



Guida per l'utente

AWS PCS



AWS PCS: Guida per l'utente

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

I marchi e l'immagine commerciale di Amazon non possono essere utilizzati in relazione a prodotti o servizi che non siano di Amazon, in una qualsiasi modalità che possa causare confusione tra i clienti o in una qualsiasi modalità che denigri o discrediti Amazon. Tutti gli altri marchi non di proprietà di Amazon sono di proprietà delle rispettive aziende, che possono o meno essere associate, collegate o sponsorizzate da Amazon.

Table of Contents

Che cos'è AWS PCS?	1
Concetti chiave	1
Configurazione	3
Registrati per un Account AWS	3
Crea un utente con accesso amministrativo	3
Installa il AWS CLI	5
Nozioni di base	6
Prerequisiti	7
Crea sottoreti e VPC	8
Trova il gruppo di sicurezza predefinito per il cluster VPC	9
Creare gruppi di sicurezza	10
Creazione dei gruppi di sicurezza	10
Creazione di un cluster	11
Crea spazio di archiviazione condiviso in Amazon EFS	12
Crea storage condiviso in FSx for Lustre	12
Creare gruppi di nodi di calcolo	14
Creazione di un profilo dell'istanza	14
Crea modelli di lancio	16
Crea un gruppo di nodi di calcolo per i nodi di accesso	17
Crea un gruppo di nodi di calcolo per i lavori	18
Crea una coda	19
Connect al cluster	20
Esplora l'ambiente del cluster	21
Cambia utente	21
Lavora con file system condivisi	22
Interagisci con Slurm	22
Esegui un processo a nodo singolo	23
Esegui un lavoro MPI multinodo con Slurm	25
Elimina le tue AWS risorse	28
Lavorare con AWS PCS	31
Cluster	31
Creazione di un cluster	32
Eliminazione di un cluster	36
Dimensione del cluster	37

Segreti del cluster	38
gruppi di nodi di calcolo	42
Creazione di un gruppo di nodi di calcolo	42
Aggiornamento di un gruppo di nodi di calcolo	48
Eliminazione di un gruppo di nodi di calcolo	51
Ricerca di istanze di gruppi di nodi di calcolo	53
Utilizzo di modelli di lancio	55
Panoramica	55
Creare un modello di avvio di base	57
Lavorare con i dati EC2 degli utenti Amazon	59
Prenotazioni della capacità	64
Parametri utili del modello di lancio	66
Queues	68
Creazione di una coda	68
Aggiornamento di una coda	70
Eliminazione di una coda	72
nodi di accesso	74
Utilizzo di un gruppo di nodi di calcolo per l'accesso	74
Utilizzo di istanze autonome come nodi di accesso	76
Rete	82
VPCe requisiti per le sottoreti	82
Creare un VPC	84
Gruppi di sicurezza	87
Interfacce di rete multiple	88
Gruppi di collocamento	90
Utilizzo di Elastic Fabric Adapter (EFA)	91
File system di rete	98
Considerazioni sull'utilizzo dei file system di rete	98
Esempi di supporti di rete	99
Immagini di macchine Amazon (AMIs)	103
Utilizzando un esempio AMIs	103
Personalizzato AMIs	105
Installatori da creare AMIs	116
Versioni Slurm	120
Domande frequenti sulle versioni di Slurm	120
Sicurezza	123

Protezione dei dati	124
Crittografia a riposo	125
Crittografia in transito	125
Gestione delle chiavi	126
Riservatezza del traffico Internet	126
Crittografia del traffico API	126
Crittografia del traffico dati	127
VPC endpoint di interfaccia (AWS PrivateLink)	127
Considerazioni	127
Creazione di un endpoint di interfaccia	128
Creazione di una policy dell'endpoint	128
Identity and Access Management	129
Destinatari	130
Autenticazione con identità	130
Gestione dell'accesso con policy	134
Come funziona AWS Parallel Computing Service con IAM	137
Esempi di policy basate su identità	143
AWS politiche gestite	147
Ruoli collegati ai servizi	153
EC2 Ruolo Spot	155
Autorizzazioni minime	155
Profili delle istanze	160
Risoluzione dei problemi	162
Convalida della conformità	164
Resilienza	165
Sicurezza dell'infrastruttura	166
Analisi e gestione delle vulnerabilità	166
Prevenzione del confused deputy tra servizi	167
IAM ruolo per EC2 le istanze Amazon fornite come parte di un gruppo di nodi di calcolo	168
Best practice di sicurezza	169
AMI-sicurezza correlata	169
Sicurezza di Slurm Workload Manager	170
Monitoraggio e registrazione	170
Sicurezza di rete	170
Registrazione di log e monitoraggio	171
AWS PCS registri dello scheduler	171

Prerequisiti	172
Configurazione dei registri dello scheduler utilizzando la console AWS PCS	172
Configurazione dei registri dello scheduler utilizzando il AWS CLI	173
Scheduler log stream, percorsi e nomi	175
Esempio di record AWS PCS di log dello scheduler	176
Monitoraggio con CloudWatch	176
Monitoraggio di parametri	177
Monitoraggio delle istanze	178
CloudTrail registri	186
AWS PCSinformazioni in CloudTrail	186
Informazioni sulle voci dei file di CloudTrail registro da AWS PCS	187
Endpoint e quote di servizio	190
Endpoint del servizio	190
Quote del servizio	191
Quote interne	192
Quote pertinenti per altri servizi AWS	192
Note di rilascio per AMIs	193
Esempio x86_64 per Slurm 23.11 () AMI AL2	193
Arm64 AMI di esempio per Slurm 23.11 () AL2	194
Cronologia dei documenti	197
AWS Glossario	198
.....	cxcix

Cos'è il servizio AWS Parallel Computing?

AWS Parallel Computing Service (AWS PCS) è un servizio gestito che semplifica l'esecuzione e la scalabilità dei carichi di lavoro di elaborazione ad alte prestazioni (HPC) e la creazione di modelli scientifici e ingegneristici AWS utilizzando Slurm. AWS PCS Utilizzalo per creare cluster di elaborazione che integrano AWS elaborazione, archiviazione, rete e visualizzazione all'avanguardia. Esegui simulazioni o crea modelli scientifici e ingegneristici. Semplifica e semplifica le operazioni del cluster utilizzando funzionalità integrate di gestione e osservabilità. Consenti ai tuoi utenti di concentrarsi sulla ricerca e l'innovazione consentendo loro di eseguire applicazioni e lavori in un ambiente familiare.

Concetti chiave

Un cluster in AWS PCS ha 1 o più code, associate ad almeno 1 gruppo di nodi di calcolo. I lavori vengono inviati alle code ed eseguiti su EC2 istanze definite da gruppi di nodi di calcolo. È possibile utilizzare queste basi per implementare architetture sofisticate. HPC

Cluster

Un cluster è una risorsa per la gestione delle risorse e l'esecuzione dei carichi di lavoro. Un cluster è una AWS PCS risorsa che definisce un insieme di configurazione di elaborazione, rete, archiviazione, identità e pianificazione dei processi. È possibile creare un cluster specificando quale job scheduler si desidera utilizzare (attualmente Slurm), quale configurazione di scheduler si desidera, quale controller di servizio si desidera gestire il cluster e in cui si VPC desidera avviare le risorse del cluster. Lo scheduler accetta e pianifica i lavori e avvia anche i nodi di calcolo (istanze) che elaborano tali lavori. EC2

Gruppo di nodi di calcolo

Un gruppo di nodi di calcolo è una raccolta di nodi di elaborazione che viene AWS PCS utilizzata per eseguire processi o fornire accesso interattivo a un cluster. Quando definisci un gruppo di nodi di calcolo, specifichi caratteristiche comuni come i tipi di EC2 istanze Amazon, il numero minimo e massimo di istanze, le VPC sottoreti di destinazione, Amazon Machine Image AMI (), l'opzione di acquisto e la configurazione di avvio personalizzata. AWS PCS utilizza queste impostazioni per avviare, gestire e terminare in modo efficiente i nodi di calcolo in un gruppo di nodi di calcolo.

