: `master`, `core` o `task`.

È possibile visualizzare e filtrare i tag che Amazon EMR aggiunge. Per ulteriori informazioni, consulta Utilizzo dei tag nella Guida per l'utente di Amazon EC2 per le istanze Linux. Poiché i tag impostati da Amazon EMR sono tag di sistema e non possono essere modificati o cancellati, le sezioni sui tag di visualizzazione e filtraggio sono le più rilevanti.

> **Note**
>
> Amazon EMR aggiunge tag all'istanza EC2 quando il suo stato viene aggiornato su In esecuzione. Se si verifica una latenza tra il momento in cui viene eseguito il provisioning dell'istanza EC2 e il momento in cui il suo stato è impostato su In esecuzione, i tag impostati da Amazon EMR non appaiono fino a quando l'istanza non viene avviata. Se non vengono visualizzati i tag, attendere qualche minuto e aggiornare la visualizzazione.

Eventi e parametri CloudWatch

Utilizza eventi e parametri per monitorare l'attività e l'integrità di un cluster Amazon EMR. Gli eventi sono utili per il monitoraggio di una specifica occorrenza in un cluster; ad esempio, quando un cluster passa dallo stato di avvio a quello di esecuzione. I parametri sono utili per monitorare uno specifico valore; ad esempio, la percentuale di spazio disponibile su disco che HDFS utilizza in un cluster.

Per ulteriori informazioni su CloudWatch Events, consulta la Guida per l'utente di Amazon CloudWatch Events. Per informazioni sui parametri CloudWatch, consulta Utilizzo dei parametri

Amazon CloudWatch e Creazione di allarmi Amazon CloudWatch nella Guida per l'utente di Amazon CloudWatch.

Argomenti

- Monitoraggio di parametri di Amazon EMR con CloudWatch
- Monitoraggio di eventi di Amazon EMR con CloudWatch
- Risposta agli eventi di CloudWatch

Monitoraggio di parametri di Amazon EMR con CloudWatch

I parametri sono aggiornati ogni cinque minuti e raccolti e trasmessi automaticamente a CloudWatch per ogni cluster Amazon EMR. Questo intervallo non è configurabile. Non è previsto alcun costo per i parametri di Amazon EMR segnalati in CloudWatch. I parametri di datapoint di cinque minuti vengono tenuti in archivio per 63 giorni, dopodiché vengono eliminati.

Come si utilizzano i parametri di Amazon EMR?

La tabella seguente mostra gli usi comuni per i parametri segnalati da Amazon EMR. Questi suggerimenti sono solo introduttivi e non costituiscono un elenco completo. Per l'elenco completo dei parametri forniti da Amazon EMR, consulta Parametri segnalati da Amazon EMR in CloudWatch.

Come...?	Parametri rilevanti
Monitorare l'avanzamento del cluster	Esaminare i parametri `RunningMapTasks` , `RemainingMapTasks` , `RunningReduceTasks` e `RemainingReduceTasks` .
Rilevare cluster inattivi	Il parametro `IsIdle` verifica se un cluster è attivo anche se non esegue attività. È possibile impostare un allarme di modo che venga attivato quando il cluster è rimasto inattivo per un determinato periodo di tempo, ad esempio 30 minuti.
Rilevare quando la capacità di storage di un nodo è esaurita	Il parametro `MRUnhealthyNodes` tiene traccia quando uno o più nodi principali o attività esauriscono la capacità di archiviaz

Come...?	Parametri rilevanti
	ione su disco locale e la transizione a uno stato YARN `UNHEALTHY` . Ad esempio, i nodi principali o attività stanno esaurendo spazio su disco e non saranno in grado di eseguire attività.
Rileva quando un cluster esaurisce la capacità d'archiviazione	Il parametro `HDFSUtilization` monitora la capacità HDFS combinata del cluster e può richiederne il ridimensionamento del cluster per aggiungere più nodi principali. Ad esempio, l'utilizzo di HDFS è elevato, il che può influire sull'integrità dei processi e sul cluster.
Rileva quando un cluster è in esecuzione a capacità ridotta	Il parametro `MRLostNodes` tiene traccia quando uno o più nodi principali o attività non sono in grado di comunicare con il nodo master. Ad esempio, il nodo principale o attività non è raggiungibile dal nodo master.

Per ulteriori informazioni, consulta Il cluster termina con NO_SLAVE_LEFT e i nodi principali FAILED_BY_MASTER e AWSSupport-AnalyzeEMRLogs.

Accesso ai parametri di CloudWatch per Amazon EMR

Puoi visualizzare i parametri che Amazon EMR segnala a CloudWatch utilizzando la console di Amazon EMR o la console di CloudWatch. Inoltre, puoi recuperare i parametri utilizzando il comando CLI CloudWatch `mon-get-stats` o l'API CloudWatch `GetMetricStatistics`. Per ulteriori informazioni su come visualizzare o recuperare i parametri di Amazon EMR utilizzando CloudWatch, consulta la Guida per l'utente di Amazon CloudWatch.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Visualizzazione dei parametri nella nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster per il quale vuoi visualizzare i parametri. Si apre la pagina dei dettagli del cluster.
3. Seleziona la scheda Monitoring (Monitoraggio) nella pagina dei dettagli del cluster. Scegli una qualsiasi delle schede Cluster status (Stato del cluster), Node status (Stato del nodo) o Inputs and outputs (Input e output) per caricare i report sull'avanzamento e sull'integrità del cluster.
4. Dopo avere scelto un parametro da visualizzare, puoi ingrandire ciascun grafico. Per filtrare l'intervallo di tempo del grafico, seleziona un'opzione precompilata o scegli Custom (Personalizzato).

Old console

Visualizzazione dei parametri nella vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/.
2. Per visualizzare i parametri per un cluster, selezionare un cluster per visualizzare il riquadro Summary (Riepilogo).
3. Scegliere Monitoring (Monitoraggio) per visualizzare informazioni su quel cluster. Scegli una qualsiasi delle schede Cluster Status (Stato cluster), Map/Reduce (Mappa/Riduci), Node Status (Stato nodo) o IO per caricare i report sull'avanzamento e sull'integrità del cluster.
4. Dopo aver scelto un parametro da visualizzare, è possibile selezionare la dimensione del grafico. Modificare i campi Start (Inizio) ed End (Fine) per filtrare i parametri in base a uno specifico intervallo di tempo.

Parametri segnalati da Amazon EMR in CloudWatch

Le tabelle seguenti elencano i parametri che Amazon EMR visualizza nella console e trasmette a CloudWatch.

Parametri di Amazon EMR

Amazon EMR invia i dati di vari parametri a CloudWatch. Tutti i cluster Amazon EMR inviano automaticamente i parametri a intervalli di cinque minuti. I parametri sono conservati per due settimane; dopo tale periodo, i dati vengono eliminati.

Il namespace `AWS/ElasticMapReduce` include i parametri descritti di seguito.

> **Note**
>
> Amazon EMR estrae i parametri da un cluster. Se un cluster diventa inaccessibile, non viene indicato alcun parametro fino a che il cluster non è di nuovo disponibile.

I seguenti parametri sono disponibili per i cluster sui quali sono in esecuzione le versioni 2.x di Hadoop.

Parametro	Descrizione
Stato del cluster	
IsIdle	Indica che un cluster non è più in esecuzione ma è ancora attivo e genera spese. È impostato su 1 se non vi sono task e processi in esecuzione, altrimenti è impostato su 0. Questo valore viene verificato a intervalli di cinque minuti e un valore 1 indica unicamente l'inattività del cluster al momento della verifica e non durante i cinque minuti. Per evitare falsi positivi, devi attivare un allarme quando questo valore è 1 durante due o più verifiche consecutive di cinque minuti. Ad esempio, puoi attivare un allarme se questo valore è 1 per trenta minuti o più. Caso d'uso: monitorare le prestazioni del cluster Unità: booleane
ContainerAllocated	Il numero di container di risorse allocati da ResourceManager. Caso d'uso: monitorare l'avanzamento del cluster

Parametro	Descrizione
	Unità: numero
ContainerReserved	Il numero di container riservati. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
ContainerPending	Il numero di container nella coda non ancora allocati. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
ContainerPendingRatio	Il rapporto tra i container in attesa e quelli allocati (ContainerPendingRatio = ContainerPending / ContainerAllocated). Se ContainerAllocated = 0, allora ContainerPendingRatio = ContainerPending. Il valore di ContainerPendingRatio rappresenta un numero, non una percentuale. Questo valore è utile per il dimensionamento delle risorse del cluster in funzione del comportamento di attribuzione dei container. Unità: numero
AppsCompleted	Il numero di applicazioni inviate a YARN che sono state completate. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
AppsFailed	Il numero di applicazioni inviate a YARN il cui completamento non è riuscito. Caso d'uso: monitorare l'avanzamento del cluster, monitorare lo stato del cluster Unità: numero

Parametro	Descrizione
AppsKilled	Il numero di applicazioni inviate a YARN che sono state interrotte. Caso d'uso: monitorare l'avanzamento del cluster, monitorare lo stato del cluster Unità: numero
AppsPending	Il numero di applicazioni inviate a YARN che sono in attesa. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
AppsRunning	Il numero di applicazioni inviate a YARN che sono in esecuzione. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
AppsSubmitted	Il numero di applicazioni inviate a YARN. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
Stato del nodo	
CoreNodesRunning	Il numero di nodi principali attivi. I punti dati per questo parametro sono indicati solo se esiste un gruppo di istanze corrispondente. Caso d'uso: monitorare lo stato del cluster Unità: numero

Parametro	Descrizione
CoreNodesPending	Il numero di nodi principali in attesa di assegnazione. È possibile che non tutti i nodi principali richiesti siano immediatamente disponibili; questo parametro indica le richieste in attesa. I punti dati per questo parametro sono indicati solo se esiste un gruppo di istanze corrispondente. Caso d'uso: monitorare lo stato del cluster Unità: numero
LiveDataNodes	La percentuale di nodi dati che ricevono attività da Hadoop. Caso d'uso: monitorare lo stato del cluster Unità: percentuale
MRTotalNodes	Il numero di nodi attualmente disponibili per processi MapReduce. Equivalente al parametro YARN `mapred.resourcemanager.TotalNodes` . Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
MRActiveNodes	Il numero di nodi in cui sono attualmente in esecuzione processi o task MapReduce. Equivalente al parametro YARN `mapred.resourcemanager.NoOfActiveNodes` . Caso d'uso: monitorare l'avanzamento del cluster Unità: numero

Parametro	Descrizione
MRLostNodes	Il numero di nodi allocati a MapReduce che sono stati contrassegnati con uno stato LOST. Equivalente al parametro YARN `mapred.resourcemanager.NoOfLostNodes` . Caso d'uso: monitorare lo stato del cluster, monitorare l'avanzamento del cluster Unità: numero
MRUnhealthyNodes	Il numero di nodi disponibili per processi MapReduce contrassegnati con uno stato UNHEALTHY. Equivalente al parametro YARN `mapred.resourcemanager.NoOfUnhealthyNodes` . Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
MRDecommissionedNodes	Il numero di nodi allocati ad applicazioni MapReduce che sono state contrassegnati con uno stato DECOMMISSIONED. Equivalente al parametro YARN `mapred.resourcemanager.NoOfDecommissionedNodes` . Caso d'uso: monitorare lo stato del cluster, monitorare l'avanzamento del cluster Unità: numero
MRRebootedNodes	Il numero di nodi disponibili per MapReduce che sono stati riavviati e contrassegnati con uno stato REBOOTED. Equivalente al parametro YARN `mapred.resourcemanager.NoOfRebootedNodes` . Caso d'uso: monitorare lo stato del cluster, monitorare l'avanzamento del cluster Unità: numero

Parametro	Descrizione
MultiMasterInstanceGroupNodesRunning	Il numero di nodi master in esecuzione. Caso d'uso: monitorare l'errore e la sostituzione del nodo master Unità: numero
MultiMasterInstanceGroupNodesRunningPercentage	La percentuale di nodi master in esecuzione sul numero dell'istanza del nodo master richiesto. Caso d'uso: monitorare l'errore e la sostituzione del nodo master Unità: percentuale
MultiMasterInstanceGroupNodesRequested	Il numero di nodi master richiesti. Caso d'uso: monitorare l'errore e la sostituzione del nodo master Unità: numero
IO	
S3BytesWritten	Il numero di byte scritti su Amazon S3. Questo parametro aggrega soltanto i processi MapReduce e non è applicabile ad altri carichi di lavoro su Amazon EMR. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero

Parametro	Descrizione
S3BytesRead	Il numero di byte letti da Amazon S3. Questo parametro aggrega soltanto i processi MapReduce e non è applicabile ad altri carichi di lavoro su Amazon EMR. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero
HDFSUtilization	La percentuale di storage HDFS attualmente utilizzato. Caso d'uso: analizzare le prestazioni del cluster Unità: percentuale
HDFSBytesRead	Il numero di byte letti da HDFS. Questo parametro aggrega soltanto i processi MapReduce e non è applicabile ad altri carichi di lavoro su Amazon EMR. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero
HDFSBytesWritten	Il numero di byte scritti su HDFS. Questo parametro aggrega soltanto i processi MapReduce e non è applicabile ad altri carichi di lavoro su Amazon EMR. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero
MissingBlocks	Il numero di blocchi in cui HDFS non ha repliche. Possono essere blocchi danneggiati. Caso d'uso: monitorare lo stato del cluster Unità: numero

Parametro	Descrizione
CorruptBlocks	Il numero di blocchi che HDFS ha indicato come danneggia ti. Caso d'uso: monitorare lo stato del cluster Unità: numero
TotalLoad	Il numero totale di trasferimenti di dati simultanei. Caso d'uso: monitorare lo stato del cluster Unità: numero
MemoryTotalMB	La quantità totale di memoria nel cluster. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
MemoryReservedMB	La quantità di memoria riservata. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
MemoryAvailableMB	La quantità di memoria disponibile da allocare. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
YARNMemoryAvailablePercentage	La percentuale di memoria rimanente disponibile per YARN (YARNMemoryAvailablePercentage = MemoryAvailableMB / MemoryTotalMB). Questo valore è utile per il dimension amento delle risorse del cluster in funzione dell'utilizzo della memoria di YARN. Unità: percentuale

Parametro	Descrizione
MemoryAllocatedMB	La quantità di memoria allocata al cluster. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
PendingDeletionBlocks	Il numero di blocchi contrassegnati per l'eliminazione. Caso d'uso: monitorare l'avanzamento del cluster, monitorare lo stato del cluster Unità: numero
UnderReplicatedBlocks	Il numero di blocchi che devono essere replicati una o più volte. Caso d'uso: monitorare l'avanzamento del cluster, monitorare lo stato del cluster Unità: numero
DfsPendingReplicationBlocks	Lo stato della replica dei blocchi: blocchi in corso di replica, età delle richieste di replica e richieste di replica non riuscite. Caso d'uso: monitorare l'avanzamento del cluster, monitorare lo stato del cluster Unità: numero
CapacityRemainingGB	La quantità di capacità rimanente del disco HDFS. Caso d'uso: monitorare l'avanzamento del cluster, monitorare lo stato del cluster Unità: numero

Di seguito sono descritti i parametri Hadoop 1:

Parametro	Descrizione
Stato del cluster	
IsIdle	Indica che un cluster non è più in esecuzione ma è ancora attivo e genera spese. È impostato su 1 se non vi sono task e processi in esecuzione, altrimenti è impostato su 0. Questo valore viene verificato a intervalli di cinque minuti e un valore 1 indica unicamente l'inattività del cluster al momento della verifica e non durante i cinque minuti. Per evitare falsi positivi, devi attivare un allarme quando questo valore è 1 durante due o più verifiche consecutive di cinque minuti. Ad esempio, puoi attivare un allarme se questo valore è 1 per trenta minuti o più. Caso d'uso: monitorare le prestazioni del cluster Unità: booleane
JobsRunning	Il numero di processi nel cluster attualmente in esecuzione. Caso d'uso: monitorare lo stato del cluster Unità: numero
JobsFailed	Il numero di processi nel cluster non riusciti. Caso d'uso: monitorare lo stato del cluster Unità: numero
Mappatura/Riduzione	
MapTasksRunning	Il numero di task di mappatura in esecuzione per ogni processo. Se hai un pianificatore installato e molteplici processi in esecuzione, vengono generati più grafici. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero

Parametro	Descrizione
MapTasksRemaining	Il numero di task di mappatura rimanenti per ogni processo. Se hai un pianificatore installato e molteplici processi in esecuzione, vengono generati più grafici. Un task di mappatura rimanente è un task il cui stato non è Running, Killed o Completed. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
MapSlotsOpen	La capacità dei task di mappatura non utilizzata. Viene calcolata come numero massimo di task di mappatura per un determinato cluster, meno il numero totale di task di mappatura attualmente in esecuzione in quel cluster. Caso d'uso: analizzare le prestazioni del cluster Unità: numero
RemainingMapTasksPerSlot	La proporzione tra i task di mappatura totali rimanenti e gli slot di mappatura totali disponibili nel cluster. Caso d'uso: analizzare le prestazioni del cluster Unità: proporzione
ReduceTasksRunning	Il numero di task di riduzione in esecuzione per ogni processo. Se hai un pianificatore installato e molteplici processi in esecuzione, vengono generati più grafici. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero

Parametro	Descrizione
ReduceTasksRemaining	Il numero di task di riduzione rimanenti per ogni processo. Se hai un pianificatore installato e molteplici processi in esecuzione, vengono generati più grafici. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
ReduceSlotsOpen	La capacità dei task di riduzione non utilizzata. Viene calcolata come capacità massima dei task di riduzione per un determinato cluster, meno il numero di task di riduzione attualmente in esecuzione in quel cluster. Caso d'uso: analizzare le prestazioni del cluster Unità: numero
Stato del nodo	
CoreNodesRunning	Il numero di nodi principali attivi. I punti dati per questo parametro sono indicati solo se esiste un gruppo di istanze corrispondente. Caso d'uso: monitorare lo stato del cluster Unità: numero
CoreNodesPending	Il numero di nodi principali in attesa di assegnazione. È possibile che non tutti i nodi principali richiesti siano immediatamente disponibili; questo parametro indica le richieste in attesa. I punti dati per questo parametro sono indicati solo se esiste un gruppo di istanze corrispondente. Caso d'uso: monitorare lo stato del cluster Unità: numero

Parametro	Descrizione
LiveDataNodes	La percentuale di nodi dati che ricevono attività da Hadoop. Caso d'uso: monitorare lo stato del cluster Unità: percentuale
TaskNodesRunning	Il numero di nodi di task attivi. I punti dati per questo parametro sono indicati solo se esiste un gruppo di istanze corrispondente. Caso d'uso: monitorare lo stato del cluster Unità: numero
TaskNodesPending	Il numero di nodi di task in attesa di assegnazione. È possibile che non tutti i nodi di task richiesti siano immediata mente disponibili; questo parametro indica le richieste in attesa. I punti dati per questo parametro sono indicati solo se esiste un gruppo di istanze corrispondente. Caso d'uso: monitorare lo stato del cluster Unità: numero
LiveTaskTrackers	La percentuale di tracker di task operativi. Caso d'uso: monitorare lo stato del cluster Unità: percentuale
IO	

Parametro	Descrizione
S3BytesWritten	Il numero di byte scritti su Amazon S3. Questo parametro aggrega soltanto i processi MapReduce e non è applicabile ad altri carichi di lavoro su Amazon EMR. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero
S3BytesRead	Il numero di byte letti da Amazon S3. Questo parametro aggrega soltanto i processi MapReduce e non è applicabile ad altri carichi di lavoro su Amazon EMR. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero
HDFSUtilization	La percentuale di storage HDFS attualmente utilizzato. Caso d'uso: analizzare le prestazioni del cluster Unità: percentuale
HDFSBytesRead	Il numero di byte letti da HDFS. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero
HDFSBytesWritten	Il numero di byte scritti su HDFS. Caso d'uso: analizzare le prestazioni del cluster, monitorare l'avanzamento del cluster Unità: numero

Parametro	Descrizione
MissingBlocks	Il numero di blocchi in cui HDFS non ha repliche. Possono essere blocchi danneggiati. Caso d'uso: monitorare lo stato del cluster Unità: numero
TotalLoad	Il numero totale corrente di lettori e writer indicato da tutti i DataNodes in un cluster. Caso d'uso: diagnosticare in quale misura un I/O elevato contribuisce al peggioramento delle prestazioni di esecuzion e dei processi. I nodi di lavoro in cui è in esecuzione il daemon DataNode devono eseguire anche task di mappatura e riduzione. Valori TotalLoad costantemente elevati possono indicare che un I/O elevato è una delle cause del peggioramento delle prestazioni. Picchi occasiona li di questo valore sono comuni e non indicano un problema. Unità: numero

Parametri della capacità del cluster

i parametri seguenti indicano le capacità correnti o di destinazione di un cluster. Queste metriche sono disponibili solo quando sono abilitati il dimensionamento gestito o la terminazione automatica.

Per i cluster composti da parchi istanze, i parametri della capacità del cluster vengono misurate in `Units`. Per i cluster composti da gruppi di istanze, i parametri della capacità del cluster vengono misurate in `Nodes` o in `VCPU` in base al tipo di unità utilizzato nella policy di dimensionamento gestito. Per ulteriori informazioni, consulta Utilizzo del dimensionamento gestito da EMR nella Guida alla gestione di Amazon EMR.

Parametro	Descrizione
• `TotalUnitsRequested` •	Numero totale di unità/nodi/vCPU di destinazione in un cluster determinato dal dimensionamento gestito.

Parametro	Descrizione
• `TotalNodesRequested` • `TotalVCPURequested`	Unità: numero
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	Numero totale di unità/nodi/VCPU correnti disponibili in un cluster in esecuzione. Quando viene richiesto il ridimensionamento di un cluster, questo parametro verrà aggiornato dopo l'aggiunta o la rimozione delle nuove istanze dal cluster. Unità: numero
• `CoreUnitsRequested` • `CoreNodesRequested` • `CoreVCPURequested`	Il numero di unità/nodi/VCPU CORE di destinazione in un cluster determinato dal dimensionamento gestito. Unità: numero
• `CoreUnitsRunning` • `CoreNodesRunning` • `CoreVCPURunning`	Il numero di unità/nodi/VCPU CORE correnti in esecuzione in un cluster. Unità: numero
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	Il numero di unità/nodi/vCPU TASK di destinazione in un cluster determinato dal dimensionamento gestito. Unità: numero

Parametro	Descrizione
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	Il numero di unità//nodi/vCPU TASK correnti in esecuzione in un cluster. Unità: numero

Amazon EMR emette le seguenti metriche con una granularità di un minuto quando abiliti la terminazione automatica utilizzando una policy di terminazione automatica. Alcune metriche sono disponibili solo per Amazon EMR versione 6.4.0 e successive. Per ulteriori informazioni sulla terminazione automatica, consulta Utilizzo di una policy di terminazione automatica.

Parametro	Descrizione
`TotalNotebookKernels`	Il numero totale di kernel notebook in esecuzione e inattivi sul cluster. Questa metrica è disponibile solo in Amazon EMR versione 6.4.0 e successive.
`AutoTerminationIsClusterIdle`	Indica se il cluster è in uso. Un valore di 0 indica che il cluster è in uso attivo da uno dei seguenti componenti: • Un'applicazione YARN • HDFS • Un notebook • Un'interfaccia utente on-cluster, come Spark History Server

Parametro	Descrizione
	Un valore di 1 indica che il cluster è inattivo. Amazon EMR verifica l'inattività continua del cluster (`AutoTerminationIsClusterIdle =1`). Quando il tempo di inattività di un cluster è uguale al valore `IdleTimeout` nella tua policy di terminazione automatica, Amazon EMR termina il cluster.

Dimensioni per i parametri Amazon EMR

I dati di Amazon EMR possono essere filtrati utilizzando una qualsiasi delle dimensioni nella tabella esposta di seguito.

Dimensione	Descrizione
JobFlowId	Uguale all'ID del cluster che è l'identificatore univoco di un cluster nel formato `j-XXXXXXXXXXXXX` . Puoi trovare questo valore facendo clic sul cluster nella console di Amazon EMR.

Monitoraggio di eventi di Amazon EMR con CloudWatch

Amazon EMR monitora gli eventi e conserva le informazioni sugli stessi fino a sette giorni nella console di Amazon AMR. Amazon EMR registra gli eventi in caso di modifica dello stato di cluster, gruppi di istanze, parchi istanze, policy di scalabilità automatica o fasi. Gli eventi registrano la data e l'ora in cui si sono verificati, i dettagli sugli elementi interessati e altri dati critici.

La tabella seguente elenca gli eventi di Amazon EMR, insieme allo stato o alla modifica dello stato che l'evento indica, la gravità dell'evento, il tipo di evento, il codice dell'evento e messaggi relativi agli eventi. Amazon EMR rappresenta gli eventi come oggetti JSON e li invia automaticamente a un flusso di eventi. L'oggetto JSON è importante quando si configurano delle regole per l'elaborazione di eventi mediante Eventi CloudWatch, in quanto le regole cercano di stabilire una corrispondenza con dei modelli nell'oggetto JSON. Per ulteriori informazioni, consulta Eventi e modelli di eventi ed Eventi Amazon EMR nella Guida per l'utente di Amazon CloudWatch Events.

> **Note**
>
> Per assicurarci di fornirti le informazioni più pertinenti, miglioriamo continuamente i nostri messaggi di errore. Per questo motivo, ti consigliamo di non ignorare il testo contenuto nei messaggi per avviare le azioni successive nel flusso di lavoro.

Eventi di avvio del cluster

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`CREATING`	`WARN`	Provisioning di parchi istanze Amazon EMR	Provisioning di EC2: capacità istanza insufficiente	Non siamo in grado di creare il tuo cluster Amazon EMR `ClusterId (ClusterName)` per il parco istanze `InstanceFleetID` Amazon EC2 ha una capacità Spot insufficiente per il tipo di istanza `[Instancetype1, Instancetype2]` e una capacità On-Demand insufficiente per il tipo di istanza `[Instancetype3,`

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
				`Instancetype4]` nella zona di disponibilità `[AvailabilityZone1, AvaliabilityZone2]` . Consulta qui la documentazione per ulteriori informazioni su come rispondere a questo evento.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`CREATING`	`WARN`	Provisioning di gruppi di istanze Amazon EMR	Provisioning di EC2: capacità istanza insufficiente	Non siamo in grado di creare il tuo cluster Amazon EMR `ClusterId (ClusterName)` per il gruppo di istanze `InstancegroupID` Amazon EC2 ha una capacità `[Spot or On-Demand]` insufficiente per il tipo di istanza `Instancetype` nella zona di disponibilità `AvailabilityZone` . Consulta qui la documentazione per ulteriori informazioni su come rispondere a questo evento.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`STARTING`	`INFO`	Modifica dello stato del cluster EMR	nessuno	Il cluster Amazon EMR `ClusterId (ClusterName)` è stato richiesto alle `Time` ed è in fase di creazione.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`STARTING`	`INFO`	Modifica dello stato del cluster EMR	nessuno	**Note** Si applica solo ai cluster con la configurazione dei parchi istanze e più zone di disponibilità selezionate in Amazon EC2. Il cluster Amazon EMR `ClusterId` `(ClusterName)` è in fase di creazione nella zona (`AvailabilityZoneID`), scelta tra le opzioni di zona

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
				di disponibilità specificate.
STARTING	INFO	Modifica dello stato del cluster EMR	nessuno	Il cluster Amazon EMR ClusterId (ClusterName) ha iniziato a eseguire le fasi alle Time.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`WAITING`	`INFO`	Modifica dello stato del cluster EMR	nessuno	Il cluster Amazon EMR `ClusterId (ClusterName)` è stato creato alle `Time` ed è pronto per l'uso. oppure Il cluster Amazon EMR `ClusterId (ClusterName)` ha completato l'esecuzione di tutte le fasi in sospeso alle `Time`. **Note** Un cluster il cui stato è `WAITING` può tuttavia elaborare

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
				dei processi.

> **Note**
>
> Gli eventi con codice evento `EC2 provisioning - Insufficient Instance Capacity` vengono emessi periodicamente quando il cluster EMR rileva un errore di capacità insufficiente da Amazon EC2 per il parco istanze o il gruppo di istanze durante la creazione o l'operazione di ridimensionamento del cluster. Per ulteriori informazioni su come rispondere a tali eventi, consulta Risposta a eventi di capacità insufficiente dell'istanza del cluster Amazon EMR.

Eventi di chiusura di un cluster

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`TERMINATED`	La gravità dipende dal motivo della modifica dello stato, come descritto di seguito: • **`CRITICAL`** se il cluster è stato terminato con uno dei seguenti motivi di modifica	Modifica dello stato del cluster EMR	nessuno	Il cluster Amazon EMR `ClusterId (ClusterName)` è stato terminato alle `Time` con un motivo di `StateChangeReason: Code` .

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
	dello stato: `INTERNAL_ERROR`, `VALIDATION_ERROR`, `INSTANCE_FAILURE`, `BOOTSTRAP_FAILURE` o `STEP_FAILURE`. • **`INFO`** se il cluster è stato terminato con uno dei seguenti motivi di modifica dello stato: `USER_REQUEST` o `ALL_STEPS_COMPLETED`.			

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`TERMINATED_WITH_ERRORS`	`CRITICAL`	Modifica dello stato del cluster EMR	nessuno	Il cluster Amazon EMR `ClusterId (ClusterName)` è stato terminato con errori alle `Time` con un motivo di `StateChangeReason: Code` .

Eventi di modifica dello stato del parco istanze

> **Note**
>
> La configurazione dei parchi istanze è disponibile solo in Amazon EMR rilasci 4.8.0 e successivi, esclusi i rilasci 5.0.0 e 5.0.3.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
Da `PROVISIONING` a `WAITING`	`INFO`		nessuno	Il provisioning per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` è

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
				completo. Il provisioning è stato avviato alle `Time` ed è durato `Num` minuti. Il parco istanze ha ora una capacità On-Demand di `Num` e una capacità Spot di `Num`. La capacità On-Demand di destinazione era `Num` e la capacità Spot di destinazione era `Num`.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
Da `WAITING` a `RESIZING`	`INFO`		nessuno	Un ridimensionamento per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` è iniziato alle `Time`. Il parco istanze è in fase di ridimensionamento da una capacità On-Demand di `Num` a una destinazione di `Num` e da una capacità Spot di `Num` a una destinazione di `Num`.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
Da `RESIZING` a `WAITING`	`INFO`		nessuno	L'operazione di ridimensionamento per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` è stata completata. Il ridimensionamento è stato avviato alle `Time` ed è durato `Num` minuti. Il parco istanze ha ora una capacità On-Demand di `Num` e una capacità Spot di `Num`. La capacità On-Demand di destinazione era `Num` e la capacità Spot di destinazione era `Num`.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
Da `RESIZING` a `WAITING`	`INFO`		nessuno	L'operazione di ridimensionamento per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` ha raggiunto il valore di timeout ed è stata interrotta. Il ridimensionamento è stato avviato alle `Time` ed è stato interrotto dopo `Num` minuti. Il parco istanze ha ora una capacità On-Demand di `Num` e una capacità Spot di `Num`. La capacità On-Demand di destinazione era `Num` e la capacità Spot di destinazione era `Num`.

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
SUSPENDED	ERROR		nessuno	Il parco istanze InstanceFleetID nel cluster Amazon EMR ClusterId (ClusterName) è stato arrestato alle Time per il seguente motivo: ReasonDesc .
RESIZING	WARNING		nessuno	L'operazione di ridimensionamento per il parco istanze InstanceFleetID nel cluster Amazon EMR ClusterId (ClusterName) è bloccata per il seguente motivo: ReasonDesc .

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
`WAITING` o `Running`	`INFO`		nessuno	Non è stato possibile completare l'operazione di dimensionamento per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` poiché Amazon EMR ha aggiunto la capacità Spot nella zona di disponibilità `AvailabilityZone` . Abbiamo annullato la richiesta di fornire capacità spot aggiuntiva. Per informazioni sulle operazioni consigliate, verifica Best practice per la flessibilità delle istanze e delle

Stato o modifica dello stato	Gravità	Tipo di evento	Codice dell'evento	Message
				zone di disponibilità e riprova.
`WAITING` o `Running`	`INFO`		nessuno	Un'operazione di ridimensionamento per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` è stata avviata da `Entity` alle `Time`.

Eventi di ridimensionamento del parco istanze

Tipo di evento	Gravità	Codice dell'evento	Message
Ridimensionamento del parco istanze Amazon EMR	`ERROR`	Timeout per il provisioning Spot	L'operazione di ridimensionamento per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` non è stata completata durante l'acquisizione della capacità Spot nella

Tipo di evento	Gravità	Codice dell'evento	Message
			zona di disponibilità `AvailabilityZone` . Abbiamo annullato la tua richiesta e non proveremo più a eseguire il provisioning della capacità Spot aggiuntiva. Il parco istanze ha eseguito il provisioning della capacità Spot di `num`. La capacità Spot di destinazione era `num`. Per ulteriori informazioni e le azioni consigliate, consulta la pagina della documentazione qui e riprova.

Tipo di evento	Gravità	Codice dell'evento	Message
Ridimensionamento del parco istanze Amazon EMR	ERROR	Timeout del provisioning On-Demand	L'operazione di ridimensionamento per il parco istanze `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` non è stata completata durante l'acquisizione della capacità On-Demand nella zona di disponibilità `AvailabilityZone` . Abbiamo annullato la tua richiesta e non proveremo più a eseguire il provisioning della capacità On-Demand aggiuntiva. Il parco istanze ha eseguito il provisioning della capacità On-Demand di `num`. La capacità On-Demand di destinazione era `num`. Per ulteriori informazioni e le azioni consigliate, consulta la pagina della documentazione qui e riprova.

Tipo di evento	Gravità	Codice dell'evento	Message
Ridimensionamento del parco istanze Amazon EMR	`WARNING`	Provisioning di EC2: capacità istanza insufficiente	Non siamo in grado di completare l'operazione di ridimensionamento per il parco istanze `InstanceFleetID` nel cluster EMR `ClusterId (ClusterName)` in quanto Amazon EC2 ha una capacità Spot insufficiente per i tipi di istanza `[Instancetype1, Instancetype2]` e una capacità On-Demand insufficiente per i tipi di istanza `[Instancetype3, Instancetype4]` nella zona di disponibilità `[AvailabilityZone1]` . Finora, il parco istanze ha eseguito il provisioning della capacità On-Demand di `num` e la capacità On-Demand di destinazione era `num`. La capacità Spot sottoposta a provisioning è `num` e la capacità Spot di destinazione era

Tipo di evento	Gravità	Codice dell'evento	Message
			num. Consulta qui la documentazione per ulteriori informazioni su come rispondere a questo evento.

Tipo di evento	Gravità	Codice dell'evento	Message
Ridimensionamento del parco istanze Amazon EMR	`WARNING`	Timeout del provisioning Spot: continuazione del ridimensionamento	Stiamo ancora eseguendo il provisioning della capacità Spot per l'operazione di ridimensionamento del parco istanze avviata alle `time` per il parco istanze con ID `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` per `[Instancetype1, Instancetype2]` nella zona di disponibilità `AvailabilityZone` . Per la precedente operazione di ridimensionamento avviata alle `time`, il periodo di timeout è scaduto, quindi Amazon EMR ha interrotto il provisioning della capacità Spot dopo l'aggiunta di `num` delle `num` istanze richieste al parco istanze. Per ulteriori informazioni, consulta la pagina di documentazione qui.

Tipo di evento	Gravità	Codice dell'evento	Message
Ridimensionamento del parco istanze Amazon EMR	`WARNING`	Timeout del provisioning On-Demand: continuazione del ridimensionamento	Stiamo ancora eseguendo il provisioning della capacità On-Demand per l'operazione di ridimensionamento del parco istanze avviata alle `time` per il parco istanze con ID `InstanceFleetID` nel cluster Amazon EMR `ClusterId (ClusterName)` per `[Instancetype1, Instancetype2]` nella zona di disponibilità `AvailabilityZone` . Per la precedente operazione di ridimensionamento avviata alle `time`, il periodo di timeout è scaduto, quindi Amazon EMR ha interrotto il provisioning della capacità On-Demand dopo l'aggiunta di `num` delle `num` istanze richieste al parco istanze. Per ulteriori informazioni,

Tipo di evento	Gravità	Codice dell'evento	Message
			consulta la pagina di documentazione qui.

> **Note**
>
> Gli eventi di timeout di provisioning vengono emessi quando Amazon EMR interrompe il provisioning della capacità Spot o On-Demand per il parco istanze dopo la scadenza del timeout. Per ulteriori informazioni su come rispondere a tali eventi, consulta Risposta agli eventi di timeout del ridimensionamento del parco istanze del cluster Amazon EMR.