Queue

Quando si desidera eseguire un processo su un cluster specifico, lo si invia a una coda particolare (a volte chiamata anche partizione). Il processo rimane in coda finché non ne AWS PCS pianifica l'esecuzione su un gruppo di nodi di calcolo. Si associano uno o più gruppi di nodi di calcolo a ciascuna coda. È necessaria una coda per pianificare ed eseguire i lavori sulle risorse del gruppo di nodi di calcolo sottostanti utilizzando varie politiche di pianificazione offerte dal job scheduler. Gli utenti non inviano i lavori direttamente a un nodo di calcolo o a un gruppo di nodi di calcolo.

Amministratore di sistema

Un amministratore di sistema distribuisce, mantiene e gestisce un cluster. Possono accedere AWS PCS tramite AWS Management Console AWS PCSAPI, e AWS SDK. Hanno accesso a cluster specifici tramite SSH o AWS Systems Manager, dove possono eseguire attività amministrative, eseguire lavori, gestire dati ed eseguire altre attività basate su shell. Per ulteriori informazioni, consulta la documentazione di [AWS Systems Manager](#).

Utente finale

Un utente finale non ha day-to-day la responsabilità di implementare o gestire un cluster. Utilizzano un'interfaccia terminale (ad esempioSSH) per accedere alle risorse del cluster, eseguire processi, gestire dati ed eseguire altre attività basate sulla shell.

Configurazione del servizio AWS Parallel Computing

Completa le seguenti attività per configurare AWS Parallel Computing Service (AWS PCS).

Argomenti

- [Registrati per un Account AWS](#)
- [Crea un utente con accesso amministrativo](#)
- [Installa il AWS CLI](#)

Registrati per un Account AWS

Se non ne hai uno Account AWS, completa i seguenti passaggi per crearne uno.

Per iscriverti a un Account AWS

1. Apri la <https://portal.aws.amazon.com/billing/registrazione>.
2. Segui 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.

Quando ti iscrivi a un Account AWS, Utente root dell'account AWS viene creato un. 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 e utilizza solo l'utente root per eseguire [attività che richiedono l'accesso di un utente root](#).

AWS ti invia un'email di conferma dopo il completamento della procedura di registrazione. In qualsiasi momento, puoi visualizzare l'attività corrente del tuo account e gestirlo accedendo a <https://aws.amazon.com/> e scegliendo Il mio account.

Crea un utente con accesso amministrativo

Dopo esserti registrato Account AWS, proteggi Utente root dell'account AWS AWS IAM Identity Center, abilita e crea un utente amministrativo in modo da non utilizzare l'utente root per le attività quotidiane.

Proteggi i tuoi Utente root dell'account AWS

1. Accedi [AWS Management Console](#) come proprietario dell'account scegliendo Utente root e inserendo il tuo indirizzo Account AWS email. Nella pagina successiva, inserisci la password.

Per informazioni sull'accesso utilizzando un utente root, consulta la pagina [Signing in as the root user](#) della Guida per l'utente di Accedi ad AWS .

2. Attiva l'autenticazione a più fattori (MFA) per il tuo utente root.

Per istruzioni, consulta [Abilitare un MFA dispositivo virtuale per l'utente Account AWS root \(console\)](#) nella Guida per l'IAMutente.

Crea un utente con accesso amministrativo

1. Abilita IAM Identity Center.

Per istruzioni, consulta [Abilitazione di AWS IAM Identity Center](#) nella Guida per l'utente di AWS IAM Identity Center .

2. In IAM Identity Center, concedi l'accesso amministrativo a un utente.

Per un tutorial sull'utilizzo di IAM Identity Center directory come fonte di identità, consulta [Configurare l'accesso utente con i valori predefiniti IAM Identity Center directory](#) nella Guida per l'AWS IAM Identity Center utente.

Accesso come utente amministratore

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

Per informazioni sull'accesso con un utente di IAM Identity Center, consulta [Accesso al portale di AWS accesso](#) nella Guida per l'Accedi ad AWS utente.

Assegna l'accesso a ulteriori utenti

1. In IAM Identity Center, crea un set di autorizzazioni che segua la migliore pratica di applicazione delle autorizzazioni con privilegi minimi.

Segui le istruzioni riportate nella pagina [Creazione di un set di autorizzazioni](#) nella Guida per l'utente di AWS IAM Identity Center .

2. Assegna al gruppo prima gli utenti e poi l'accesso con autenticazione unica (Single Sign-On).

Per istruzioni, consulta [Aggiungere gruppi](#) nella Guida per l'utente di AWS IAM Identity Center .

Installa il AWS CLI

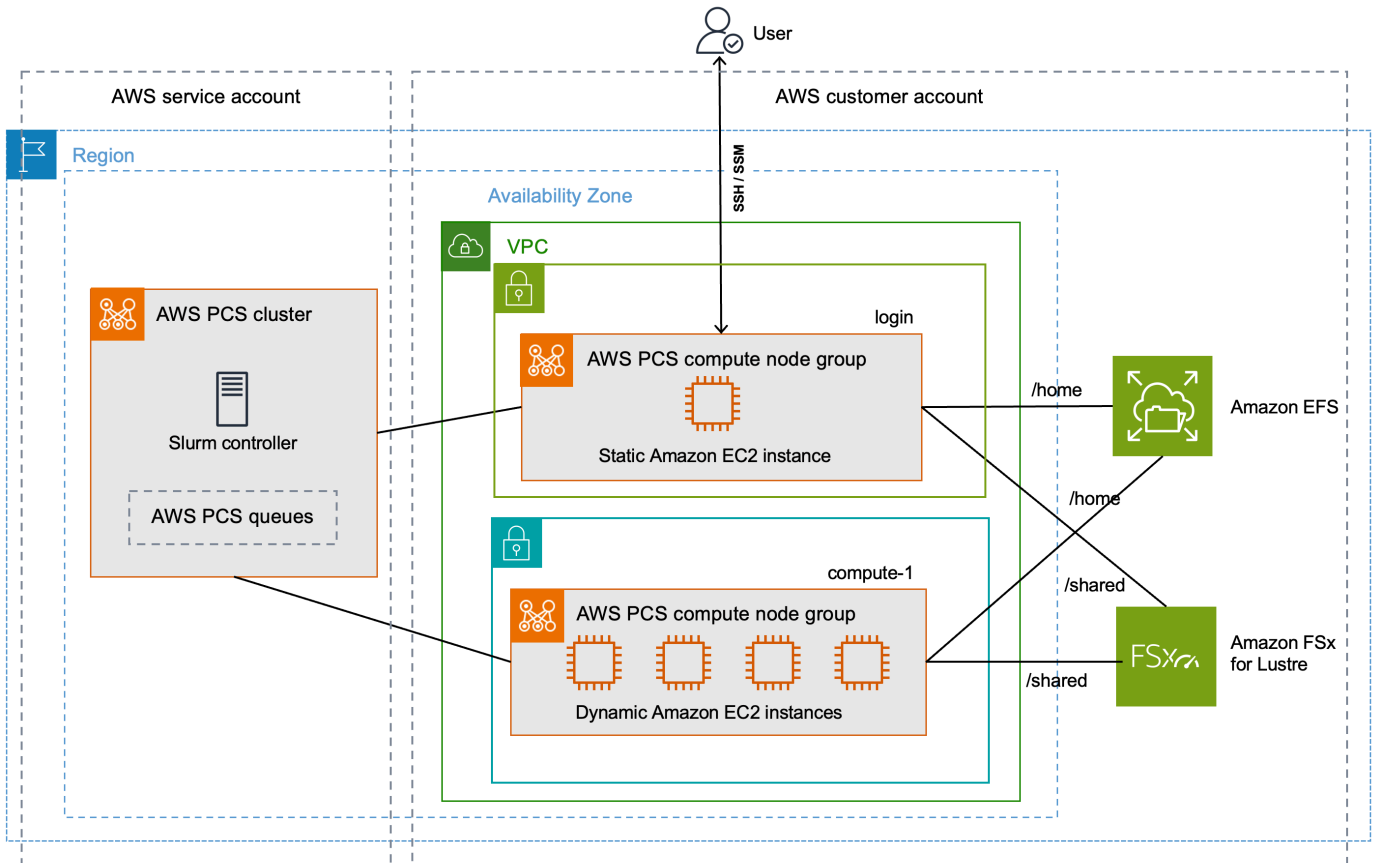
È necessario utilizzare la versione più recente di AWS CLI. Per informazioni, consulta [Installare o aggiornare alla versione più recente di AWS CLI nella](#) Guida per l'AWS Command Line Interface utente della versione 2.