Eventi del gruppo di istanze

Tipo di evento	Gravità	Codice dell'evento	Message
Da `RESIZING` a `Running`	`INFO`	nessuno	L'operazione di ridimensionamento per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (ClusterName)` è stata completata. Il numero di istanze è ora di `Num`. Il ridimensionamento è stato avviato alle `Time` ed è stato completato in `Num` minuti.
Da `RUNNING` a `RESIZING`	`INFO`	nessuno	Un ridimensionamento per il gruppo di istanze `InstanceGroupID`

Tipo di evento	Gravità	Codice dell'evento	Message
			nel cluster Amazon EMR `ClusterId (ClusterName)` è iniziato alle `Time`. In seguito al ridimensionamento il numero di istanze passa da `Num` a `Num`.
`SUSPENDED`	`ERROR`	nessuno	Il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (ClusterName)` è stato arrestato alle `Time` per il seguente motivo: `ReasonDesc` .
`RESIZING`	`WARNING`	nessuno	L'operazione di ridimensionamento per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (ClusterName)` è bloccata per il seguente motivo: `ReasonDesc` .

Tipo di evento	Gravità	Codice dell'evento	Message
Ridimensionamento del gruppo di istanze Amazon EMR	`WARNING`	Provisioning di EC2: capacità istanza insufficiente	Non siamo in grado di completare l'operazione di ridimensionamento iniziata alle `time` per il gruppo di istanze `InstanceGroupID` nel cluster EMR `ClusterId (ClusterName)` poiché Amazon EC2 ha una capacità `Spot/On Demand` insufficiente per il tipo di istanza `[Instancetype]` nella zona di disponibilità `[AvailabilityZone1]` . Finora, il gruppo di istanze ha un numero di istanze in esecuzione di `num` e il numero di istanze richiesto era `num`. Consulta qui la documentazione per ulteriori informazioni su come rispondere a questo evento.

Tipo di evento	Gravità	Codice dell'evento	Message
Da `RUNNING` a `RESIZING`	`INFO`	nessuno	Un ridimensionamento per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId` `(ClusterName)` è stato avviato da `Entity` alle `Time`.

> **Note**
>
> Con Amazon EMR versione 5.21.0 e successive, puoi sovrascrivere le configurazioni del cluster e specificare classificazioni di configurazione aggiuntive per ogni gruppo di istanze in un cluster in esecuzione. A questo scopo, utilizza la console di Amazon EMR, la AWS Command Line Interface (AWS CLI) o l'SDK AWS. Per ulteriori informazioni, consulta Specifica di una configurazione per un gruppo di istanze in un cluster in esecuzione.

La tabella seguente elenca gli eventi di Amazon EMR per l'operazione di riconfigurazione, insieme allo stato o alla modifica dello stato indicato dall'evento, la gravità dell'evento e messaggi relativi agli eventi.

Stato o modifica dello stato	Gravità	Message
`RUNNING`	`INFO`	Una riconfigurazione per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId` `(ClusterName)` è stata avviata dall'utente alle `Time`. La versione di configurazione richiesta è `Num`.

Stato o modifica dello stato	Gravità	Message
Da RECONFIGURING a Running	INFO	La riconfigurazione per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (ClusterName) è stata completata. La riconfigurazione è stata avviata alle Time e ha richiesto Num minuti per il completamento. La versione di configurazione corrente è Num.
Da RUNNING a RECONFIGURING in	INFO	Un riconfigurazione per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (ClusterName) è iniziata alle Time. Viene eseguita la configurazione dal numero di versione Num al numero di versione Num.
RESIZING	INFO	L'operazione di riconfigurazione verso la versione di configurazione Num per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (ClusterName) viene temporaneamente bloccata alle Time perché il gruppo di istanze si trova in State.

Stato o modifica dello stato	Gravità	Message
RECONFIGURING	INFO	L'operazione di ridimensionamento verso il numero di istanze Num per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (ClusterName) viene temporaneamente bloccata alle Time perché il gruppo di istanze si trova in State.
RECONFIGURING	WARNING	L'operazione di riconfigurazione per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (ClusterName) non è riuscita alle Time e ha richiesto Num minuti per la mancata esecuzione. La versione di configurazione non riuscita è Num.
RECONFIGURING	INFO	Viene eseguito il ripristino delle configurazioni al numero di versione riuscita precedente Num per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (ClusterName) alle Time. La nuova versione di configurazione è Num.

Stato o modifica dello stato	Gravità	Message
Da `RECONFIGURING` a `Running`	`INFO`	È stata ripristinata la versione riuscita precedente `Num` delle configurazioni per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (ClusterName)` alle `Time`. La nuova versione di configurazione è `Num`.
Da `RECONFIGURING` a `SUSPENDED`	`CRITICAL`	Impossibile ripristinare la versione riuscita precedente `Num` per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (ClusterName)` alle `Time`.

Eventi di policy di Auto Scaling (scalabilità automatica)

Stato o modifica dello stato	Gravità	Message
`PENDING`	`INFO`	Una policy di dimensionamento automatico è stata aggiunta al gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (ClusterName)` alle `Time`. La policy è in attesa di essere collegata. oppure La policy di dimensionamento automatico per il gruppo di

Stato o modifica dello stato	Gravità	Message
		istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (Cluster Name)` è stata aggiornata alle `Time`. La policy è in attesa di essere collegata.
`ATTACHED`	`INFO`	La policy di dimensionamento automatico per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (Cluster Name)` è stata collegata alle `Time`.
`DETACHED`	`INFO`	La policy di dimensionamento automatico per il gruppo di istanze `InstanceGroupID` nel cluster Amazon EMR `ClusterId (Cluster Name)` è stata scollegata alle `Time`.

Stato o modifica dello stato	Gravità	Message
FAILED	ERROR	Non è stato possibile collegare la policy di dimensionamento automatico per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (Cluster Name) e la policy ha dato esito negativo alle Time. oppure Non è stato possibile scollegar e la policy di dimensionamento automatico per il gruppo di istanze InstanceGroupID nel cluster Amazon EMR ClusterId (Cluster Name) e la policy ha dato esito negativo alle Time.

Eventi di fase

Stato o modifica dello stato	Gravità	Message
PENDING	INFO	La fase StepID (StepName) è stata aggiunta al cluster Amazon EMR ClusterId (ClusterName) alle Time ed è in attesa di esecuzione.
CANCEL_PENDING	WARN	La fase StepID (StepName) nel cluster Amazon EMR ClusterId (ClusterN ame) è stata annullata

Stato o modifica dello stato	Gravità	Message
		alle Time ed è in attesa di annullamento.
RUNNING	INFO	La fase StepID (StepName) nel cluster Amazon EMR ClusterId (ClusterName) ha iniziato l'esecuzione alle Time.
COMPLETED	INFO	La fase StepID (StepName) nel cluster Amazon EMR ClusterId (ClusterName) ha completato l'esecuzione alle Time. L'esecuzione della fase è stata avviata alle Time ed il completamento ha richiesto Num minuti.
CANCELLED	WARN	La richiesta di annullamento è riuscita per la fase del cluster StepID (StepName) nel cluster Amazon EMR ClusterId (ClusterName) alle Time e la fase è ora annullata.
FAILED	ERROR	La fase StepID (StepName) nel cluster Amazon EMR ClusterId (ClusterName) non è riuscita alle Time.

Visualizzazione degli eventi con la console Amazon EMR

Per ogni cluster, puoi visualizzare un semplice elenco di eventi nel riquadro dei dettagli, che enumera gli eventi in ordine di occorrenza decrescente. Puoi inoltre visualizzare tutti gli eventi di tutti i cluster in una regione in ordine di occorrenza decrescente.

Se non desideri che un utente visualizzi tutti gli eventi di cluster per una regione, aggiungi un'istruzione che nega l'autorizzazione (`"Effect": "Deny"`) per l'operazione `elasticmapreduce:ViewEventsFromAllClustersInConsole` a una policy collegata all'utente.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza nella console, consulta la sezione Novità della console.

New console

Visualizzazione degli eventi per tutti i cluster in una Regione con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Events (Eventi).

Visualizzazione degli eventi per un determinato cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster), quindi seleziona un cluster.
3. Per visualizzare tutti gli eventi, seleziona la scheda Events (Eventi) nella pagina dei dettagli del cluster.

Old console

Visualizzazione degli eventi per tutti i cluster in una Regione con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/.
2. Scegliere Events (Eventi).

Visualizzazione degli eventi per un determinato cluster con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce/.
2. Scegliere Cluster List (Elenco cluster), selezionare un cluster e quindi scegliere View details (Visualizza dettagli).
3. Scegliere Events (Eventi) nel riquadro dei dettagli del cluster.

Risposta agli eventi di CloudWatch

Questa sezione descrive vari modi in cui puoi rispondere agli eventi utilizzabili emessi da Amazon EMR come messaggi di eventi di CloudWatch.

Argomenti

- Creazione di regole per gli eventi Amazon EMR con CloudWatch
- Impostazione di allarmi su parametri di CloudWatch
- Risposta a eventi di capacità insufficiente dell'istanza del cluster Amazon EMR
- Risposta agli eventi di timeout del ridimensionamento del parco istanze del cluster Amazon EMR

Creazione di regole per gli eventi Amazon EMR con CloudWatch

Amazon EMR invia automaticamente gli eventi a un flusso di eventi di CloudWatch. Puoi creare regole che corrispondono a eventi in base a un modello specificato e instradare gli eventi verso target allo scopo di intraprendere delle azioni, ad esempio inviare un'e-mail di notifica. I modelli sono confrontati con l'oggetto JSON dell'evento. Per ulteriori informazioni sui dettagli degli eventi Amazon EMR, consulta Eventi Amazon EMR nella Guida per l'utente di Amazon CloudWatch Events.

Per informazioni su come creare una regola per eventi CloudWatch, consulta Creazione di una regola CloudWatch che si attiva in base a un evento.

Impostazione di allarmi su parametri di CloudWatch

Amazon EMR invia i parametri ad Amazon CloudWatch. In risposta, puoi utilizzare CloudWatch per impostare gli allarmi sui parametri di Amazon EMR. Ad esempio, puoi configurare un allarme in CloudWatch per inviare un'e-mail quando l'utilizzo di HDFS supera l'80%. Per istruzioni dettagliate, consulta Creazione o modifica di un allarme CloudWatch nella Guida per l'utente di Amazon CloudWatch.

Risposta a eventi di capacità insufficiente dell'istanza del cluster Amazon EMR

Panoramica

I cluster Amazon EMR restituiscono il codice evento `EC2 provisioning - Insufficient Instance Capacity` quando la zona di disponibilità selezionata non ha una capacità sufficiente per soddisfare la richiesta di avvio o ridimensionamento del cluster. L'evento viene emesso periodicamente sia con i gruppi di istanze che con i parchi istanze se Amazon EMR incontra ripetutamente eccezioni di capacità insufficiente e non è in grado di soddisfare la richiesta di provisioning per un'operazione di avvio o ridimensionamento del cluster.

Questa pagina descrive come rispondere al meglio a questo tipo di evento quando si verifica per il cluster EMR.

Risposta consigliata a un evento di capacità insufficiente

Ti consigliamo di rispondere a un evento di capacità insufficiente in uno dei seguenti modi:

- Attendi il ripristino della capacità. La capacità cambia frequentemente, quindi un'eccezione di capacità insufficiente può essere ripristinata autonomamente. Il ridimensionamento dei cluster inizierà o terminerà non appena la capacità di Amazon EC2 sarà disponibile.
- In alternativa, puoi terminare il cluster, modificare le configurazioni del tipo di istanza e creare un nuovo cluster con la richiesta di configurazione del cluster aggiornata. Per ulteriori informazioni, consulta Best practice per la flessibilità delle istanze e delle zone di disponibilità.

Puoi impostare regole o risposte automatiche a un evento di capacità insufficiente, come descritto nella sezione successiva.

Ripristino automatico da un evento di capacità insufficiente

Puoi creare automazione in risposta a eventi di Amazon EMR come quelli con codice evento `EC2 provisioning - Insufficient Instance Capacity`. Ad esempio, la seguente funzione

AWS Lambda termina un cluster EMR con un gruppo di istanze che utilizza istanze On-Demand, quindi crea un nuovo cluster EMR con un gruppo di istanze che contiene tipi di istanze diversi rispetto alla richiesta originale.

Le seguenti condizioni attivano il processo automatizzato:

- L'evento di capacità insufficiente viene emesso per i nodi primari o principali da più di 20 minuti.
- Il cluster non è in uno stato READY o WAITING. Per ulteriori informazioni sugli stati del cluster EMR, consulta Comprensione del ciclo di vita del cluster.

> **Note**
>
> Quando crei un processo automatizzato per un'eccezione di capacità insufficiente, devi considerare che l'evento di capacità insufficiente è recuperabile. La capacità cambia spesso e i cluster riprenderanno il ridimensionamento o inizieranno a funzionare non appena la capacità di Amazon EC2 sarà disponibile.

Example funzione per rispondere a un evento di capacità insufficiente

```
// 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")
```

Risposta agli eventi di timeout del ridimensionamento del parco istanze del cluster Amazon EMR

Panoramica

I cluster Amazon EMR emettono eventi durante l'esecuzione dell'operazione di ridimensionamento per i cluster del parco istanze. Gli eventi di timeout di provisioning vengono emessi quando Amazon EMR interrompe il provisioning della capacità Spot o On-Demand per il parco istanze dopo la scadenza del timeout. La durata del timeout può essere configurata dall'utente come parte delle specifiche di ridimensionamento per i parchi istanze. In scenari di ridimensionamento consecutivo per lo stesso parco istanze, Amazon EMR emette gli eventi `Spot provisioning timeout - continuing resize` o `On-Demand provisioning timeout - continuing resize` alla scadenza del timeout per l'operazione di ridimensionamento corrente. Quindi inizia il provisioning della capacità per la successiva operazione di ridimensionamento del parco istanze.

Risposta agli eventi di timeout del ridimensionamento del parco istanze

Ti consigliamo di rispondere a un evento di timeout del provisioning in uno dei seguenti modi:

- Rivedi le specifiche di ridimensionamento e riprova a eseguire l'operazione di ridimensionamento. Poiché la capacità cambia frequentemente, i cluster verranno ridimensionati correttamente non appena la capacità di Amazon EC2 sarà disponibile. Consigliamo ai clienti di configurare valori inferiori per la durata del timeout per i lavori che richiedono SLA più rigorosi.
- In alternativa, puoi:
  - Avviare un nuovo cluster con tipi di istanze diversificati in base alle best practice per la flessibilità dell'istanza e della zona di disponibilità oppure
  - Avviare un cluster con la capacità On-demand
- Per il timeout del provisioning, l'evento di ridimensionamento continua, puoi attendere anche l'elaborazione delle operazioni di ridimensionamento. Amazon EMR continuerà a elaborare in

sequenza le operazioni di ridimensionamento attivate per il parco istanze, rispettando le specifiche di ridimensionamento configurate.

Puoi anche impostare regole o risposte automatiche a questo evento come descritto nella sezione successiva.

Ripristino automatico da un evento di timeout del provisioning

Puoi creare automazione in risposta agli eventi di Amazon EMR con il codice evento `Spot Provisioning timeout`. Ad esempio, la seguente funzione AWS Lambda arresta un cluster EMR con un parco istanze che utilizza istanze Spot per i nodi Attività, quindi crea un nuovo cluster EMR con un parco istanze che contiene più tipi di istanze diversificati rispetto alla richiesta originale. In questo esempio, l'evento `Spot Provisioning timeout` emesso per i nodi attività attiverà l'esecuzione della funzione Lambda.

Example Funzione di esempio per rispondere all'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")
```

Visualizzazione dei parametri dell'applicazione cluster con Ganglia

Ganglia è disponibile con Amazon EMR versioni 4.2 e successive. Ganglia è un progetto open source per un sistema scalabile e distribuito, progettato per monitorare i cluster e le griglie riducendo al minimo l'impatto sulle loro prestazioni. Quando abiliti Ganglia sul cluster, puoi generare report e visualizzare le prestazioni del cluster nel suo insieme, nonché esaminare le prestazioni di singole istanze nodi. Ganglia è anche configurato per acquisire e visualizzare i parametri Hadoop e Spark. Per ulteriori informazioni, consulta la sezione Ganglia nella Guida ai rilasci di Amazon EMR.

Registrazione delle chiamate API di Amazon EMR in AWS CloudTrail

Amazon EMR è integrato con AWS CloudTrail, un servizio che offre un registro delle operazioni eseguite da un utente, un ruolo o un servizio AWS in Amazon EMR. CloudTrail acquisisce tutte le chiamate API per Amazon EMR come eventi. Le chiamate acquisite includono le chiamate dalla console di Amazon EMR e le chiamate di codice alle operazioni delle API di Amazon EMR. Se crei un percorso, puoi abilitare la distribuzione continua di eventi CloudTrail in un bucket Amazon S3, inclusi gli eventi per Amazon EMR. Se non si configura un trail, è comunque possibile visualizzare gli eventi più recenti nella console di CloudTrail in Event history (Cronologia eventi). Le informazioni raccolte da CloudTrail consentono di determinare la richiesta effettuata ad Amazon EMR, l'indirizzo IP da cui è partita la richiesta, l'autore della richiesta, il momento in cui è stata eseguita e altri dettagli.

Per ulteriori informazioni su CloudTrail, consultare la Guida per l'utente di AWS CloudTrail.

Informazioni su Amazon EMR in CloudTrail

CloudTrail è abilitato sull'account AWS al momento della sua creazione. Quando si verifica un'attività in Amazon EMR, questa viene registrata in un evento CloudTrail insieme ad altri eventi di servizio AWS nella Event History (Cronologia degli eventi). È possibile visualizzare, cercare e scaricare

gli eventi recenti nell'account AWS. Per ulteriori informazioni, consultare Visualizzazione di eventi mediante la cronologia eventi di CloudTrail.

Per una registrazione continua degli eventi nell'account AWS, inclusi gli eventi per Amazon EMR, crea un percorso. Un trail consente a CloudTrail di distribuire i file di log in un bucket Amazon S3. Per impostazione predefinita, quando si crea un trail nella console, il trail sarà valido in tutte le regioni AWS. Il trail registra gli eventi di tutte le Regioni nella partizione AWS e distribuisce i file di log nel bucket Amazon S3 specificato. Inoltre, è possibile configurare altri servizi AWS per analizzare con maggiore dettaglio e usare i dati evento raccolti nei log CloudTrail. Per ulteriori informazioni, consulta gli argomenti seguenti:

- Panoramica della creazione di un percorso
- Servizi e integrazioni CloudTrail supportati
- Configurazione delle notifiche Amazon SNS per CloudTrail
- Ricezione di file di log CloudTrail da più regioni e Ricezione di file di log CloudTrail da più account

Tutte le operazioni Amazon EMR vengono registrate da CloudTrail e sono documentate nella Documentazione di riferimento delle API di Amazon EMR. Ad esempio, le chiamate alle operazioni `RunJobFlow`, `ListCluster` e `DescribeCluster` generano voci nei file di log di CloudTrail.

Ogni evento o voce di log contiene informazioni sull'utente che ha generato la richiesta. Le informazioni di identità consentono di determinare quanto segue:

- Se la richiesta è stata effettuata con credenziali utente root o AWS Identity and Access Management (IAM).
- Se la richiesta è stata effettuata con le credenziali di sicurezza temporanee per un ruolo o un utente federato.
- Se la richiesta è stata effettuata da un altro servizio AWS.

Nel caso in cui un processo, anziché un utente, crei un cluster, è possibile utilizzare l'identificatore `principalId` per determinare l'utente associato alla creazione del cluster. Per ulteriori informazioni, consulta Elemento CloudTrail userIdentity.

Esempio: voci del file di log di Amazon EMR

Un percorso è una configurazione che consente la distribuzione di eventi come i file di log in un bucket Amazon S3 specificato. I file di log di CloudTrail possono contenere una o più voci di log.

Un evento rappresenta una singola richiesta da un'origine e include informazioni sull'operazione richiesta, sulla data e sull'ora dell'operazione, sui parametri richiesti e così via. I file di log CloudTrail non sono una traccia di pila ordinata delle chiamate API pubbliche e di conseguenza non devono apparire in base a un ordine specifico.

L'esempio seguente mostra una voce di log di CloudTrail che illustra l'operazione 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
   ]
}
```


Uso del dimensionamento del cluster

È possibile regolare il numero di istanze Amazon EC2 disponibili per un cluster Amazon EMR automaticamente o manualmente in risposta a carichi di lavoro con richieste variabili. Per utilizzare il dimensionamento automatico, sono disponibili due opzioni. Puoi abilitare il dimensionamento gestito da Amazon EMR o creare una policy di dimensionamento automatico personalizzato. La tabella seguente descrive le differenze tra le due opzioni.

	Dimensionamento gestito da Amazon EMR	Dimensionamento automatico personalizzato
Policy e regole di dimension amento	Non è richiesta alcuna policy. Amazon EMR gestisce l'attività di dimensionamento automatic o valutando continuam ente i parametri del cluster e prendendo decisioni di dimensionamento ottimizzate.	È necessario definire e gestire le policy e le regole di dimensionamento automatic o, ad esempio le condizion i specifiche che attivano le attività di dimensionamento, i

	Dimensionamento gestito da Amazon EMR	Dimensionamento automatico personalizzato
		periodi di valutazione, i periodi di attesa e così via.
Versioni di Amazon EMR supportate	Amazon EMR versione 5.30.0 e successive (tranne Amazon EMR versione 6.0.0)	Amazon EMR 4.0.0 e versioni successive
Composizione cluster supportata	Gruppi di istanze o parchi istanze	Solo gruppi di istanze
Configurazione dei limiti di dimensionamento	I limiti di dimensionamento sono configurati per l'intero cluster.	I limiti di dimensionamento possono essere configura ti solo per ogni gruppo di istanze.
Frequenza di valutazione dei parametri	Ogni 5-10 secondi Una valutazione più frequente dei parametri consente ad Amazon EMR di prendere decisioni di dimensionamento più precise.	È possibile definire i periodi di valutazione solo in incrementi di cinque minuti.
Applicazioni supportate	Sono supportate solo le applicazioni YARN, come Spark, Hadoop, Hive, Flink. Il dimensionamento gestito da Amazon EMR non supporta applicazioni non basate su YARN, come Presto o HBase.	È possibile scegliere quali applicazioni sono supportat e quando si definiscono le regole di dimensionamento automatico.

Considerazioni

- Un cluster Amazon EMR è sempre costituito da uno o tre nodi primari. Una volta configurato inizialmente il cluster, è possibile dimensionare solo i nodi principali e i nodi attività. Non è possibile dimensionare il numero di nodi primari per il cluster.
- Per i gruppi di istanze, le operazioni di riconfigurazione e le operazioni di ridimensionamento avvengono consecutivamente e non contemporaneamente. Se si avvia una riconfigurazione durante il ridimensionamento di un gruppo di istanze, la riconfigurazione inizia quando il gruppo di istanze completa il ridimensionamento in corso. Al contrario, se si avvia un'operazione di ridimensionamento mentre un gruppo di istanze ne esegue la riconfigurazione.

Utilizzo del dimensionamento gestito in Amazon EMR

> **Important**
>
> Per il dimensionamento gestito si consiglia di utilizzare i rilasci più recenti di Amazon EMR (Amazon EMR 6.14.0). In alcune versioni preliminari di Amazon EMR, potresti riscontrare esiti negativi intermittenti delle applicazioni o ritardi nel dimensionamento. Amazon EMR. ha risolto questo problema con le versioni 5.x 5.30.2, 5.31.1, 5.32.1, 5.33.1 e successive e con le versioni 6.x 6.1.1, 6.2.1, 6.3.1 e successive. Per ulteriori informazioni sulla regione e sulla disponibilità del rilascio, consulta Disponibilità di dimensionamento gestito.

Panoramica

Con Amazon EMR versioni 5.30.0 e successive (tranne Amazon EMR 6.0.0), puoi abilitare il dimensionamento gestito da Amazon EMR. Il dimensionamento gestito ti permette di aumentare o diminuire automaticamente il numero di istanze o unità nel cluster in base al carico di lavoro. Amazon EMR valuta continuamente i parametri dei cluster per prendere decisioni di dimensionamento che ottimizzano i cluster in termini di costi e velocità. Il dimensionamento gestito è disponibile per i cluster composti da gruppi di istanze o parchi istanze.

Disponibilità di dimensionamento gestito

- In Asia Pacific (Giacarta), la scalabilità gestita di Amazon EMR è disponibile con Amazon EMR 6.14.0 e versioni successive.

- Nelle seguenti Regioni AWS,la scalabilità gestita di Amazon EMR è disponibile con Amazon EMR 5.30.0, 6.1.0 e versioni successive:

  Stati Uniti orientali (Virginia settentrionale e Ohio), Stati Uniti occidentali (Oregon e California settentrionale), Sud America (San Paolo), Europa (Francoforte, Irlanda, Londra, Milano, Parigi e Stoccolma), Canada (Centrale), Asia Pacifico (Hong Kong), Mumbai, Seoul, Singapore, Sydney e Tokyo), Medio Oriente (Bahrein), Africa (Città del Capo), AWS GovCloud (Stati Uniti orientali), AWS GovCloud (Stati Uniti occidentali),Cina (Pechino) gestito da Sinnet, e Cina (Ningxia) gestito da NWCD.
- Il dimensionamento gestito da Amazon EMR funziona solo con applicazioni YARN, come Spark, Hadoop, Hive e Flink. Non sono supportate le applicazioni non basate su YARN, come Presto e HBase.

Parametri del dimensionamento gestito

È necessario configurare i seguenti parametri per il dimensionamento gestito. Il limite si applica solo ai nodi core e task. Il nodo primario non può essere dimensionato dopo la configurazione iniziale.

- Minimum (Minimo) (`MinimumCapacityUnits`): il limite minimo della capacità EC2 consentita in un cluster. Si misura in core vCPU (Virtual Central Processing Unit) o istanze per gruppi di istanze. Si misura in unità per parchi istanze.
- Maximum (Massimo) (`MaximumCapacityUnits`): il limite massimo della capacità EC2 consentita in un cluster. Si misura in core vCPU (Virtual Central Processing Unit) o istanze per gruppi di istanze. Si misura in unità per parchi istanze.
- On-Demand limit (Limite on demand) (`MaximumOnDemandCapacityUnits`) (facoltativo): limite massimo della capacità EC2 consentita per il tipo di mercato on demand in un cluster. Se questo parametro non è specificato, il valore predefinito viene impostato su `MaximumCapacityUnits`.
  - Questo parametro viene utilizzato per dividere l'allocazione della capacità tra istanze on demand e Spot. Ad esempio, se imposti 2 istanze come parametro minimo, il parametro massimo sarà 100 istanze e il limite on demand sarà 10 istanze, pertanto il dimensionamento gestito da Amazon EMR aumenta fino a 10 istanze on demand e assegna la capacità rimanente alle istanze spot. Per ulteriori informazioni, consulta Scenari di assegnazione dei nodi.
- Maximum core nodes (Numero massimo di nodi principali) (`MaximumCoreCapacityUnits`) (facoltativo): il limite massimo della capacità EC2 consentita per il tipo di nodo principale in un cluster. Se questo parametro non è specificato, il valore predefinito viene impostato su `MaximumCapacityUnits`.

- Questo parametro viene utilizzato per dividere l'allocazione della capacità tra nodi principali e attività. Ad esempio, se imposti 2 istanze come parametro minimo, il parametro massimo sarà 100 istanze e il nodo core massimo sarà 17 istanze, pertanto il dimensionamento gestito da Amazon EMR aumenta fino a 17 nodi core e assegna le 83 istanze rimanenti ai nodi attività. Per ulteriori informazioni, consulta Scenari di assegnazione dei nodi.

Per ulteriori informazioni sui parametri del dimensionamento gestito, consulta `ComputeLimits`.

Considerazioni sul dimensionamento gestito di Amazon EMR

- Il dimensionamento gestito è supportato nelle versioni limitate di Regioni AWS e di Amazon EMR. Per ulteriori informazioni, consulta Disponibilità di dimensionamento gestito.
- È necessario configurare i parametri richiesti per il dimensionamento gestito da Amazon EMR. Per ulteriori informazioni, consulta Parametri del dimensionamento gestito.
- Per utilizzare il dimensionamento gestito in API Gateway, il processo di raccolta dei parametri deve essere in grado di connettersi all'endpoint API pubblico. Se utilizzi un nome DNS privato con Amazon Virtual Private Cloud, dimensionamento gestito non funzionerà correttamente. Per garantire che il dimensionamento gestito funzioni, consigliamo di eseguire una delle seguenti operazioni:
  - Rimuovi l'endpoint VPC dell'interfaccia API Gateway dal tuo Amazon VPC.
  - Segui le istruzioni in Why do I get an HTTP 403 Forbidden error when connecting to my API Gateway APIs from a VPC? (Perché visualizzo un errore HTTP 403 Forbidden durante la connessione alle API di API Gateway da un VPC?) per disabilitare l'impostazione del nome DNS privato.
  - In alternativa, avvia il cluster in una sottorete privata. Per ulteriori informazioni, consulta l'argomento in Sottoreti private.
- Se i processi YARN sono rallentati in modo intermittente durante la riduzione e i log di YARN Resource Manager mostrano che la maggior parte dei nodi sono stati elencati come negati durante quel periodo, puoi regolare la soglia di timeout di disattivazione.

  Riduci il `spark.blacklist.decommissioning.timeout` da un'ora a un minuto per rendere disponibile il nodo per altri container in sospeso per continuare l'elaborazione del processo.

  Inoltre, è necessario impostare `YARN.resourcemanager.nodemanager-graceful-decommission-timeout-secs` su un valore maggiore per garantire che Amazon EMR non forzi la terminazione del nodo mentre il "Processo Spark" più lungo è ancora in esecuzione sul

nodo. Il valore predefinito attuale è 60 minuti, il che significa che la YARN forza la terminazione del container dopo 60 minuti una volta che il nodo entra nello stato di disattivazione.

Il seguente esempio di riga del log Resource Manager di YARN mostra i nodi aggiunti allo stato di disattivazione:

```
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: []
```

Scopri maggiori dettagli su come Amazon EMR si integra con l'elenco rifiutati di YARN durante la disattivazione dei nodi, i casi in cui i nodi in Amazon EMR possono essere aggiunti all'elenco rifiutati e la configurazione del comportamento di disattivazione dei nodi di Spark.

- L'utilizzo eccessivo dei volumi EBS può causare problemi di dimensionamento gestito. Si consiglia di mantenere il volume EBS inferiore al 90% di utilizzo. Per ulteriori informazioni, consulta Archiviazione dell'istanza.
- I parametri di Amazon CloudWatch sono fondamentali per il funzionamento del dimensionamento gestito di Amazon EMR. Si consiglia di monitorare attentamente i parametri di Amazon CloudWatch per assicurarsi che i dati non manchino. Per ulteriori informazioni su come configurare gli allarmi CloudWatch per rilevare i parametri mancanti, consulta Utilizzo degli allarmi Amazon CloudWatch.
- Le operazioni di dimensionamento gestito su cluster 5.30.0 e 5.30.1 senza Presto installato possono causare errori delle applicazioni o far sì che un gruppo di istanze o un parco istanze uniforme mantenga lo stato `ARRESTED`, in particolare quando un'operazione di dimensionamento verso il basso è seguita rapidamente da un'operazione di dimensionamento verso l'alto.

  Come soluzione alternativa, scegli Presto come applicazione da installare quando crei un cluster con Amazon EMR rilasci 5.30.0 e 5.30.1, anche se il tuo processo non richiede Presto.
- Quando imposti il nodo core massimo e il limite on demand per il dimensionamento gestito da Amazon EMR, considera le differenze tra gruppi di istanze e parchi istanze. Ogni gruppo di istanze è costituito dallo stesso tipo di istanza e dalla stessa opzione di acquisto per istanze: on demand o Spot. Per ogni parco istanze, puoi specificare fino a cinque tipi di istanze, che possono essere assegnate come istanze on demand e Spot. Per ulteriori informazioni, consulta Creazione di un

cluster con parchi istanze o gruppi di istanze uniformi, Opzioni dei parchi istanze e Scenari di assegnazione dei nodi.

- Con Amazon EMR rilascio 5.30.0 e successivi, se si rimuove la regola predefinita Allow All (Consenti tutto) in uscita su 0.0.0.0/ nel gruppo di sicurezza principale è necessario aggiungere una regola per consentire la connettività TCP in uscita al gruppo di sicurezza di accesso al servizio sulla porta 9443. Inoltre, il gruppo di sicurezza di accesso al servizio deve consentire il traffico TCP in ingresso sulla porta 9443 dal gruppo di sicurezza principale. Per ulteriori informazioni sulla configurazione dei gruppi di sicurezza, consulta la sezione Amazon EMR-managed security group for the primary instance (private subnets) (Gruppo di sicurezza gestito da Amazon EMR per l'istanza primaria [sottoreti private]).
- Il dimensionamento gestito attualmente non supporta la caratteristica YARN node labels (Etichette di nodi YARN). Evita di utilizzare le etichette dei nodi sui cluster con scalabilità gestita. Ad esempio, non consentire agli esecutori di funzionare solo sui nodi delle attività. Quando utilizzi le etichette dei nodi nei tuoi cluster Amazon EMR, potresti scoprire che il cluster non sta aumentando, il che può portare a un rallentamento dell'applicazione.
- Puoi utilizzare AWS CloudFormation per configurare il dimensionamento gestito da Amazon EMR. Per ulteriori informazioni, consulta AWS::EMR::Cluster nella Guida per l'utente di AWS CloudFormation.

Cronologia delle funzionalità

Questa tabella elenca gli aggiornamenti della funzionalità di dimensionamento gestito da Amazon EMR.

Data di rilascio	Funzionalità	Versioni di Amazon EMR
10 ottobre 2023	Il dimensionamento gestito è disponibile nella regione Asia Pacific (Giacarta) `ap-southeast-3` .	6.14.0 e versioni successive
28 luglio 2023	Dimensionamento gestito migliorato per passare a gruppi di istanze di attività diversi su scala verticale quando Amazon EMR subisce	5.34.0 e versioni successive, 6.4.0 e versioni successive

Data di rilascio	Funzionalità	Versioni di Amazon EMR
	un ritardo nel dimension amento verticale rispetto al gruppo di istanze corrente.	
16 giugno 2023	Dimensionamento gestito migliorato per rilevare i nodi che eseguono il master dell'applicazione in modo che tali nodi non vengano dimensionati ridotti. Per ulteriori informazioni, consulta Informazioni su strategia e scenari di assegnazione dei nodi.	5.34.0 e versioni successive, 6.4.0 e versioni successive

Data di rilascio	Funzionalità	Versioni di Amazon EMR
21 marzo 2022	Aggiunta la consapevolezza dei dati di shuffle di Spark utilizzata durante la riduzione verticale dei cluster. Per i cluster Amazon EMR con Apache Spark e la caratteristica di dimensionamento gestito abilitato, Amazon EMR monitora continuamente gli esecutori Spark e le posizioni intermedie dei dati di shuffle. Utilizzando queste informazioni, Amazon EMR riduce verticalmente solo le istanze sottoutilizzate che non contengono dati di shuffle utilizzati attivamente. Ciò impedisce il ricalcolo dei dati di shuffle persi, contribuendo a ridurre i costi e migliorare le prestazioni dei processi. Per ulteriori informazioni, consulta la Guida di programmazione Spark.	5.34.0 e versioni successive, 6.4.0 e versioni successive

Configurazione del dimensionamento gestito per Amazon EMR

Le sezioni seguenti spiegano come avviare un cluster EMR che utilizza il dimensionamento gestito con la AWS Management Console, il AWS SDK for Java, o la AWS Command Line Interface.

Argomenti

- Utilizzo della AWS Management Console per configurare il dimensionamento gestito
- Utilizzo della AWS CLI per configurare il dimensionamento gestito
- Usare AWS SDK for Java per configurare il dimensionamento gestito

Utilizzo della AWS Management Console per configurare il dimensionamento gestito

Puoi utilizzare la console Amazon EMR per configurare il dimensionamento gestito quando crei un cluster o per modificare una policy di dimensionamento gestita per un cluster in esecuzione.

New console

Configurazione del dimensionamento gestito durante la creazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. Scegli una versione di Amazon EMR emr-5.30.0 o successiva, tranne quella emr-6.0.0.
4. In Cluster scaling and provisioning option (Opzione di dimensionamento e provisioning del cluster), scegli Use EMR-managed scaling (Utilizza dimensionamento gestito da EMR). Specifica il numero Minimum (Minimo) e Maximum (Massimo) di istanze, il numero di istanze Maximum core node (Massimo del nodo core) e il numero di istanze Maximum On-Demand (Massimo on demand).
5. Scegli qualsiasi altra opzione applicabile al cluster.
6. Per avviare il cluster, scegli Create cluster (Crea cluster).

Configurazione del dimensionamento gestito su un cluster esistente con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare.
3. Nella scheda Instances (Istanze) della pagina dei dettagli del cluster, cerca la sezione Instance group settings (Impostazioni del gruppo di istanze). Seleziona Edit cluster scaling (Modifica dimensionamento del cluster) per specificare nuovi valori per il numero Minimum (Minimo) e Maximum (Massimo) di istanze e per il limite On-Demand (On demand).

Old console

Quando crei un cluster con la vecchia console, puoi configurare il dimensionamento gestito utilizzando le opzioni di configurazione rapida o le opzioni di configurazione avanzata del cluster. Inoltre puoi creare o modificare una policy di dimensionamento gestito per un cluster in esecuzione modificando le impostazioni di Managed Scaling (Dimensionamento gestito) nella pagina Summary (Riepilogo) o Hardware.

Utilizzo delle opzioni di configurazione rapida per configurare il dimensionamento gestito durante la creazione di un cluster con la vecchia console

1. Apri la console di Amazon EMR, scegli Create cluster (Crea cluster) e apri Create Cluster - Quick options (Crea un cluster - Opzioni rapide).
2. Nella sezione Hardware configuration (Configurazione hardware) accanto a Cluster scaling and provisioning option (Opzione di dimensionamento e provisioning del cluster), seleziona la casella di controllo per abilitare Scale cluster nodes based on workload (Dimensiona i nodi cluster in base al carico di lavoro).
3. Sotto Core and task units (Unità principali e attività), specifica il numero Minimum (Minimo) e Maximum (Massimo) di istanze principali e attività.

Utilizzo delle opzioni avanzate per configurare il dimensionamento gestito durante la creazione di un cluster con la vecchia console

1. Nella console di Amazon EMR, seleziona Create cluster (Crea cluster), poi Go to advanced options (Vai alle opzioni avanzate), quindi scegli Step 1: Software and Steps (Fase 1: Software e fasi) e infine Step 2: Hardware Configuration (Fase 2: Configurazione hardware).
2. Nella sezione Cluster composition (Composizione cluster) selezionare Parchi di istanze o Gruppi di istanze uniformi.
3. In Cluster scaling and provisioning option (Opzione di dimensionamento e provisioning del cluster)m seleziona Enable cluster scaling (Abilita dimensionamento del cluster). Quindi seleziona Use EMR managed scaling (Usa dimensionamento gestito da EMR). In Core and task units (Unità principali e attività), specifica il numero di istanze o unità di parchi istanze Minimum (Minimo) e Maximum (Massimo), l'On-Demand limit (Limite on demand) e il Maximum Core Node (Numero massimo di nodi principali).

   Per i cluster composti da gruppi di istanze, è inoltre possibile scegliere Create a custom automatic scaling policy (Crea una policy di dimensionamento automatico personalizzato se

si desidera definire le policy di dimensionamento automatico personalizzate per ogni gruppo di istanze. Per ulteriori informazioni, consulta Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze .

Modifica del dimensionamento gestito di un cluster esistente con la vecchia console

1. Apri la console di Amazon EMR, seleziona il cluster dall'elenco dei cluster e scegli la scheda Hardware.
2. Nella sezione Cluster scaling and provisioning option (Opzione di dimensionamento e provisioning del cluster), seleziona Edit (Modifica) per il dimensionamento gestito da Amazon EMR.
3. Nella sezione Cluster scaling and provisioning option (Opzione di dimensionamento e provisioning del cluster), specifica nuovi valori per il numero Minimum (Minimo) e Maximum (Massimo) di istanze e per On-Demand limit (Limite on demand).

Utilizzo della AWS CLI per configurare il dimensionamento gestito

Puoi utilizzare i comandi della AWS CLI per Amazon EMR per configurare il dimensionamento gestito quando crei un cluster. È possibile utilizzare una sintassi abbreviata, specificando la configurazione JSON inline all'interno dei relativi comandi, oppure si può fare riferimento a un file contenente la configurazione JSON. Inoltre puoi applicare una policy di dimensionamento gestito a un cluster esistente e rimuovere una policy di dimensionamento gestito applicata in precedenza. Inoltre, è possibile recuperare i dettagli di una configurazione di policy di dimensionamento da un cluster in esecuzione.

Abilitazione del dimensionamento gestito durante l'avvio del cluster

Puoi abilitare il dimensionamento gestito durante l'avvio del cluster, come illustrato nell'esempio seguente.

```
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}'
```

Inoltre puoi specificare una configurazione di policy gestite utilizzando l'opzione --managed-scaling-policy quando usi `create-cluster`.

Applicazione di una policy di dimensionamento gestito a un cluster esistente

Puoi applicare una policy di dimensionamento gestito a un cluster esistente come illustrato nell'esempio seguente.

```
aws emr put-managed-scaling-policy
--cluster-id j-123456
--managed-scaling-policy ComputeLimits='{MinimumCapacityUnits=1,
MaximumCapacityUnits=10,  MaximumOnDemandCapacityUnits=10, UnitType=Instances}'
```

Inoltre puoi applicare una policy di dimensionamento gestito a un cluster esistente utilizzando il comando `aws emr put-managed-scaling-policy`. L'esempio seguente utilizza un riferimento a un file JSON, `managedscaleconfig.json`, che specifica la configurazione della policy di dimensionamento gestito.

```
aws emr put-managed-scaling-policy --cluster-id j-123456 --managed-scaling-policy
 file://./managedscaleconfig.json
```

Nell'esempio seguente viene illustrato il contenuto del file `managedscaleconfig.json` che definisce la policy di dimensionamento gestito.

```
{
    "ComputeLimits": {
        "UnitType": "Instances",
        "MinimumCapacityUnits": 1,
        "MaximumCapacityUnits": 10,
        "MaximumOnDemandCapacityUnits": 10
    }
}
```

Recupero di una configurazione di policy di dimensionamento gestito

Il comando `GetManagedScalingPolicy` recupera la configurazione della policy. Ad esempio, il seguente comando recupera la configurazione per il cluster con un ID cluster pari a `j-123456`.

```
aws emr get-managed-scaling-policy --cluster-id j-123456
```

Il comando produce il seguente esempio di output.

```
{
   "ManagedScalingPolicy": {
      "ComputeLimits": {
         "MinimumCapacityUnits": 1,
         "MaximumOnDemandCapacityUnits": 10,
         "MaximumCapacityUnits": 10,
         "UnitType": "Instances"
      }
   }
}
```

Per ulteriori informazioni sull'utilizzo dei comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Rimozione della policy di dimensionamento gestito

Il comando `RemoveManagedScalingPolicy` rimuove la configurazione della policy. Ad esempio, il seguente comando recupera la configurazione per il cluster con ID cluster `j-123456`.

```
aws emr remove-managed-scaling-policy --cluster-id j-123456
```

Usare AWS SDK for Java per configurare il dimensionamento gestito

Il seguente estratto di programma mostra come configurare il dimensionamento gestito utilizzando 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() { }
}
```

Informazioni su strategia e scenari di assegnazione dei nodi

In questa sezione viene fornita una panoramica della strategia di allocazione dei nodi e degli scenari di dimensionamento comuni che è possibile utilizzare con il dimensionamento gestito da Amazon EMR.

Strategia di assegnazione dei nodi

Il dimensionamento gestito da Amazon EMR assegna nodi core e attività in base alle seguenti strategie di aumento e riduzione:

Strategia di dimensionamento verso l'alto

- Il dimensionamento gestito da Amazon EMR aggiunge capacità prima ai nodi core e poi ai nodi attività fino al raggiungimento della capacità massima consentita o fino al raggiungimento della capacità target di aumento desiderata.

- Quando Amazon EMR subisce un ritardo nel dimensionamento verticale rispetto al gruppo di istanze corrente, i cluster che utilizzano il dimensionamento gestito passano automaticamente a un gruppo di istanze di attività diverso.
- Se hai impostato il parametro `MaximumCoreCapacityUnits`, Amazon EMR ridimensiona i nodi principali fino a quando le unità principali non raggiungono il limite massimo consentito. La capacità rimanente viene aggiunta ai nodi attività.
- Se hai impostato il parametro `MaximumOnDemandCapacityUnits`, Amazon EMR ridimensiona il cluster utilizzando le istanze on demand fino a quando le unità on demand non raggiungono il limite massimo consentito. La capacità rimanente viene aggiunta utilizzando istanze Spot.
- Se hai impostato entrambi i parametri `MaximumCoreCapacityUnits` e `MaximumOnDemandCapacityUnits`, Amazon EMR considera entrambi i limiti durante il dimensionamento.

  Ad esempio, se l'opzione `MaximumCoreCapacityUnits` è minore di `MaximumOnDemandCapacityUnits`, Amazon EMR ridimensiona innanzitutto i nodi principali fino al raggiungimento del limite di capacità principale. Per la capacità rimanente, Amazon EMR utilizza innanzitutto le istanze on demand per ridimensionare i nodi attività fino al raggiungimento del limite on demand e, in seguito, utilizza le istanze Spot per i nodi attività.

Strategia di riduzione verticale

- Amazon EMR versioni 5.34.0 e successive e Amazon EMR versioni 6.4.0 e successive, supportano il dimensionamento gestito a conoscenza dei dati di shuffle di Spark (dati che Spark ridistribuisce tra le partizioni per eseguire operazioni specifiche). Per ulteriori informazioni sulle operazioni di shuffle, consulta la Guida di programmazione Spark. Il dimensionamento gestito riduce verticalmente solo le istanze sottoutilizzate e che non contengono dati di shuffle utilizzati attivamente. Questo dimensionamento intelligente previene la perdita di dati di shuffle, evitando la necessità di nuovi tentativi di eseguire processi e calcolo dei dati intermedi.
- Il dimensionamento gestito da Amazon EMR rimuove prima i nodi attività e poi i nodi core fino al raggiungimento della capacità target di riduzione desiderata. Il cluster non ridimensiona mai al di sotto dei vincoli minimi nella policy di dimensionamento gestito.
- All'interno di ogni tipo di nodo (nodi core o nodi attività), il dimensionamento gestito da Amazon EMR rimuove prima le istanze spot e poi le istanze on demand.
- Per i cluster avviati con Amazon EMR 5.x rilasci 5.34.0 e successivi e 6.x rilasci 6.4.0 e successivi, il dimensionamento gestito da Amazon EMR non riduce i nodi su cui è in esecuzione `ApplicationMaster` per Apache Spark. Ciò riduce al minimo gli errori e i nuovi tentativi del

processo, il che aiuta a migliorare le prestazioni lavorative e a ridurre i costi. Per confermare su quali nodi del cluster è in esecuzione `ApplicationMaster`, visita Spark History Server e filtra i driver nella scheda Esecutori dell'ID applicazione Spark.

Se il cluster non ha alcun carico, Amazon EMR annulla l'aggiunta di nuove istanze da una valutazione precedente ed esegue operazioni di dimensionamento verso il basso. Se il cluster presenta un carico pesante, Amazon EMR annulla la rimozione delle istanze ed esegue operazioni di dimensionamento verso l'alto.

Considerazioni sull'assegnazione

È consigliabile utilizzare l'opzione di acquisto on demand per i nodi principali al fine di evitare la perdita di dati HDFS in caso di recupero Spot. È possibile utilizzare l'opzione di acquisto Spot per i nodi attività al fine di ridurre i costi e ottenere un'esecuzione più rapida dei processi quando vengono aggiunte più istanze Spot ai nodi attività.

Scenari di assegnazione dei nodi

È possibile creare vari scenari di dimensionamento in base alle proprie esigenze impostando i parametri massimo, minimo, limite on demand e nodo principale massimo in diverse combinazioni.

Scenario 1: Dimensionare i soli nodi principali

Per ridimensionare solo i nodi principali, i parametri di dimensionamento gestiti devono soddisfare i seguenti requisiti:

- Il limite on demand è uguale al limite massimo.
- Il nodo principale massimo è uguale al limite massimo.

Quando non vengono specificati i parametri limite on demand e nodo principale massimo, entrambi i parametri impostano il limite massimo.

I seguenti esempi dimostrano lo scenario di dimensionamento dei nodi principali.

Stato iniziale del cluster	Parametri di dimensionamento	Comportamento del dimension amento
Gruppi di istanze	`UnitType`: Istanze	

Stato iniziale del cluster	Parametri di dimensionamento	Comportamento del dimensionamento
Principale: 1 on demand Attività: 1 on demand e 1 Spot	`MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	Dimensionamento da 1 a 20 istanze o unità di parchi istanze sui nodi principali utilizzando il tipo on demand. Nessun dimensionamento sui nodi attività.
Parchi istanze Principale: 1 on demand Attività: 1 on demand e 1 Spot	Tipo di unità: Unità di parchi istanza `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 20	

Scenario 2: Dimensionamento dei soli nodi attività

Per ridimensionare solo i nodi attività, i parametri di dimensionamento gestiti devono soddisfare il seguente requisito:

- Il nodo principale massimo è uguale al limite minimo.

I seguenti esempi dimostrano lo scenario di dimensionamento dei nodi attività.

Stato iniziale del cluster	Parametri di dimensionamento	Comportamento del dimensionamento
Gruppi di istanze	`UnitType`: Istanze	

Stato iniziale del cluster	Parametri di dimensionamento	Comportamento del dimensionamento
Principale: 2 on demand Attività: 1 Spot	`MinimumCapacityUnits` 2: `MaximumCapacityUnits` : 20 `MaximumCoreCapacityUnits` 2:	Mantenimento dei nodi principali a 2 e dimensionamento dei soli nodi attività da 0 a 18 istanze o unità di parchi istanze. La capacità tra i limiti minimo e massimo viene aggiunta solo ai nodi attività.
Parchi istanze Principale: 2 on demand Attività: 1 Spot	`UnitType`: Unità di parchi istanza `MinimumCapacityUnits` 2: `MaximumCapacityUnits` : 20 `MaximumCoreCapacityUnits` 2:	

Scenario 3: Solo istanze On Demand nel cluster

Per avere solo istanze on demand, il cluster e i parametri di dimensionamento gestito devono soddisfare i requisiti seguenti:

- Il limite on demand è uguale al limite massimo.

  Quando il limite on demand non è specificato, il valore del parametro viene impostato di default sul limite massimo. Il valore di default indica che Amazon EMR ridimensiona solo le istanze on demand.

Se il nodo principale massimo è inferiore al limite massimo, è possibile utilizzare il parametro nodo principale massimo per dividere l'assegnazione della capacità tra nodi principali e nodi attività.

Per abilitare questo scenario in un cluster composto da gruppi di istanze, tutti i gruppi di nodi nel cluster devono utilizzare il tipo di mercato on demand durante la configurazione iniziale.

Negli esempi seguenti viene illustrato lo scenario che presenta istanze on demand nell'intero cluster.

Stato iniziale del cluster	Parametri di dimensionamento	Comportamento del dimension amento
Gruppi di istanze Principale: 1 on demand Attività: 1 on demand	`UnitType`: Istanze `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	Dimension amento da 1 a 12 istanze o unità di parchi istanze sui nodi principal i utilizzando il tipo on demand. Dimension amento della capacità rimanente utilizzando l'opzione on demand sui nodi attività. Nessun dimension amento utilizzan do istanze Spot.
Parchi istanze Principale: 1 on demand Attività: 1 on demand	`UnitType`: Unità di parchi istanza `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 20 `MaximumCoreCapacityUnits` : 12	

Scenario 4: Solo istanze Spot nel cluster

Per avere solo istanze Spot, i parametri di dimensionamento gestito devono soddisfare i requisiti seguenti:

- Il limite on demand è impostato su 0.

Se il nodo principale massimo è inferiore al limite massimo, è possibile utilizzare il parametro nodo principale massimo per dividere l'assegnazione della capacità tra nodi principali e nodi attività.

Per abilitare questo scenario in un cluster composto da gruppi di istanze, il gruppo di istanze principale deve utilizzare l'opzione acquisto Spot durante la configurazione iniziale. Se non è

presente alcuna istanza spot nel gruppo di istanze attività, il dimensionamento gestito da Amazon EMR crea un gruppo di attività utilizzando le istanze spot quando necessario.

Negli esempi seguenti viene illustrato lo scenario che presenta istanze Spot nell'intero cluster.

Stato iniziale del cluster	Parametri di dimensionamento	Comportamento del dimensionamento
Gruppi di istanze Principale: 1 Spot Attività: 1 Spot	`UnitType`: Istanze `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	Dimensionamento da 1 a 20 istanze o unità di parchi istanze sui nodi principali utilizzando il tipo Spot. Nessun dimensionamento utilizzando il tipo on demand.
Parchi istanze Principale: 1 Spot Attività: 1 Spot	`UnitType`: Unità di parchi istanza `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 0	

Scenario 5: Dimensionamento delle istanze On Demand sui nodi principali e delle istanze Spot sui nodi attività

Per dimensionare istanze On Demand sui nodi principali e istanze Spot sui nodi attività, i parametri di ridimensionamento gestiti devono soddisfare i seguenti requisiti:

- Il limite on demand deve essere uguale al nodo principale massimo.
- Sia il limite on demand che il nodo principale massimo devono essere inferiori al limite massimo.

Per abilitare questo scenario in un cluster composto da gruppi di istanze, il gruppo di nodi principale deve utilizzare l'opzione di acquisto on demand.

Negli esempi seguenti viene illustrato lo scenario di dimensionamento delle istanze on demand sui nodi principali e delle istanze Spot sui nodi attività.

Stato iniziale del cluster	Parametri di dimensionamento	Comportamento del dimensionamento
Gruppi di istanze Principale: 1 on demand Attività: 1 on demand e 1 Spot	`UnitType`: Istanze `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	Dimensionamento verso l'alto a 6 unità on demand sul nodo principale poiché nel nodo attività è già presente 1 unità on demand e il limite massimo per il tipo on demand è 7. In seguito, dimensionamento verso l'alto a 13 unità Spot sui nodi attività.
Parchi istanze Principale: 1 on demand Attività: 1 on demand e 1 Spot	`UnitType`: Unità di parchi istanza `MinimumCapacityUnits` : 1 `MaximumCapacityUnits` : 20 `MaximumOnDemandCapacityUnits` : 7 `MaximumCoreCapacityUnits` : 7	

Informazioni sui parametri di dimensionamento gestito

Amazon EMR pubblica i parametri di alta risoluzione con dati a una granularità di un minuto quando il dimensionamento gestito è abilitato per un cluster. È possibile visualizzare gli eventi su ogni iniziazione e completamento del ridimensionamento controllati dal dimensionamento gestito con la console Amazon EMR o la console Amazon CloudWatch. I parametri di CloudWatch sono fondamentali per il funzionamento del dimensionamento gestito di Amazon EMR. Si consiglia di monitorare attentamente i parametri di CloudWatch per assicurarsi che i dati non manchino. Per ulteriori informazioni su come configurare gli allarmi CloudWatch per rilevare i parametri mancanti,

consulta Utilizzo degli allarmi Amazon CloudWatch. Per ulteriori informazioni sul utilizzo degli eventi CloudWatch con Amazon EMR, consulta Monitoraggio di CloudWatch Events.

i parametri seguenti indicano le capacità correnti o di destinazione di un cluster. Questi parametri sono disponibili solo quando è abilitata il dimensionamento gestito. Per i cluster composti da parchi istanze, i parametri della capacità del cluster vengono misurate in `Units`. Per i cluster composti da gruppi di istanze, i parametri della capacità del cluster vengono misurate in `Nodes` o in `vCPU` in base al tipo di unità utilizzato nella policy di dimensionamento gestito.

Parametro	Descrizione
• `TotalUnitsRequested` • `TotalNodesRequested` • `TotalVCPURequested`	Numero totale di unità/nodi/vCPU di destinazione in un cluster determinato dal dimensionamento gestito. Unità: numero
• `TotalUnitsRunning` • `TotalNodesRunning` • `TotalVCPURunning`	Numero totale di unità/nodi/VCPU correnti disponibili in un cluster in esecuzione. Quando viene richiesto il ridimensionamento di un cluster, questo parametro verrà aggiornato o dopo l'aggiunta o la rimozione delle nuove istanze dal cluster. Unità: numero
• `CoreUnitsRequested` • `CoreNodesRequested` • `CoreVCPURequested`	Il numero di unità/nodi/VCPU CORE di destinazione in un cluster determinato dal dimensionamento gestito. Unità: numero
• `CoreUnitsRunning` • `CoreNodesRunning`	Il numero di unità/nodi/VCPU CORE correnti in esecuzione in un cluster. Unità: numero

Parametro	Descrizione
• `CoreVCPURunning`	
• `TaskUnitsRequested` • `TaskNodesRequested` • `TaskVCPURequested`	Il numero di unità/nodi/vCPU TASK di destinazione in un cluster determinato dal dimensionamento gestito. Unità: numero
• `TaskUnitsRunning` • `TaskNodesRunning` • `TaskVCPURunning`	Il numero di unità//nodi/vCPU TASK correnti in esecuzione in un cluster. Unità: numero

I parametri seguenti indicano lo stato di utilizzo del cluster e delle applicazioni. Questi parametri sono disponibili per tutte le caratteristiche Amazon EMR, ma vengono pubblicati a una risoluzione più elevata con dati a una granularità di un minuto quando il dimensionamento gestito è abilitato per un cluster. È possibile correlare i parametri seguenti con i parametri della capacità del cluster nella tabella precedente per comprendere le decisioni relative al dimensionamento gestito.

Parametro	Descrizione
`AppsCompleted`	Il numero di applicazioni inviate a YARN che sono state completate. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
`AppsPending`	Il numero di applicazioni inviate a YARN che sono in attesa. Caso d'uso: monitorare l'avanzamento del cluster

Parametro	Descrizione
	Unità: numero
AppsRunning	Il numero di applicazioni inviate a YARN che sono in esecuzione. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
ContainerAllocated	Il numero di container di risorse allocati da ResourceManager. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
ContainerPending	Il numero di container nella coda non ancora allocati. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
ContainerPendingRatio	Il rapporto tra i container in attesa e quelli allocati (ContainerPendingRatio = ContainerPending / ContainerAllocated). Se ContainerAllocated = 0, allora ContainerPendingRatio = ContainerPending. Il valore di ContainerPendingRatio rappresenta un numero, non una percentuale. Questo valore è utile per il dimensionamento delle risorse del cluster in funzione del comportamento di attribuzione dei container. Unità: numero

Parametro	Descrizione
`HDFSUtilization`	La percentuale di storage HDFS attualmente utilizzato. Caso d'uso: analizzare le prestazioni del cluster Unità: percentuale
`IsIdle`	Indica che un cluster non è più in esecuzione ma è ancora attivo e genera spese. È impostato su 1 se non vi sono task e processi in esecuzione, altrimenti è impostato su 0. Questo valore viene verificato a intervalli di cinque minuti e un valore 1 indica unicamente l'inattività del cluster al momento della verifica e non durante i cinque minuti. Per evitare falsi positivi, devi attivare un allarme quando questo valore è 1 durante due o più verifiche consecutive di cinque minuti. Ad esempio, puoi attivare un allarme se questo valore è 1 per trenta minuti o più. Caso d'uso: monitorare le prestazioni del cluster Unità: booleane
`MemoryAvailableMB`	La quantità di memoria disponibile da allocare. Caso d'uso: monitorare l'avanzamento del cluster Unità: numero
`MRActiveNodes`	Il numero di nodi in cui sono attualmente in esecuzione processi o task MapReduce. Equivalente al parametro YARN `mapred.resourcemanager.NoOfActiveNodes` . Caso d'uso: monitorare l'avanzamento del cluster Unità: numero

Parametro	Descrizione
YARNMemoryAvailablePercentage	La percentuale di memoria rimanente disponibile per YARN (YARNMemoryAvailablePercentage = MemoryAvailableMB / MemoryTotalMB). Questo valore è utile per il dimensionamento delle risorse del cluster in funzione dell'utilizzo della memoria di YARN. Unità: percentuale

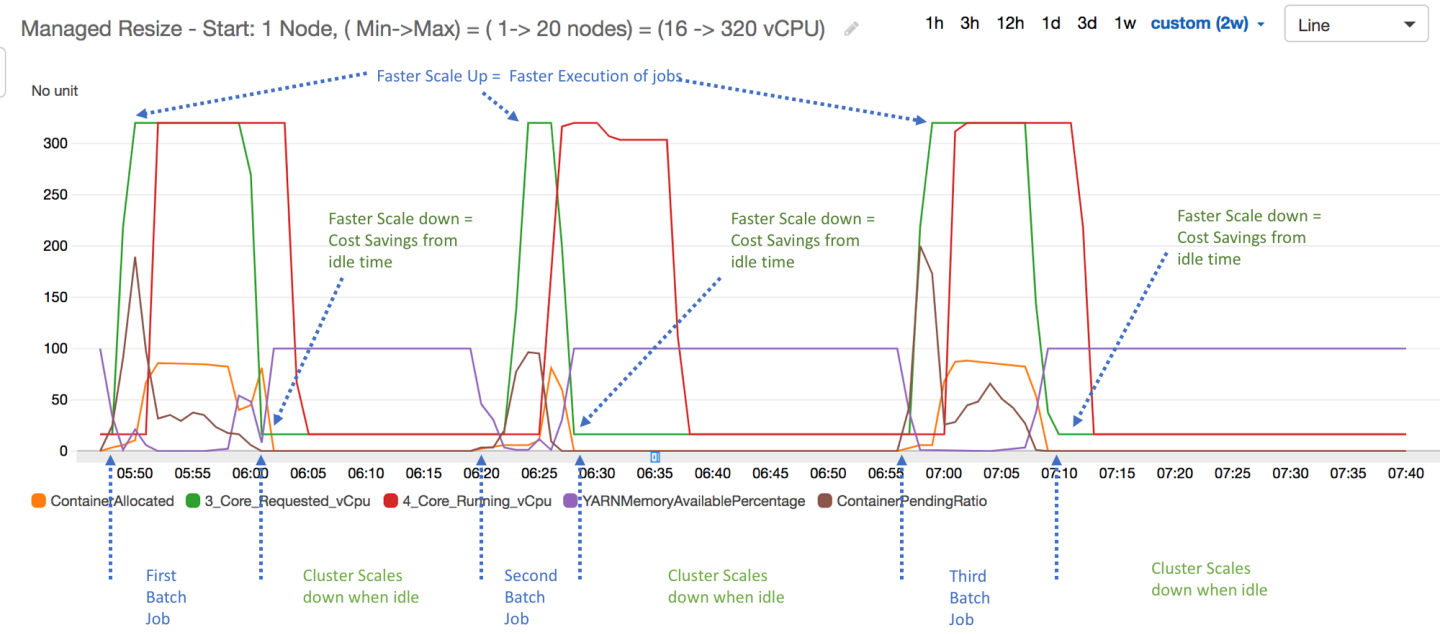
Grafici dei parametri di dimensionamento gestito

Puoi visualizzare in grafico i parametri per vedere i modelli di carico di lavoro del cluster e le corrispondenti decisioni di dimensionamento adottate dal dimensionamento gestito da Amazon EMR come illustrato nella procedura riportata di seguito.

Per visualizzare in grafico i parametri di dimensionamento gestito nella console CloudWatch

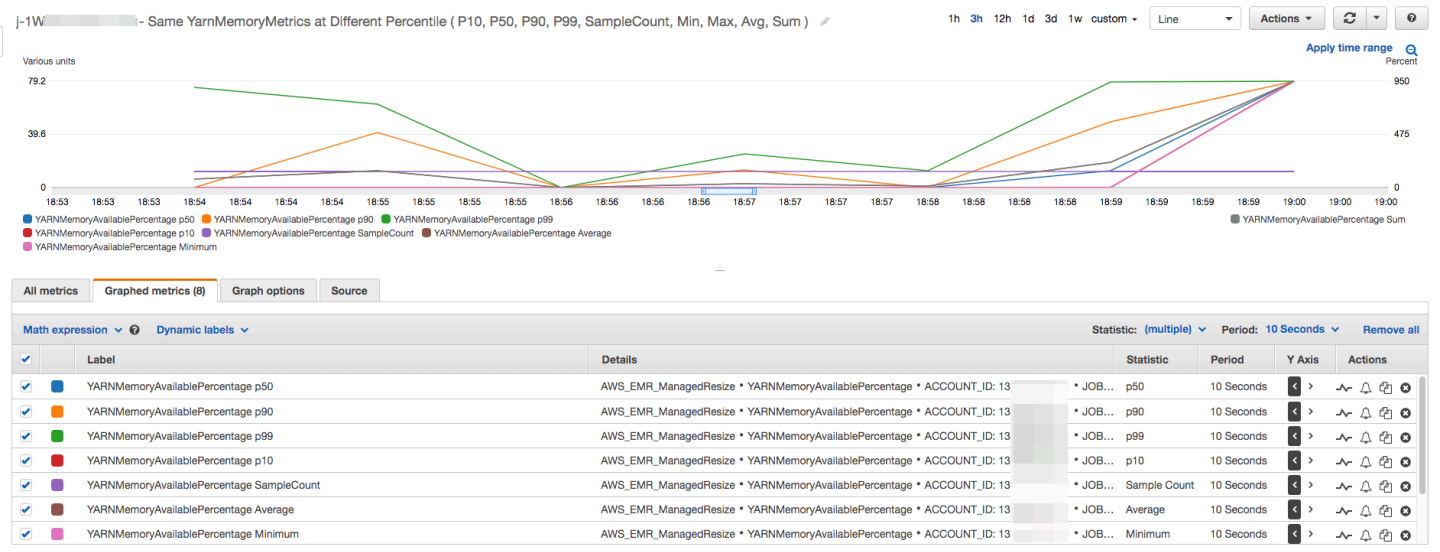
1. Apri la CloudWatch console (Console CloudWatch).
2. Nel riquadro di navigazione, seleziona Amazon EMR. È possibile cercare il cluster da monitorare in base al relativo identificatore.
3. Scorrere fino al parametro da rappresentare graficamente. Aprire un parametro per visualizzare il grafico.
4. Per rappresentare graficamente uno o più parametri, seleziona la casella di controllo accanto a ciascun parametro.

Nell'esempio seguente viene illustrata l'attività di dimensionamento gestito da Amazon EMR di un cluster. Il grafico mostra tre periodi di dimensionamento automatico, che consentono di risparmiare sui costi quando è presente un carico di lavoro meno attivo.



Tutti i parametri relativi alla capacità e all'utilizzo del cluster vengono pubblicati a intervalli di un minuto. Anche altre informazioni statistiche sono associate a ogni dato di un minuto, permettendo il monitoraggio di varie funzioni come `Percentiles`, `Min`, `Max`, `Sum`, `Average`, `SampleCount`.

Ad esempio, il grafico seguente traccia lo stesso parametro `YARNMemoryAvailablePercentage` in percentili diversi, P10, P50, P90, P99, insieme a `Sum`, `Average`, `Min`, `SampleCount`.



Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze

Il dimensionamento automatico con una policy personalizzata in Amazon EMR rilascio 4.0 e successivi consente di impiegare la scalabilità orizzontale e ridimensionare in modo programmatico i nodi principali e nodi attività basati su un parametro CloudWatch e su altri parametri specificati in una policy di dimensionamento. Il dimensionamento automatico con una policy personalizzata è disponibile con la configurazione dei gruppi di istanze e non è disponibile quando si utilizzano i parchi istanze. Per ulteriori informazioni sui gruppi di istanze e i parchi di istanze, vedere Creazione di un cluster con parchi istanze o gruppi di istanze uniformi.

> **Note**
>
> Per utilizzare la scalabilità automatica con una caratteristica di policy personalizzata in Amazon EMR, è necessario impostare `true` per il parametro `VisibleToAllUsers` quando si crea un cluster. Per ulteriori informazioni, consulta SetVisibleToAllUsers.

La policy di dimensionamento è parte di un gruppo di istanze di configurazione. È possibile specificare una policy durante la configurazione iniziale di un gruppo di istanze o modificando un gruppo di istanze in un cluster esistente, anche quando tale gruppo di istanze è attivo. Ogni gruppo di istanze in un cluster, eccetto il gruppo di istanze primarie, può avere la propria policy di dimensionamento, che consiste in regole di aumento e riduzione orizzontali. Le regole di scalabilità verticale e orizzontale possono essere configurate in modo indipendente, con parametri diversi per ciascuna regola.

Puoi configurare le policy di dimensionamento con la AWS Management Console, la AWS CLI oppure l'API di Amazon EMR. Quando si utilizza la AWS CLI o l'API Amazon EMR, è necessario specificare la policy di scalabilità in formato JSON. Inoltre, quando utilizzi la AWS CLI o l'API di Amazon EMR, puoi specificare parametri CloudWatch personalizzati. I parametri personalizzati non sono selezionabili con la AWS Management Console. Quando crei inizialmente una policy di dimensionamento con la console, viene preconfigurata una policy predefinita adatta a molte applicazioni per facilitare l'avvio. È possibile eliminare o modificare le regole predefinite.

Anche se la scalabilità automatica consente di regolare la capacità del cluster EMR in tempo reale, è necessario considerare i requisiti del carico di lavoro della baseline e pianificare le configurazioni del nodo e del gruppo di istanze. Per ulteriori informazioni, consulta Linee guida di configurazione del cluster.

> **Note**
>
> Per la maggior parte dei carichi di lavoro, è auspicabile impostare regole di scalabilità orizzontale e verticale per ottimizzare l'utilizzo delle risorse. L'impostazione di una regola senza l'altra significa che è necessario ridimensionare manualmente il conteggio delle istanze dopo un'attività di dimensionamento. In altre parole, questo imposta una policy di scalabilità orizzontale o verticale automatica "unidirezionale" o con un reset manuale.

Creazione del ruolo IAM per la scalabilità automatica

La scalabilità automatica in Amazon EMR richiede un ruolo IAM con autorizzazioni per aggiungere e terminare le istanze quando vengono attivate le attività di dimensionamento. A questo scopo è disponibile un ruolo predefinito configurato con la policy di ruolo e di attendibilità più appropriata, ovvero `EMR_AutoScaling_DefaultRole`. Quando crei un cluster con una policy di dimensionamento per la prima volta con la AWS Management Console, Amazon EMR crea il ruolo predefinito e collega la policy gestita predefinita per le autorizzazioni, `AmazonElasticMapReduceforAutoScalingRole`.

Quando crei un cluster con una policy di dimensionamento automatico con la AWS CLI, prima devi assicurarti che il ruolo IAM predefinito esista o di disporre di un ruolo IAM personalizzato collegato a una policy che fornisce le autorizzazioni appropriate. Per creare il ruolo predefinito, è possibile eseguire il comando `create-default-roles` prima di creare un cluster. È quindi possibile specificare l'opzione `--auto-scaling-role EMR_AutoScaling_DefaultRole` al momento della creazione di un cluster. In alternativa, è possibile creare un ruolo di scalabilità automatica personalizzato, quindi specificarlo quando si crea un cluster, ad esempio `--auto-scaling-role` *`MyEMRAutoScalingRole`*. Se si crea un ruolo di scalabilità automatica personalizzato per Amazon EMR, si consiglia di basare le policy di autorizzazione per il proprio ruolo personalizzato sulla base della policy gestita. Per ulteriori informazioni, consulta Configurazione dei ruoli di servizio IAM per le autorizzazioni di Amazon EMR per i servizi e risorse AWS.

Comprensione delle regole di scalabilità automatica

Quando una regola di scalabilità orizzontale attiva un'attività di dimensionamento per un gruppo di istanze, le istanze Amazon EC2 vengono aggiunte al gruppo di istanze alle regole correnti. Nuovi nodi possono essere utilizzati da applicazioni come Apache Spark, Apache Hive e Presto non appena l'istanza Amazon EC2 passa allo stato `InService`. È anche possibile impostare una regola di scalabilità verticale che chiude le istanze e rimuove i nodi. Per ulteriori informazioni sul ciclo di vita

delle istanze Amazon EC2 soggette a scalabilità automatica, consulta Auto Scaling Lifecycle (Ciclo di vita della scalabilità automatica) in Guida per l'utente di Amazon EC2 Auto Scaling.

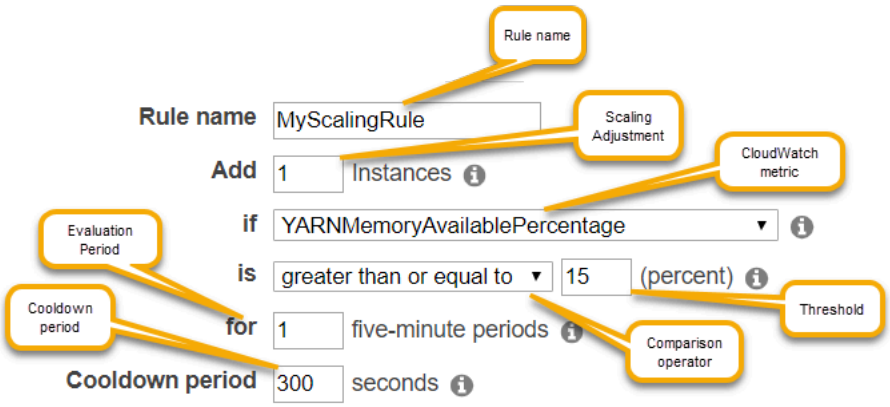
È possibile configurare il modo in cui un cluster termina le istanze Amazon EC2. È possibile scegliere di terminare l'istanza allo scadere dell'ora dell'istanza Amazon EC2 per la fatturazione o al termine dell'attività. Questa impostazione si applica sia alla scalabilità automatica che alle operazioni di ridimensionamento manuale. Per ulteriori informazioni su questa configurazione, consulta Ridimensionamento del cluster.

I seguenti parametri per ciascuna regola di una policy determinano il comportamento di scalabilità automatica.

> **Note**
>
> I parametri elencati in questa pagina sono basati sulla AWS Management Console per Amazon EMR. Quando si utilizza la AWS CLI o l'API Amazon EMR, sono disponibili ulteriori opzioni di configurazione avanzate. Per ulteriori informazioni sulle opzioni avanzate, consulta SimpleScalingPolicyConfiguration nella Guida di riferimento alle API di Amazon EMR.

- Il numero massimo di istanze e istanze minime. Il vincolo Maximum instances (Istanze massime) specifica il numero massimo di istanze Amazon EC2 presenti nel gruppo di istanze e si applica a tutte le regole di scalabilità orizzontale. Allo stesso modo, il vincolo Minimum instances (Istanze minime) specifica il numero minimo di istanze Amazon EC2 e si applica a tutte le regole di scalabilità verticale.
- Il Rule name (Nome della regola), che deve essere univoco all'interno della policy.
- La Scaling adjustment (Regolazione di dimensionamento), che determina il numero di istanze EC2 da aggiungere (per le regole di scalabilità orizzontale) o terminare (per le regole di scalabilità verticale) durante l'operazione di scalabilità attivata dalla regola.
- Il CloudWatch metric (Parametro CloudWatch), che viene visualizzato per una condizione di allarme.
- Un Comparison operator (Operatore di confronto), che viene utilizzato per confrontare il parametro del CloudWatch con il valore Threshold (Soglia) e determinare una condizione di trigger.
- Un Evaluation period (Periodo di valutazione), in incrementi di cinque minuti, per il quale il parametro CloudWatch deve essere in una condizione di trigger prima che venga attivata l'attività di dimensionamento.

- Un Cooldown period (Periodo di attesa), in secondi, che determina la quantità di tempo che deve intercorrere tra un'attività di dimensionamento avviato da una regola e l'inizio dell'attività di dimensionamento successiva, indipendentemente dalla regola che la attiva. Quando un gruppo di istanze ha terminato un'attività di dimensionamento e ha raggiunto il suo stato post-dimensionamento, il periodo di attesa fornisce un'opportunità per i parametri CloudWatch che potrebbero innescare le successive attività di dimensionamento per stabilizzarsi. Per ulteriori informazioni, consulta Periodi di attesa di Auto Scaling nella Guida per l'utente di Amazon EC2 Auto Scaling.



Considerazioni e limitazioni

- I parametri di Amazon CloudWatch sono fondamentali per il funzionamento di scalabilità automatica di Amazon EMR. Si consiglia di monitorare attentamente i parametri di Amazon CloudWatch per assicurarsi che i dati non manchino. Per ulteriori informazioni su come configurare gli allarmi Amazon CloudWatch per rilevare i parametri mancanti, consulta Utilizzo degli allarmi Amazon CloudWatch.
- L'utilizzo eccessivo dei volumi EBS può causare problemi di dimensionamento gestito. Si consiglia di monitorare attentamente l'utilizzo del volume EBS per assicurarsi che il volume EBS sia inferiore al 90% di utilizzo. Consulta Archiviazione dell'istanza per informazioni su come specificare volumi EBS aggiuntivi.
- Il dimensionamento automatico con una policy personalizzata nei rilasci da 5.18 a 5.28 di Amazon EMR potrebbe riscontrare un errore di dimensionamento causato da dati mancati a intermittenza nei parametri di Amazon CloudWatch. Si consiglia di utilizzare le versioni più recenti di Amazon EMR per migliorare la scalabilità automatica. Se desideri utilizzare un rilascio di Amazon EMR compreso tra 5.18 e 5.28, puoi anche contattare il Supporto AWS per richiedere una patch.

Uso di AWS Management Console per configurare la scalabilità automatica

Quando crei un cluster, configuri una policy di dimensionamento per i gruppi di istanze mediante le opzioni di configurazione avanzate del cluster. È anche possibile creare o modificare una policy di dimensionamento per un gruppo di istanze attive modificando i gruppi di istanze nelle impostazioni Hardware di un cluster esistente.

> **Note**
>
> La nuova console Amazon EMR (https://console.aws.amazon.com/emr) utilizza il dimensionamento gestito anziché il dimensionamento automatico. Per utilizzare il dimensionamento automatico, assicurati di aver effettuato l'accesso alla vecchia console all'indirizzo https://console.aws.amazon.com/elasticmapreduce.

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Se stai creando un cluster, nella console di Amazon EMR seleziona Create Cluster (Crea cluster), quindi Go to advanced options (Vai alle opzioni avanzate), scegli l'opzione per Step 1: Software and Steps (Fase 1: software e fasi) e procedi a Step 2: Hardware Configuration (Fase 2: configurazione hardware).

   - o -

   Se si modifica un gruppo di istanze in un cluster in esecuzione, selezionare il cluster dall'elenco del cluster, quindi espandere la sezione Hardware.
3. In Cluster scaling and provisioning option (Opzione di dimensionamento e provisioning del cluster), seleziona Enable cluster scaling (Abilita dimensionamento del cluster). Quindi selezionare Create a custom automatic scaling policy (Crea policy di dimensionamento automatico personalizzato).

   Nella tabella Custom automatic scaling policies (Policy di dimensionamento automatico personalizzato), fare clic sull'icona a forma di matita visualizzata nella riga del gruppo di istanze che si desidera configurare. Si apre la schermata delle regole di Auto Scaling.
4. Digitare il Maximum instances (Numero massimo di istanze) che si desidera che il gruppo di istanze contenga dopo che è stato dimensionato orizzontalmente e digitare il Minimum instances

(Numero di istanze minimo) che si desidera che il gruppo di istanze contenga dopo che è stato dimensionato verticalmente.

5. Fare clic sulla matita per modificare i parametri della regola, fare clic sulla X per rimuovere una regola dalla policy e fare clic su Add rule (Aggiungi regola) per aggiungere ulteriori regole.
6. Scegliere i parametri delle regole come descritto in precedenza in questo argomento. Per le descrizioni dei parametri CloudWatch disponibili per Amazon EMR, consulta Parametri e dimensioni di Amazon EMR nella Guida per l'utente di Amazon CloudWatch.

Uso di AWS CLI per configurare la scalabilità automatica

È possibile utilizzare i comandi della AWS CLI per Amazon EMR per configurare la scalabilità automatica quando si crea un cluster e un gruppo di istanze. È possibile utilizzare una sintassi abbreviata, specificando la configurazione JSON inline all'interno dei relativi comandi, oppure si può fare riferimento a un file contenente la configurazione JSON. È inoltre possibile applicare una policy di scalabilità automatica a un gruppo di istanze esistente e rimuovere una policy di scalabilità automatica precedentemente applicata. Inoltre, è possibile recuperare i dettagli di una configurazione di policy di dimensionamento da un cluster in esecuzione.

> **Important**
>
> Quando si crea un cluster basato su una policy di scalabilità automatica, è necessario utilizzare il comando `--auto-scaling-role` *`MyAutoScalingRole`* per specificare il ruolo IAM per la scalabilità automatica. Il ruolo predefinito è *`EMR_AutoScaling_DefaultRole`* e può essere creato con il comando `create-default-roles`. Il ruolo può essere aggiunto solo quando il cluster viene creato e non può essere aggiunto a un cluster esistente.

Per una descrizione dettagliata dei parametri disponibili durante la configurazione di una policy di scalabilità automatica, consulta PutAutoScalingPolicy nella Guida di riferimento alle API di Amazon EMR.

Creazione di un cluster con una policy di scalabilità automatica applicata a un gruppo di istanze

È possibile specificare una configurazione di scalabilità automatica all'interno dell'opzione `--instance-groups` del comando `aws emr create-cluster`. L'esempio seguente mostra un comando di creazione di cluster in cui viene fornita una policy di scalabilità automatica per il gruppo di istanze principale. Il comando crea una configurazione di dimensionamento equivalente alla policy di

aumento orizzontale predefinita che appare quando crei una policy di dimensionamento automatico con la AWS Management Console per Amazon EMR. Per brevità, non viene mostrata una policy di scalabilità verticale. Non è consigliabile creare una regola di scalabilità orizzontale senza una regola di scalabilità verticale.