Digitate il seguente comando al prompt dei comandi per verificarlo AWS CLI; dovrebbe mostrare informazioni di aiuto.

```
aws pcs help
```

Guida introduttiva con AWS PCS

Questo è un tutorial per creare un cluster semplice che puoi usare per provare AWS PCS. La figura seguente mostra il design del cluster.



Il tutorial sulla progettazione del cluster ha i seguenti componenti chiave:

- A VPC e sottoreti che soddisfano i requisiti [AWS PCS di rete](#).
- Un EFS file system Amazon, che verrà utilizzato come home directory condivisa.
- Un file system Amazon FSx for Lustre, che fornisce una directory condivisa ad alte prestazioni.
- Un AWS PCS cluster, che fornisce un controller Slurm.
- 2 gruppi di nodi di calcolo.
 - Il gruppo di `login` nodi, che fornisce un accesso interattivo basato su shell al sistema.
 - Il gruppo di `compute-1` nodi fornisce istanze con scalabilità elastica per eseguire i processi.
- 1 coda che invia i lavori alle istanze del gruppo di nodi. EC2 `compute-1`

Il cluster richiede AWS risorse aggiuntive, come gruppi di sicurezza, IAM ruoli e modelli di EC2 avvio, che non sono mostrati nel diagramma.

Argomenti

- [Prerequisiti per iniziare con AWS PCS](#)
- [Crea sottoreti VPC e sottoreti per AWS PCS](#)
- [Creare gruppi di sicurezza per AWS PCS](#)
- [Crea un cluster in AWS PCS](#)
- [Crea storage condiviso per AWS PCS Amazon Elastic File System](#)
- [Crea storage condiviso per AWS PCS Amazon FSx for Lustre](#)
- [Crea gruppi di nodi di calcolo in AWS PCS](#)
- [Crea una coda per gestire i lavori in AWS PCS](#)
- [Connect al AWS PCS cluster](#)
- [Esplora l'ambiente del cluster in AWS PCS](#)
- [Esegui un processo a nodo singolo in AWS PCS](#)
- [Esegui un MPI lavoro multinodo con Slurm in AWS PCS](#)
- [Elimina le tue AWS risorse per AWS PCS](#)

Prerequisiti per iniziare con AWS PCS

Prima di iniziare questo tutorial, installa e configura i seguenti strumenti e risorse necessari per creare e gestire un AWS PCS cluster.

- AWS CLI— Uno strumento da riga di comando per lavorare con AWS i servizi, tra cui AWS PCS. Per ulteriori informazioni, vedere [Installazione o aggiornamento alla versione più recente di AWS CLI nella Guida per l'AWS Command Line Interface utente della versione 2](#). Dopo aver installato AWS CLI, ti consigliamo di configurarlo anche. Per ulteriori informazioni, consulta [Configurare il AWS CLI nella Guida per AWS Command Line Interface l'utente della versione 2](#).
- IAM Autorizzazioni richieste: il responsabile della IAM sicurezza che stai utilizzando deve disporre delle autorizzazioni per lavorare con AWS PCS IAM ruoli, ruoli collegati al servizio AWS CloudFormation VPC, a e risorse correlate. Per ulteriori informazioni [Servizio di Identity and Access Management per AWS Parallel Computing](#), consulta la sezione [Creazione di un ruolo collegato ai servizi nella Guida per l'utente AWS Identity and Access Management](#). È necessario che tutti

i passaggi di questa guida siano completati dallo stesso utente. Esegui il comando seguente per controllare l'utente corrente:

```
aws sts get-caller-identity
```

- Ti consigliamo di completare i passaggi della riga di comando descritti in questo argomento in una shell Bash. In alternativa, puoi apportare alcune modifiche alla tua shell per alcuni comandi di script, come i caratteri di continuazione della riga, e per il modo in cui le variabili vengono impostate e utilizzate. Inoltre, le regole di escape e di utilizzo delle virgolette per la shell (interprete di comandi) potrebbero essere diverse. Per ulteriori informazioni, consulta [Virgolette e lettere con stringhe nella Guida per l' AWS CLI](#) [AWS Command Line Interface](#) utente della versione 2.

Crea sottoreti VPC e sottoreti per AWS PCS

È possibile creare sottoreti VPC e utilizzando un modello. CloudFormation Usa quanto segue URL per scaricare il CloudFormation modello, quindi carica il modello nella [AWS CloudFormation console](#) per creare un nuovo CloudFormation stack. Per ulteriori informazioni, consulta [Uso della AWS CloudFormation console](#) nella Guida per l'AWS CloudFormation utente.

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/net/hpc_large_scale/assets/main.yaml
```

Con il modello aperto nella AWS CloudFormation console, inserisci le seguenti opzioni. Puoi utilizzare i valori predefiniti forniti nel modello.

- In Fornisci un nome per lo stack:
 - In Nome dello stack, inserisci:

```
hpc-networking
```

- In Parametri:
 - Sotto VPC:
 - Sotto CidrBlock, inserisci:

```
10.3.0.0/16
```

- In Sottoreti A:
 - In CidrPublicSubnetA, inserisci:

10.3.0.0/20

- In CidrPrivateSubnetA, inserisci:

10.3.128.0/20

- In Sottoreti B:
 - In CidrPublicSubnetB, inserisci:

10.3.16.0/20

- In CidrPrivateSubnetB, inserisci:

10.3.144.0/20

- In Sottoreti C:
 - Per ProvisionSubnetsC, seleziona True
 - In CidrPublicSubnetC, inserisci:

10.3.32.0/20

- In CidrPrivateSubnetC, inserisci:

10.3.160.0/20

- In Capacità:
 - Seleziona la casella Riconosco che AWS CloudFormation potrebbe creare IAM risorse.

Monitora lo stato dello CloudFormation stack. Quando raggiunge CREATE_COMPLETE, trova l'ID per il gruppo di sicurezza predefinito nel nuovo VPC. L'ID verrà utilizzato più avanti nel tutorial.

Trova il gruppo di sicurezza predefinito per il cluster VPC

Per trovare l'ID del gruppo di sicurezza predefinito nel nuovo VPC, segui questa procedura:

- Accedi alla [VPCconsole Amazon](#).
- Nella VPCdashboard, seleziona Filtra per VPC.
 - Scegli da VPC dove inizia il nome hpc-networking.

- In Sicurezza, scegli Gruppi di sicurezza.
- Trova l'ID del gruppo di sicurezza per il gruppo denominato default. Ha la descrizione default VPC security group. L'ID verrà utilizzato in un secondo momento per configurare i modelli di EC2 avvio.

Creare gruppi di sicurezza per AWS PCS

AWS PCS si affida a gruppi di sicurezza per gestire il traffico di rete in entrata e in uscita da un cluster e dai relativi gruppi di nodi di calcolo. Per informazioni dettagliate su questo argomento, vedere.

[Requisiti e considerazioni sui gruppi di sicurezza](#)

In questo passaggio, utilizzerai un CloudFormation modello per due gruppi di sicurezza.

- Un gruppo di sicurezza del cluster, che consente le comunicazioni tra AWS PCS controller, nodi di elaborazione e nodi di accesso.
- Un gruppo SSH di sicurezza in entrata, che puoi aggiungere facoltativamente ai tuoi nodi di accesso per supportare l'accesso SSH

Crea i gruppi di sicurezza per AWS PCS

È possibile creare sottoreti VPC e con questo CloudFormation modello. Usa quanto segue URL per scaricare il CloudFormation modello, quindi carica il modello nella [AWS CloudFormation console](#) per creare un nuovo CloudFormation stack. Per ulteriori informazioni, consulta [Uso della AWS CloudFormation console](#) nella Guida per l'AWS CloudFormation utente.

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/getting_started/assets/pcs-cluster-sg.yaml
```

Con il modello aperto nella AWS CloudFormation console, inserisci le seguenti opzioni. Tieni presente che alcune opzioni saranno precompilate nel modello: puoi semplicemente lasciarle come valori predefiniti.

- In Fornisci un nome per lo stack
 - In Nome dello stack, inserisci:

```
getstarted-sg
```


- In Parametri
 - In VpcId, scegli da VPC dove inizia il nome hpc-networking.
 - (Facoltativo) In ClientIpCidr, inserisci un intervallo IP più restrittivo per il gruppo di SSH sicurezza in entrata. Ti consigliamo di limitarlo con il tuo IP/subnet (x.x.x.x/32 per il tuo IP o x.x.x.x/24 per l'intervallo. PUBLICSostituisci x.x.x.x con il tuo IP. [Puoi ottenere il tuo IP pubblico utilizzando strumenti come https://ifconfig.co/](https://ifconfig.co/))

Monitora lo stato dello CloudFormation stack. Quando raggiunge CREATE_COMPLETE il gruppo di sicurezza, le risorse sono pronte.

Sono stati creati due gruppi di sicurezza, con i seguenti nomi:

- cluster-getstarted-sg— questo è il gruppo di sicurezza del cluster
- inbound-ssh-getstarted-sg— questo è un gruppo di sicurezza per consentire l'accesso in entrata SSH

Crea un cluster in AWS PCS

In AWS PCS, un cluster è una risorsa persistente per la gestione delle risorse e l'esecuzione dei carichi di lavoro. Si crea un cluster per uno scheduler specifico (AWS PCS attualmente supporta Slurm) in una sottorete di uno strumento nuovo o esistente. VPC Il cluster accetta e pianifica i lavori e avvia anche i nodi di calcolo (EC2istanze) che elaborano tali lavori.

Creazione di un cluster

1. Apri la [AWS PCS console](#) e scegli Crea cluster.
2. Nella sezione Configurazione del cluster, inserisci i seguenti campi:
 - Nome del cluster: immettere get-started
 - Dimensioni del controller: seleziona Small
3. Nella sezione Rete, selezionate i valori per i seguenti campi:
 - VPC— Scegli il VPC nome hpc-networking:Large-Scale-HPC
 - Subnet: seleziona la sottorete con cui inizia il nome hpc-networking:PrivateSubnetA
 - Gruppi di sicurezza: selezionare il gruppo di sicurezza del cluster denominato cluster-getstarted-sg

4. Scegli Create cluster (Crea cluster).

Note

Il campo Stato mostra Creazione durante il provisioning del cluster. La creazione del cluster può richiedere diversi minuti.

Crea storage condiviso per AWS PCS Amazon Elastic File System

Amazon Elastic File System (AmazonEFS) è un AWS servizio che fornisce uno storage di file senza server e completamente elastico in modo da poter condividere i dati dei file senza fornire o gestire capacità e prestazioni di storage. Per ulteriori informazioni, consulta [Cos'è Amazon Elastic File System?](#) nella Amazon Elastic File System User Guide.

Il cluster AWS PCS dimostrativo utilizza un EFS file system per fornire una home directory condivisa tra i nodi del cluster. Crea un EFS file system VPC uguale al tuo cluster.

Per creare il tuo EFS file system Amazon

1. Vai alla [EFSconsole Amazon](#).
2. Assicurati che sia impostato sullo stesso Regione AWS punto in cui lo proverai AWS PCS.
3. Scegliere Create file system (Crea file system).
4. Nella pagina Crea file system, imposta i seguenti parametri:
 - Per Nome immetti `getstarted-efs`.
 - In Virtual Private Cloud (VPC), scegliete il VPC nome `hpc-networking:Large-Scale-HPC`
 - Scegli Create (Crea) . Verrà visualizzata di nuovo la pagina dei file system.
5. Prendi nota dell'ID del file system per il `getstarted-efs` file system. Queste informazioni serviranno in seguito.

Crea storage condiviso per AWS PCS Amazon FSx for Lustre

Amazon FSx for Lustre semplifica ed economica l'avvio e l'esecuzione del popolare file system Lustre ad alte prestazioni. Usi Lustre per carichi di lavoro in cui la velocità è importante, come l'apprendimento automatico, l'elaborazione ad alte prestazioni (HPC), l'elaborazione video e la

modellazione finanziaria. Per ulteriori informazioni, consulta [Cos'è Amazon FSx for Lustre?](#) nella Guida per l'utente di Amazon FSx for Lustre.

Il cluster AWS PCS dimostrativo può utilizzare un file system FSx for Lustre per fornire una directory condivisa ad alte prestazioni tra i nodi del cluster. Crea un file system FSx for Lustre nello stesso VPC del tuo cluster.

Per creare il tuo file FSx system for Lustre

1. Vai alla [FSxconsole Amazon](#).
2. Assicurati che la console sia impostata per l'utilizzo Regione AWS come il cluster.
3. Scegliere Create file system (Crea file system).
 - Per Seleziona il tipo di file system, scegli Amazon FSx for Lustre, quindi scegli Avanti.
4. Nella pagina Specificare i dettagli del file system, imposta i seguenti parametri:
 - In Dettagli del file system
 - Per Nome immetti `getstarted-fsx`.
 - Per il tipo di distribuzione e archiviazione, scegli Persistente, SSD
 - Per Throughput per unità di storage, scegli 125 MB/s/TiB
 - Per Capacità di archiviazione, immettere 1,2 TiB
 - Per Configurazione dei metadati, scegliete Automatico
 - Per Tipo di compressione dei dati, scegli LZ4
 - In Rete e sicurezza
 - Per Virtual Private Cloud (VPC), scegli il VPC nome `hpc-networking:Large-Scale-HPC`
 - Per i gruppi VPC di sicurezza, lascia il nome al gruppo di sicurezza default
 - Per Subnet, scegli la sottorete con cui inizia il nome `hpc-networking:PrivateSubnetA`
 - Lasciate le altre opzioni impostate sui valori predefiniti.
 - Scegli Next (Successivo).
5. Nella pagina Rivedi e crea, scegli Crea file system. Verrà visualizzata di nuovo la pagina File system.
6. Vai alla pagina dei dettagli del file system FSx for Lustre che hai creato.
7. Prendi nota dell'ID del file system e del nome del montaggio. Queste informazioni serviranno in seguito.