```
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
```

Il comando seguente illustra come utilizzare la riga di comando per fornire la definizione della policy di dimensionamento automatico come parte di un file di configurazione del gruppo di istanze denominato *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 il contenuto del file di configurazione come segue:

```
[
{
  "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}"
             }
          ]
        }
       }
      }
     ]
   }
}
]
```

Aggiunta di un gruppo di istanze con policy di scalabilità automatica a un cluster

Puoi specificare una configurazione della policy di dimensionamento utilizzando l'opzione `--instance-groups` con il comando `add-instance-groups` nello stesso modo in cui puoi farlo quando utilizzi `create-cluster`. L'esempio seguente utilizza un riferimento a un file JSON, *`instancegroupconfig.json`*, con la configurazione di un gruppo di istanze.

```
aws emr add-instance-groups --cluster-id j-1EKZ3TYEVF1S2 --instance-groups file://your/
path/to/instancegroupconfig.json
```

Applicazione di una policy di scalabilità automatica a un gruppo di istanze esistenti o modifica di una policy applicata

Utilizzare il comando `aws emr put-auto-scaling-policy` per applicare una policy di scalabilità automatica a un gruppo di istanze esistente. Il gruppo di istanze deve essere parte di un cluster che utilizza il ruolo IAM di scalabilità automatica. L'esempio seguente utilizza un riferimento a un file JSON, *`autoscaleconfig.json`*, che specifica la configurazione automatica della policy di scalabilità automatica.

```
aws emr put-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07 --auto-scaling-policy file://your/path/to/autoscaleconfig.json
```

Il contenuto del file `autoscaleconfig.json`, che definisce la stessa regola di scalabilità orizzontale mostrata nell'esempio precedente, è mostrato di seguito.

```
{
          "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"
                        }
                }
        }]
  }
```

Rimozione di una policy di scalabilità automatica da un gruppo di istanze

```
aws emr remove-auto-scaling-policy --cluster-id j-1EKZ3TYEVF1S2 --instance-group-
id ig-3PLUZBA6WLS07
```

Recupero di una configurazione di policy di scalabilità automatica

Il comando `describe-cluster` recupera la configurazione della policy nel blocco InstanceGroup. Ad esempio, il seguente comando recupera la configurazione per il cluster con un ID cluster pari a `j-1CWOHP4PI30VJ`.

```
aws emr describe-cluster --cluster-id j-1CWOHP4PI30VJ
```

Il comando produce il seguente esempio di output.

```
{
    "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."
            }
        }
    }
}
```

Ridimensionamento manuale di un cluster in esecuzione

È possibile aggiungere e rimuovere istanze da parchi istanze e gruppi di istanze core e attività in un cluster in esecuzione con la AWS Management Console, la AWS CLI o l'API di Amazon EMR. Se un cluster utilizza gruppi di istanze, deve essere cambiato esplicitamente il conteggio delle istanze. Se il cluster utilizza parchi di istanze, è possibile modificare le unità di destinazione per le istanze on demand e le istanze Spot. Il parco di istanze aggiunge e rimuove le istanze per soddisfare la nuova destinazione. Per ulteriori informazioni, consulta Opzioni del parco istanze. Le applicazioni possono utilizzare le nuove istanze Amazon EC2 fornite per ospitare i nodi non appena sono disponibili. Quando le istanze vengono rimosse, Amazon EMR chiude le attività in modo da non interrompere i lavori e proteggerli dalla perdita di dati. Per ulteriori informazioni, consulta Terminazione al completamento dell'attività.

Ridimensionamento di un cluster con la console

È possibile utilizzare la console di Amazon EMR per ridimensionare un cluster in esecuzione.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Modifica del conteggio delle istanze di un cluster esistente con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster) e seleziona il cluster da aggiornare. Il cluster deve essere in esecuzione; non è possibile ridimensionare un cluster di provisioning o terminato.
3. Nella scheda Instances (Istanze) della pagina dei dettagli del cluster, visualizza il pannello Instance groups (Gruppi di istanze).
4. Per ridimensionare un gruppo di istanze esistente, utilizza il pulsante di opzione accanto al gruppo di istanze core o attività che desideri ridimensionare, quindi scegli Resize instance group (Ridimensiona gruppo di istanze). Specifica il nuovo numero di istanze per il gruppo di istanze, quindi seleziona Resize (Ridimensiona).

   > **Note**
   >
   > Se scegli di ridurre le dimensioni di un gruppo di istanze in esecuzione, Amazon EMR selezionerà in modo intelligente le istanze da rimuovere dal gruppo per ridurre al minimo la perdita di dati. Per un controllo più granulare dell'operazione di ridimensionamento, puoi selezionare l'ID del gruppo di istanze, scegliere le istanze che desideri rimuovere e quindi utilizzare l'opzione Terminate (Termina). Per ulteriori informazioni sul comportamento di riduzione intelligente, consulta la sezione Ridimensionamento del cluster.

5. Se desideri annullare l'operazione di ridimensionamento, puoi utilizzare il pulsante di opzione per un gruppo di istanze con lo stato Resizing (In fase di ridimensionamento) e quindi scegliere Stop resize (Interrompi il ridimensionamento) dall'elenco delle operazioni.

6. Per aggiungere uno o più gruppi di istanze attività al cluster in risposta all'aumento del carico di lavoro, scegli Add task instance group (Aggiungi gruppo di istanze attività) dall'elenco delle operazioni. Scegli il tipo di istanza Amazon EC2, inserisci il numero di istanze per il gruppo di attività, quindi seleziona Add task instance group (Aggiungi gruppo di istanze attività) per tornare al pannello Instance groups (Gruppi di istanze) del cluster.

Old console

Modifica del conteggio delle istanze di un cluster esistente con la vecchia console

1. Nella pagina Cluster List (Elenco cluster), fare clic su un cluster da ridimensionare.
2. Nella pagina Cluster Details (Dettagli cluster), scegliere Hardware.
3. Se il cluster utilizza gruppi di istanze, scegliere Resize (Ridimensiona) nella colonna Instance count (Numero di istanze) per il gruppo di istanze che si desidera ridimensionare, digitare un nuovo numero di istanze, quindi selezionare il segno di spunta verde.

   –OPPURE–

   Se il cluster utilizza parchi istanze, seleziona Resize (Ridimensiona) nella colonna Provisioned capacity (Capacità con provisioning), digita nuovi valori per On-demand units (Unità on demand) e Spot units (Unità Spot) e scegli Resize (Ridimensiona).

Quando si modifica il numero di nodi, Status (Stato) del gruppo di istanze viene aggiornato. Quando la modifica richiesta è completata, lo Status (Stato) è Running (In esecuzione).

Ridimensionamento di un cluster con la AWS CLI

È possibile utilizzare la AWS CLI per ridimensionare un cluster in esecuzione. Si può aumentare o diminuire il numero di nodi di attività e si può aumentare il numero di nodi principali in un cluster in esecuzione. È anche possibile chiudere un'istanza nel gruppo di istanze core con la AWS CLI o l'API. Questo deve essere fatto con cautela. La chiusura di un'istanza nel gruppo di istanze core comporta la perdita di dati e l'istanza non viene sostituita automaticamente.

Oltre a ridimensionare i gruppi core e attività, è inoltre possibile aggiungere uno o più gruppi di istanze attività a un cluster in esecuzione con il comando AWS CLI.

Ridimensionamento di un cluster modificando il conteggio delle istanze con la AWS CLI

È possibile aggiungere istanze al gruppo core o al gruppo attività ed è possibile rimuoverle dal gruppo attività utilizzando il sottocomando `modify-instance-groups` di AWS CLI con il parametro `InstanceCount`. Per aggiungere istanze ai gruppi principali o di attività, aumentare il `InstanceCount`. Per ridurre il numero di istanze nel gruppo di attività, diminuire `InstanceCount`. La modifica del conteggio delle istanze del gruppo di attività a 0 rimuove tutte le istanze ma non il gruppo di istanze.

- Per aumentare il numero di istanze nel gruppo di istanze dell'attività da 3 a 4, digitare il seguente comando e sostituire *ig-31JXXXXXXBT0* con l'ID del gruppo di istanze.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-31JXXXXXXBT0,InstanceCount=4
```

  Per recuperare `InstanceGroupId`, utilizzare il sottocomando `describe-cluster`. L'output è un oggetto JSON chiamato `Cluster` contenente l'ID dell'istanza di ciascun gruppo. Per utilizzare questo comando, è necessario l'ID del cluster (che è possibile recuperare con il comando `aws emr list-clusters` o la console). Per recuperare l'ID del gruppo di istanze, digitare il comando seguente e sostituire *j-2AXXXXXXGAPLF* con l'ID del cluster.

```
aws emr describe-cluster --cluster-id j-2AXXXXXXGAPLF
```

  Con il comando AWS CLI, è possibile anche terminare un'istanza nel gruppo di istanze core con il sottocomando `--modify-instance-groups`.

  > **Warning**
  >
  > La specifica di `EC2InstanceIdsToTerminate` deve essere eseguita con cautela. Le istanze vengono terminate immediatamente, indipendentemente dallo stato delle applicazioni in esecuzione su di esse, e l'istanza non viene sostituita automaticamente. Questo è vero indipendentemente dalla configurazione del Scale down behavior (Comportamento di ridimensionamento) del cluster. La cessazione di un'istanza in questo modo comporta il rischio di perdita di dati e di un comportamento imprevedibile del cluster.

Per terminare una specifica istanza è necessario l'ID del gruppo di istanze (restituito dal sottocomando `aws emr describe-cluster --cluster-id`) e l'ID dell'istanza (restituito dal sottocomando `aws emr list-instances --cluster-id`), quindi digitare il seguente comando, sostituire *ig-6RXXXXXX07SA* con l'ID del gruppo di istanze e sostituire *i-f9XXXXf2* con l'ID dell'istanza.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-6RXXXXXX07SA,EC2InstanceIdsToTerminate=i-f9XXXXf2
```

Per ulteriori informazioni sull'utilizzo dei comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Ridimensionamento di un cluster aggiungendo gruppi di istanze attività con la AWS CLI

Con la AWS CLI è possibile aggiungere a un cluster da 1 a 48 gruppi di istanze attività con il sottocomando `--add-instance-groups`. I gruppi di istanze attività possono essere aggiunti solo a un cluster contenente un gruppo di istanze primarie e un gruppo di istanze core. Quando si utilizza la AWS CLI, è possibile aggiungere fino a cinque gruppi di istanze di attività ogni volta che si utilizza il sottocomando `--add-instance-groups`.

1. Per aggiungere un singolo gruppo di istanze di attività a un cluster, digitare il seguente comando e sostituire *j-JXBXXXXXX37R* con l'ID del cluster.

   ```
   aws emr add-instance-groups --cluster-id j-JXBXXXXXX37R --instance-groups
    InstanceCount=6,InstanceGroupType=task,InstanceType=m5.xlarge
   ```

2. Per aggiungere più gruppi di istanze di attività a un cluster, digitare il seguente comando e sostituire *j-JXBXXXXXX37R* con l'ID del cluster. È possibile aggiungere fino a cinque gruppi di istanze di attività in un unico 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
   ```

   Per ulteriori informazioni sull'utilizzo dei comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Interruzione di un ridimensionamento

Utilizzando Amazon EMR versione 4.1.0 o successive, è possibile emettere un ridimensionamento nel corso di un'operazione di ridimensionamento. Inoltre, è possibile interrompere una richiesta di ridimensionamento precedentemente inviata o inviare una nuova richiesta per sovrascrivere una richiesta precedente senza attendere che termini. È possibile arrestare un ridimensionamento esistente anche dalla console o con la chiamata API `ModifyInstanceGroups` con il conteggio corrente come conteggio target del cluster.

Il seguente screenshot mostra un gruppo di istanze di attività di ridimensionamento, ma possono essere arrestate scegliendo Stop (Interrompi).

Interruzione del ridimensionamento con la AWS CLI

È possibile utilizzare AWS CLI per arrestare un ridimensionamento con il sottocomando `modify-instance-groups`. Supponiamo di avere sei istanze nel proprio gruppo di istanze e di volerle portare a 10. In seguito si decide di annullare la richiesta:

- La richiesta iniziale:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=10
```

  La seconda richiesta di bloccare la prima richiesta:

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-myInstanceGroupId,InstanceCount=6
```

> **Note**
>
> Poiché questo processo è asincrono, è possibile vedere i conteggi delle istanze cambiare rispetto alle richieste API precedenti prima che le richieste successive vengano onorate. In caso di riduzione, è possibile che, se si ha un lavoro in corso sui nodi, il gruppo di istanze non si riduca fino a quando i nodi non hanno completato il loro lavoro.

Stato sospeso

Un gruppo di istanze entra in uno stato di sospensione se incontra troppi errori durante il tentativo di avviare i nuovi nodi del cluster. Ad esempio, se i nuovi nodi hanno esito negativo durante l'esecuzione delle operazioni di bootstrap, il gruppo di istanze passa allo stato SUSPENDED (SOSPESO) piuttosto che al provisioning continuo dei nuovi nodi. Dopo aver risolto il problema sottostante, reimpostare il numero desiderato di nodi sul gruppo di istanze del cluster, quindi il gruppo di istanze riprende l'allocazione dei nodi. La modifica di un gruppo di istanze istruisce Amazon EMR a tentare di fornire nuovamente i nodi. Nessun nodo in esecuzione viene riavviato o terminato.

In AWS CLI, il sottocomando `list-instances` restituisce tutte le istanze e i relativi stati come il sottocomando `describe-cluster`. Se Amazon EMR rileva un guasto in un gruppo di istanze, modifica lo stato del gruppo su `SUSPENDED`.

Ripristino di un cluster in stato SUSPENDED con la AWS CLI

Digitare il sottocomando `describe-cluster` con il parametro `--cluster-id` per visualizzare lo stato delle istanze nel cluster.

- Per visualizzare informazioni su tutte le istanze e i gruppi di istanza in un cluster, digitare il comando seguente e sostituire *`j-3KVXXXXXXY7UG`* con l'ID del cluster.

  ```
  aws emr describe-cluster --cluster-id j-3KVXXXXXXY7UG
  ```

  L'output mostra le informazioni sui gruppi di istanze e lo stato delle istanze:

  ```
  {
      "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
            }
```

```
...
}
```

Per visualizzare le informazioni su un particolare gruppo di istanze, digitare il sottocomando `list-instances` con i parametri `--cluster-id` e `--instance-group-types`. È possibile visualizzare informazioni per i gruppi primari, core o attività.

```
aws emr list-instances --cluster-id j-3KVXXXXXXY7UG --instance-group-types "CORE"
```

Usare il sottocomando `modify-instance-groups` con il parametro `--instance-groups` per resettare un cluster nello stato SUSPENDED. Il sottocomando `describe-cluster` restituisce l'id del gruppo di istanze.

```
aws emr modify-instance-groups --instance-groups
 InstanceGroupId=ig-3SUXXXXXXQ9ZM,InstanceCount=3
```

Considerazioni sulla riduzione delle dimensioni del cluster

Se scegli di ridurre le dimensioni di un cluster in esecuzione, considera il comportamento di Amazon EMR e le best practice seguenti:

- Per ridurre l'impatto sui processi in corso, Amazon EMR seleziona in modo intelligente le istanze da rimuovere. Per maggiori informazioni sul comportamento di riduzione del cluster, consulta Terminazione al completamento dell'attività nella Guida alla gestione di Amazon EMR.
- Quando riduci le dimensioni di un cluster, Amazon EMR copia i dati dalle istanze rimosse alle istanze rimanenti. Verifica che vi sia una capacità di archiviazione sufficiente per questi dati nelle istanze che rimangono nel gruppo.
- Amazon EMR tenta di rimuovere le autorizzazioni HDFS sulle istanze del gruppo. Prima di ridurre le dimensioni di un cluster, ti consigliamo di ridurre al minimo le operazioni I/O di scrittura HDFS.
- Per un controllo più granulare quando si riducono le dimensioni di un cluster, puoi visualizzare il cluster nella console e accedere alla scheda Instances (Istanze). Seleziona l'ID per il gruppo di istanze che desideri ridimensionare. Quindi utilizza l'opzione Terminate (Termina) per le istanze specifiche che desideri rimuovere.

Configurazione dei timeout per il provisioning della capacità

Quando si utilizzano parchi istanze, è possibile configurare i timeout di provisioning. Un timeout di provisioning indica ad Amazon EMR di interrompere il provisioning della capacità dell'istanza se il cluster supera una soglia temporale specifica durante l'avvio o le operazioni di dimensionamento del cluster . Gli argomenti seguenti illustrano come configurare un timeout di provisioning per l'avvio del cluster e per le operazioni di dimensionamento del cluster.

Argomenti

Configurare i timeout di provisioning per l'avvio del cluster in Amazon EMR

Puoi definire un periodo di timeout per il provisioning delle istanze Spot per ogni parco istanze del tuo cluster. Se Amazon EMR non è in grado di eseguire il provisioning della capacità Spot, puoi scegliere di terminare il cluster o di effettuare il provisioning della capacità on demand. Se il periodo di timeout termina durante il processo di ridimensionamento del cluster, Amazon EMR annulla le richieste Spot non sottoposte a provisioning. Le istanze Spot non sottoposte a provisioning non vengono trasferite alla capacità on demand.

> **Note**
>
> Non è possibile personalizzare il periodo di timeout di provisioning nella vecchia console. Per scoprire le differenze tra la vecchia e la nuova esperienza nella console, consulta la sezione Novità della console.

Esegui i seguenti passaggi per personalizzare un periodo di timeout di provisioning per l'avvio del cluster con la console Amazon EMR.

New console

Configurazione del timeout durante la creazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.

2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. Nella pagina Crea cluster, vai alla Configurazione del cluster e seleziona Parchi istanze.
4. In Opzione di dimensionamento e provisioning del cluster, specifica la dimensione Spot per i parchi istanze principali e di attività.
5. In Configurazione del timeout Spot, seleziona Termina il cluster dopo il timeout Spot o Passa a on demand dopo il timeout Spot. Quindi, specifica il periodo di timeout per il provisioning delle istanze Spot. Il valore predefinito è 1 ora.
6. Scegli qualsiasi altra opzione applicabile al tuo cluster.
7. Per avviare il tuo cluster con il timeout configurato, scegli Crea cluster.

AWS CLI

Per specificare un timeout di provisioning con il 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"}]'
```

Personalizzare un periodo di timeout di provisioning per il ridimensionamento del cluster in Amazon EMR

È possibile definire un periodo di timeout per il provisioning delle istanze spot per ogni parco istanze nel tuo cluster. Se Amazon EMR non è in grado di fornire la capacità Spot, annulla la richiesta di ridimensionamento e interrompe i tentativi di fornire capacità Spot aggiuntiva. Quando crei un cluster, puoi configurare il timeout. Per un cluster in esecuzione, puoi aggiungere o aggiornare un timeout.

Alla scadenza del periodo di timeout, Amazon EMR invia automaticamente gli eventi a uno stream Eventi Amazon CloudWatch. Con CloudWatch, è possibile creare regole che corrispondono agli eventi in base a un modello specificato e quindi instradare gli eventi verso gli obiettivi per intraprendere azioni. Ad esempio, è possibile configurare una regola per inviare una notifica via email. Per ulteriori informazioni su come creare regole, consulta Creazione di regole per gli eventi Amazon EMR con CloudWatch. Per ulteriori informazioni sui dati dei diversi eventi, consulta Eventi di modifica dello stato del parco istanze.

Esempi di timeout di provisioning per il ridimensionamento del cluster

Per specificare un timeout di provisioning per il ridimensionamento con la AWS CLI

L'esempio seguente utilizza il comando `create-cluster` per aggiungere un timeout di provisioning per il ridimensionamento.

```
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"}]'
```

L'esempio seguente utilizza il comando `modify-instance-fleet` per aggiungere un timeout di provisioning per il ridimensionamento.

```
aws emr modify-instance-fleet \
--cluster-id j-XXXXXXXXXXXX \
--instance-fleet '{"InstanceFleetId":"if-XXXXXXXXXXX","ResizeSpecifications":
{"SpotResizeSpecification":{"TimeoutDurationMinutes":30},"OnDemandResizeSpecification":
{"TimeoutDurationMinutes":60}}}' \
--region us-east-1
```

L'esempio seguente utilizza il `add-instance-fleet-command` per aggiungere un timeout di provisioning per il ridimensionamento.

```
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
```

Specificare un timeout di provisioning per il ridimensionamento e l'avvio con la AWS CLI

L'esempio seguente utilizza il comando `create-cluster` per aggiungere un timeout di provisioning per il ridimensionamento e l'avvio.

```
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"}]'
```

Considerazioni sui timeout del provisioning del ridimensionamento

Quando configuri i timeout di provisioning del cluster per i tuoi parchi istanze, considera i seguenti comportamenti.

- È possibile configurare i timeout di provisioning sia per le istanze spot che on demand. Il timeout minimo di provisioning è di 5 minuti. Il timeout massimo di provisioning è di 7 giorni.
- È possibile configurare solo i timeout di provisioning per un cluster EMR che utilizza parchi istanze. È necessario configurare separatamente ciascun parco istanze principale e dell'attività.
- Quando crei un cluster, puoi configurare i timeout di provisioning. È possibile aggiungere un timeout o aggiornare un timeout esistente per un cluster in esecuzione.
- Se invii più operazioni di ridimensionamento, Amazon EMR tiene traccia dei timeout di provisioning per ogni operazione di ridimensionamento. Ad esempio, imposta il timeout di provisioning su un cluster su *60* minuti. Quindi, invia un'operazione di ridimensionamento *R1* all'orario *T1*. Invia una seconda operazione di ridimensionamento *R2* all'orario *T2*. Il timeout di provisioning per R1 scade a *T1 + 60 minuti*. Il timeout di provisioning per R2 scade a *T2 + 60 minuti*.
- Se invii una nuova operazione di ridimensionamento scalabile prima della scadenza del timeout, Amazon EMR continua a tentare di fornire capacità per il tuo cluster EMR.

Ridimensionamento del cluster

> **Note**
>
> A partire da Amazon EMR rilascio 5.10.0, le opzioni relative al comportamento di riduzione non sono più supportate. A causa dell'introduzione della fatturazione al secondo in Amazon

EC2, adesso il comportamento di ridimensionamento predefinito per i cluster Amazon EMR è terminare al completamento dell'attività.

Con i rilasci da 5.1.0 a 5.9.1 di Amazon EMR sono disponibili due opzioni per il comportamento di riduzione: terminazione allo scadere del limite di ore-istanze per la fatturazione Amazon EC2 o terminazione al completamento dell'attività. A partire da Amazon EMR rilascio 5.10.0, l'impostazione per la terminazione allo scadere del limite di ore-istanze è obsoleta a causa dell'introduzione della fatturazione al secondo in Amazon EC2. Si sconsiglia di specificare la terminazione allo scadere dell'ora dell'istanza nelle versioni in cui l'opzione è disponibile.

Warning

Se si utilizza AWS CLI per emettere un `modify-instance-groups` con `EC2InstanceIdsToTerminate`, queste istanze vengono terminate immediatamente, senza considerare queste impostazioni e indipendentemente dallo stato delle applicazioni in esecuzione su di esse. La cessazione di un'istanza in questo modo comporta il rischio di perdita di dati e di un comportamento imprevedibile del cluster.

Quando la terminazione al completamento dell'attività è specificata, Amazon EMR elenca come negati e scarica i processi dai nodi prima di terminare le istanze Amazon EC2. Con entrambi i comportamenti specificati, Amazon EMR non termina le istanze Amazon EC2 nei gruppi di istanze principali se questo potrebbe portare alla corruzione di HDFS.

Terminazione al completamento dell'attività

Amazon EMR consente di ridimensionare il cluster senza influenzare il carico di lavoro. Amazon EMR normalmente disattiva YARN, HDFS e altri daemon sui nodi principali e attività durante un'operazione di ridimensionamento senza perdere dati o interrompere i processi. Amazon EMR riduce le dimensioni dei gruppi di istanze solo se il lavoro assegnato ai gruppi è stato completato e questi sono inattivi. Per la disattivazione con tolleranza di YARN NodeManager, è possibile regolare manualmente il tempo di attesa per la disattivazione di un nodo.

In questo momento è impostato utilizzando una proprietà nella classificazione di configurazione `YARN-site`. Se utilizzi il rilascio 5.12.0 e successivi di Amazon EMR, specifica la proprietà `YARN.resourcemanager.nodemanager-graceful-decommission-`

`timeout-secs`. Se utilizzi rilasci precedenti di Amazon EMR, specifica la proprietà `YARN.resourcemanager.decommissioning.timeout`.

Se ci sono ancora contenitori o applicazioni YARN in esecuzione al superamento del tempo di disattivazione, il nodo è costretto a essere disattivato e YARN ricondiziona i contenitori interessati su altri nodi. Il valore predefinito è 3600 secondi (1 ora). Puoi impostare questo timeout con un valore arbitrariamente alto per forzare la riduzione graduale ad attendere più a lungo. Per ulteriori informazioni, consulta la sezione Graceful Decommission of YARN nodes (Disattivazione normale dei nodi YARN) nella documentazione di Apache Hadoop.

Gruppi di nodi attività

Amazon EMR seleziona in modo intelligente le istanze che non presentano attività in esecuzione su qualsiasi fase o applicazione e rimuove prima tali istanze da un cluster. Se tutte le istanze del cluster sono in uso, Amazon EMR attende il completamento delle attività su un'istanza prima di rimuoverla dal cluster. Il tempo di attesa predefinito è 1 ora. Questo valore può essere modificato con l'impostazione `YARN.resourcemanager.decommissioning.timeout`. Amazon EMR utilizza dinamicamente la nuova impostazione. Puoi impostarlo su un numero arbitrariamente elevato per assicurarti che Amazon EMR non termini alcuna attività riducendo al contempo le dimensioni del cluster.

Gruppi di nodi principali

Sui nodi principali, i daemons YARN NodeManager e HDFS DataNode devono essere rimossi affinché il gruppo di istanze possa essere ridotto. Per quanto riguarda YARN, la riduzione graduale garantisce che un nodo contrassegnato per la disattivazione passi allo stato `DECOMMISSIONED` solo se non vi sono contenitori o applicazioni in attesa o incompleti. La disattivazione termina immediatamente se non vi sono contenitori in funzione sul nodo all'inizio della disattivazione.

Per gli HDFS, le riduzioni graduali assicurano che la capacità nominale degli HDFS sia sufficientemente grande da adattarsi a tutti i blocchi esistenti. Se la capacità target non è sufficientemente grande, solo un numero parziale di istanze principali viene disattivato in modo che i nodi rimanenti possano gestire i dati correnti che risiedono negli HDFS. È necessario garantire una capacità HDFS aggiuntiva per consentire un'ulteriore disattivazione. Dovresti anche cercare di ridurre al minimo le operazioni I/O di scrittura prima di tentare di ridurre i gruppi di istanze. Un numero eccessivo di I/O di scrittura potrebbe ritardare il completamento dell'operazione di ridimensionamento.

Un altro limite è il fattore di replica predefinito, `dfs.replication` all'interno di `/etc/hadoop/conf/hdfs-site`. Durante la creazione di un cluster, Amazon EMR configura il valore in base al

numero di istanze del cluster: 1 con 1-3 istanze, 2 per i cluster con 4-9 istanze e 3 per i cluster con oltre 10 istanze.

Warning

1. L'impostazione di `dfs.replication` su 1 per i cluster con meno di quattro nodi può causare la perdita di dati HDFS in caso di disattivazione anche di un singolo nodo. Ti consigliamo di utilizzare un cluster con almeno quattro nodi principali per i carichi di lavoro di produzione.
2. Amazon EMR non consente ai cluster di dimensionare i nodi principali al di sotto di `dfs.replication`. Ad esempio, se `dfs.replication = 2`, il numero minimo di nodi principali è 2.
3. Quando utilizzi il dimensionamento gestito, il dimensionamento automatico o scegli di dimensionare manualmente il cluster, ti consigliamo di impostare `dfs.replication` su 2 o su un valore superiore.

Una riduzione graduale non consente di ridurre i nodi principali al di sotto del fattore di replica HDFS. Questo per consentire a HDFS di chiudere i file a causa di repliche insufficienti. Per aggirare questo limite, riduci il fattore di replica e riavvia il daemon NameNode.

Configura il comportamento di dimensionamento verso il basso per Amazon EMR.

Note

L'opzione di comportamento di riduzione con terminazione allo scadere del limite di ore-istanze non è più supportata per Amazon EMR rilascio 5.10.0 e successivi. Le seguenti opzioni relative alla riduzione vengono visualizzate nella console Amazon EMR solo per i rilasci da 5.1.0 a 5.9.1.

È possibile utilizzare la AWS Management Console, la AWS CLI o l'API Amazon EMR per configurare il comportamento di dimensionamento verso il basso al momento della creazione di un cluster.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. Per scoprire le differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Configurazione del comportamento di riduzione con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR su EC2, nel riquadro di navigazione a sinistra, scegli Cluster e quindi seleziona Crea cluster.
3. Nella sezione Cluster scaling and provisioning option (Opzione di dimensionamento e provisioning del cluster), cerca Cluster termination (Terminazione del cluster) e scegli di terminare manualmente il cluster o chiedi ad Amazon EMR di terminare il cluster dopo un determinato periodo di inattività. Facoltativamente, attiva la protezione da terminazione da bug o errori.
4. Scegli qualsiasi altra opzione applicabile al cluster.
5. Per avviare il cluster, scegli Create cluster (Crea cluster).

Old console

Configurazione del comportamento di riduzione con la vecchia console

1. Apri la console di Amazon EMR all'indirizzo https://console.aws.amazon.com/elasticmapreduce.
2. Scegli Crea cluster. Vai a Advanced options (Opzioni avanzate) e scegli le impostazioni di configurazione in Step 1: Software and Steps (Passaggio 1: Software e fasi) e Step 2: Hardware (Passaggio 2: Hardware).
3. Nel Step 3: General Cluster Settings (Passaggio 3: Impostazioni generali del cluster), seleziona il comportamento di riduzione preferito. Completa le configurazioni rimanenti e crea il cluster.

AWS CLI

Configurazione del comportamento di riduzione con la AWS CLI

- Per l'opzione `--scale-down-behavior` puoi specificare `TERMINATE_AT_INSTANCE_HOUR` o `TERMINATE_AT_TASK_COMPLETION`.

Terminazione di un cluster

Questa sezione descrive i metodi per terminare un cluster. Per informazioni sull'abilitazione della protezione da cessazione e sulla cessazione automatica dei cluster, consulta Controllo della terminazione di un cluster. Puoi terminare i cluster il cui stato è `STARTING`, `RUNNING` o `WAITING`. Un cluster il cui stato è `WAITING` deve essere terminato altrimenti viene eseguito all'infinito generando spese sul tuo account. Puoi terminare un cluster che non riesce a passare dallo stato `STARTING` a un altro stato o che non è in grado di completare una fase.

Se desideri terminare un cluster per il quale la protezione da cessazione è attivata, devi disattivare tale protezione per poter terminare il cluster. I cluster possono essere terminati tramite la console, AWS CLI o, a livello di codice, mediante l'API `TerminateJobFlows`.

A seconda della configurazione del cluster, l'operazione di terminazione del cluster e di rilascio delle risorse allocate, come le istanze EC2, può durare da 5 a 20 minuti.

> **Note**
>
> Non è possibile riavviare un cluster terminato, ma è possibile clonare un cluster terminato per riutilizzarne la configurazione per un nuovo cluster. Per ulteriori informazioni, consulta Clonazione di un cluster mediante la console.

> **Important**
>
> Amazon EMR utilizza il ruolo di servizio Amazon EMR e il ruolo `AWSServiceRoleForEMRCleanup` per pulire le risorse del cluster del tuo account che non usi più, come le istanze Amazon EC2. È necessario includere azioni relative alle policy del ruolo per eliminare o terminare le risorse. Altrimenti, Amazon EMR non può eseguire queste azioni di pulizia e potresti incorrere in costi per le risorse inutilizzate che rimangono nel cluster.

Terminazione di un cluster con la console

Puoi terminare uno o più cluster utilizzando la console di Amazon EMR. La procedura per terminare un cluster nella console varia a seconda dell'attivazione o meno della protezione da cessazione. Per terminare un cluster protetto, devi dapprima disattivare la protezione da cessazione.

New console

Terminazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. Seleziona Cluster, quindi scegli il cluster da terminare.
3. Nel menu a discesa Actions (Azioni), scegli Terminate cluster (Termina cluster) per aprire il prompt Terminate cluster (Termina cluster).
4. Al prompt, scegli Terminate (Termina). A seconda della configurazione del cluster, la terminazione potrebbe richiedere da 5 a 10 minuti. Per ulteriori informazioni sui cluster Amazon EMR, consulta la sezione Terminazione di un cluster.

Old console

Terminazione di un cluster con la protezione da terminazione disattivata con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Selezionare il cluster da terminare. È possibile selezionare più cluster e terminarli contemporaneamente.
3. Scegliere Terminate (Termina).
4. Quando richiesto, scegliere Terminate (Termina).

   Amazon EMR termina le istanze nel cluster e disattiva il salvataggio dei dati di log.

Terminazione di un cluster con la protezione da terminazione attivata con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Nella pagina Cluster List (Elenco cluster), selezionare il cluster da terminare. È possibile selezionare più cluster e terminarli contemporaneamente.
3. Scegliere Terminate (Termina).
4. Quando richiesto, scegliere Change (Cambia) per disattivare la protezione da cessazione. Se sono stati selezionati più cluster, scegliere Turn off all (Disattiva tutto) per disattivare la protezione da cessazione per tutti i cluster contemporaneamente.
5. Nella finestra di dialogo Terminate clusters (Termina cluster) per Termination Protection (Protezione da cessazione), scegliere Off (Disattivato) e quindi fare clic sul segno di spunta per confermare.
6. Fare clic su Terminate (Termina).

   Amazon EMR termina le istanze nel cluster e disattiva il salvataggio dei dati di log.

Terminazione di un cluster mediante la AWS CLI

Per terminare un cluster non protetto utilizzando l'AWS CLI

Per terminare un cluster non protetto mediante AWS CLI, utilizzare il sottocomando `terminate-clusters` con il parametro --cluster-ids.

- Digitare il comando seguente per terminare un singolo cluster e sostituire *j-3KVXXXXXXX7UG* con l'ID del cluster.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
  ```

  Per terminare più cluster, digitare il comando seguente e sostituire *j-3KVXXXXXXX7UG* e *j-WJ2XXXXXX8EU* con gli ID dei cluster.

  ```
  aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
  ```

  Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Per terminare un cluster protetto mediante l'AWS CLI

Per terminare un cluster protetto mediante AWS CLI, disattivare dapprima la protezione da cessazione utilizzando il sottocomando `modify-cluster-attributes` con il parametro `--no-termination-protected`. Utilizzare quindi il sottocomando `terminate-clusters` con il parametro `--cluster-ids` per terminare il cluster.

1. Digitare il comando seguente per disattivare la protezione da cessazione e sostituire *`j-3KVTXXXXXX7UG`* con l'ID del cluster.

   ```
   aws emr modify-cluster-attributes --cluster-id j-3KVTXXXXXX7UG --no-termination-protected
   ```

2. Per terminare il cluster, digitare il comando seguente e sostituire *`j-3KVXXXXXXX7UG`* con l'ID del cluster.

   ```
   aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG
   ```

   Per terminare più cluster, digitare il comando seguente e sostituire *`j-3KVXXXXXXX7UG`* e *`j-WJ2XXXXXX8EU`* con gli ID dei cluster.

   ```
   aws emr terminate-clusters --cluster-ids j-3KVXXXXXXX7UG j-WJ2XXXXXX8EU
   ```

   Per ulteriori informazioni sull'utilizzo di comandi Amazon EMR in AWS CLI, consulta https://docs.aws.amazon.com/cli/latest/reference/emr.

Terminazione di un cluster mediante l'API

L'operazione `TerminateJobFlows` arresta l'elaborazione della fase, carica tutti i dati di log da Amazon EC2 in Amazon S3 (se configurato) e termina il cluster Hadoop. Un cluster viene terminato automaticamente anche se `KeepJobAliveWhenNoSteps` è impostato su `False` in una richiesta `RunJobFlows`.

Puoi utilizzare questa operazione per terminare un singolo cluster o un elenco di cluster in base ai relativi ID.

Per ulteriori informazioni sui parametri di input univoci per `TerminateJobFlows`, consulta TerminateJobFlows. Per ulteriori informazioni sui parametri generici nella richiesta, consulta la sezione relativa ai Parametri di richiesta comuni.

Clonazione di un cluster mediante la console

La console di Amazon EMR consente di clonare un cluster, ossia di creare una copia della configurazione del cluster originale da utilizzare come base per un nuovo cluster.

> **Note**
>
> Abbiamo riprogettato la console Amazon EMR per facilitarne l'utilizzo. È possibile clonare cluster che utilizzano il dimensionamento automatico nella nuova console, ma è possibile creare nuovi cluster solo se si desidera utilizzare il dimensionamento manuale o gestito. Per maggiori informazioni sulle differenze tra la vecchia e la nuova esperienza sulla console, consulta la sezione Novità della console.

New console

Clonazione di un cluster con la nuova console

1. Accedi alla AWS Management Console e apri la console Amazon EMR all'indirizzo https://console.aws.amazon.com/emr.
2. In EMR on EC2 (EMR su EC2), nel riquadro di navigazione a sinistra, scegli Clusters (Cluster).
3. Clonazione di un cluster dall'elenco dei cluster
   a. Utilizza le opzioni di ricerca e filtro per trovare il cluster che desideri clonare nella visualizzazione a elenco.
   b. Seleziona la casella di controllo a sinistra della riga relativa al cluster da clonare.
   c. L'opzione Clone (Clona) sarà ora disponibile nella parte superiore della visualizzazione a elenco. Seleziona Clone (Clona) per avviare il processo di clonazione. Se il cluster ha delle fasi configurate, scegli Include steps (Includi fasi) e Continue (Continua) se desideri clonare le fasi insieme alle altre configurazioni del cluster.
   d. Esamina le impostazioni del nuovo cluster, che sono state copiate dal cluster clonato. Se necessario, modifica le impostazioni. Quando reputi la configurazione del nuovo cluster soddisfacente, seleziona Create cluster (Crea cluster) per avviare il nuovo cluster.
4. Clonazione di un cluster da una pagina di dettaglio del cluster

a. Per accedere alla pagina dei dettagli del cluster che desideri clonare, seleziona il relativo Cluster ID (ID cluster dalla visualizzazione a elenco dei cluster.

b. Nella parte superiore della pagina dei dettagli del cluster, seleziona Clone cluster (Clona cluster) dal menu Actions (Operazioni) per avviare il processo di clonazione. Se il cluster ha delle fasi configurate, scegli Include steps (Includi fasi) e Continue (Continua) se desideri clonare le fasi insieme alle altre configurazioni del cluster.

c. Esamina le impostazioni del nuovo cluster, che sono state copiate dal cluster clonato. Se necessario, modifica le impostazioni. Quando reputi la configurazione del nuovo cluster soddisfacente, seleziona Create cluster (Crea cluster) per avviare il nuovo cluster.

Old console

Clonazione di un cluster con la vecchia console

1. Passa alla nuova console Amazon EMR e seleziona Passa alla vecchia console dalla barra di navigazione laterale. Per ulteriori informazioni su cosa aspettarti quando passi alla vecchia console, consulta Utilizzo della vecchia console.
2. Scegli Crea cluster.
3. Nella pagina Cluster List (Elenco cluster), fare clic su un cluster da clonare.
4. Nella parte superiore della pagina Cluster Details (Dettagli del cluster), fare clic su Clone (Clona).

   Nella finestra di dialogo, scegliere Yes (Sì) per includere le fasi del cluster originale nel cluster clonato. Scegliere No per clonare la configurazione del cluster originale senza includere le fasi.

> **Note**
>
> Se si clona un cluster creato utilizzando AMI 3.1.1 e versioni successive (Hadoop 2.x) o AMI 2.4.8 e versioni successive (Hadoop 1.x) e si includono le fasi, tutte le fasi di sistema (come la configurazione di Hive) sono clonate insieme alle fasi inviate dall'utente, fino a un totale di 1.000 fasi. Qualsiasi fase meno recente non più visualizzata nella cronologia delle fasi della console non può essere clonata. Per le AMI precedenti, è possibile clonare solo 256 fasi (incluse quelle di sistema). Per ulteriori informazioni, consulta Invio di lavoro a un cluster.

5. Viene visualizzata la pagina Create Cluster (Crea cluster) con una copia della configurazione del cluster originale. Esaminare la configurazione, apportare le modifiche necessarie, quindi fare clic su Create Cluster (Crea cluster).

Automatizzazione di cluster ricorrenti con AWS Data Pipeline

AWS Data Pipeline è un servizio che consente di automatizzare il trasferimento e la trasformazione di dati. Puoi utilizzare questo metodo per programmare il trasferimento di dati di input a Amazon S3 e l'avvio di cluster per elaborare quei dati. Ad esempio, immaginiamo che disponi di un server Web che registra log di traffico. Se si desidera eseguire un cluster settimanale per analizzare i dati di traffico, è possibile utilizzare AWS Data Pipeline per programmare tali cluster. AWS Data Pipeline è un flusso di lavoro basato sui dati, in modo che un'attività (avviando il cluster) possa essere dipendente da un'altra attività (trasferendo i dati di input ad Amazon S3). Dispone inoltre di una potente caratteristica di ripetizione tentativi.

Per ulteriori informazioni su AWS Data Pipeline, consulta la Guida per gli sviluppatori di AWS Data Pipeline, in particolare i tutorial relativi ad Amazon EMR:

- Tutorial: Avvio di un flusso di lavoro di Amazon EMR
- Nozioni di base sull'elaborazione di log Web con AWS Data Pipeline, Amazon EMR e Hive
- Tutorial: Importazione ed esportazione di dati di Amazon DynamoDB tramite AWS Data Pipeline

Risoluzione dei problemi dei cluster

Un cluster EMR viene eseguito in un ecosistema complesso costituito da software open-source, codice applicazione personalizzato e Servizi AWS. Quando si verifica un problema con una di queste parti, l'esecuzione del cluster potrebbe non riuscire o impiegare più tempo del previsto per il completamento. Gli argomenti seguenti possono aiutarti a identificare e a risolvere i problemi del cluster.

Argomenti

Quando si sviluppa una nuova applicazione Hadoop, si consiglia di abilitare il debug ed elaborare un sottoinsieme piccolo ma rappresentativo dei dati per testare l'applicazione. È inoltre possibile eseguire l'applicazione passo dopo passo per testare ogni passaggio separatamente. Per ulteriori informazioni, consulta Configurazione della registrazione e del debug di cluster e Fase 5: Test dettagliato del cluster.

Quali strumenti sono disponibili per la risoluzione dei problemi?

Per identificare e correggere gli errori del cluster, puoi utilizzare gli strumenti descritti in questa pagina. Potrebbe essere necessario inizializzare alcuni strumenti all'avvio del cluster. Per impostazione predefinita, sono disponibili altri strumenti per ogni cluster.

Argomenti

- Monitoraggio delle prestazioni del cluster EMR

Visualizzazione dei dettagli del cluster EMR

È possibile utilizzare AWS Management Console, AWS CLI o l'API EMR per recuperare informazioni dettagliate su un cluster EMR e sull'esecuzione di un lavoro. Per ulteriori informazioni sull'utilizzo di AWS Management Console e AWS CLI, consulta Visualizzazione dello stato e dei dettagli di un cluster.

Riquadro dei dettagli della console di Amazon EMR

Nell'elenco Cluster sulla console di Amazon EMR puoi visualizzare le informazioni di alto livello sullo stato di ogni cluster nel tuo account e Regione AWS. L'elenco mostra tutti i cluster attivi e terminati che hai avviato negli ultimi due mesi. Dall'elenco Clusters (Cluster) è possibile selezionare un Name (Nome) di cluster per visualizzare i dettagli del cluster. Queste informazioni sono organizzate in diverse categorie per facilitarne la navigazione.

La funzionalità Interfacce utente dell'applicazione, disponibile nella pagina dei dettagli del cluster, può essere utile per risolvere i problemi dei cluster. Fornisce lo stato delle applicazioni YARN e, per alcune di esse, come ad esempio le applicazioni Spark, puoi esplorare diversi parametri e sfaccettature, come ad esempio processi, fasi ed esecutori. Per ulteriori informazioni, consulta Visualizzazione della cronologia dell'applicazione. Questa funzionalità è disponibile solo per Amazon EMR versione 5.8.0 e successive.

Interfaccia a riga di comando di Amazon EMR

Puoi individuare i dettagli su un cluster dalla AWS CLI utilizzando l'argomento `--describe` .

API Amazon EMR

È possibile individuare i dettagli su un cluster dall'API utilizzando l'azione `DescribeJobFlows`.

Visualizzazione dei dettagli degli errori del cluster EMR

Quando un cluster EMR termina con un errore, le API `DescribeCluster` e `ListClusters` restituiscono un codice di errore e un messaggio di errore. Per alcuni errori del cluster, l'array di dati `ErrorDetail` può aiutarti a risolvere l'errore.

Per un elenco dei codici di errore che includono dati `ErrorDetail`, consulta Codici di errore con informazioni ErrorDetail.

> **Note**
>
> Per assicurarci di fornirti le informazioni più recenti e pertinenti, miglioriamo continuamente i nostri messaggi di errore. Non è consigliabile analizzare il testo di `ErrorMessage` perché è soggetto a modifiche.

Esecuzione di script e configurazione di processi Amazon EMR

Come parte del processo di risoluzione dei problemi, potrebbe essere utile eseguire script personalizzati sul cluster o visualizzare e configurare i processi del cluster.

Visualizzare e riavviare i processi di applicazione

Può essere utile visualizzare i processi in esecuzione sul cluster per diagnosticare potenziali problemi. È possibile arrestare e riavviare i processi del cluster effettuando la connessione al nodo principale del cluster. Per ulteriori informazioni, consulta Visualizzazione e riavvio di Amazon EMR e dei processi applicativi (daemon).

Esegui comandi e script senza connessione SSH

Per eseguire un comando o uno script sul cluster come fase, è possibile utilizzare gli strumenti `command-runner.jar` o `script-runner.jar` senza stabilire una connessione SSH al nodo principale. Per ulteriori informazioni, consulta Esegui comandi e script su un cluster Amazon EMR.

Visualizzare file di log di

Amazon EMR e Hadoop generano entrambi file di log quando il cluster è in esecuzione. Puoi accedere a questi file di log da diversi strumenti, in base alla configurazione specificata quando hai avviato il cluster. Per ulteriori informazioni, consulta Configurazione della registrazione e del debug di cluster.

File di log sul nodo master

Ogni cluster pubblica i file di log nella directory /mnt/var/log/ del nodo master. Questi file di log sono disponibili solo quando il cluster è in esecuzione.

File di log archiviati in Amazon S3

Se si avvia il cluster e si specifica un percorso di log Amazon S3, il cluster copia i file di log memorizzati in /mnt/var/log/ sul nodo master in Amazon S3 a intervalli di 5 minuti. Questo assicura l'accesso ai file di log anche dopo la chiusura del cluster. Poiché i file vengono archiviati a intervalli di 5 minuti, gli ultimi minuti di un cluster terminato improvvisamente potrebbero non essere disponibili.

Monitoraggio delle prestazioni del cluster EMR

Amazon EMR fornisce diversi strumenti per controllare le prestazioni del cluster.

Interfacce Web Hadoop

Ogni cluster pubblica una serie di interfacce web sul nodo master che contengono informazioni sul cluster. È possibile accedere a queste pagine Web utilizzando un tunnel SSH per collegarle al nodo master. Per ulteriori informazioni, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

Metriche di CloudWatch

Ogni cluster trasmette dei parametri a CloudWatch. CloudWatch è un servizio Web che monitora i parametri, e che è possibile utilizzare per impostare allarmi in questi parametri. Per ulteriori informazioni, consulta Monitoraggio di parametri di Amazon EMR con CloudWatch.

Visualizzazione e riavvio di Amazon EMR e dei processi applicativi (daemon)

Durante la risoluzione dei problemi relativi a un cluster, è possibile che tu voglia elencare i processi in esecuzione. Puoi anche interrompere o riavviare i processi. Ad esempio, puoi riavviare il processo dopo aver modificato una configurazione o notato un problema con un determinato processo dopo l'analisi di file di log e messaggi di errore.

I tipi di processi che vengono eseguiti su un cluster sono due: i processi Amazon EMR (ad esempio, il controller dell'istanza e Log Pusher) e i processi associati alle applicazioni installate nel cluster (ad esempio, hadoop-hdfs-namenode e hadoop-yarn-resourcemanager).

Per utilizzare i processi direttamente su un cluster, è necessario innanzitutto collegare al nodo principale. Per ulteriori informazioni, consulta Connessione a un cluster.

Visualizzazione dei processi in esecuzione

Il metodo utilizzato per visualizzare i processi in esecuzione su un cluster differisce a seconda della versione di Amazon EMR utilizzata.

EMR 5.30.0 and 6.0.0 and later

Example : elenca tutti i processi in esecuzione

Nell'esempio seguente viene utilizzato `systemctl` e specifica `--type` per visualizzare tutti i processi.

```
systemctl --type=service
```

Example : elenca processi specifici

Nell'esempio seguente sono elencati tutti i processi con nomi che contengono `hadoop`.

```
systemctl --type=service | grep -i hadoop
```

Output di esempio:

```
 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 : consulta un report dettagliato sullo stato per un processo specifico

Nell'esempio seguente viene visualizzato un report dettagliato sullo stato del servizio `hadoop-hdfs-namenode`.

```
sudo systemctl status hadoop-hdfs-namenode
```

Output di esempio:

```
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 : elenca tutti i processi in esecuzione

Nell'esempio seguente sono elencati tutti i processi in esecuzione.

```
initctl list
```

EMR 2.x - 3.x

Example : elenca tutti i processi in esecuzione

Nell'esempio seguente sono elencati tutti i processi in esecuzione.

```
ls /etc/init.d/
```

Arresto e riavvio di processi

Dopo aver determinato quali processi sono in esecuzione, puoi interromperli e riavviarli se necessario.

EMR 5.30.0 and 6.0.0 and later

Example : arresta un processo

L'esempio seguente arresta il processo `hadoop-hdfs-namenode`.

```
sudo systemctl stop hadoop-hdfs-namenode
```

Puoi eseguire una query sul `status` per verificare che il processo sia stato arrestato.

```
sudo systemctl status hadoop-hdfs-namenode
```

Output di esempio:

```
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 : avvia un processo

Nell'esempio seguente viene avviato il processo `hadoop-hdfs-namenode`.

```
sudo systemctl start hadoop-hdfs-namenode
```

È possibile eseguire una query sullo stato per verificare che il processo sia in esecuzione.

```
sudo systemctl status hadoop-hdfs-namenode
```

Output di esempio:

```
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 : arresta un processo in esecuzione

L'esempio seguente arresta il servizio `hadoop-hdfs-namenode`.

```
sudo stop hadoop-hdfs-namenode
```

Example : riavvia un processo arrestato

L'esempio seguente riavvia il servizio `hadoop-hdfs-namenode`. È necessario utilizzare il comando `start` e non `restart`.

```
sudo start hadoop-hdfs-namenode
```

Example : verifica lo stato del processo

Quanto segue recupera lo stato di `hadoop-hdfs-namenode`. Puoi utilizzare il comando `status` per verificare che il processo sia stato arrestato o avviato.

```
sudo status hadoop-hdfs-namenode
```

EMR 2.x - 3.x

Example : arresta un processo di applicazione

L'esempio seguente arresta il servizio `hadoop-hdfs-namenode`, associato alla versione di Amazon EMR installata sul cluster.

```
sudo /etc/init.d/hadoop-hdfs-namenode stop
```

Example : riavvia un processo di applicazione

Il seguente comando di esempio riavvia il processo `hadoop-hdfs-namenode`:

```
sudo /etc/init.d/hadoop-hdfs-namenode start
```

Example : arresta un processo Amazon EMR

L'esempio seguente arresta un processo, come il controller di istanza, che non è associato alla versione di Amazon EMR sul cluster.

```
sudo /sbin/stop instance-controller
```

Example : riavvia un processo Amazon EMR

L'esempio seguente riavvia un processo, come il controller di istanza, che non è associato alla versione di Amazon EMR sul cluster.

```
sudo /sbin/start instance-controller
```

> **Note**
>
> I comandi `/sbin/start, stop` e `restart` sono symlinks a `/sbin/intictl`. Per ulteriori informazioni su `initctl`, consulta la pagina initctl man digitando `man initctl` al prompt dei comandi.

Errori comuni in Amazon EMR

A volte, l'elaborazione dei dati da parte dei cluster non riesce oppure è lenta. Le sezioni seguenti elencano alcuni problemi comuni relativi ai cluster con suggerimenti su come risolverli.

Argomenti

- Ricerca di un cluster mancante

Codici di errore con informazioni ErrorDetail

Quando un cluster EMR termina con un errore, le API `DescribeCluster` e `ListClusters` restituiscono un codice di errore e un messaggio di errore. Per alcuni errori del cluster, l'array di dati `ErrorDetail` può aiutarti a risolvere l'errore.

Gli errori che includono un array `ErrorDetail` forniscono i seguenti dettagli:

`ErrorCode`

Un codice di errore univoco che puoi utilizzare per l'accesso programmatico.

`ErrorData`

Un elenco di identificatori in coppie chiave-valore che puoi utilizzare per la programmazione o la ricerca manuale. Per le descrizioni dei valori `ErrorData` inclusi in un codice di errore, consulta la pagina di risoluzione dei problemi relativa al codice di errore.

`ErrorMessage`

Descrizione dell'errore con un collegamento a ulteriori informazioni nella documentazione di Amazon EMR.

> **Note**
>
> Non è consigliabile analizzare il testo di `ErrorMessage` perché è soggetto a modifiche.

Codici di errore per categoria

- Codici di errore per errori di bootstrap
- Codici di errori interni
- Codici di errore relativi agli errori di convalida

Codici di errore per errori di bootstrap

Le seguenti sezioni forniscono informazioni sulla risoluzione dei problemi relativi ai codici di errore per gli errori di bootstrap.

Argomenti

BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE

Panoramica

Quando un cluster termina con un errore `BOOTSTRAP_FAILURE_PRIMARY_WITH_NON_ZERO_CODE`, un'azione di bootstrap non è riuscita nell'istanza primaria. Per ulteriori informazioni sulle operazioni di bootstrap, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo.

Risoluzione

Per risolvere questo errore, esamina i dettagli restituiti nell'errore dell'API, modifica lo script dell'operazione di bootstrap e crea un nuovo cluster con l'operazione di bootstrap aggiornata.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`primary-instance-id`

L'ID dell'istanza primaria in cui l'azione di bootstrap non è riuscita.

`bootstrap-action`

Il numero ordinale dell'operazione di bootstrap che non è riuscita. Uno script con un valore `bootstrap-action` di `1` è la prima azione di bootstrap da eseguire sull'istanza.

`return-code`

Il codice restituito per l'operazione di bootstrap non riuscita.

`amazon-s3-path`

La posizione Amazon S3 dell'azione di bootstrap non riuscita.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere la causa principale dell'errore dell'operazione di bootstrap. Quindi avvia un nuovo cluster.

1. Esamina i file di log dell'operazione di bootstrap in Amazon S3 per identificare la causa principale dell'errore. Per ulteriori informazioni su come visualizzare i log di Amazon EMR, consulta Visualizzare file di log di .
2. Se hai attivato i log del cluster quando hai creato l'istanza, consulta il log `stdout` per ulteriori informazioni. Puoi trovare il log `stdout` per l'operazione di bootstrap in questa posizione Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

   Per ulteriori informazioni sui log del cluster, consulta Configurazione della registrazione e del debug di cluster.
3. Per determinare l'errore dell'operazione di bootstrap, esamina le eccezioni nei log `stdout` e il valore in `return-code` in `ErrorData`.
4. Usa i risultati del passaggio precedente per rivedere l'operazione di bootstrap in modo da evitare le eccezioni o gestirle correttamente quando si verificano.
5. Avvia un nuovo cluster con l'operazione di bootstrap aggiornata.

BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY

Panoramica

Un cluster termina con l'errore `BOOTSTRAP_FAILURE_BA_DOWNLOAD_FAILED_PRIMARY` quando l'istanza primaria non è in grado di scaricare lo script di un'operazione di bootstrap dalla posizione Amazon S3 specificata. Le cause potenziali includono:

- Il file script dell'operazione di bootstrap non si trova nella posizione Amazon S3 specificata.
- Il ruolo di servizio per le istanze Amazon EC2 sul cluster (chiamato anche profilo di istanza EC2 per Amazon EMR) non dispone delle autorizzazioni per accedere al bucket Amazon S3 in cui risiede lo script dell'operazione di bootstrap. Per ulteriori informazioni sui ruoli di servizio, consulta Ruolo di servizio per istanze EC2 del cluster (profilo istanza EC2).

Per ulteriori informazioni sulle operazioni di bootstrap, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo.

Risoluzione

Per risolvere questo errore, assicurati che l'istanza primaria disponga dell'accesso appropriato allo script dell'operazione di bootstrap.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`primary-instance-id`

L'ID dell'istanza primaria in cui l'azione di bootstrap non è riuscita.

`bootstrap-action`

Il numero ordinale dell'operazione di bootstrap che non è riuscita. Uno script con un valore `bootstrap-action` di `1` è la prima azione di bootstrap da eseguire sull'istanza.

`amazon-s3-path`

La posizione Amazon S3 dell'azione di bootstrap non riuscita.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere la causa principale dell'errore dell'operazione di bootstrap. Quindi avvia un nuovo cluster.

Fasi per la risoluzione dei problemi

1. Utilizza il valore `amazon-s3-path` dell'array `ErrorData` per trovare lo script dell'operazione di bootstrap pertinente in Amazon S3.
2. Se hai attivato i log del cluster quando hai creato l'istanza, consulta il log `stdout` per ulteriori informazioni. Puoi trovare il log `stdout` per l'operazione di bootstrap in questa posizione Amazon S3:

```
s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-
actions/Failed_Bootstrap_Action_Number/stdout.gz
```

Per ulteriori informazioni sui log del cluster, consulta Configurazione della registrazione e del debug di cluster.

3. Per determinare l'errore dell'operazione di bootstrap, esamina le eccezioni nei log `stdout` e il valore in `return-code` in `ErrorData`.
4. Usa i risultati del passaggio precedente per rivedere l'operazione di bootstrap in modo da evitare le eccezioni o gestirle correttamente quando si verificano.
5. Avvia un nuovo cluster con l'operazione di bootstrap aggiornata.

BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY

Panoramica

L'errore `BOOTSTRAP_FAILURE_FILE_NOT_FOUND_PRIMARY` indica che l'istanza primaria non riesce a trovare lo script dell'operazione di bootstrap che l'istanza ha appena scaricato dal bucket Amazon S3 specificato.

Risoluzione

Per risolvere questo errore, verifica che l'istanza primaria disponga dell'accesso appropriato allo script dell'operazione di bootstrap.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`primary-instance-id`

L'ID dell'istanza primaria in cui l'azione di bootstrap non è riuscita.

`bootstrap-action`

Il numero ordinale dell'operazione di bootstrap che non è riuscita. Uno script con un valore `bootstrap-action` di `1` è la prima azione di bootstrap da eseguire sull'istanza.

`amazon-s3-path`

La posizione Amazon S3 dell'azione di bootstrap non riuscita.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere la causa principale dell'errore dell'operazione di bootstrap. Quindi avvia un nuovo cluster.

1. Per trovare lo script dell'operazione di bootstrap pertinente in Amazon S3, utilizza il valore `amazon-s3-path` dell'array `ErrorData`.
2. Esamina i file di log dell'operazione di bootstrap in Amazon S3 per identificare la causa principale dell'errore. Per ulteriori informazioni su come visualizzare i log di Amazon EMR, consulta Visualizzare file di log di .

   > **Note**
   >
   > Se non hai attivato i log per il tuo cluster, devi creare un nuovo cluster con le stesse configurazioni e operazioni di bootstrap. Per garantire che i log del cluster siano stati attivati, consulta Configurazione della registrazione e del debug di cluster.

3. Esamina il log `stdout` per le operazioni di bootstrap e conferma che non vi siano processi personalizzati che eliminano i file nella cartella `/emr/instance-controller/lib/bootstrap-actions` sulle istanze primarie. Puoi trovare il log `stdout` per l'operazione di bootstrap in questa posizione Amazon S3:

   ```
   s3://EXAMPLE-BUCKET/logs/Your_Cluster_Id/node/Primary_Instance_Id/bootstrap-actions/Failed_Bootstrap_Action_Number/stdout.gz
   ```

4. Avvia un nuovo cluster con l'operazione di bootstrap aggiornata.

Codici di errori interni

Le seguenti sezioni forniscono informazioni sulla risoluzione dei problemi relativi ai codici di errore interni.

Argomenti

INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ

Panoramica

Un cluster termina con un errore `INTERNAL_ERROR_EC2_INSUFFICIENT_CAPACITY_AZ` quando la zona di disponibilità selezionata non ha una capacità sufficiente per soddisfare la richiesta del tipo di istanza di Amazon EC2. La sottorete selezionata per un cluster determina la zona di disponibilità. Per ulteriori informazioni sulle sottoreti per Amazon EMR, consulta Configurazione delle reti.

Risoluzione

Per risolvere questo errore, modifica le configurazioni del tipo di istanza e crea un nuovo cluster con la richiesta aggiornata.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`instance-type`

Il tipo di istanza che ha esaurito la capacità.

`availability-zone`

La zona di disponibilità in cui si risolve la sottorete.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere la causa principale dell'errore di configurazione del cluster:

- Rivedi le best practice per la configurazione dei cluster. Consulta Best practice per la configurazione dei cluster nella Guida alla gestione di Amazon EMR.
- Risolvi i problemi di avvio e rivedi la configurazione. Consulta Risoluzione dei problemi di avvio delle istanze nella Guida per l'utente di Amazon EC2 per le istanze Linux.
- Avvia un nuovo cluster con la configurazione aggiornata.

INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY

Panoramica

Un cluster termina con un errore `INTERNAL_ERROR_SPOT_PRICE_INCREASE_PRIMARY` quando Amazon EMR non è in grado di soddisfare la richiesta di istanza Spot per il nodo primario perché le istanze non sono disponibili al prezzo Spot massimo o inferiore. Per ulteriori informazioni, consulta Istanze Spot nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Risoluzione

Per risolvere questo errore, specifica i tipi di istanza per il cluster che rientrano nel tuo target di prezzo o aumenta il limite di prezzo per lo stesso tipo di istanza.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`primary-instance-id`

L'ID per l'istanza primaria del cluster che ha avuto esito negativo.

`instance-type`

Il tipo di istanza che ha esaurito la capacità.

`availability-zone`

La zona di disponibilità in cui risiede la sottorete.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per risolvere i problemi relativi alla tua strategia di configurazione del cluster, quindi avvia un nuovo cluster:

1. Esamina le best practice per le istanze Spot di Amazon EC2 e rivedi la tua strategia di configurazione del cluster. Per ulteriori informazioni, consulta Best practice per Spot EC2 nella Guida per l'utente Amazon EC2 per le istanze Linux e Best practice per la configurazione dei cluster.
2. Modifica le configurazioni del tipo di istanza o la zona di disponibilità e crea un nuovo cluster con la richiesta aggiornata.
3. Se il problema persiste, utilizza la capacità On-Demand per l'istanza primaria.

INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY

Panoramica

Un cluster termina con un errore `INTERNAL_ERROR_SPOT_NO_CAPACITY_PRIMARY` quando la capacità non è sufficiente per soddisfare una richiesta di istanza Spot per il nodo primario. Per ulteriori informazioni, consulta Istanze Spot nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Risoluzione

Per risolvere questo errore, specifica i tipi di istanza per il cluster che rientrano nel tuo target di prezzo o aumenta il limite di prezzo per lo stesso tipo di istanza.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`primary-instance-id`

L'ID per l'istanza primaria del cluster che ha avuto esito negativo.

`instance-type`

Il tipo di istanza che ha esaurito la capacità.

availability-zone

La zona di disponibilità in cui si risolve la sottorete.

public-doc

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per risolvere i problemi relativi alla tua strategia di configurazione del cluster, quindi avvia un nuovo cluster:

1. Esamina le best practice per le istanze Spot di Amazon EC2 e rivedi la tua strategia di configurazione del cluster. Per ulteriori informazioni, consulta Best practice per Spot EC2 nella Guida per l'utente Amazon EC2 per le istanze Linux e Best practice per la configurazione dei cluster.
2. Modifica le configurazioni dei tipi di istanza e crea un nuovo cluster con la richiesta aggiornata.
3. Se il problema persiste, utilizza la capacità On-Demand per l'istanza primaria.