Note

Il campo Stato mostra Creazione durante il provisioning del file system. La creazione del file system può richiedere diversi minuti. Attendi il completamento prima di procedere con il resto del tutorial.

Crea gruppi di nodi di calcolo in AWS PCS

Un gruppo di nodi di calcolo è una raccolta virtuale di nodi di calcolo (EC2istanze) che AWS PCS viene avviata e gestita. Quando si definisce un gruppo di nodi di calcolo, si specificano caratteristiche comuni come i tipi di istanze, il numero minimo e massimo di EC2 istanze, le VPC sottoreti di destinazione, l'opzione di acquisto preferita e la configurazione di avvio personalizzata. AWS PCSavvia, gestisce e termina in modo efficiente i nodi di calcolo in un gruppo di nodi di calcolo, in base a queste impostazioni. Il cluster dimostrativo utilizza un gruppo di nodi di calcolo per fornire nodi di accesso per l'accesso degli utenti e un gruppo di nodi di calcolo separato per elaborare i lavori. I seguenti argomenti descrivono le procedure per configurare questi gruppi di nodi di calcolo nel cluster.

Argomenti

- [Creare un profilo di istanza per AWS PCS](#)
- [Crea modelli di lancio per AWS PCS](#)
- [Crea un gruppo di nodi di calcolo per i nodi di accesso in AWS PCS](#)
- [Crea un gruppo di nodi di calcolo per eseguire lavori di elaborazione in AWS PCS](#)

Creare un profilo di istanza per AWS PCS

I gruppi di nodi di calcolo richiedono un profilo di istanza al momento della creazione. Se utilizzi il per AWS Management Console creare un ruolo per AmazonEC2, la console crea automaticamente un profilo di istanza e gli assegna lo stesso nome del ruolo. Per ulteriori informazioni, consulta [Usare i profili di istanza](#) nella Guida AWS Identity and Access Management per l'utente.

Nella procedura seguente, usi per creare un ruolo per AmazonEC2, che crea anche il profilo di istanza per i tuoi gruppi di nodi di calcolo. AWS Management Console

Per creare il ruolo e il profilo dell'istanza

- Passare alla [console IAM](#).
- In Gestione accessi scegli Policy.
 - Seleziona Create Policy (Crea policy).
 - In Specificare le autorizzazioni, per Policy editor, scegli JSON.
 - Sostituire il contenuto dell'editor di testo con quanto segue:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "pcs:RegisterComputeNodeGroupInstance"
      ],
      "Resource": "*",
      "Effect": "Allow"
    }
  ]
}
```

- Scegli Next (Successivo).
- In Rivedi e crea, per Nome della politica, inserisci `AWSPCS-getstarted-policy`.
- Scegli Create Policy (Crea policy).
- In Access management (Gestione accessi), scegli Roles (Ruoli).
- Scegliere Crea ruolo.
- In Seleziona entità attendibile:
 - Per il tipo di entità affidabile, seleziona AWS servizio
 - In Caso d'uso, seleziona EC2.
 - Quindi, in Scegli un caso d'uso per il servizio specificato, scegli EC2.
 - Scegli Next (Successivo).
- In Aggiungi autorizzazioni:
 - In Politiche di autorizzazione, cerca `AWSPCS-getstarted-policy`.
 - Seleziona la casella accanto a `AWSPCS-getstarted-policy` per aggiungerlo al ruolo.
 - Nelle politiche di autorizzazione, cerca `A. mazonSSMManaged InstanceCore`
 - Seleziona la casella accanto `mazonSSMManaged InstanceCore` a `A` per aggiungerlo al ruolo.

- Scegli Next (Successivo).
- In Nome, rivedi e crea:
 - In Dettagli del ruolo:
 - Per Nome ruolo, inserisci `AWSPCS-getstarted-role`.
- Scegliere Crea ruolo.

Crea modelli di lancio per AWS PCS

Quando crei un gruppo di nodi di calcolo, fornisci un modello di EC2 avvio da AWS PCS utilizzare per configurare EC2 le istanze da avviare. Ciò include impostazioni come gruppi di sicurezza e script che vengono eseguiti all'avvio dell'istanza.

In questo passaggio, verrà utilizzato un CloudFormation modello per creare due modelli di EC2 avvio. Un modello verrà utilizzato per creare nodi di accesso e l'altro verrà utilizzato per creare nodi di calcolo. La differenza fondamentale tra loro è che i nodi di accesso possono essere configurati per consentire l'accesso in entrataSSH.

Accedi al modello CloudFormation

Usa quanto segue URL per scaricare il CloudFormation modello, quindi carica il modello nella [AWS CloudFormation console](#) per creare un nuovo CloudFormation stack. Per ulteriori informazioni, consulta [Uso della AWS CloudFormation console](#) nella Guida per l'AWS CloudFormation utente.

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/getting_started/assets/pcs-1t-efs-fsx1.yaml
```

Usa il CloudFormation modello per creare modelli di EC2 lancio

Utilizza la procedura seguente per completare il CloudFormation modello nella AWS CloudFormation console

- In Fornisci un nome per lo stack:
 - In Nome dello stack, inserisci `getstarted-1t`
- In Parametri:
 - In Sicurezza
 - Per `VpcSecurityGroupId`, seleziona il gruppo di sicurezza denominato `default` nel tuo `clusterVPC`.

- Per `ClusterSecurityGroupId`, seleziona il gruppo denominato `cluster-getstarted-sg`
- Per `SshSecurityGroupId`, seleziona il gruppo denominato `inbound-ssh-getstarted-sg`
- Per `SshKeyName`, seleziona la tua SSH key pair preferita.
- In File system
 - Per `EfsFilesystemId`, inserisci l'ID del file system del EFS file system che hai creato in precedenza nel tutorial.
 - Ad `FSxLustreFilesystemIdesempio`, inserisci l'ID del file system del file system FSx for Lustre che hai creato in precedenza nel tutorial.
 - Per `FSxLustreFilesystemMountName`, inserisci il nome di montaggio corrispondente per il file system Lustre. FSx
- Scegliete Avanti, quindi scegliete nuovamente Avanti.
- Scegli Invia.

Monitora lo stato dello CloudFormation stack. Quando raggiunge `CREATE_COMPLETE` il modello di lancio è pronto per essere utilizzato.

Note

Per vedere tutte le risorse create dal CloudFormation modello, apri la [AWS CloudFormation console](#). Scegli lo stack `getstarted-1t`, quindi la scheda Resources (Risorse).

Crea un gruppo di nodi di calcolo per i nodi di accesso in AWS PCS

Un gruppo di nodi di calcolo è una raccolta virtuale di nodi di calcolo (EC2istanze) che AWS PCS viene avviata e gestita. Quando si definisce un gruppo di nodi di calcolo, si specificano caratteristiche comuni come i tipi di istanze, il numero minimo e massimo di EC2 istanze, le VPC sottoreti di destinazione, l'opzione di acquisto preferita e la configurazione di avvio personalizzata. AWS PCSavvia, gestisce e termina in modo efficiente i nodi di calcolo in un gruppo di nodi di calcolo, in base a queste impostazioni.

In questo passaggio, lancerai un gruppo di nodi di calcolo statici che fornisce l'accesso interattivo al cluster. Puoi utilizzare SSH o Amazon EC2 Systems Manager (SSM) per accedere, quindi eseguire comandi shell e gestire i job Slurm.