Codici di errore relativi agli errori di convalida

Le seguenti sezioni forniscono informazioni sulla risoluzione dei problemi relativi ai codici di errore per gli errori di convalida.

Argomenti

- 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

Panoramica

Quando il cluster e le sottoreti a cui si fa riferimento per il cluster appartengono a cloud privati virtuali (VPC) diversi, il cluster termina con un errore `VALIDATION_ERROR_SUBNET_NOT_FROM_ONE_VPC`.

Puoi avviare i cluster con Amazon EMR con la configurazione dei parchi istanze tra le sottoreti in un VPC. Per ulteriori informazioni sui parchi istanze, consulta Configurazione di parchi istanze nella Guida alla gestione di Amazon EMR.

Risoluzione

Per risolvere questo errore, utilizza sottoreti che appartengono allo stesso VPC del cluster.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`vpc`

Per ogni coppia subnet:VPC, l'ID del VPC a cui appartiene la sottorete.

`subnet`

Per ogni coppia subnet:VPC, l'ID della sottorete.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere l'errore:

1. Controlla gli ID delle sottoreti elencati nell'array `ErrorData` e verifica che appartengano al VPC in cui desideri avviare il cluster EMR.
2. Modifica le configurazioni delle sottoreti. Puoi utilizzare uno dei metodi seguenti per trovare tutte le sottoreti pubbliche e private disponibili in un VPC.

   - Passa alla console di Amazon VPC. Scegli Sottoreti ed elenca tutte le sottoreti che risiedono all'interno della Regione AWS per il tuo cluster. Per trovare solo sottoreti pubbliche o private, applica il filtro Assegna automaticamente gli indirizzi IPv4 pubblici. Per trovare e selezionare le sottoreti nel VPC utilizzato dal cluster, utilizza l'opzione Filtra per VPC. Per ulteriori informazioni su come creare sottoreti, consulta Creazione di una sottorete nella Guida per l'utente di Amazon Virtual Private Cloud.

- Utilizza la AWS CLI per trovare tutte le sottoreti pubbliche e private disponibili nel VPC utilizzato dal cluster. Per ulteriori informazioni, consulta l'API describe-subnets. Per creare nuove sottoreti in un VPC, consulta l'API create-subnet.

3. Avvia un nuovo cluster con sottoreti dallo stesso VPC del cluster.

VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC

Panoramica

Quando il cluster e i gruppi di sicurezza assegnati al tuo cluster appartengono a cloud privati virtuali (VPC) diversi, il cluster termina con un errore `VALIDATION_ERROR_SECURITY_GROUP_NOT_FROM_ONE_VPC`. Per ulteriori informazioni sui gruppi di sicurezza, consulta Specifica dei gruppi di sicurezza gestiti da Amazon EMR e aggiuntivi e Controllo del traffico di rete con gruppi di sicurezza.

Risoluzione

Per risolvere questo errore, utilizza gruppi di sicurezza che appartengono allo stesso VPC del cluster.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`vpc`

Per ogni coppia security-group:VPC, l'ID del VPC a cui appartiene il gruppo di sicurezza.

`security-group`

Per ogni coppia security-group:VPC, l'ID del gruppo di sicurezza.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere l'errore:

1. Controlla gli ID dei gruppi di sicurezza elencati nell'array `ErrorData` e verifica che appartengano al VPC in cui desideri avviare il cluster EMR.

2. Passa alla console di Amazon VPC. Scegli Gruppi di sicurezza per elencare tutti i gruppi di sicurezza all'interno della regione selezionata. Trova i gruppi di sicurezza provenienti dallo stesso VPC del cluster, quindi modifica la configurazione del gruppo di sicurezza.
3. Avvia un nuovo cluster con gruppi di sicurezza provenienti dallo stesso VPC del cluster.

VALIDATION_ERROR_INVALID_SSH_KEY_NAME

Panoramica

Un cluster termina con un errore `VALIDATION_ERROR_INVALID_SSH_KEY_NAME` quando utilizzi una coppia di chiavi Amazon EC2 non valida per SSH nell'istanza primaria. Il nome della coppia di chiavi potrebbe non essere corretto o la coppia di chiavi potrebbe non esistere nella Regione AWS richiesta. Per ulteriori informazioni sulle coppie di chiavi, consulta Coppie di chiavi Amazon EC2 e istanze Linux nella Guida per l'utente di Amazon EC2 per le istanze Linux.

Risoluzione

Per risolvere questo errore, crea un nuovo cluster con un nome valido per la coppia di chiavi SSH.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`ssh-key`

Il nome della coppia di chiavi SSH che hai fornito quando hai creato il cluster.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere l'errore:

1. Controlla il file *`keypair`*.pem e verifica che corrisponda al nome della chiave SSH che vedi nella console Amazon EMR.
2. Passa alla console Amazon EC2. Verifica che il nome della chiave SSH che hai usato sia disponibile nella Regione AWS utilizzata dal cluster. Puoi trovare la tua Regione AWS accanto al tuo ID account nella parte superiore della AWS Management Console.

3. Avvia un nuovo cluster con un nome valido per la chiave SSH.

VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED

Panoramica

Un cluster termina con un errore `VALIDATION_ERROR_INSTANCE_TYPE_NOT_SUPPORTED` quando la Regione AWS e le zone di disponibilità del cluster non supportano il tipo di istanza specificato per uno o più gruppi di istanze. Amazon EMR è in grado di supportare un tipo di istanza in una zona di disponibilità all'interno di una Regione ma non in un'altra. La sottorete selezionata per un cluster determina la zona di disponibilità all'interno della regione. Per un elenco dei tipi di istanza e delle regioni supportati da Amazon EMR, consulta Tipi di istanze supportati.

Risoluzione

Per risolvere questo errore, specifica i tipi di istanza per il tuo cluster supportati da Amazon EMR nella regione e nella zona di disponibilità in cui richiedi il cluster.

Per risolvere il problema del cluster EMR non riuscito, consulta le informazioni su `ErrorDetail` restituite dalle API `DescribeCluster` e `ListClusters`. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail. L'array `ErrorData` all'interno di `ErrorDetail` restituisce le seguenti informazioni per questo codice di errore:

`instance-types`

L'elenco dei tipi di istanza non supportati.

`availability-zones`

L'elenco delle zone di disponibilità in cui viene risolta la sottorete.

`public-doc`

L'URL pubblico della documentazione per il codice di errore.

Passaggi da completare

Esegui i passaggi seguenti per identificare e correggere l'errore:

1. Utilizza la AWS CLI per recuperare i tipi di istanza disponibili in una zona di disponibilità. A tale scopo, puoi utilizzare il comando `ec2 describe-instance-type-offerings` per filtrare i tipi di istanze disponibili in base alla posizione (Regione AWS o zona di disponibilità).

Ad esempio, il comando seguente restituisce i tipi di istanza offerti nella zona di disponibilità specificata, *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
```

Per ulteriori informazioni su come scoprire i tipi di istanza disponibili, consulta Individuazione di un tipo di istanza Amazon EC2.

2. Dopo aver determinato i tipi di istanza disponibili nella stessa regione e zona di disponibilità del cluster, scegli una delle seguenti risoluzioni per continuare:

   a. Crea un nuovo cluster e scegli una sottorete per il cluster in una zona di disponibilità in cui il tipo di istanza è disponibile e supportato da Amazon EMR.

   b. Crea un nuovo cluster nella stessa regione e nella stessa sottorete Amazon EC2 del cluster in cui si è verificato l'errore, ma con un tipo di istanza supportato in quella posizione da Amazon EMR.

Per un elenco dei tipi di istanza e delle regioni supportati da Amazon EMR, consulta Tipi di istanze supportati. Per confrontare le funzionalità dei tipi di istanza, consulta Tipi di istanza Amazon EC2.

Errori nelle risorse

I seguenti errori sono comunemente causati da risorse vincolate sul cluster.

Argomenti

- Il cluster termina con NO_SLAVE_LEFT e i nodi principali FAILED_BY_MASTER
- Impossibile replicare il blocco, gestito solo per la replica su zero nodi.
- QUOTA EC2 SUPERATA
- Troppi errori di recupero
- Il file può essere replicato solo su 0 nodi invece di 1
- Nodi elencati come negati
- Errori di limitazione
- Tipo di istanza non supportato
- Capacità di EC2 esaurita

Il cluster termina con NO_SLAVE_LEFT e i nodi principali FAILED_BY_MASTER

In genere, ciò accade perché la protezione da cessazione è disabilitata e tutti i nodi principali superano la capacità di storage su disco come specificato da una soglia di utilizzo massimo nella classificazione di configurazione `yarn-site`, che corrisponde al file `yarn-site.xml`. Per impostazione predefinita, questo valore è 90%. Quando l'utilizzo del disco per un nodo principale supera la soglia di utilizzo, il servizio di stato NodeManager di YARN segnala il nodo come `UNHEALTHY`. Mentre è in questo stato, Amazon EMR elenca come negato il nodo e non assegna i container YARN a tale nodo. Se il nodo rimane nello stato non integro per 45 minuti, Amazon EMR contrassegna l'istanza Amazon EC2 associata per la terminazione come `FAILED_BY_MASTER`. Quando tutte le istanze Amazon EC2 associate con i nodi principali sono contrassegnate per la terminazione, il cluster viene terminato con lo stato `NO_SLAVE_LEFT` perché non vi sono risorse per l'esecuzione dei processi.

Il superamento dell'utilizzo del disco su un nodo principali potrebbe causare una reazione a catena. Se un singolo nodo supera la soglia di utilizzo del disco a causa di HDFS, è probabile che anche altri nodi siano vicini alla soglia. Il primo nodo supera la soglia di utilizzo del disco, quindi Amazon EMR lo elenca come negato. Ciò aumenta il carico di utilizzo del disco per i nodi rimanenti perché questi iniziano a replicare tra loro i dati HDFS che hanno perso sul nodo elencato come negato. Di conseguenza, ogni nodo diventa `UNHEALTHY` nello stesso modo e infine il cluster viene terminato.

Best practice e raccomandazioni

Configurazione di hardware cluster con archiviazione adeguata

Quando crei un cluster, accertati che vi sia un numero sufficiente di nodi principali e che ogni nodo disponga di volumi adeguati di storage EBS e instance store per HDFS. Per ulteriori informazioni, consulta Calcolo della capacità HDFS richiesta di un cluster. Puoi inoltre aggiungere istanze principali ai gruppi di istanze esistenti manualmente o utilizzando il dimensionamento automatico. Le nuove istanze hanno la stessa configurazione di storage delle altre istanze nel gruppo di istanze. Per ulteriori informazioni, consulta Uso del dimensionamento del cluster.

Abilitare la protezione da cessazione

Abilita la protezione da cessazione. In questo modo, se un nodo principale è elencato come negato, puoi connetterti all'istanza Amazon EC2 associata utilizzando SSH per risolvere i problemi relativi ai dati e recuperare i dati. Se abiliti la protezione da cessazione, ricorda che Amazon EMR non sostituisce l'istanza Amazon EC2 con una nuova istanza. Per ulteriori informazioni, consulta Utilizzo della protezione da cessazione.

Crea un allarme per il parametro CloudWatch MRUnhealthyNodes

Questo parametro indica il numero di nodi con stato `UNHEALTHY`. È equivalente al parametro YARN `mapred.resourcemanager.NoOfUnhealthyNodes`. Puoi impostare una notifica per questo allarme in modo che ti vengano segnalati i nodi non integri prima che venga raggiunto il timeout di 45 minuti. Per ulteriori informazioni, consulta Monitoraggio di parametri di Amazon EMR con CloudWatch.

Impostazioni Tweak mediante yarn-site

Le impostazioni riportate di seguito possono essere regolate in base ai requisiti dell'applicazione. Ad esempio, potresti voler aumentare la soglia di utilizzo del disco in base alla quale un nodo segnala lo stato `UNHEALTHY` aumentando il valore di `yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`.

Puoi impostare questi valori quando crei un cluster utilizzando la classificazione di configurazione `yarn-site`. Per ulteriori informazioni, consulta Configurazione delle applicazioni nella Guida ai rilasci di Amazon EMR. Puoi inoltre connetterti alle istanze Amazon EC2 associate ai nodi principali tramite SSH, quindi aggiungere i valori in `/etc/hadoop/conf.empty/yarn-site.xml` utilizzando un editor di testo. Dopo aver apportato la modifica, devi riavviare hadoop-yarn-nodemanager come riportato di seguito.

> **Important**
>
> Quando riavvii il servizio NodeManager, i container YARN attivi vengono terminati a meno che `yarn.nodemanager.recovery.enabled` non sia impostato su `true` mediante la classificazione di configurazione `yarn-site` quando crei il cluster. Devi inoltre specificare la directory in cui memorizzare lo stato del container utilizzando la proprietà `yarn.nodemanager.recovery.dir`.

```
sudo /sbin/stop hadoop-yarn-nodemanager
sudo /sbin/start hadoop-yarn-nodemanager
```

Per ulteriori informazioni sulle proprietà `yarn-site` correnti e i valori predefiniti, consulta la pagina delle impostazioni predefinite di YARN nella documentazione di Apache Hadoop.

Proprietà	Valore predefinito	Descrizione
yarn.nodemanager.disk-health-checker.interval-ms	120000	La frequenza (in secondi) con cui viene eseguito il controllo dello stato del disco.
yarn.nodemanager.disk-health-checker.min-healthy-disks	0.25	La frazione minima del numero di dischi che devono essere integri affinché NodeManager avvii nuovi container. Ciò corrisponde sia a yarn.nodemanager.local-dirs (per impostazione predefinita `/mnt/yarn` in Amazon EMR) che a yarn.nodemanager.log-dirs (per impostazione predefinita `/var/log/hadoop-yarn/containers`, del quale è eseguito il symlink a `mnt/var/log/hadoop-yarn/containers` in Amazon EMR).
`yarn.nodemanager.disk-health-checker.max-disk-utilization-per-disk-percentage`	90,0	La percentuale massima di utilizzo dello spazio su disco consentita dopo che un disco viene contrassegnato come difettoso. I valori possono andare da 0,0 a 100,0. Se il valore è maggiore o uguale a 100, NodeManager verifica la presenza di un disco intero. Ciò è valido per `yarn-nodemanager.local-dirs`

Proprietà	Valore predefinito	Descrizione
		e `yarn.nodemanager.log-dirs` .
`yarn.nodemanager.disk-health-checker.min-free-space-per-disk-mb`	0	Lo spazio minimo che deve essere disponibile su un disco affinché possa essere utilizzato. Ciò è valido per `yarn-nodemanager.local-dirs` e `yarn.nodemanager.log-dirs` .

Impossibile replicare il blocco, gestito solo per la replica su zero nodi.

Generalmente, l'errore "Cannot replicate block, only managed to replicate to zero nodes (Impossibile replicare il blocco, gestito solo per la replica su zero nodi)." si verifica quando un cluster non dispone di sufficiente spazio di archiviazione HDFS. Questo errore si verifica quando la quantità di dati generata nel cluster è superiore alla capacità di archiviazione di HDFS. Questo errore viene visualizzato solo durante l'esecuzione del cluster in quanto quando termina il processo rilascia lo spazio HDFS che stava utilizzando.

La quantità di spazio HDFS disponibile per un cluster dipende dal numero e dal tipo di istanze Amazon EC2 che vengono utilizzate come nodi principali. I nodi di task non vengono utilizzati per lo storage HDFS. L'intero spazio su disco di ciascuna istanza Amazon EC2, inclusi i volumi di archiviazione EBS associati, è disponibile per HDFS. Per ulteriori informazioni sulla quantità di archiviazione locale per ciascun tipo di istanza EC2, consulta la sezione relativa a Tipi e famiglie di istanze nella Guida per l'utente di Amazon EC2 per le istanze Linux.

L'altro fattore che può influenzare la quantità di spazio HDFS disponibile è il fattore di replica, ovvero il numero di copie di ciascun blocco di dati che viene archiviato in HDFS per la ridondanza. Il fattore di replica aumenta con il numero di nodi nel cluster: sono disponibili 3 copie di ogni blocco di dati per un cluster con 10 o più nodi, 2 copie di ogni blocco per un cluster con un numero di nodi da 4 a 9 e 1 copia (nessuna ridondanza) per i cluster con al massimo 3 nodi. Lo spazio HDFS totale disponibile è diviso per il fattore di replica. In alcuni casi, ad esempio incrementando il numero di nodi da 9 a 10, l'aumento del fattore di replica può causare effettivamente la diminuzione della quantità di spazio HDFS disponibile.

Ad esempio, per un cluster con 10 nodi principali di tipo m1.xlarge sarebbero disponibili 2833 GB di spazio in HDFS ((10 nodi x 850 GB per nodo)/fattore di replica 3).

Se le dimensioni del cluster superano la quantità di spazio disponibile per HDFS, puoi aggiungere altri nodi principali nel cluster o utilizzare la compressione dati per creare più spazio HDFS. Se il cluster può essere interrotto e riavviato, puoi valutare il ricorso ai nodi principali di un tipo di istanza Amazon EC2 più grande. Puoi anche valutare la modifica del fattore di replica. Tiene presente, tuttavia, che la riduzione del fattore di replica riduce la ridondanza dei dati HDFS e la capacità del cluster di recuperare da blocchi HDFS persi o danneggiati.

QUOTA EC2 SUPERATA

Se viene visualizzato un messaggio `EC2 QUOTA EXCEEDED`, le cause potrebbero essere diverse. A seconda della differenze di configurazione, potrebbero essere necessari tra i 5 e i 20 minuti affinché i cluster precedenti vengano terminati e le risorse allocate rilasciate. Se visualizzi un errore `EC2 QUOTA EXCEEDED` al tentativo di avvio di un cluster, potrebbe essere dovuto al rilascio non ancora avvenuto delle risorse di un terminato recentemente. Questo messaggio può anche essere causato dal ridimensionamento di un gruppo di istanze o di un parco istanze in una dimensione di destinazione maggiore della quota di istanza corrente per l'account. Questa operazione può essere eseguita manualmente o automaticamente attraverso il dimensionamento automatico.

Considera le seguenti opzioni per risolvere il problema:

- Segui le istruzioni in Quote di servizio di AWS in Riferimenti generali di Amazon Web Services per richiedere un aumento dei limiti del servizio. Per alcune API, l'impostazione di un evento CloudWatch potrebbe rappresentare un'opzione migliore rispetto all'aumento dei limiti. Per ulteriori dettagli, consulta Quando impostare gli eventi EMR in CloudWatch.
- Se uno o più cluster in esecuzione non hanno capacità, ridimensiona i gruppi di istanze o riduci le capacità di destinazione sui parchi istanze per l'esecuzione di cluster.
- Crea i cluster con un numero minore di istanze EC2 o una capacità di destinazione ridotta.

Troppi errori di recupero

La presenza di messaggi di errore "Too many fetch-failures (Troppi errori fetch)" o "Error reading task output (Errore di lettura output attività)" nella fase o nei log del tentativo di attività indica che l'attività in esecuzione dipende dall'output di un'altra attività. Ciò si verifica spesso quando un'attività di riduzione viene accodata per l'esecuzione e richiede l'output di una o più attività di mappatura e tale output non è ancora disponibile.

L'output potrebbe non essere disponibile per diversi motivi:

- L'attività preliminare è ancora in fase di elaborazione. Questa è spesso un'attività di mappatura.
- I dati potrebbero non essere disponibili a causa di scarsa connettività di rete se i dati si trovano su un'istanza diversa.
- Se si utilizza HDFS per recuperare l'output, potrebbe verificarsi un problema con HDFS.

La causa più comune di questo errore è che l'attività precedente è ancora in corso di elaborazione. Ciò è probabile soprattutto se gli errori si verificano durante il primo tentativo di esecuzione delle attività di riduzione. Puoi verificare se questo è il motivo esaminando il log syslog per la fase di cluster che restituisce l'errore. Se il syslog mostra che entrambe le attività di mappatura e riduzione stanno avanzando, significa che la fase di riduzione è iniziata mentre ci sono attività di mappatura non ancora completate.

Ti consigliamo di cercare nei log una percentuale di avanzamento della mappatura che raggiunge il 100% e quindi scende nuovamente a un valore più basso. Quando la percentuale di mappatura è pari al 100%, questo non significa che tutte le attività di mappatura sono state completate, ma semplicemente che Hadoop sta eseguendo tutte le attività di mappatura. Se questo valore scende nuovamente sotto il 100%, significa che un'attività di mappatura non è riuscita e, a seconda della configurazione, Hadoop potrebbe tentare di pianificare di nuovo l'attività. Se la percentuale di mappatura rimane al 100% nei log, esamina i parametri CloudWatch, in particolare `RunningMapTasks`, per controllare se l'attività di mappatura è ancora in corso di elaborazione. Puoi anche trovare queste informazioni utilizzando l'interfaccia Web Hadoop sul nodo master.

Se si verifica questo problema, puoi provare a eseguire una serie di operazioni:

- Configura la fase di riduzione per aspettare più a lungo prima dell'avvio. A questo scopo, puoi modificare l'impostazione di configurazione Hadoop mapred.reduce.slowstart.completed.maps su un tempo più lungo. Per ulteriori informazioni, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo.
- Associa il conteggio del riduttore alla funzionalità del riduttore totale del cluster. A questo scopo, modifica l'impostazione di configurazione Hadoop mapred.reduce.tasks per il processo.
- Utilizza un codice classe combiner per ridurre al minimo la quantità di output da recuperare.
- Verifica che non ci siano problemi con il servizio Amazon EC2 che influenzano le prestazioni di rete del cluster. A questo scopo, puoi utilizzare il Service Health Dashboard.

- Esamina le risorse di CPU e della memoria delle istanze nel cluster per verificare che l'elaborazione dei dati non stia sovraccaricando le risorse dei nodi. Per ulteriori informazioni, consulta Configurazione di hardware e reti cluster.
- Controlla la versione di Amazon Machine Image (AMI) utilizzata nel cluster Amazon EMR. Se la versione è compresa tra 2.3.0 e 2.4.4, esegui l'aggiornamento a una versione più recente. Le versioni AMI nell'intervallo specificato utilizzano una versione di Jetty che potrebbe non essere in grado di fornire l'output dalla fase di mappatura. L'errore di recupero si verifica quando i riduttori non sono in grado di ottenere l'output dalla fase di mappatura.

  Jetty è un server HTTP open source utilizzato per le comunicazioni da macchina a macchina all'interno di un cluster Hadoop.

Il file può essere replicato solo su 0 nodi invece di 1

Un file scritto in HDFS viene replicato in più nodi principali. Se si verifica questo errore, significa che il daemon NameNode non dispone di istanze DataNode disponibili per scrivere dati in HDFS. In altre parole, la replica dei blocchi non viene eseguita. Questo errore può essere causato da diversi problemi:

- Possibile mancanza di spazio libero per il filesystem HDFS. Questa è la causa più probabile.
- Istanze DataNode non disponibili quando il processo è stato eseguito.
- Comunicazione bloccata tra le istanze DataNode con il nodo master.
- Istanze non disponibili nel gruppo di istanze principale.
- Possibile mancanza di autorizzazioni. Ad esempio, il daemon JobTracker potrebbe non disporre delle autorizzazioni per creare informazioni sul job tracker.
- Spazio riservato insufficiente per un'istanza DataNode. Controlla se questo è il caso verificando l'impostazione di configurazione dfs.datanode.du.reserved.

Per verificare se il problema è causato dall'esaurimento dello spazio su disco per HDFS, esamina il parametro `HDFSUtilization` in CloudWatch. Se questo valore è troppo elevato, puoi aggiungere altri nodi principali al cluster. Se disponi di un cluster che pensi potrebbe esaurire lo spazio su disco HDFS, puoi impostare un allarme in CloudWatch per avvisare quando il valore di `HDFSUtilization` supera un livello specificato. Per ulteriori informazioni, consulta Ridimensionamento manuale di un cluster in esecuzione e Monitoraggio di parametri di Amazon EMR con CloudWatch.

Se il problema non è l'esaurimento dello spazio su disco per HDFS, controlla i log DataNode, i log NameNode e la connettività di rete per altri problemi che potrebbero impedire ad HDFS di replicare i dati. Per ulteriori informazioni, consulta Visualizzare file di log di .

Nodi elencati come negati

Il daemon NodeManager è responsabile dell'avvio e della gestione di container su nodi principali e di task. I contenitori vengono allocati nel daemon NodeManager dal daemon ResourceManager in esecuzione sul nodo master. Il ResourceManager monitora il nodo NodeManager tramite un heartbeat.

Di seguito sono riportati un paio di casi in cui il daemon ResourceManager elenca come negato un NodeManager rimuovendolo dal pool di nodi disponibili per elaborare i processi:

- Il nodo NodeManager non ha inviato un heartbeat al daemon ResourceManager negli ultimi 10 minuti (600.000 millisecondi). Questo periodo di tempo può essere configurato utilizzando l'impostazione di configurazione `yarn.nm.liveness-monitor.expiry-interval-ms`. Per ulteriori informazioni sulla modifica delle classificazioni di configurazione Yarn, consulta la sezione relativa alla Configurazione delle applicazioni nella Guida ai rilasci di Amazon EMR.
- NodeManager controlla l'integrità dei dischi determinati da `yarn.nodemanager.local-dirs` e `yarn.nodemanager.log-dirs`. I controlli includono le autorizzazioni e lo spazio libero su disco (< 90%). Se un disco non supera il controllo, il NodeManager smette di utilizzare quel particolare disco ma segnala ancora lo stato del nodo come integro. Se diversi dischi non superano il controllo, il nodo viene segnalato come non integro al ResourceManager e non vengono assegnati nuovi container al nodo.

Il master dell'applicazione può anche elencare come negato un nodo NodeManager se il numero di processi non riusciti è superiore a tre. Puoi modificare questa impostazione su un valore più alto utilizzando il parametro di configurazione `mapreduce.job.maxtaskfailures.per.tracker`. Altre impostazioni di configurazione che è possibile modificare sono il numero di tentativi di esecuzione di una task prima che venga contrassegnata come non riuscita: `mapreduce.map.max.attempts` per task di mappatura e `mapreduce.reduce.maxattempts` per task di riduzione. Per ulteriori informazioni sulla modifica delle classificazioni di configurazione, consulta la sezione relativa alla Configurazione delle applicazioni nella Guida ai rilasci di Amazon EMR.

Errori di limitazione

Gli errori "Throttled from *Amazon EC2* while launching cluster (Limitato da Amazon EC2 durante l'avvio del cluster)" e "Failed to provision instances due to throttling from *Amazon EC2* (Impossibile effettuare il provisioning delle istanze a causa della limitazione di Amazon EC2)" si verificano quando Amazon EMR non è in grado di completare una richiesta in quanto un altro servizio ha limitato l'attività. Amazon EC2 è la fonte più comune di errori di limitazione, ma non è l'unica. I limiti del servizio AWS sono specifici in base alla Regione e servono a migliorare le prestazioni; infatti, un errore di limitazione indica a un utente che ha superato i limiti del servizio per il suo account in una specifica Regione.

Possibili cause

L'origine più comune degli errori di limitazione di Amazon EC2 è il numero elevato di istanze cluster avviate che porta al superamento del limite del servizio per le istanze EC2. Le istanze cluster possono essere avviate per i seguenti motivi:

- Vengono creati nuovi cluster.
- I cluster vengono ridimensionati manualmente. Per ulteriori informazioni, consulta Ridimensionamento manuale di un cluster in esecuzione.
- I gruppi di istanze in un cluster aggiungono istanze (ridimensionamento) come risultato di una regola di ridimensionamento automatico. Per ulteriori informazioni, consulta Comprensione delle regole di scalabilità automatica.
- I parchi istanze in un cluster aggiungono le istanze per soddisfare una maggiore capacità di destinazione. Per ulteriori informazioni, consulta Configurazione di parchi istanze.

È anche possibile che la frequenza o il tipo di richiesta API ad Amazon EC2 causi errori di limitazione. Per ulteriori informazioni su come Amazon EC2 limita le richieste API, consulta Tasso delle richieste API sulle query nella Guida di riferimento alle API di Amazon EC2.

Soluzioni

Prendi in considerazione le soluzioni seguenti:

- Segui le istruzioni in Quote di servizio di AWS in Riferimenti generali di Amazon Web Services per richiedere un aumento dei limiti del servizio. Per alcune API, l'impostazione di un evento CloudWatch potrebbe rappresentare un'opzione migliore rispetto all'aumento dei limiti. Per ulteriori dettagli, consulta Quando impostare gli eventi EMR in CloudWatch.

- Se disponi di cluster che si avviano con la stessa pianificazione, ad esempio all'inizio di ogni ora, puoi aspettarti tempi di avvio straordinari.
- Se disponi di cluster dimensionati per la domanda di picco e hai periodicamente la capacità di istanza, considera la possibilità di specificare il ridimensionamento automatico per aggiungere e rimuovere le istanze on demand. In questo modo, le istanze vengono utilizzate in modo più efficiente e, in base al profilo della domanda, è possibile richiedere un numero inferiore di istanze in un dato momento in un account. Per ulteriori informazioni, consulta Utilizzo del dimensionamento automatico con una policy personalizzata per i gruppi di istanze .

Tipo di istanza non supportato

Se cerchi di creare un cluster ma ricevi il messaggio di errore "The requested instance type *InstanceType* is not supported in the requested Availability Zone (Il tipo di istanza richiesto InstanceType non è supportato nella zona di disponibilità richiesta)", significa che hai creato il cluster e hai specificato un tipo di istanza per uno o più gruppi di istanze che non sono supportati da Amazon EMR nella Regione e nella zona di disponibilità in cui è stato creato il cluster. Amazon EMR è in grado di supportare un tipo di istanza in una zona di disponibilità all'interno di una Regione ma non in un'altra. La sottorete selezionata per un cluster determina la zona di disponibilità all'interno della regione.

Soluzione

Determinare i tipi di istanza disponibili in una zona di disponibilità utilizzando la AWS CLI

- Utilizzare il comando `ec2 run-instances` con l'opzione `--dry-run`. Nell'esempio seguente, sostituisci *m5.xlarge* con il tipo di istanza da utilizzare, *ami-035be7bafff33b6b6* con l'AMI associata a quel tipo di istanza e la *subnet-12ab3c45* con una sottorete nella zona di disponibilità su cui intendi eseguire la query.

  ```
  aws ec2 run-instances --instance-type m5.xlarge --dry-run --image-
  id ami-035be7bafff33b6b6 --subnet-id subnet-12ab3c45
  ```

  Per istruzioni su come trovare un ID AMI, consulta Ricerca di un'AMI Linux. Per trovare un ID di sottorete, puoi utilizzare il comando describe-subnets.

Per ulteriori informazioni su come scoprire i tipi di istanza disponibili, consulta Individuazione di un tipo di istanza Amazon EC2.

Dopo aver determinato i tipi di istanza disponibili, è possibile eseguire una delle operazioni seguenti:

- Crea il cluster nella stessa regione e sottorete EC2 e seleziona un tipo di istanza diverso con funzionalità simili a quelle della scelta iniziale. Per una lista di tipi di istanze supportate, consulta Tipi di istanze supportati. Per confrontare le caratteristiche dei tipi di istanza EC2, consulta la sezione sui tipi di istanza Amazon EC2.
- Scegli una sottorete per il cluster in una zona di disponibilità in cui il tipo di istanza è disponibile e supportato da Amazon EMR.

Capacità di EC2 esaurita

Un errore "EC2 is out of capacity for *`InstanceType`* (EC2 fuori capacità per InstanceType" si verifica quando si tenta di creare un cluster o aggiungere istanze a un cluster in una zona di disponibilità che non ha più il tipo di istanza EC2 specificato. La sottorete selezionata per un cluster determina la zona di disponibilità.

Per creare un cluster, effettua una delle operazioni indicate di seguito:

- Specifica un diverso tipo di istanza con funzionalità simili
- Crea il cluster in una Regione diversa
- Seleziona una sottorete in una zona di disponibilità in cui potrebbe essere disponibile il tipo di istanza che desideri.

Per aggiungere istanze a un cluster in esecuzione, effettua una delle seguenti operazioni:

- Modifica le configurazioni dei gruppi di istanze o le configurazioni dei parchi istanze per aggiungere tipi di istanze disponibili con capacità simili. Per una lista di tipi di istanze supportate, consulta Tipi di istanze supportati. Per confrontare le caratteristiche dei tipi di istanza EC2, consulta la sezione sui tipi di istanza Amazon EC2.
- Termina il cluster e ricrealo in una Regione e in una zona di disponibilità in cui è disponibile il tipo di istanza.

Errori di input e output

I seguenti errori sono comuni nelle operazioni di input e output del cluster.

Argomenti

Il percorso per Amazon Simple Storage Service (Amazon S3) presenta almeno tre barre?

Quando si specifica un bucket Amazon S3, è necessario includere una barra di terminazione alla fine dell'URL. Ad esempio, invece di fare riferimento a un bucket come "s3n://*DOC-EXAMPLE-BUCKET1*", si dovrebbe utilizzare "s3n://*DOC-EXAMPLE-BUCKET1*/", altrimenti Hadoop ha esito negativo sul cluster nella maggior parte dei casi.

Stai cercando di attraversare le directory di input in modo ricorsivo?

Hadoop non cerca i file nelle directory di input in modo ricorsivo. Se si dispone di una struttura di directory come /corpus/01/01.txt, /corpus/01/02.txt, /corpus/02/01.txt, ecc. e si specifica /corpus/ come parametro di input per il cluster, Hadoop non trova alcun file di input perché la directory /corpus/ è vuota e Hadoop non controlla il contenuto delle sottodirectory. Allo stesso modo, Hadoop non controlla ricorsivamente le sottodirectory bucket Amazon S3.

I file di input devono trovarsi direttamente nella directory di input o nel bucket Amazon S3 specificato, non nelle sottocartelle.

La directory di output esiste già?

Se si specifica un percorso di output che esiste già, Hadoop farà fallire il cluster nella maggior parte dei casi. Questo significa che se si esegue un cluster una volta e poi lo si esegue di nuovo con esattamente gli stessi parametri, probabilmente funzionerà la prima volta e poi mai più; dopo la prima esecuzione, il percorso di output esiste e quindi causa il fallimento di tutte le esecuzioni successive.

Si sta tentando di specificare una risorsa utilizzando un URL HTTP?

Hadoop non accetta le posizioni di risorsa specificate utilizzando il prefisso http:// Non è possibile fare riferimento a una risorsa utilizzando un URL HTTP. Ad esempio, se si passa http://mysite/myjar.jar come parametro JAR il cluster fallisce.

Si sta facendo riferimento a un bucket Amazon S3 che utilizza un formato di nome non valido?

Se si tenta di utilizzare un nome di bucket come "*DOC-EXAMPLE-BUCKET1*.1" con Amazon EMR, il cluster avrà esito negativo perché Amazon EMR richiede che i nomi del bucket siano nomi host RFC 2396 validi; il nome non può terminare con un numero. Inoltre, a causa dei requisiti di Hadoop, i nomi dei bucket Amazon S3 utilizzati con Amazon EMR devono contenere solo lettere minuscole, numeri, punti (.) e trattini alti (-). Per ulteriori informazioni su come formattare i nomi dei bucket Amazon S3, consulta Restrizioni e limitazioni dei bucket nella Guida per l'utente di Amazon Simple Storage Service.

Si stanno verificando problemi di caricamento dei dati in e da Amazon S3?

Amazon S3 è la più popolare fonte di input e output per Amazon EMR. Un errore comune è trattare Amazon S3 come un normale file system. Occorre tener conto delle differenze tra Amazon S3 e un file system quando si esegue il cluster.

- Se si verifica un errore interno in Amazon S3, l'applicazione deve gestirlo in modo semplice e ripetere l'operazione.
- Se le chiamate ad Amazon S3 impiegano troppo tempo per tornare indietro, l'applicazione potrebbe dover ridurre la frequenza con cui chiama Amazon S3.
- L'elenco di tutti gli oggetti in un bucket Amazon S3 costituisce una chiamata costosa. L'applicazione dovrebbe ridurre al minimo il numero di volte che ciò accade.

Ci sono diversi modi per migliorare il modo in cui il cluster interagisce con Amazon S3.

- Avviare il cluster utilizzando la versione più recente di Amazon EMR.
- Utilizzare S3DistCp per spostare gli oggetti all'interno e all'esterno di Amazon S3. S3DistCp implementa la gestione degli errori, i tentativi e i back-off per soddisfare i requisiti di Amazon S3. Per ulteriori informazioni, consulta Copie distribuite utilizzando S3DistCp.
- Progettare l'applicazione con la consistenza finale in mente. Utilizza HDFS per la memorizzazione dei dati intermedi mentre il cluster è in esecuzione e Amazon S3 solo per inserire i dati iniziali e produrre i risultati finali.
- Se i cluster impegneranno 200 o più transazioni al secondo in Amazon S3, contatta il supporto per preparare il bucket per più transazioni al secondo e prendI in considerazione l'utilizzo delle strategie di partizione chiave descritte in Consigli per le prestazioni in Amazon S3.

- Impostare l'impostazione di configurazione Hadoop io.file.buffer.size su 65536. In questo modo Hadoop trascorre meno tempo a cercare tra gli oggetti Amazon S3.
- Considera la possibilità di disabilitare la funzione di esecuzione speculativa di Hadoop se il cluster riscontra problemi di concomitanza con Amazon S3. Questo è utile anche quando si esegue la risoluzione dei problemi di un cluster lento. È possibile eseguire questa operazione impostando le proprietà `mapreduce.map.speculative` e `mapreduce.reduce.speculative` su `false`. Quando si avvia un cluster, è possibile impostare questi valori utilizzando la classificazione di configurazione `mapred-env`. Per ulteriori informazioni, consulta Configurazione delle applicazioni nella Guida alle versioni di Amazon EMR.
- Se è in esecuzione un cluster Hive, consulta Si stanno verificando problemi di caricamento dei dati in e da Amazon S3 in Hive?.

Per ulteriori informazioni, consulta Best practice per gli errori di Amazon S3 nella Guida per l'utente di Amazon Simple Storage Service.

Errori di autorizzazioni

Di seguito sono descritti alcuni degli errori comuni relativi all'utilizzo di autorizzazioni o credenziali.

Argomenti

- Le credenziali specificate per SSH sono corrette?
- Se utilizzi IAM, hai impostato le policy Amazon EC2 appropriate?

Le credenziali specificate per SSH sono corrette?

Se non sei in grado di utilizzare SSH per eseguire la connessione al nodo master, il problema è probabilmente legato alle tue credenziali di sicurezza.

In primo luogo, verifica che il file .pem contenente la chiave SSH disponga delle autorizzazioni appropriate. Puoi utilizzare chmod per modificare le autorizzazioni per il file .pem come illustrato nell'esempio seguente, dove devi sostituire mykey.pem con il nome del tuo file .pem.