Per creare il gruppo di nodi di calcolo

- Apri la [AWS PCSconsole](#) e vai a Clusters.
- Seleziona il cluster denominato `get-started`
- Passa ai gruppi di nodi di calcolo e scegli Crea.
- Nella sezione Configurazione del gruppo di nodi di calcolo, fornisci quanto segue:
 - Nome del gruppo di nodi di calcolo: immettere. `login`
- In Configurazione informatica, inserisci o seleziona questi valori:
 - EC2modello di lancio: scegli il modello di lancio in cui si trova il nome `login-getstarted-1t`
 - IAMprofilo di istanza: scegli il profilo di istanza denominato `AWSPCS-getstarted-role`
 - Sottoreti: seleziona la sottorete da cui inizia il nome. `hpc-networking:PublicSubnetA`
 - Istanze: seleziona. `c6i.xlarge`
 - Configurazione di scalabilità: per il numero minimo di istanze, immettere. `1` Per Numero massimo di istanze, immettete. `1`
- In Impostazioni aggiuntive, specificate quanto segue:
 - AMIID: seleziona il AMI punto con cui inizia il nome `aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11`
- Scegli Crea gruppo di nodi di calcolo.

Il campo Stato mostra Creazione durante il provisioning del gruppo di nodi di calcolo. Puoi procedere al passaggio successivo del tutorial mentre è in corso.

Crea un gruppo di nodi di calcolo per eseguire lavori di elaborazione in AWS PCS


In questo passaggio, lancerai un gruppo di nodi di calcolo con scalabilità elastica per eseguire i lavori inviati al cluster.

Per creare il gruppo di nodi di calcolo

- Apri la [AWS PCSconsole](#) e vai a Clusters.
- Seleziona il cluster denominato `get-started`
- Vai ai gruppi di nodi di calcolo e scegli Crea.
- Nella sezione Configurazione del gruppo di nodi di calcolo, fornisci quanto segue:

- Nome del gruppo di nodi di calcolo: immettere. `compute-1`
- In Configurazione informatica, inserisci o seleziona questi valori:
 - EC2 modello di lancio: scegli il modello di lancio in cui si trova il nome `compute-getstarted-1t`
 - IAM profilo di istanza: scegli il profilo di istanza denominato `AWSPCS-getstarted-role`
 - Sottoreti: seleziona la sottorete da cui inizia il nome. `hpc-networking:PrivateSubnetA`
 - Istanze: seleziona. `c6i.xlarge`
 - Configurazione di scalabilità: per il numero minimo di istanze, immettere. `0` Per Numero massimo di istanze, immettete. `4`
- In Impostazioni aggiuntive, specificate quanto segue:
 - AMIID: seleziona il AMI punto con cui inizia il nome `aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11`.
- Scegli Crea gruppo di nodi di calcolo.

Il campo Stato mostra Creazione durante il provisioning del gruppo di nodi di calcolo.

 Important

Attendi che il campo Stato mostri Attivo prima di procedere al passaggio successivo di questo tutorial.

Crea una coda per gestire i lavori in AWS PCS

Si invia un lavoro a una coda per eseguirlo. Il lavoro rimane in coda finché non ne AWS PCS pianifica l'esecuzione su un gruppo di nodi di calcolo. Ogni coda è associata a uno o più gruppi di nodi di calcolo, che forniscono le EC2 istanze necessarie per eseguire l'elaborazione.

In questo passaggio, creerai una coda che utilizza il gruppo di nodi di calcolo per elaborare i lavori.

Per creare una coda

- [Apri la console AWS PCS](#)
- Seleziona il cluster denominato `get-started`.
- Passa ai gruppi di nodi di calcolo e assicurati che lo stato del `compute-1` gruppo sia Attivo.

⚠ Important

Lo stato del compute-1 gruppo deve essere Attivo prima di procedere al passaggio successivo.

- Vai a Code e scegli Crea coda.
 - Nella sezione Configurazione della coda, fornisci i seguenti valori:
 - Nome della coda: inserisci quanto segue: demo
 - Gruppi di nodi di calcolo: seleziona il gruppo di nodi di calcolo denominato. compute-1
- Scegliere Crea coda.

Il campo Stato mostra Creazione durante la creazione della coda.

⚠ Important

Attendi che il campo Stato mostri Attivo prima di procedere al passaggio successivo di questo tutorial.

Connect al AWS PCS cluster

Dopo che lo stato del gruppo di nodi di login calcolo diventa Attivo, puoi connetterti all'EC2istanza che ha creato.

Per connettersi al nodo di accesso

- Apri la [AWS PCSconsole](#) e vai a Clusters.
- Seleziona il cluster denominato get-started.
- Scegli Gruppi di nodi Compute.
- Passa al gruppo di nodi di calcolo denominato. login
- Trova l'ID del gruppo di nodi Compute.
- In un'altra finestra o scheda del browser, apri la [EC2console Amazon](#).
 - Seleziona Instances (Istanze).
 - Cerca le EC2 istanze con il tag seguente. Replace (Sostituisci) *node-group-id* con il valore dell'ID del gruppo di nodi Compute del passaggio precedente. Dovrebbe esserci 1 istanza.

```
aws:pcs:compute-node-group-id=node-group-id
```

- Connect all'EC2istanza. È possibile utilizzare Session Manager oSSH.

Session Manager

- Selezionare l'istanza.
- Scegli Connetti.
- In Connect to instance, seleziona Session Manager.
- Scegli Connetti.
- Scegli Connetti. Nel browser viene avviato un terminale interattivo.

SSH

- Selezionare l'istanza.
- Scegli Connetti.
- In Connect to instance, seleziona SSHclient.
- Segui le istruzioni fornite dalla console.

Note

Il nome utente dell'istanza **ec2-user** non lo è **root**.

Esplora l'ambiente del cluster in AWS PCS

Dopo aver effettuato l'accesso al cluster, puoi eseguire i comandi della shell. Ad esempio, puoi cambiare utente, lavorare con i dati su file system condivisi e interagire con Slurm.

Cambia utente

Se hai effettuato l'accesso al cluster utilizzando Session Manager, potresti essere connesso come **comessm-user**. Si tratta di un utente speciale creato per Session Manager. Passa all'utente predefinito su Amazon Linux 2 utilizzando il seguente comando. Non sarà necessario farlo se ti sei connesso tramiteSSH.

```
sudo su - ec2-user
```

Lavora con file system condivisi

È possibile confermare che il EFS filesystem e i file system FSx for Lustre sono disponibili con il comando. `df -h` L'output sul cluster dovrebbe essere simile al seguente:

```
[ec2-user@ip-10-3-6-103 ~]$ df -h
Filesystem                Size      Used Avail Use% Mounted on
devtmpfs                  3.8G         0  3.8G   0% /dev
tmpfs                     3.9G         0  3.9G   0% /dev/shm
tmpfs                     3.9G   556K  3.9G   1% /run
tmpfs                     3.9G         0  3.9G   0% /sys/fs/cgroup
/dev/nvme0n1p1            24G       18G   6.6G  73% /
127.0.0.1:/                8.0E         0  8.0E   0% /home
10.3.132.79@tcp:/z1shxbev  1.2T     7.5M  1.2T   1% /shared
tmpfs                     780M         0  780M   0% /run/user/0
tmpfs                     780M         0  780M   0% /run/user/1000
```

Il `/home` filesystem monta `127.0.0.1` e ha una capacità molto grande. Questo è il EFS file system che avete creato in precedenza nel tutorial. Tutti i file scritti qui saranno disponibili `/home` in tutti i nodi del cluster.

Il `/shared` filesystem monta un IP privato e ha una capacità di 1,2 TB. Questo è il file system FSx for Lustre creato in precedenza nel tutorial. Tutti i file scritti qui saranno disponibili `/shared` in tutti i nodi del cluster.

Interagisci con Slurm

Argomenti

- [Elenca code e nodi](#)
- [Mostra offerte di lavoro](#)

Elenca code e nodi

È possibile elencare le code e i nodi a cui sono associate. `sinfo` L'output del cluster dovrebbe essere simile al seguente:

```
[ec2-user@ip-10-3-6-103 ~]$ sinfo
PARTITION AVAIL  TIMELIMIT  NODES  STATE NODELIST
demo      up    infinite     4  idle~ compute-1-[1-4]
```

```
[ec2-user@ip-10-3-6-103 ~]$
```

Notate la partizione denominata. demo Il suo stato è up e ha un massimo di 4 nodi. È associato ai nodi del gruppo di compute-1 nodi. Se modifichi il gruppo di nodi di calcolo e aumenti il numero massimo di istanze a 8, verrà letto il numero di nodi 8 e verrà letto l'elenco dei nodi. compute-1- [1-8] Se creassi un secondo gruppo di nodi di calcolo denominato test con 4 nodi e lo aggiungessi alla demo coda, tali nodi verranno visualizzati anche nell'elenco dei nodi.

Mostra offerte di lavoro

Puoi elencare tutti i lavori, in qualsiasi stato, sul sistema consequue. L'output del cluster dovrebbe essere simile al seguente:

```
[ec2-user@ip-10-3-6-103 ~]$ squeue
JOBID PARTITION NAME USER ST TIME NODES NODELIST(REASON)
```

Prova a eseguire squeue di nuovo più tardi, quando hai un job Slurm in sospeso o in esecuzione.

Esegui un processo a nodo singolo in AWS PCS

Per eseguire un lavoro utilizzando Slurm, si prepara uno script di invio che specifica i requisiti del lavoro e lo si invia a una coda con il comando. sbatch In genere, questa operazione viene eseguita da una directory condivisa in modo che i nodi di accesso e di calcolo abbiano uno spazio comune per l'accesso ai file.

Connect al nodo di login del cluster ed esegui i seguenti comandi al prompt della shell.

- Diventa l'utente predefinito. Passa alla directory condivisa.

```
sudo su - ec2-user
cd /shared
```

- Utilizzate i seguenti comandi per creare uno script di lavoro di esempio:

```
cat << EOF > job.sh
#!/bin/bash
#SBATCH -J single
#SBATCH -o single.%j.out
#SBATCH -e single.%j.err
```

```
echo "This is job \${SLURM_JOB_NAME} [\${SLURM_JOB_ID}] running on \
\${SLURMD_NODENAME}, submitted from \${SLURM_SUBMIT_HOST}" && sleep 60 && echo "Job
complete"
EOF
```

- Invia lo script di lavoro allo scheduler Slurm:

```
sbatch -p demo job.sh
```

- Quando il lavoro viene inviato, restituirà un ID del lavoro come numero. Usa quell'ID per controllare lo stato del lavoro. Replace (Sostituisci) *job-id* nel comando seguente con il numero restituito da `sbatch`.

```
squeue --job job-id
```

Example

```
squeue --job 1
```

Il `squeue` comando restituisce un output simile al seguente:

```
JOBID PARTITION NAME USER      ST TIME NODES NODELIST(REASON)
1      demo      test ec2-user CF 0:47 1      compute-1
```

- Continuare a controllare lo stato del processo finché non raggiunge lo stato R (in esecuzione). Il lavoro è terminato quando `squeue` non restituisce nulla.
- Ispeziona il contenuto della `/shared` directory.

```
ls -alth /shared
```

L'output del comando è simile al seguente:

```
-rw-rw-r- 1 ec2-user ec2-user 107 Mar 19 18:33 single.1.out
-rw-rw-r- 1 ec2-user ec2-user 0 Mar 19 18:32 single.1.err
-rw-rw-r- 1 ec2-user ec2-user 381 Mar 19 18:29 job.sh
```

I file `single.1.err` denominati `single.1.out` e scritti da uno dei nodi di calcolo del cluster. Poiché il processo è stato eseguito in una directory condivisa (`/shared`), sono disponibili anche

nel nodo di accesso. Questo è il motivo per cui hai configurato un file system FSx for Lustre per questo cluster.

- Ispeziona il contenuto del `single.1.out` file.

```
cat /shared/single.1.out
```

L'output è simile a quello riportato di seguito:

```
This is job test [1] running on compute-1, submitted from ip-10-3-13-181
Job complete
```

Esegui un MPI lavoro multinodo con Slurm in AWS PCS

Queste istruzioni dimostrano l'utilizzo di Slurm per eseguire un job di invio di messaggi all'interfaccia (). MPI AWS PCS

Esegui i seguenti comandi al prompt della shell del tuo nodo di accesso.

- Diventa l'utente predefinito. Passa alla sua home directory.

```
sudo su - ec2-user
cd ~/
```

- Crea codice sorgente nel linguaggio di programmazione C.

```
cat > hello.c << EOF
// * mpi-hello-world - https://www.mpitutorial.com
// Released under MIT License
//
// Copyright (c) 2014 MPI Tutorial.
//
// Permission is hereby granted, free of charge, to any person obtaining a copy
// of this software and associated documentation files (the "Software"), to
// deal in the Software without restriction, including without limitation the
// rights to use, copy, modify, merge, publish, distribute, sublicense, and/or
// sell copies of the Software, and to permit persons to whom the Software is
// furnished to do so, subject to the following conditions:
// The above copyright notice and this permission notice shall be included in
// all copies or substantial portions of the Software.
//
```

```
// THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
// IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
// FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
// AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
// LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING
// FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER
// DEALINGS IN THE SOFTWARE.

#include <mpi.h>
#include <stdio.h>
#include <stddef.h>

int main(int argc, char** argv) {
    // Initialize the MPI environment. The two arguments to MPI Init are not
    // currently used by MPI implementations, but are there in case future
    // implementations might need the arguments.
    MPI_Init(NULL, NULL);

    // Get the number of processes
    int world_size;
    MPI_Comm_size(MPI_COMM_WORLD, &world_size);

    // Get the rank of the process
    int world_rank;
    MPI_Comm_rank(MPI_COMM_WORLD, &world_rank);

    // Get the name of the processor
    char processor_name[MPI_MAX_PROCESSOR_NAME];
    int name_len;
    MPI_Get_processor_name(processor_name, &name_len);

    // Print off a hello world message
    printf("Hello world from processor %s, rank %d out of %d processors\n",
           processor_name, world_rank, world_size);

    // Finalize the MPI environment. No more MPI calls can be made after this
    MPI_Finalize();
}
EOF
```

- Caricate il MPI modulo Open.


```
module load openmpi
```

- Compila il programma C.

```
mpicc -o hello hello.c
```

- Scrivi uno script di invio di lavori Slurm.

```
cat > hello.sh << EOF
#!/bin/bash
#SBATCH -J multi
#SBATCH -o multi.out
#SBATCH -e multi.err
#SBATCH --exclusive
#SBATCH --nodes=4
#SBATCH --ntasks-per-node=1

srun $HOME/hello
EOF
```

- Passa alla directory condivisa.

```
cd /shared
```

- Invia lo script del lavoro.

```
sbatch -p demo ~/hello.sh
```

- Utilizzatelo squeue per monitorare il lavoro fino al termine.
- Controlla il contenuto di `multi.out`:

```
cat multi.out
```

L'output è simile a quello riportato di seguito. Nota che ogni rank ha il proprio indirizzo IP perché è stato eseguito su un nodo diverso.

```
Hello world from processor ip-10-3-133-204, rank 0 out of 4 processors
Hello world from processor ip-10-3-128-219, rank 2 out of 4 processors
Hello world from processor ip-10-3-141-26, rank 3 out of 4 processors
Hello world from processor ip-10-3-143-52, rank 1 out of 4 processor
```

Elimina le tue AWS risorse per AWS PCS

Dopo aver finito con i gruppi di cluster e nodi che hai creato per questo tutorial, dovresti eliminare le risorse che hai creato.