```
chmod og-rwx mykey.pem
```

La seconda possibilità è che non stai utilizzando la coppia di chiavi che hai specificato alla creazione del cluster. Si tratta di un errore comune se hai creato più coppie di chiavi. Verifica i dettagli del cluster nella console di Amazon EMR (o utilizza l'opzione `--describe` nella CLI) per determinare il nome della coppia di chiavi specificato alla creazione del cluster.

Dopo aver verificato di utilizzare la coppia di chiavi corretta e che le autorizzazioni siano impostate correttamente nel file .pem, puoi utilizzare il comando seguente per utilizzare SSH per la connessione al nodo master, sostituendo mykey.pem con il nome del file .pem e `hadoop@ec2-01-001-001-1.compute-1.amazonaws.com` con il nome DNS pubblico del nodo master (disponibile mediante l'opzione `--describe` nella CLI o tramite la console di Amazon EMR).

> **Important**
>
> Dovrai utilizzare il nome di login `hadoop` quando esegui la connessione a un nodo di cluster Amazon EMR, altrimenti potrebbe verificarsi un errore simile a `Server refused our key`.

```
ssh -i mykey.pem hadoop@ec2-01-001-001-1.compute-1.amazonaws.com
```

Per ulteriori informazioni, consulta Connessione al nodo primario tramite SSH.

Se utilizzi IAM, hai impostato le policy Amazon EC2 appropriate?

Poiché Amazon EMR utilizza le istanze EC2 come nodi, è necessario che determinate policy Amazon EC2 siano impostate per gli utenti di Amazon EMR per consentire ad Amazon EMR di gestire tali istanze per conto dell'utente. Se non disponi delle autorizzazioni necessarie, Amazon EMR restituisce l'errore: "Account utente non autorizzato a chiamare EC2."

Per ulteriori informazioni sulle policy Amazon EC2 che il tuo account IAM deve impostare per eseguire Amazon EMR, consulta Funzionamento di Amazon EMR con IAM.

Errori del cluster Hive

In genere, puoi trovare la causa di un errore Hive nel file `syslog`, a cui ti colleghi dal riquadro Steps (Fasi). Se non riesci a determinare il problema, controlla nel messaggio di errore del tentativo di attività Hadoop tramite il riquadro Task Attempts (Tentativi attività).

I seguenti errori sono comuni ai cluster Hive.

Argomenti

Stai usando la versione più recente di Hive?

La versione più recente di Hive contiene tutte le patch e le correzioni dei bug correnti e può risolvere il problema.

Hai rilevato un errore di sintassi nello script Hive?

Se una fase non riesce, cerca nel file `stdout` dei log la fase che ha eseguito lo script Hive. Se l'errore non esiste, cerca nel file `syslog` dei log dei tentativi di attività il tentativo di attività non riuscito. Per ulteriori informazioni, consulta Visualizzare file di log di .

Un processo non è riuscito durante l'esecuzione in modalità interattiva?

Se stai eseguendo Hive in maniera interattiva sul nodo master e il cluster non è riuscito, nel log dei tentativi di attività cerca le voci `syslog` relative al tentativo di attività non riuscito. Per ulteriori informazioni, consulta Visualizzare file di log di .

Si stanno verificando problemi di caricamento dei dati in e da Amazon S3 in Hive?

Se si verificano problemi di accesso ai dati in Amazon S3, controlla innanzitutto le possibili cause elencate in Si stanno verificando problemi di caricamento dei dati in e da Amazon S3?. Se nessuno di questi problemi è la causa, considera le seguenti opzioni specifiche di Hive.

- Assicurati di utilizzare la versione più recente di Hive che contiene tutte le patch correnti e le correzioni dei bug che possono risolvere il problema. Per ulteriori informazioni, consulta Apache Hive.
- L'uso di `INSERT OVERWRITE` richiede l'elenco dei contenuti della cartella o del bucket Amazon S3. Questa è un'operazione costosa. Se possibile, elimina manualmente il percorso anziché lasciare che sia Hive a elencare ed eliminare gli oggetti esistenti.

- Con le versioni di Amazon EMR precedenti alla 5.0, è possibile utilizzare il comando seguente in HiveQL per memorizzare anticipatamente nella cache i risultati di un'operazione di elenco Amazon S3 localmente sul cluster:

```
set hive.optimize.s3.query=true;
```

- Ove possibile, utilizza partizioni statiche.
- In alcune versioni di Hive e Amazon EMR, potrebbe non essere possibile utilizzare ALTER TABLE (ALTERA TABELLA) perché la tabella è archiviata in un percorso diverso da quello previsto da Hive. La soluzione è aggiungere o aggiornare seguendo in `/home/hadoop/conf/core-site.xml`:

```
<property>
    <name>fs.s3n.endpoint</name>
    <value>s3.amazonaws.com</value>
</property>
```

Errori VPC

Gli errori indicati di seguito sono comuni nella configurazione VPC in Amazon EMR.

Argomenti

Configurazione sottorete non valida

Nella pagina Cluster Details (Dettagli cluster), nel campo Status (Stato), verrà visualizzato un errore simile al seguente:

`The subnet configuration was invalid: Cannot find route to InternetGateway in main RouteTable` *`rtb-id`* `for vpc` *`vpc-id`*.

Per risolvere questo problema, devi creare un gateway Internet e collegarlo al tuo VPC. Per ulteriori informazioni, consulta la pagina relativa all'Aggiunta di un gateway Internet al VPC.

In alternativa, verifica di aver configurato il tuo VPC con Enable DNS resolution (Abilita risoluzione DNS) e Enable DNS hostname support (Abilita supporto nome host DNS). Per ulteriori informazioni, consulta Utilizzo del DNS con il tuo VPC.

Set opzioni DHCP mancante

Riscontri un errore di fase nel log di sistema del cluster (syslog) simile al seguente:

```
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.
```

oppure

```
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.
```

Per risolvere il problema, devi configurare un VPC che include un set di opzioni DHCP con parametri impostati sui seguenti valori:

> **Note**
>
> Se utilizzi la Regione AWS GovCloud (Stati Uniti occidentali), imposta il nome di dominio su **`us-gov-west-1.compute.internal`** anziché sul valore utilizzato nell'esempio seguente.

- domain-name = **`ec2.internal`**

  Utilizza **`ec2.internal`** se la regione è Stati Uniti orientali (Virginia settentrionale). Per le altre regioni, utilizza *`region-name`***`.compute.internal`**. Ad esempio, in us-west-2, utilizza domain-name=**`us-west-2.compute.internal`**.
- domain-name-servers = **`AmazonProvidedDNS`**

Per ulteriori informazioni, consulta la pagina relativa ai Set di opzioni DHCP.

Errori di autorizzazioni

Un errore nel log `stderr` per una fase indica che una risorsa Amazon S3 non dispone delle autorizzazioni appropriate. Si tratta di un errore 403 simile a questo:

```
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
```

Se ActionOnFailure è impostato su `TERMINATE_JOB_FLOW`, di conseguenza il cluster verrebbe chiuso con lo stato `SHUTDOWN_COMPLETED_WITH_ERRORS`.

Alcuni modi per risolvere questo problema sono:

- Se utilizzi una policy di bucket Amazon S3 all'interno di un VPC, assicurati di concedere l'accesso a tutti i bucket creando un endpoint VPC e selezionando Allow all (Consenti tutti) in corrispondenza dell'opzione Policy durante la creazione dell'endpoint.
- Assicurati che le policy associate alle risorse S3 includano il VPC in cui avvii il cluster.
- Prova a eseguire il seguente comando dal cluster per verificare che sia possibile accedere al bucket

  ```
  hadoop fs -copyToLocal s3://path-to-bucket /tmp/
  ```

- Per ottenere informazioni di debug più specifiche, imposta il parametro `log4j.logger.org.apache.http.wire` su `DEBUG` nel file `/home/hadoop/conf/log4j.properties` nel cluster. Puoi controllare il file di log `stderr` dopo aver provato ad accedere al bucket dal cluster. Il file di log fornirò informazioni più dettagliate:

  ```
  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]"
  ```

Errori che generano **`START_FAILED`**

Prima dell'AMI 3.7.0, per i VPC in cui il nome host è specificato, Amazon EMR mappa i nomi host interni della sottorete con gli indirizzi di dominio personalizzati come segue:

`ip-X.X.X.X.customdomain.com.tld`. Ad esempio, se il nome host fosse `ip-10.0.0.10` e il VPC avesse l'opzione del nome di dominio impostata su customdomain.com, il nome host risultante mappato da Amazon EMR sarebbe `ip-10.0.1.0.customdomain.com`. Viene aggiunta una voce in `/etc/hosts` per risolvere il nome host in 10.0.0.10. Questo comportamento è stato modificato con l'AMI 3.7.0 e ora Amazon EMR mantiene la configurazione DHCP del VPC in modo completo. In passato, i clienti potevano anche specificare la mappatura del nome host tramite un'operazione di bootstrap.

Se preferisci mantenere questo comportamento, devi fornire la configurazione di risoluzione per DNS e inoltro necessaria per il dominio personalizzato.

`Terminated with errors` per il cluster e mancato avvio di NameNode

Quando avvii un cluster EMR in un VPC che utilizza un nome di dominio DNS personalizzato, potrebbero verificarsi errori sul cluster con il seguente messaggio nella console:

```
Terminated with errors  On the master instance(instance-id), bootstrap action 1
 returned a  non-zero return code
```

L'errore è causato dal mancato avvio di NameNode. Di conseguenza, i log NameNode riporteranno il seguente errore, il cui URI Amazon S3 ha questo 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)
```

La causa di questo errore è una potenziale situazione problematica, in cui un'istanza EC2 può avere più set di nomi di dominio completo all'avvio di cluster EMR in un VPC, con presenza sia di un server DNS fornito da AWS, sia di un server DNS personalizzato fornito dall'utente. Se il server DNS fornito dall'utente non prevede alcun record puntatore (PTR) per eventuali record A utilizzati per designare i nodi in un cluster EMR, il cluster non potrà avviarsi se configurato in questo modo. La soluzione è aggiungere 1 record PTR per ogni record A creato all'avvio di un'istanza EC2 nelle sottoreti del VPC.

Errori relativi ai cluster di streaming

In genere, puoi individuare la causa di un errore di streaming in un file `syslog`. Un collegamento a questo file è visualizzato nel riquadro Steps (Fasi).

Gli errori esposti di seguito sono comuni ai cluster di streaming.

Argomenti

- I dati sono inviati al mappatore in un formato errato?
- Si è verificato il timeout dello script?
- Stai passando argomenti di streaming non validi?
- Lo script è terminato con un errore?

I dati sono inviati al mappatore in un formato errato?

Per determinare se è il caso, cerca un messaggio di errore nel file `syslog` di un tentativo di attività nei log dei tentativi di attività. Per ulteriori informazioni, consulta Visualizzare file di log di .

Si è verificato il timeout dello script?

Il timeout predefinito per uno script mappatore o riduttore è di 600 secondi. Se lo script richiede più tempo, il tentativo di attività non riuscirà. Per determinare se si è verificato tale problema, controlla il file `syslog` di un tentativo di attività non riuscito nei log dei tentativi di attività. Per ulteriori informazioni, consulta Visualizzare file di log di .

Puoi modificare il limite di tempo impostando un nuovo valore per l'impostazione di configurazione `mapred.task.timeout`. Questa impostazione specifica il numero di millisecondi dopo i quali Amazon EMR terminerà un'attività che non ha letto l'input, scritto l'output o aggiornato la stringa di stato. Puoi aggiornare questo valore passando un ulteriore argomento di streaming `-jobconf mapred.task.timeout=800000`.

Stai passando argomenti di streaming non validi?

Lo streaming di Hadoop supporta solo gli argomenti elencati di seguito. Se passi degli argomenti che non sono tra questi, si verificherà un errore nel cluster.

```
-blockAutoGenerateCacheFiles
-cacheArchive
-cacheFile
-cmdenv
-combiner
-debug
-input
-inputformat
-inputreader
-jobconf
-mapper
-numReduceTasks
-output
-outputformat
-partitioner
-reducer
-verbose
```

Inoltre, lo streaming Hadoop riconosce solo gli argomenti passati utilizzando la sintassi Java, ovvero preceduti da un solo trattino. Se passi argomenti preceduti da un doppio trattino, si verificherà un errore nel cluster.

Lo script è terminato con un errore?

Se lo script mappatore o riduttore termina con un errore, puoi individuare l'errore nel file `stderr` relativo al tentativo di attività non riuscito nei log dei tentativi di attività. Per ulteriori informazioni, consulta Visualizzare file di log di .

Errori del cluster JAR personalizzato

I seguenti errori sono comuni ai cluster JAR personalizzati.

Argomenti

Il cluster JAR genera un'eccezione prima di creare un processo?

Se il programma principale del cluster JAR personalizzato genera un'eccezione durante la creazione del processo Hadoop, si consiglia di cercare nel file `syslog` del log delle fasi. Per ulteriori informazioni, consulta Visualizzare file di log di .

Il cluster JAR genera un errore all'interno di un'attività di mappatura?

Se il cluster JAR personalizzato e il mappatore generano un'eccezione durante l'elaborazione dei dati di input, si consiglia di cercare nel file `syslog` dei log del tentativo di attività. Per ulteriori informazioni, consulta Visualizzare file di log di .

Errori della Regione AWS GovCloud (Stati Uniti occidentali)

La Regione AWS GovCloud (Stati Uniti occidentali) è diversa dalle altre in termini di sicurezza, configurazione e impostazioni predefinite. Di conseguenza, puoi avvalerti della seguente checklist per risolvere gli eventuali errori di Amazon EMR specifici della Regione AWS GovCloud (Stati Uniti occidentali), prima di utilizzare raccomandazioni di natura più generale per la risoluzione dei problemi.

- Verifica che i ruoli IAM siano configurati correttamente. Per ulteriori informazioni, consulta Configurazione dei ruoli di servizio IAM per le autorizzazioni di Amazon EMR per i servizi e risorse AWS.
- Assicurarsi che nella configurazione VPC siano correttamente impostati i parametri relativi a risoluzione/supporto nome host DNS, Internet gateway e set di opzioni DHCP. Per ulteriori informazioni, consulta Errori VPC.

Se questa procedura non risolve il problema, continua con la procedura per la risoluzione degli errori più comuni in Amazon EMR. Per ulteriori informazioni, consulta Errori comuni in Amazon EMR.

Ricerca di un cluster mancante

Se il cluster non è presente nell'elenco delle console o nell'API `ListClusters`, controlla quanto segue:

- Verifica che l'età del cluster dal momento del completamento sia inferiore a due mesi. Amazon EMR conserva le informazioni sui metadati relativi ai cluster completati per due mesi senza alcun costo aggiuntivo. Non puoi eliminare i cluster completati dalla console, ma Amazon EMR elimina automaticamente i cluster completati dopo due mesi.
- Conferma che disponi delle autorizzazioni di ruolo per visualizzare il cluster.
- Conferma che stai visualizzando la stessa Regione AWS in cui si trova il cluster.

Risoluzione dei problemi di un cluster con errori

In questa sezione viene descritto il processo di risoluzione dei problemi di un cluster in cui si sono verificati errori. Ciò significa che il cluster è stato terminato con un codice di errore.

> **Note**
>
> Quando un cluster EMR termina con un errore, le API `DescribeCluster` e `ListClusters` restituiscono un codice di errore e un messaggio di errore. Per alcuni errori del cluster, l'array di dati `ErrorDetail` può anche aiutarti a risolvere l'errore. Per ulteriori informazioni, consulta Codici di errore con informazioni ErrorDetail.

Se il cluster viene eseguito ma impiega molto tempo per restituire risultati, consulta Risoluzione dei problemi che rallentano un cluster.

Argomenti

Fase 1: Raccolta dei dati relativi al problema

Il primo passaggio per la risoluzione dei problemi di un cluster consiste nel raccogliere informazioni su ciò che non ha funzionato, nonché sullo stato corrente e sulla configurazione del cluster. Queste informazioni verranno utilizzate nei passaggi seguenti per confermare o escludere possibili cause del problema.

Definizione del problema

Una chiara definizione del problema è il primo punto da cui iniziare. Alcune domande da porsi:

- Cosa mi aspettavo che succedesse? Che cos'è successo invece?
- Quando si è verificato questo problema per la prima volta? Quante volte è successo da allora?
- È cambiato qualcosa nel modo in cui configuro o eseguo il cluster?

Dettagli del cluster

I seguenti dettagli del cluster sono utili per individuare i problemi. Per ulteriori informazioni su come raccogliere tali informazioni, consulta Visualizzazione dello stato e dei dettagli di un cluster.

- Identificatore del cluster (chiamato anche identificatore del flusso di lavoro).
- Regione AWS e zona di disponibilità in cui è stato avviato il cluster.
- Stato del cluster, inclusi i dettagli dell'ultima modifica dello stato.
- Tipo e numero di istanze EC2 specificate per i nodi master, principali e attività.

Fase 2: Controllo dell'ambiente

Amazon EMR funziona come parte di un ecosistema di servizi Web e software open source. Fattori che influenzano tali dipendenze possono influire sulle prestazioni di Amazon EMR.

Argomenti

- Verifica della presenza di interruzioni del servizio
- Controllo dei limiti di utilizzo
- Controllo della versione di rilascio
- Controllo della configurazione della sottorete Amazon VPC

Verifica della presenza di interruzioni del servizio

Amazon EMR utilizza diversi servizi Amazon Web Services al suo interno. Esegue server virtuali su Amazon EC2, memorizza dati e script su Amazon S3 e segnala i parametri a CloudWatch. Gli eventi che disturbano questi servizi sono rari, ma quando si verificano possono causare problemi in Amazon EMR.

Prima di proseguire, controllare il Service Health Dashboard. Controlla la Regione in cui è stato avviato il cluster per verificare se esistono eventi di interruzione in uno di questi servizi.

Controllo dei limiti di utilizzo

Se avvii un cluster di grandi dimensioni, hai avviato più cluster contemporaneamente o sei un utente che condivide un Account AWS con altri utenti, il cluster potrebbe avere esito negativo poiché hai superato un limite del servizio AWS.

Amazon EC2 limita il numero di istanze server virtuali in esecuzione su una singola Regione AWS a 20 istanze su richiesta o riservate. Se avvii un cluster con più di 20 nodi o avvii un cluster che fa sì che il numero totale di istanze EC2 attive sul tuo Account AWS superi 20, il cluster non sarà in grado di avviare tutte le istanze EC2 necessarie e potrebbe avere esito negativo. Quando ciò si verifica, Amazon EMR restituisce un errore `EC2 QUOTA EXCEEDED`. Puoi richiedere ad AWS l'aumento del numero di istanze EC2 eseguibili sul tuo account inviando una domanda di Request to Increase Amazon EC2 Instance Limit (Richiesta di aumento del limite di istanze Amazon EC2).

Un altro fattore che può causare il superamento dei limiti di utilizzo è il ritardo tra il momento in cui un cluster viene terminato e quello in cui rilascia tutte le sue risorse. A seconda della configurazione, potrebbero essere necessari tra i 5 e i 20 minuti affinché un cluster venga terminato e le risorse allocate rilasciate. Se visualizzi un errore `EC2 QUOTA EXCEEDED` al tentativo di avvio di un cluster, potrebbe essere dovuto al rilascio non ancora avvenuto delle risorse di un cluster terminato recentemente. In questo caso, puoi richiedere l'aumento della quota Amazon EC2 o attendere 20 minuti prima di riavviare il cluster.

Amazon S3 limita il numero di bucket creati su un account a 100. Se il cluster crea un nuovo bucket che supera questo limite, la creazione del bucket avrà esito negativo e potrebbe causare l'esito negativo del cluster.

Controllo della versione di rilascio

Confronta l'etichetta di rilascio utilizzata per avviare il cluster con il rilascio di Amazon EMR più recente. Ogni rilascio di Amazon EMR include miglioramenti quali nuove applicazioni, caratteristiche,

patch e correzioni di bug. Il problema che si presenta nel cluster potrebbe essere già stato corretto nella versione del rilascio più recente. Se possibile, riavvia il cluster utilizzando la versione più recente.

Controllo della configurazione della sottorete Amazon VPC

Se il cluster è stato avviato in una sottorete Amazon VPC, la sottorete deve essere configurata come descritto in Configurazione delle reti. Occorre inoltre verificare che la sottorete in cui si avvia il cluster disponga di indirizzi IP elastici liberi sufficienti per assegnarne uno a ciascun nodo del cluster.

Fase 3: Esame dell'ultima modifica dello stato

L'ultima modifica dello stato fornisce informazioni su cosa si è verificato l'ultima volta che lo stato del cluster è cambiato: Questo spesso può fornire un'indicazione su cosa non ha funzionato quando lo stato del cluster cambia in `FAILED`. Ad esempio, se si avvia un cluster di streaming e si specifica un percorso di output che esiste già in Amazon S3, il cluster restituisce un errore con un'ultima modifica di stato di "Streaming output directory already exists (La directory di output dello streaming esiste già)".

Puoi individuare il valore dell'ultima modifica dello stato dalla console visualizzando il riquadro dei dettagli per il cluster, dalla CLI utilizzando gli argomenti `list-steps` o `describe-cluster`, oppure dall'API utilizzando le azioni `DescribeCluster` e `ListSteps`. Per ulteriori informazioni, consulta Visualizzazione dello stato e dei dettagli di un cluster.

Fase 4: Esame dei file di log

Il passaggio successivo consiste nell'esaminare i file di log per individuare un codice di errore o altra indicazione relativa al problema riscontrato dal cluster. Per informazioni sui file di log disponibili, su dove trovarli e su come visualizzarli, consulta Visualizzare file di log di .

Potrebbe essere necessaria qualche indagine per determinare l'accaduto. Hadoop esegue il lavoro dei processi nei tentativi di attività su vari nodi del cluster. Amazon EMR può avviare tentativi di attività speculative, terminando gli altri tentativi di attività che non vengono completati per primi. Questo genera un'attività significativa che viene registrata nei file di log del controller, stderr e syslog mentre si verifica. Inoltre, anche se più tentativi di attività vengono eseguiti contemporaneamente, un file di log può visualizzare i risultati solo in modo lineare.

Inizia controllando i log delle operazioni di bootstrap per individuare errori o modifiche impreviste alla configurazione durante l'avvio del cluster. Successivamente, cerca nei log di fase per identificare i

processi Hadoop avviati come parte di una fase con errori. Esamina i log dei processi Hadoop per identificare i tentativi di attività non riusciti. Il log dei tentativi di attività conterrà dettagli su ciò che ha causato l'esito negativo di un tentativo di attività.

Le sezioni seguenti descrivono come utilizzare i vari file di log per identificare gli errori nel cluster.

Controllo dei log delle operazioni di bootstrap

Le operazioni di bootstrap eseguono script sul cluster mentre quest'ultimo viene avviato. Vengono comunemente utilizzate per installare software aggiuntivi nel cluster o per modificare le impostazioni di configurazione rispetto ai valori predefiniti. Il controllo di questi log può fornire informazioni dettagliate sugli errori verificatisi durante la configurazione del cluster e sulle modifiche alle impostazioni di configurazione che potrebbero influire sulle prestazioni.

Controllo dei log della fase

Esistono quattro tipi di log di fase.

- controller: contiene i file generati da Amazon EMR che derivano da errori riscontrati durante il tentativo di eseguire la fase. Se la fase ha esito negativo durante il caricamento, è possibile trovare la traccia di stack in questo log. Gli errori che si verificano durante il caricamento o l'accesso all'applicazione sono spesso descritti nel log, mentre mancano gli errori relativi ai file del mappatore.
- stderr: contiene messaggi di errore che si sono presentati durante l'elaborazione della fase. Gli errori di caricamento delle applicazioni sono spesso descritti in questo log. Talvolta, questo log contiene una traccia di stack.
- stdout: contiene lo stato generato dagli eseguibili del mappatore e del riduttore. Gli errori di caricamento delle applicazioni sono spesso descritti in questo log. Talvolta, questo log contiene messaggi di errore dell'applicazione.
- syslog: contiene log generati da software non Amazon, come Apache e Hadoop. Gli errori di streaming sono spesso descritti in questo log.

Controlla stderr per rilevare errori evidenti. Se stderr visualizza un breve elenco di errori, la fase è giunta a un rapido arresto generando un errore. Questo è spesso causato da un errore nelle applicazioni di mappatore e riduttore eseguite nel cluster.

Esamina le ultime righe di controller e syslog per cercare avvisi di errore o guasti. Segui tutte le notifiche relative alle attività non riuscite, in particolare "Job Failed (Processo non riuscito)".

Controllo dei log del tentativo di attività

Se l'analisi precedente dei log di fase ha rilevato una o più attività non riuscite, esamina i log dei tentativi di attività corrispondenti per ottenere informazioni più dettagliate sugli errori.

Fase 5: Test dettagliato del cluster

Una tecnica utile quando cerchi di rintracciare l'origine di un errore è riavviare il cluster e inviare le fasi una alla volta. Ciò consente di controllare i risultati di ciascuna fase prima di elaborare quella successiva e offre l'opportunità di correggere ed eseguire nuovamente una fase non riuscita. Inoltre ha il vantaggio di poter caricare i dati di input una sola volta.

Per eseguire il test dettagliato di un cluster

1. Avviare un nuovo cluster con keep-alive e la protezione da cessazione abilitati. Keep-alive consente di mantenere il cluster in esecuzione dopo che tutte le fasi in sospeso sono state elaborate. La protezione da cessazione impedisce l'arresto di un cluster in caso di errore. Per ulteriori informazioni, consulta Configurazione di un cluster per continuare o terminare dopo l'esecuzione della fase e Utilizzo della protezione da cessazione.
2. Inviare una fase al cluster. Per ulteriori informazioni, consulta Invio di lavoro a un cluster.
3. Al termine dell'elaborazione della fase, verificare la presenza di errori nei file di log della fase. Per ulteriori informazioni, consulta Fase 4: Esame dei file di log. Per individuare in modo veloce questi file di log, collegati al nodo master e visualizza qui i file di log. I file di log vengono visualizzati solo se la fase rimane in esecuzione per qualche tempo, termina o non riesce.
4. Se la fase viene completata senza errori, eseguire la fase successiva. In caso di errori, analizzare l'errore nei file di log. Se si è verificato un errore nel codice, apportare la correzione ed eseguire nuovamente la fase. Continuare finché tutte le fasi non vengono eseguite senza errore.
5. Al termine del debug del cluster, se desideri terminarlo, devi farlo manualmente. Questo è necessario perché il cluster è stato avviato con la protezione da cessazione abilitata. Per ulteriori informazioni, consulta Utilizzo della protezione da cessazione.

Risoluzione dei problemi che rallentano un cluster

Questa sezione descrive la procedura di risoluzione dei problemi relativi a un cluster in esecuzione che impiega troppo tempo per restituire i risultati. Per ulteriori informazioni sulle soluzioni da adottare nel caso in cui il cluster venga terminato con un codice di errore, consulta Risoluzione dei problemi di un cluster con errori

Amazon EMR ti consente di specificare il numero e il tipo di istanze nel cluster. Queste specifiche sono quelle che influiscono maggiormente sulla velocità di elaborazione dei dati. Una soluzione potrebbe quindi essere quella di eseguire di nuovo il cluster, questa volta specificando le istanze EC2 con maggiori risorse oppure specificando un maggior numero di istanze nel cluster. Per ulteriori informazioni, consulta Configurazione di hardware e reti cluster.

I seguenti argomenti illustrano la procedura di identificazione delle altre cause che possono rallentare un cluster.

Argomenti

Fase 1: Raccolta dei dati relativi al problema

Il primo passaggio per la risoluzione dei problemi di un cluster consiste nel raccogliere informazioni su ciò che non ha funzionato, nonché sullo stato corrente e sulla configurazione del cluster. Queste informazioni verranno utilizzate nei passaggi seguenti per confermare o escludere possibili cause del problema.

Definizione del problema

Una chiara definizione del problema è il primo punto da cui iniziare. Alcune domande da porsi:

- Cosa mi aspettavo che succedesse? Che cos'è successo invece?
- Quando si è verificato questo problema per la prima volta? Quante volte è successo da allora?
- È cambiato qualcosa nel modo in cui configuro o eseguo il cluster?

Dettagli del cluster

I seguenti dettagli del cluster sono utili per individuare i problemi. Per ulteriori informazioni su come raccogliere tali informazioni, consulta Visualizzazione dello stato e dei dettagli di un cluster.

- Identificatore del cluster (chiamato anche identificatore del flusso di lavoro).
- Regione AWS e zona di disponibilità in cui è stato avviato il cluster.
- Stato del cluster, inclusi i dettagli dell'ultima modifica dello stato.
- Tipo e numero di istanze EC2 specificate per i nodi master, principali e attività.

Fase 2: Controllo dell'ambiente

Argomenti

Verifica della presenza di interruzioni del servizio

Amazon EMR utilizza diversi servizi Amazon Web Services al suo interno. Esegue server virtuali su Amazon EC2, memorizza dati e script su Amazon S3 e segnala i parametri a CloudWatch. Gli eventi che disturbano questi servizi sono rari, ma quando si verificano possono causare problemi in Amazon EMR.

Prima di proseguire, controllare il Service Health Dashboard. Controlla la Regione in cui è stato avviato il cluster per verificare se esistono eventi di interruzione in uno di questi servizi.

Controllo dei limiti di utilizzo

Se avvii un cluster di grandi dimensioni, hai avviato più cluster contemporaneamente o sei un utente che condivide un Account AWS con altri utenti, il cluster potrebbe avere esito negativo poiché hai superato un limite del servizio AWS.

Amazon EC2 limita il numero di istanze server virtuali in esecuzione su una singola Regione AWS a 20 istanze su richiesta o riservate. Se avvii un cluster con più di 20 nodi o avvii un cluster che fa sì che il numero totale di istanze EC2 attive sul tuo Account AWS superi 20, il cluster non sarà in grado

di avviare tutte le istanze EC2 necessarie e potrebbe avere esito negativo. Quando ciò si verifica, Amazon EMR restituisce un errore `EC2 QUOTA EXCEEDED`. Puoi richiedere ad AWS l'aumento del numero di istanze EC2 eseguibili sul tuo account inviando una domanda di Request to Increase Amazon EC2 Instance Limit (Richiesta di aumento del limite di istanze Amazon EC2).

Un altro fattore che può causare il superamento dei limiti di utilizzo è il ritardo tra il momento in cui un cluster viene terminato e quello in cui rilascia tutte le sue risorse. A seconda della configurazione, potrebbero essere necessari tra i 5 e i 20 minuti affinché un cluster venga terminato e le risorse allocate rilasciate. Se visualizzi un errore `EC2 QUOTA EXCEEDED` al tentativo di avvio di un cluster, potrebbe essere dovuto al rilascio non ancora avvenuto delle risorse di un cluster terminato recentemente. In questo caso, puoi richiedere l'aumento della quota Amazon EC2 o attendere 20 minuti prima di riavviare il cluster.

Amazon S3 limita il numero di bucket creati su un account a 100. Se il cluster crea un nuovo bucket che supera questo limite, la creazione del bucket avrà esito negativo e potrebbe causare l'esito negativo del cluster.

Controllo della configurazione della sottorete Amazon VPC

Se il cluster è stato avviato in una sottorete Amazon VPC, la sottorete deve essere configurata come descritto in Configurazione delle reti. Occorre inoltre verificare che la sottorete in cui si avvia il cluster disponga di indirizzi IP elastici liberi sufficienti per assegnarne uno a ciascun nodo del cluster.

Riavvio del cluster

Il rallentamento dell'elaborazione può essere causato da una condizione transitoria. Valuta la possibilità di terminare e riavviare il cluster per verificare se tale operazione migliora le prestazioni.

Fase 3: Analisi dei file di log

Il passaggio successivo consiste nell'esaminare i file di log per individuare un codice di errore o altra indicazione relativa al problema riscontrato dal cluster. Per informazioni sui file di log disponibili, su dove trovarli e su come visualizzarli, consulta Visualizzare file di log di .

Potrebbe essere necessaria qualche indagine per determinare l'accaduto. Hadoop esegue il lavoro dei processi nei tentativi di attività su vari nodi del cluster. Amazon EMR può avviare tentativi di attività speculative, terminando gli altri tentativi di attività che non vengono completati per primi. Questo genera un'attività significativa che viene registrata nei file di log del controller, stderr e syslog mentre si verifica. Inoltre, anche se più tentativi di attività vengono eseguiti contemporaneamente, un file di log può visualizzare i risultati solo in modo lineare.

Inizia controllando i log delle operazioni di bootstrap per individuare errori o modifiche impreviste alla configurazione durante l'avvio del cluster. Successivamente, cerca nei log di fase per identificare i processi Hadoop avviati come parte di una fase con errori. Esamina i log dei processi Hadoop per identificare i tentativi di attività non riusciti. Il log dei tentativi di attività conterrà dettagli su ciò che ha causato l'esito negativo di un tentativo di attività.

Le sezioni seguenti descrivono come utilizzare i vari file di log per identificare gli errori nel cluster.

Controllo dei log delle operazioni di bootstrap

Le operazioni di bootstrap eseguono script sul cluster mentre quest'ultimo viene avviato. Vengono comunemente utilizzate per installare software aggiuntivi nel cluster o per modificare le impostazioni di configurazione rispetto ai valori predefiniti. Il controllo di questi log può fornire informazioni dettagliate sugli errori verificatisi durante la configurazione del cluster e sulle modifiche alle impostazioni di configurazione che potrebbero influire sulle prestazioni.

Controllo dei log della fase

Esistono quattro tipi di log di fase.

- controller: contiene i file generati da Amazon EMR che derivano da errori riscontrati durante il tentativo di eseguire la fase. Se la fase ha esito negativo durante il caricamento, è possibile trovare la traccia di stack in questo log. Gli errori che si verificano durante il caricamento o l'accesso all'applicazione sono spesso descritti nel log, mentre mancano gli errori relativi ai file del mappatore.
- stderr: contiene messaggi di errore che si sono presentati durante l'elaborazione della fase. Gli errori di caricamento delle applicazioni sono spesso descritti in questo log. Talvolta, questo log contiene una traccia di stack.
- stdout: contiene lo stato generato dagli eseguibili del mappatore e del riduttore. Gli errori di caricamento delle applicazioni sono spesso descritti in questo log. Talvolta, questo log contiene messaggi di errore dell'applicazione.
- syslog: contiene log generati da software non Amazon, come Apache e Hadoop. Gli errori di streaming sono spesso descritti in questo log.

Controlla stderr per rilevare errori evidenti. Se stderr visualizza un breve elenco di errori, la fase è giunta a un rapido arresto generando un errore. Questo è spesso causato da un errore nelle applicazioni di mappatore e riduttore eseguite nel cluster.

Esamina le ultime righe di controller e syslog per cercare avvisi di errore o guasti. Segui tutte le notifiche relative alle attività non riuscite, in particolare "Job Failed (Processo non riuscito)".

Controllo dei log del tentativo di attività

Se l'analisi precedente dei log di fase ha rilevato una o più attività non riuscite, esamina i log dei tentativi di attività corrispondenti per ottenere informazioni più dettagliate sugli errori.

Verifica dei log dei daemon Hadoop

In rari casi, Hadoop stesso potrebbe non funzionare. Per comprendere la natura del problema, è necessario esaminare i log di Hadoop. Questi log si trovano all'indirizzo `/var/log/hadoop/` su ogni nodo.

È possibile utilizzare i log di JobTracker per mappare un tentativo di attività non riuscito al nodo su cui tale attività è stata eseguita. Una volta rilevato il nodo associato al tentativo di attività, è possibile verificare lo stato dell'istanza EC2 che ospita tale nodo per verificare se si sono verificati problemi quali l'esaurimento della CPU o della memoria.

Fase 4: Verifica dell'integrità del cluster e delle istanze

Un cluster Amazon EMR è composto di nodi in esecuzione su istanze Amazon EC2. Se tali istanze sono limitate dalle risorse (ad esempio, CPU o memoria esaurita), presentano problemi di connettività di rete o vengono terminate, la velocità di elaborazione del cluster diminuisce.

In un cluster sono presenti fino a tre tipi di nodi:

- nodo master: gestisce il cluster. Eventuali problemi di prestazioni nel nodo si ripercuotono all'intero cluster.
- nodi principali: elaborano le attività map-reduce e gestiscono il file system distribuito Hadoop (HDFS). Eventuali problemi di prestazioni in uno di questi nodi possono provocare il rallentamento delle operazioni HDFS nonché dell'elaborazione map-reduce. Puoi aggiungere dei nodi principali a un cluster per migliorare le prestazioni, ma non puoi rimuoverne. Per ulteriori informazioni, consulta Ridimensionamento manuale di un cluster in esecuzione.
- nodi attività: elaborano le attività map-reduce. Sono esclusivamente risorse di calcolo che non archiviano dati. Puoi aggiungere nodi di task a un cluster per migliorare le prestazioni oppure rimuovere quelli non necessari. Per ulteriori informazioni, consulta Ridimensionamento manuale di un cluster in esecuzione.

Quando esamini lo stato di un cluster, devi considerare le prestazioni globali del cluster nonché quelle delle singole istanze. Sono disponibili vari strumenti che puoi utilizzare:

Verifica dell'integrità del cluster con CloudWatch

Ogni cluster Amazon EMR trasmette dei parametri a CloudWatch. Questi parametri forniscono informazioni di riepilogo sulle prestazioni del cluster, ad esempio il carico totale, l'utilizzo di HDFS, le attività in esecuzione, le attività rimanenti, i blocchi danneggiati e altro ancora. I parametri di CloudWatch forniscono quindi informazioni complete sullo stato del cluster e possono risultare utili per identificare il problema all'origine del rallentamento dell'elaborazione. Oltre a utilizzare CloudWatch per analizzare un problema di prestazioni, puoi impostare allarmi mediante i quali CloudWatch segnala un problema di prestazioni. Per ulteriori informazioni, consulta Monitoraggio di parametri di Amazon EMR con CloudWatch.

Verifica dell'integrità del processo e di HDFS

Utilizzare Cronologia applicazione nella pagina dei dettagli del cluster per visualizzare informazioni sull'applicazione YARN. Per alcune applicazioni, puoi esaminare ulteriori dettagli e accedere direttamente ai log. Ciò è particolarmente utile per le applicazioni Spark. Per ulteriori informazioni, consulta Visualizzazione della cronologia dell'applicazione.

Hadoop fornisce una serie di interfacce Web che puoi utilizzare per visualizzare le informazioni. Per ulteriori informazioni su come accedere a queste interfacce Web, consulta Visualizzazione di interfacce Web ospitate su cluster Amazon EMR.

- JobTracker: fornisce informazioni sull'avanzamento del processo elaborato dal cluster. Puoi utilizzare questa interfaccia per identificare quando un processo si blocca.
- HDFS NameNode: fornisce informazioni sulla percentuale di utilizzo di HDFS e sullo spazio disponibile in ogni nodo. Puoi utilizzare questa interfaccia per identificare quando HDFS è limitato dalle risorse e richiede capacità supplementare.
- TaskTracker: fornisce informazioni sulle attività del processo elaborato dal cluster. Puoi utilizzare questa interfaccia per identificare quando un'attività si blocca.

Verifica dell'integrità delle istanze con Amazon EC2

Un altro modo di ottenere informazioni sullo stato delle istanze nel cluster consiste nell'utilizzare la console di Amazon EC2. Poiché ogni nodo nel cluster viene eseguito su un'istanza EC2, puoi

utilizzare gli strumenti forniti da Amazon EC2 per verificarne lo stato. Per ulteriori informazioni, consulta Visualizzazione di istanze cluster in Amazon EC2.

Fase 5: Verifica della presenza di gruppi sospesi

Un gruppo di istanze è considerato sospeso quando si verificano troppi errori durante il tentativo di avvio di nodi. Ad esempio, in caso di errori ripetuti in nuovi nodi durante l'esecuzione di operazioni di bootstrap, il gruppo di istanze passerà, dopo un determinato periodo di tempo, allo stato `SUSPENDED` anziché tentare reiteratamente di effettuare il provisioning di nuovi nodi.

L'avvio di un nodo potrebbe non riuscire se:

- Hadoop o il cluster è in qualche modo danneggiato e non accetta un nuovo nodo nel cluster.
- Un'operazione di bootstrap non riesce sul nuovo nodo.
- Il nodo non funziona correttamente e la sessione Hadoop non riesce.

Se lo stato di un gruppo di istanze è `SUSPENDED` e quello del cluster è `WAITING`, puoi aggiungere una fase di cluster per ripristinare il numero desiderato di nodi di task e principali. Con l'aggiunta di una fase l'elaborazione del cluster riprende e il gruppo di istanze ritorna allo stato `RUNNING`.

Per ulteriori informazioni su come ripristinare un cluster con stato sospeso, consulta Stato sospeso.

Fase 6: Controllo delle impostazioni di configurazione

Le impostazioni di configurazione specificano dettagli relativi all'esecuzione di un cluster, come il numero di nuovi tentativi per un'attività e la quantità di memoria disponibile per l'ordinamento. Quando avvii un cluster utilizzando Amazon EMR, oltre alle impostazioni di configurazione di Hadoop standard sono disponibili anche impostazioni specifiche di Amazon EMR. Le impostazioni di configurazione sono memorizzate nel nodo master del cluster. Puoi verificare le impostazioni di configurazione per assicurarti che il cluster disponga delle risorse necessarie per una corretta esecuzione.

Amazon EMR definisce le impostazioni di configurazione di Hadoop di default che utilizza per avviare un cluster. I valori sono basati sull'AMI e sul tipo di istanza specificato per il cluster. Puoi modificare i valori di default delle impostazioni di configurazione mediante un'operazione di bootstrap o specificando nuovi valori nei parametri di esecuzione dei processi. Per ulteriori informazioni, consulta Creazione di operazioni di bootstrap per l'installazione di software aggiuntivo. Per determinare se un'operazione di bootstrap ha modificato le impostazioni di configurazione, controlla i log delle operazioni di bootstrap.

Amazon EMR registra le impostazioni Hadoop utilizzate per eseguire ogni processo. I dati di log sono archiviati in un file denominato `job_job-id_conf.xml` nella directory `/mnt/var/log/hadoop/history/` del nodo master, dove *job id* viene sostituito con l'identificatore del processo. Se hai attivato l'archiviazione dei log, questi dati sono copiati in Amazon S3 nella cartella `logs/date/jobflow-id/jobs`, dove *date* è la data in cui è stato eseguito il processo e *jobflow-id* è l'identificatore del cluster.

Le seguenti impostazioni di configurazione di processi Hadoop sono particolarmente utili per la risoluzione dei problemi di prestazioni. Per ulteriori informazioni sulle impostazioni di configurazione di Hadoop e su come queste influiscono sul comportamento di Hadoop, visita la pagina all'indirizzo http://hadoop.apache.org/docs/.

Warning

1. L'impostazione di `dfs.replication` su 1 per i cluster con meno di quattro nodi può causare la perdita di dati HDFS in caso di disattivazione anche di un singolo nodo. Ti consigliamo di utilizzare un cluster con almeno quattro nodi principali per i carichi di lavoro di produzione.
2. Amazon EMR non consente ai cluster di dimensionare i nodi principali al di sotto di `dfs.replication`. Ad esempio, se `dfs.replication = 2`, il numero minimo di nodi principali è 2.
3. Quando utilizzi il dimensionamento gestito, il dimensionamento automatico o scegli di dimensionare manualmente il cluster, ti consigliamo di impostare `dfs.replication` su 2 o su un valore superiore.

Impostazione di configurazione	Descrizione
dfs.replication	Il numero di nodi HDFS in cui un singolo blocco (ad esempio un blocco di disco rigido) viene copiato per generare un ambiente di tipo RAID. Determina il numero di nodi HDFS che contengono una copia del blocco.
io.sort.mb	La memoria totale disponibile per l'ordinamento. Il valore deve essere 10x io.sort.factor. Questa impostazione può anche essere utilizzata per calcolare la memoria totale

Impostazione di configurazione	Descrizione
	utilizzata dal nodo di task moltiplicando io.sort.mb per mapred.tasktracker.ap.tasks.maximum.
io.sort.spill.percent	Utilizzata durante l'ordinamento. In tal caso il disco viene utilizzato in quanto la memoria di ordinamento assegnata è esaurita.
mapred.child.java.opts	Obsoleta. Utilizzare mapred.map.child.java.opts e mapred.reduce.child.java.opts. Le opzioni Java che TaskTracker utilizza all'avvio di una JVM in cui eseguire un'attività. "-Xmx" è un parametro comune per l'imposta zione della dimensione massima della memoria.
mapred.map.child.java.opts	Le opzioni Java che TaskTracker utilizza all'avvio di una JVM in cui eseguire un'attività di mappatura. "-Xmx" è un parametro comune per l'impostazione della dimensione massima dell'heap di memoria.
mapred.map.tasks.speculativ e.execution	Determina se i tentativi di attività di mappatura della stessa attività possono essere avviati in parallelo.
mapred.reduce.tasks.specula tive.execution	Determina se i tentativi di attività di riduzione della stessa attività possono essere avviati in parallelo.
mapred.map.max.attempts	Il numero massimo di tentativi per un'attività di mappatura . Se tutti i tentativi non riescono, l'attività di mappatura viene contrassegnata come non riuscita.
mapred.reduce.child.java.opts	Le opzioni Java che TaskTracker utilizza all'avvio di una JVM in cui eseguire un'attività di riduzione. "-Xmx" è un parametro comune per l'impostazione della dimensione massima dell'heap di memoria.
mapred.reduce.max.attempts	Il numero massimo di tentativi per un'attività di riduzione . Se tutti i tentativi non riescono, l'attività di mappatura viene contrassegnata come non riuscita.

Impostazione di configurazione	Descrizione
mapred.reduce.slowstart.completed.maps	Il numero di attività di mappatura che devono essere completate prima dell'avvio delle attività di riduzione. Se il tempo di attesa è troppo breve, è possibile che vengano generati errori "Too many fetch-failure (Troppi errori di recupero)" nei tentativi.
mapred.reuse.jvm.num.tasks	Un'attività viene eseguita in una singola JVM. Specifica il numero di attività che possono riutilizzare la stessa JVM.
mapred.tasktracker.map.tasks.maximum	Il numero massimo di attività che possono essere eseguite in parallelo per nodo di task durante la mappatura.
mapred.tasktracker.reduce.tasks.maximum	Il numero massimo di attività che possono essere eseguite in parallelo per nodo di task durante la riduzione.

Se le attività del cluster utilizzano molta memoria, è possibile migliorare le prestazioni utilizzando un minor numero di attività per nodo principale e riducendo la dimensione heap del job tracker.

Fase 7: Analisi dei dati di input

Esamina i dati di input. Sono distribuiti in modo uniforme tra i valori chiave? Se i dati sono ripartiti soprattutto su uno o soltanto alcuni valori chiave, il carico di elaborazione può essere mappato a un numero ridotto di nodi, mentre altri sono inattivi. Questa distribuzione squilibrata del lavoro può comportare tempi di elaborazione più elevati.

Un esempio di set di dati squilibrati sarebbe un cluster eseguito per ordinare alfabeticamente le parole, ma con un set di dati che contiene unicamente parole che iniziano con la lettera "a". Dopo la mappatura del lavoro, il nodo che elabora i valori che iniziano con "a" risulterebbe sovraccaricato, mentre i nodi che elaborano le parole che iniziano con altre lettere sarebbero inattivi.

Risoluzione dei problemi in un cluster Lake Formation

Questa sezione descrive il processo di risoluzione dei problemi comuni relativi all'utilizzo di Amazon EMR con AWS Lake Formation.

Accesso al data lake non consentito

Prima di poter analizzare ed elaborare i dati nel data lake, devi attivare esplicitamente il filtro dei dati sui cluster Amazon EMR. Quando l'accesso ai dati ha esito negativo, verrà visualizzato un messaggio `Access is not allowed` generico nell'output delle voci del notebook.

Per istruzioni sull'attivazione del filtro dei dati su Amazon EMR, consulta Attivazione del filtro dei dati su Amazon EMR nella Guida per gli sviluppatori di AWS Lake Formation.

Scadenza della sessione

Il timeout della sessione per gli EMR Notebooks e Zeppelin è controllato dall'impostazione `Maximum CLI/API session duration` del ruolo IAM per Lake Formation. Il valore predefinito per questa impostazione è un'ora. Quando si verifica un timeout di sessione, verrà visualizzato il seguente messaggio nell'output delle voci del notebook quando si tenta di eseguire i comandi SQL Spark.

```
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…
```

Per convalidare la sessione, aggiorna la pagina. Verrà richiesto di eseguire nuovamente l'autenticazione utilizzando il provider di identità e di essere reindirizzati al notebook. Puoi continuare a eseguire query dopo la riautenticazione.

Nessuna autorizzazione per l'utente sulla tabella richiesta

Quando tenti di accedere a una tabella a cui non hai accesso, visualizzerai la seguente eccezione nell'output delle voci del notebook quando tenti di eseguire i comandi SQL Spark.

```
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: …
```

Per accedere alla tabella, devi concedere l'accesso all'utente aggiornando le autorizzazioni associate a questa tabella in Lake Formation.

Interrogazione dei dati tra account condivisi con Lake Formation

Quando utilizzi Amazon EMR per accedere ai dati condivisi con te da un altro account, alcune librerie Spark tenteranno di chiamare l'operazione API `Glue:GetUserDefinedFunctions`. Poiché le versioni 1 e 2 delle autorizzazioni gestite da AWS RAM non supportano questa operazione, viene visualizzato il seguente messaggio di errore:

`"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"`

Per risolvere questo errore, l'amministratore del data lake che ha creato la condivisione di risorse deve aggiornare le autorizzazioni gestite da AWS RAM collegate alla condivisione di risorse. La versione 3 delle autorizzazioni gestite da AWS RAM consente ai responsabili di eseguire l'operazione `glue:GetUserDefinedFunctions`.

Se crei una nuova condivisione di risorse, Lake Formation applica la versione più recente dell'autorizzazione gestita da AWS RAM per impostazione predefinita senza richiedere alcuna azione da parte tua. Per abilitare l'accesso ai dati tra account per le condivisioni di risorse esistenti, è necessario aggiornare le autorizzazioni gestite da AWS RAM alla versione 3.

Puoi visualizzare le autorizzazioni AWS RAM assegnate alle risorse condivise con te in AWS RAM. Le autorizzazioni seguenti sono incluse nella versione 3:

```
Databases
  AWSRAMPermissionGlueDatabaseReadWriteForCatalog
  AWSRAMPermissionGlueDatabaseReadWrite

Tables
  AWSRAMPermissionGlueTableReadWriteForCatalog
  AWSRAMPermissionGlueTableReadWriteForDatabase

AllTables
  AWSRAMPermissionGlueAllTablesReadWriteForCatalog
  AWSRAMPermissionGlueAllTablesReadWriteForDatabase
```

Aggiornamento della versione delle autorizzazioni gestite da AWS RAM delle condivisioni di risorse esistenti

L'utente (amministratore del data lake) può aggiornare le autorizzazioni gestite da AWS RAM a una versione più recente seguendo le istruzioni riportate nella Guida per l'utente di AWS RAM oppure revocare tutte le autorizzazioni esistenti per il tipo di risorsa e concederle nuovamente. Se revochi le autorizzazioni, AWS RAM elimina la condivisione di risorse AWS RAM associata al tipo di risorsa. Quando concedi nuovamente le autorizzazioni, AWS RAM crea nuove condivisioni di risorse collegando la versione più recente delle autorizzazioni gestite da AWS RAM.

Inserimento, creazione e modifica di tabelle

L'inserimento, la creazione e la modifica di tabelle nei database protetti da policy Lake Formation non sono supportati. Quando esegui queste operazioni, visualizzerai la seguente eccezione nell'output delle voci del notebook quando tenti di eseguire i comandi SQL Spark:

```
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: …
```

Per ulteriori informazioni, consulta Limitazioni dell'integrazione di Amazon EMR con AWS Lake Formation.

Scrittura di applicazioni per l'avvio e la gestione di cluster

Argomenti

È possibile accedere alle funzionalità fornite dall'API di Amazon EMR chiamando funzioni wrapper in uno degli SDK AWS. Gli SDK AWS forniscono funzioni specifiche del linguaggio che eseguono il wrapping dell'API del servizio Web e semplificano la connessione al servizio Web, gestendo molti dettagli di connessione per conto dell'utente. Per ulteriori informazioni su come chiamare Amazon EMR utilizzando uno degli SDK, consulta Utilizzo di SDK per chiamare le API di Amazon EMR.

> **Important**
>
> Il tasso massimo di richieste per Amazon EMR è una richiesta ogni dieci secondi.

Esempio di codice sorgente Java Amazon EMR end-to-end

Gli sviluppatori possono chiamare l'API di Amazon EMR utilizzando codice Java personalizzato per eseguire le stesse operazioni possibili con la console Amazon EMR o la CLI. In questa sezione vengono illustrate le fasi end-to-end necessarie per installare AWS Toolkit for Eclipse ed eseguire un esempio di codice sorgente Java pienamente funzionale che aggiunge fasi a un cluster Amazon EMR.

> **Note**
>
> Questo esempio si basa su Java, ma Amazon EMR supporta anche diversi linguaggi di programmazione con una raccolta di SDK Amazon EMR. Per ulteriori informazioni, consulta Utilizzo di SDK per chiamare le API di Amazon EMR.

Questo esempio di codice sorgente Java illustra come eseguire le seguenti attività utilizzando l'API di Amazon EMR:

- Recupero delle credenziali di AWS e invio delle stesse ad Amazon EMR per effettuare chiamate API
- Configurare una nuova fase personalizzata e una nuova fase predefinita
- Aggiunta di nuove fasi a un cluster Amazon EMR esistente
- Recuperare ID delle fasi del cluster da un cluster in esecuzione

> **Note**
>
> In questo esempio viene illustrato come aggiungere fasi a un cluster esistente e, pertanto, richiede che sia disponibile un cluster attivo sull'account.

Prima di iniziare, installare una versione di Eclipse IDE per Java EE Developers che corrisponde alla piattaforma del computer. Per ulteriori informazioni, vai a Eclipse Downloads (Download di Eclipse).

Quindi, installa il plug-in Database Development per Eclipse.

Installazione del plug-in Database Development per Eclipse

1. Aprire l'IDE Eclipse.
2. Scegliere Help (Guida) e Install New Software (Installa nuovo software).
3. Nel campo Work with: (Utilizza con:), digitare **`http://download.eclipse.org/releases/kepler`** o il percorso che corrisponde al numero di versione dell'IDE Eclipse.
4. Nell'elenco di elementi, scegliere Database Development (Sviluppo di database) e Finish (Fine).
5. Riavviare Eclipse quando richiesto.

Quindi, installa il Toolkit for Eclipse per rendere disponibili i modelli di progetto del codice sorgente preconfigurati.

Installazione del Toolkit for Eclipse

1. Aprire l'IDE Eclipse.
2. Scegliere Help (Guida) e Install New Software (Installa nuovo software).
3. Nel campo Utilizza con:, digitare **`https://aws.amazon.com/eclipse`**.
4. Nell'elenco di elementi, scegli AWS Toolkit for Eclipse e Finish (Fine).
5. Riavviare Eclipse quando richiesto.

In seguito, crea un nuovo progetto Java AWS ed esegui il codice sorgente Java di esempio.

Creazione di un nuovo progetto Java AWS

1. Aprire l'IDE Eclipse.
2. Scegliere File, New (Nuovo)e Other (Altro).
3. Nella finestra di dialogo Select a wizard (Seleziona una procedura guidata), scegli AWS Java Project (Progetto Java e Next (Successivo).
4. Nella finestra di dialogo New AWS Java Project (Nuovo progetto Java), nel campo **`Project name:`**, immetti il nome del nuovo progetto, ad esempio **`EMR-sample-code`**.
5. Scegli Configure AWS accounts... (Configura account AWS...), immetti le chiavi di accesso pubblico e privato e scegli Finish (Fine). Per ulteriori informazioni sulla creazione di chiavi di accesso, consulta la sezione Come ottenere le credenziali di sicurezza in Riferimenti generali di Amazon Web Services.

   > **Note**
   >
   > Si consiglia di non incorporare le chiavi di accesso direttamente nel codice. L'SDK Amazon EMR consente di inserire le chiavi di accesso in percorsi noti in modo da evitare di conservarle nel codice.

6. Nel nuovo progetto Java, fare clic con il pulsante destro del mouse sulla cartella src, quindi scegliere New (Nuovo) e Class (Classe).
7. Nella finestra di dialogo Java Class (Classe Java), nel campo Name (Nome), immettere un nome per la nuova classe, ad esempio **`main`**.
8. Nella sezione Which method stubs would you like to create? (Quali stub del metodo si desidera creare?), scegliere public static void main(String[] args) e Finish (Fine).
9. Immettere il codice sorgente Java all'interno della nuova classe e aggiungere le istruzioni import appropriate per le classi e metodi nell'esempio. Per praticità, il codice sorgente completo è riportato di seguito.

   > **Note**
   >
   > Nel seguente codice di esempio, sostituisci l'ID cluster di esempio (JobFlowId), *`j-xxxxxxxxxxxx`*, con un ID cluster valido nell'account trovato nella AWS Management Console o utilizzando il seguente comando della AWS CLI:

```
aws emr list-clusters --active | grep "Id"
```

Inoltre, sostituisci il percorso Amazon S3 di esempio, *s3://path/to/my/jarfolder*, con un percorso valido per il JAR. Infine, sostituisci il nome classe di esempio, *com.my.Main1*, con il nome corretto della classe nel tuo JAR, se applicabile.

```
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. Scegli Run (Esegui), Run As (Esegui come) e Java Application (Applicazione Java).
11. Se l'esempio viene eseguito correttamente, nella finestra di console IDE Eclipse viene visualizzato un elenco di ID per le nuove fasi. L'output corretto sarà simile al seguente:

```
[s-39BLQZRJB2E5E, s-1L6A4ZU2SAURC]
```

Concetti comuni per le chiamate API

Argomenti

- Endpoint per Amazon EMR
- Specifica dei parametri del cluster in Amazon EMR
- Zone di disponibilità in Amazon EMR
- Come utilizzare file e librerie aggiuntivi in cluster Amazon EMR

Quando scrivi un'applicazione che chiama l'API di Amazon EMR, ci sono diversi concetti da tenere a mente quando si chiama una delle funzioni wrapper di un SDK.

Endpoint per Amazon EMR

Un endpoint è un URL che rappresenta il punto di partenza per un servizio Web. Ogni richiesta di servizio Web deve contenere un endpoint. L'endpoint specifica la Regione AWS in cui i cluster vengono creati, descritti o terminati. Il suo formato è `elasticmapreduce.`*`regionname`*`.amazonaws.com`. Se specifichi l'endpoint generale (`elasticmapreduce.amazonaws.com`), la richiesta viene indirizzata da Amazon EMR a un endpoint nella regione predefinita. Per account creati a partire dall'8 marzo 2013, la regione predefinita è us-west-2; per i vecchi account, la regione predefinita è us-east-1.

Per ulteriori informazioni sugli endpoint per Amazon EMR, consulta la sezione Regioni ed endpoint in Riferimenti generali di Amazon Web Services.

Specifica dei parametri del cluster in Amazon EMR

I parametri `Instances` consentono di configurare il tipo e il numero di istanze EC2 per creare nodi per elaborare i dati. Hadoop distribuisce l'elaborazione dei dati su più nodi del cluster. Il nodo master serve a monitorare l'integrità dei nodi principali e di task ed esegue il polling dei nodi per lo stato del risultato del processo. I nodi principali e di task eseguono l'elaborazione effettiva dei dati. Se si dispone di un cluster a nodo singolo, il nodo svolge la funzione di nodo master e principale.

Il parametro `KeepJobAlive` in una richiesta `RunJobFlow` determina se terminare il cluster quando esaurisce le fasi del cluster da eseguire. Impostare questo valore su `False` quando l'esecuzione del cluster è quella prevista. Durante la risoluzione dei problemi del flusso di elaborazione e l'aggiunta di fasi mentre l'esecuzione del cluster è sospesa, è opportuno impostare il valore su `True`. Questo consente di ridurre il tempo e le spese di caricamento dei risultati in Amazon Simple Storage Service (Amazon S3), solo per ripetere il processo dopo la modifica di una fase per riavviare il cluster.

Se `KeepJobAlive` è `true`, dopo che il cluster è stato configurato per completare il lavoro, invia una richiesta `TerminateJobFlows` per evitare che il cluster attivo continui a generare costi AWS.

Per ulteriori informazioni sui parametri che sono univoci per `RunJobFlow`, consulta RunJobFlow. Per ulteriori informazioni sui parametri generici nella richiesta, consulta la sezione relativa ai Parametri di richiesta comuni.

Zone di disponibilità in Amazon EMR

Amazon EMR usa istanze EC2 come nodi per elaborare cluster. Le ubicazioni di queste istanze EC2 sono composte da zone di disponibilità e Regioni. Le regioni sono disperse e situate in aree geografiche separate. Le zone di disponibilità sono ubicazioni distinte all'interno di una Regione isolata dai guasti che si verificano in altre zone di disponibilità. Ogni zona di disponibilità offe una connettività di rete economica, a bassa latenza ad altre zone di disponibilità nella stessa Regione. Per un elenco delle regioni e degli endpoint per Amazon EMR, consulta la sezione Regioni ed endpoint in Riferimenti generali di Amazon Web Services.

Il parametro `AvailabilityZone` specifica il percorso generale del cluster. Questo parametro è facoltativo e, in generale, ne sconsigliamo l'utilizzo. Quando `AvailabilityZone` non è specificato, Amazon EMR sceglie automaticamente il valore `AvailabilityZone` ottimale per il cluster. Questo parametro può essere utile se desideri co-individuare le tue istanze con altre istanze in esecuzione esistenti e il cluster deve leggere o scrivere dati di tali istanze. Per ulteriori informazioni, consulta la Guida per l'utente di Amazon EC2 per le istanze Linux.

Come utilizzare file e librerie aggiuntivi in cluster Amazon EMR

Talvolta potrebbe essere necessario utilizzare file aggiuntivi o librerie personalizzate con applicazioni mappatore o riduttore. Ad esempio, potrebbe essere necessario utilizzare una libreria che consente di convertire un file PDF in testo normale.

Per memorizzare nella cache un file utilizzato dal mappatore o riduttore durante lo streaming Hadoop

- Nel campo `args` JAR, aggiungere il seguente argomento:

  ```
  -cacheFile s3://bucket/path_to_executable#local_path
  ```

  Il file, `local_path`, si trova nella directory di lavoro del mappatore, che potrebbe fare riferimento al file.

Utilizzo di SDK per chiamare le API di Amazon EMR

Argomenti

- Utilizzo di AWS SDK for Java per creare un cluster Amazon EMR

Gli SDK AWS forniscono funzioni che eseguono il wrapping dell'API e tengono conto di molti dettagli di connessione, ad esempio il calcolo di firme, la gestione dei tentativi di richiesta e la gestione degli errori. Gli SDK contengono anche codice di esempio, tutorial e altre risorse per aiutarti a iniziare a scrivere applicazioni che chiamano AWS. La chiamata di funzioni wrapper in un SDK può semplificare notevolmente il processo di scrittura di un'applicazione AWS.

Per ulteriori informazioni su come scaricare e utilizzare gli SDK AWS, consulta SDK in Strumenti per Amazon Web Services.

Utilizzo di AWS SDK for Java per creare un cluster Amazon EMR

AWS SDK for Java offre tre pacchetti con funzionalità Amazon EMR:

- com.amazonaws.services.elasticmapreduce
- com.amazonaws.services.elasticmapreduce.model
- com.amazonaws.services.elasticmapreduce.util

Per ulteriori informazioni su questi pacchetti, consulta la Guida di riferimento alle API di AWS SDK for Java.

L'esempio seguente mostra in che modo i kit SDK possono semplificare la programmazione con Amazon EMR. Il codice di esempio sottostante utilizza l'oggetto `StepFactory`, una classe helper per la creazione di tipi di fase Amazon EMR comuni, per creare un cluster Hive interattivo con il debug abilitato.

```
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());

    }

}
```

Come minimo, devi passare un ruolo del servizio e un ruolo jobflow corrispondenti, rispettivamente, a EMR_DefaultRole ed EMR_EC2_DefaultRole. A questo scopo, puoi invocare questo comando della AWS CLI per lo stesso account. Innanzitutto, verifica se i ruoli esistono già:

```
aws iam list-roles | grep EMR
```

Il profilo dell'istanza (EMR_EC2_DefaultRole) e il ruolo del servizio (EMR_DefaultRole) saranno entrambi visualizzati se esistono:

```
"RoleName": "EMR_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_DefaultRole"
            "RoleName": "EMR_EC2_DefaultRole",
            "Arn": "arn:aws:iam::AccountID:role/EMR_EC2_DefaultRole"
```

Se i ruoli predefiniti non esistono, puoi utilizzare il seguente comando per crearli:

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

Gestione delle Service Quotas di Amazon EMR

Argomenti

- Cosa sono le Service Quotas di Amazon EMR
- Come gestire le Service Quotas di Amazon EMR
- Quando impostare gli eventi EMR in CloudWatch

Gli argomenti di questa sezione descrivono le Service Quotas (precedentemente definite "limiti del servizio") di Amazon EMR, come gestirle nella AWS Management Console e quando è più vantaggioso utilizzare gli eventi CloudWatch anziché le quote di servizio per monitorare i cluster e attivare le operazioni.

Cosa sono le Service Quotas di Amazon EMR

L'account AWS include quote di servizio predefinite, note anche come limiti, per ogni servizio AWS. Il servizio EMR presenta due tipi di limiti:

- Limiti delle risorse: è possibile utilizzare EMR per creare risorse EC2. Tuttavia, queste risorse CE2 sono soggette a quote di servizio. I limiti delle risorse in questa categoria sono:
  - Il numero massimo di cluster attivi che possono essere eseguiti nello stesso momento.
  - Il numero massimo di istanze attive per gruppo di istanze.
- Limiti delle API: quando si utilizzano le API EMR, i due tipi di limitazioni sono:
  - Limite di frammentazione: questo è il numero massimo di chiamate API che puoi effettuare contemporaneamente. Ad esempio, il numero massimo di richieste API AddInstanceFleet che è possibile effettuare al secondo è impostato su 5 chiamate/secondo come impostazione predefinita. Ciò implica che il limite di frammentazione dell'API AddInstanceFleet è di 5 chiamate/secondo o che, in qualsiasi momento, è possibile effettuare al massimo 5 chiamate API AddInstanceFleet. Tuttavia, dopo aver utilizzato il limite di frammentazione, le chiamate successive sono limitate dal limite di frequenza.
  - Limite di frequenza: questa è la velocità di rifornimento della capacità di espansione dell'API. Ad esempio, la frequenza di rifornimento delle chiamate AddInstanceFleet è impostata su 0,5 chiamate/secondo come impostazione predefinita. Ciò significa che dopo aver raggiunto il limite di frammentazione, è necessario attendere almeno 2 secondi (0,5 chiamate/secondo x 2 secondi = 1 chiamata) per effettuare la chiamata API. Se si effettua una chiamata prima di questo valore temporale, si è limitati dal servizio Web EMR. In qualsiasi momento, è possibile effettuare solo un numero di chiamate pari alla capacità di espansione senza alcuna limitazione. Per ogni ulteriore secondo di attesa, la capacità di espansione aumenta di 0,5 chiamate fino a raggiungere il limite massimo di 5, ossia il limite di frammentazione.

Come gestire le Service Quotas di Amazon EMR

Service Quotas è una funzionalità AWS utile a visualizzare e gestire le quote di servizio o i limiti di Amazon EMR da una posizione centrale utilizzando la AWS Management Console, l'API o la CLI. Per

ulteriori informazioni sulla visualizzazione delle quote e sulla richiesta di un aumento, consulta Quote di servizio di AWS in Riferimenti generali di Amazon Web Services.

Per alcune API, l'impostazione di un evento CloudWatch potrebbe rivelarsi un'opzione migliore rispetto all'aumento delle quote di servizio. Per risparmiare tempo, è anche possibile usare CloudWatch per impostare gli allarmi e attivare le richieste di aumento in modo proattivo, prima di raggiungere la quota di servizio. Per ulteriori dettagli, consulta Quando impostare gli eventi EMR in CloudWatch.

Quando impostare gli eventi EMR in CloudWatch

Per alcune API di polling, ad esempio DescribeCluster, DescribeStep e ListClusters, l'impostazione di un evento CouldWatch può ridurre il tempo di risposta alle modifiche e liberare quote di servizio. Se hai impostato una funzione Lambda che si attiva al variare dello stato di un cluster, ad esempio quando una fase viene completata o il cluster viene terminato, puoi utilizzare tale attivazione per avviare l'operazione successiva nel flusso di lavoro anziché attendere il polling successivo. In caso contrario, se hai impostato specifiche istanze Amazon EC2 o funzioni Lambda che eseguono costantemente il polling dell'API EMR per le modifiche, oltre a sprecare risorse di calcolo potresti anche raggiungere la tua quota di servizio.

Di seguito sono riportati alcuni casi che mostrano come trarre vantaggio dal passaggio a un'architettura basata su eventi.

Caso 1: Polling di EMR utilizzando le chiamate API DescribeCluster per il completamento della fase

Example Polling di EMR utilizzando le chiamate API DescribeCluster per il completamento della fase

Un modello comune consiste nell'inviare una fase a un cluster in esecuzione ed eseguire il polling di Amazon EMR per lo stato relativo alla fase, in genere utilizzando le API DescribeCluster o DescribeStep. Questa attività può anche essere eseguita con un ritardo minimo collegandosi all'evento di modifica dello stato della fase Amazon EMR.

Questo evento include le seguenti informazioni nel relativo payload.

```
{
  "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."
  }
}
```

Nella mappa di dettaglio, una funzione Lambda potrebbe analizzare "state", "stepId" o "clusterId" per trovare informazioni pertinenti.

Caso 2: Polling di EMR per i cluster disponibili per l'esecuzione di flussi di lavoro

Example Polling di EMR per i cluster disponibili per l'esecuzione di flussi di lavoro

Un modello utile per i clienti che eseguono più cluster consiste nell'eseguire flussi di lavoro su cluster non appena sono disponibili. Se sono presenti molti cluster in esecuzione e necessiti di eseguire un flusso di lavoro su un cluster in attesa, potresti eseguire il polling EMR utilizzando le chiamate API DescribeCluster o ListClusters per i cluster disponibili. Un altro modo per ridurre il ritardo nel sapere quando un cluster è pronto per una fase è elaborare l'evento di modifica dello stato del cluster Amazon EMR.

Questo evento include le seguenti informazioni nel relativo payload.

```
{
  "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 ..."
  }
}
```

Per questo evento, potresti impostare una funzione Lambda per inviare immediatamente un flusso di lavoro in attesa a un cluster non appena il suo stato cambia in WAITING (IN ATTESA).

Caso 3: Polling di EMR per la terminazione del cluster

Example Polling di EMR per la terminazione del cluster

Un modello comune tra i clienti che eseguono molti cluster EMR è il polling di Amazon EMR per cluster terminati in modo che il lavoro non venga più inviato a esso. È possibile implementare questo modello con le chiamate API DescribeCluster e ListClusters o utilizzando l'evento di modifica dello stato del cluster Amazon EMR.

In seguito alla terminazione del cluster, l'evento emesso ha un aspetto simile all'esempio seguente.

```
{
  "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 sezione "Detail (Dettaglio)" del payload include il clusterId e lo stato su cui è possibile agire.

Glossario per AWS

Per la terminologia AWS più recente, consultare il glossario AWS nella documentazione di riferimento per Glossario AWS.