Important

Ti verranno addebitati i costi di fatturazione per tutte le risorse in esecuzione nel Account AWS

Per eliminare AWS PCS le risorse che hai creato per questo tutorial

- Apri la [AWS PCSconsole](#).
- Passa al cluster denominato get-started.
- Vai alla sezione Code.
- Seleziona la coda denominata demo.
- Scegli Elimina.

Important


Attendi che la coda sia stata eliminata prima di procedere.

- Vai alla sezione Compute node groups.
- Seleziona il gruppo di nodi di calcolo denominato compute-1.
- Scegli Elimina.
- Seleziona il gruppo di nodi di calcolo denominato login.
- Scegli Elimina.

Important

Attendi che entrambi i gruppi di nodi di calcolo siano stati eliminati prima di procedere.

- Nella pagina dei dettagli del cluster per iniziare, scegli Elimina.

 Important

Attendi che il cluster sia stato eliminato prima di procedere con i passaggi successivi.


Per eliminare altre AWS risorse che hai creato per questo tutorial

- Apri la [IAMconsole](#).
 - Scegli Ruoli.
 - Seleziona il ruolo denominato AWSPCS-getstarted-role, quindi scegli Elimina.
 - Dopo che il ruolo è stato eliminato, scegli Politiche.
 - Seleziona la politica denominata AWSPCS-getstarted-policy, quindi scegli Elimina.
- Apri la [AWS CloudFormation console](#).
 - Seleziona lo stack denominato getstarted-It.
 - Scegli Elimina.

 Important


Attendi che lo stack venga eliminato prima di procedere.

- Apri la [EFSconsole Amazon](#).
 - Seleziona File system.
 - Seleziona il file system denominato getstarted-efs.
 - Scegli Elimina.

 Important

Attendi che il file system venga eliminato prima di procedere.

- Apri la [FSxconsole Amazon](#).
 - Seleziona File system.
 - Seleziona il file system denominato getstarted-fsx.
 - Scegli Elimina.

 Important

Attendi che il file system venga eliminato prima di procedere.

- Apri la [AWS CloudFormation console](#).
 - Seleziona lo stack denominato getstarted-sg.
 - Scegli Elimina.
- Apri la [AWS CloudFormation console](#).
 - Seleziona lo stack denominato hpc-networking.
 - Scegli Delete (Elimina).

Lavorare con AWS PCS

Questo capitolo fornisce informazioni e linee guida per aiutarti a utilizzare AWS PCS.

Argomenti

- [AWS PCSgrappoli](#)
- [AWS PCSgruppi di nodi di calcolo](#)
- [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#)
- [AWS PCScode](#)
- [AWS PCSnodi di accesso](#)
- [AWS PCSRete](#)
- [Utilizzo di file system di rete con AWS PCS](#)
- [Amazon Machine Images \(AMIs\) per AWS PCS](#)
- [Versioni Slurm in AWS PCS](#)

AWS PCSgrappoli

Un AWS PCS cluster è costituito dai seguenti componenti:

- Istanze gestite del software di pianificazione del HPC sistema, come il daemon di controllo Slurm (`slurmctld`)
- Componenti che si integrano con lo scheduler di HPC sistema per il provisioning e la gestione delle EC2 istanze Amazon.
- Componenti che si integrano con lo scheduler di HPC sistema per trasmettere log e metriche ad Amazon. CloudWatch

Questi componenti vengono eseguiti in un account gestito da AWS Collaborano per gestire le EC2 istanze Amazon nel tuo account cliente. AWS PCSfornisce interfacce di rete elastiche nella tua VPC sottorete Amazon per fornire la connettività dal software di pianificazione alle EC2 istanze Amazon (ad esempio, per supportare la pianificazione di lavori in batch su di esse e consentire agli utenti di eseguire comandi di pianificazione per elencare e gestire tali lavori).

Argomenti

- [Creazione di un cluster in AWS Parallel Computing Service](#)

- [Eliminazione di un cluster in AWS PCS](#)
- [Scelta della dimensione del AWS PCS cluster](#)
- [Utilizzo dei segreti del cluster in AWS PCS](#)

Creazione di un cluster in AWS Parallel Computing Service

Questo argomento fornisce una panoramica delle opzioni disponibili e descrive gli elementi da considerare quando si crea un cluster in AWS Parallel Computing Service (AWS PCS). Se è la prima volta che crei un AWS PCS cluster, ti consigliamo di seguirlo [Guida introduttiva con AWS PCS](#). Il tutorial può aiutarti a creare un HPC sistema funzionante senza approfondire tutte le opzioni disponibili e le architetture di sistema possibili.

Prerequisiti

- Una sottorete esistente VPC e che soddisfa i requisiti. [AWS PCS Rete](#) Prima di implementare un cluster per l'uso in produzione, è consigliabile avere una conoscenza approfondita dei requisiti VPC e delle sottoreti. Per creare una sottorete VPC e, vedi. [Creare un file VPC per il tuo AWS PCS cluster](#)
- Un [IAM preside](#) con i permessi per creare e gestire AWS PCS risorse. Per ulteriori informazioni, consulta [Servizio di Identity and Access Management per AWS Parallel Computing](#).

Crea un cluster AWS PCS

È possibile utilizzare AWS Management Console o AWS CLI per creare un cluster.

AWS Management Console

Come creare un cluster

1. Apri la AWS PCS console in <https://console.aws.amazon.com/pcs/home#/clusters> e scegli Crea cluster.
2. Nella sezione Configurazione del cluster, inserisci i seguenti campi:
 - Nome del cluster: un nome per il cluster. Il nome può contenere solo caratteri alfanumerici (con distinzione tra lettere maiuscole e minuscole) e trattini. Deve iniziare con un carattere alfabetico e non può superare i 40 caratteri. Il nome deve essere univoco all'interno del Regione AWS e in Account AWS cui si sta creando il cluster.

- Scheduler: scegli uno scheduler e una versione. AWS PCS attualmente supporta Slurm 23.11. Per ulteriori informazioni, consulta [Versioni Slurm in AWS PCS](#).
 - Dimensioni del controller: scegli una dimensione per il controller. Ciò determina il numero di processi e nodi di elaborazione simultanei che il AWS PCS cluster può gestire. È possibile impostare la dimensione del controller solo al momento della creazione del cluster. Per ulteriori informazioni sul dimensionamento, vedere [Scelta della dimensione del AWS PCS cluster](#).
3. Nella sezione Rete, selezionate i valori per i seguenti campi:
- VPC— Scegliete un sistema esistente VPC che soddisfi AWS PCS i requisiti. Per ulteriori informazioni, consulta [AWS PCS VPCe requisiti e considerazioni sulla sottorete](#). Dopo aver creato il cluster, non è possibile modificarlo VPC. Se non VPCs ne è elencato nessuno, devi prima crearne uno.
 - Subnet: vengono elencate tutte le sottoreti disponibili nell'area selezionata VPC. Scegline due in zone di disponibilità diverse. Ogni sottorete deve soddisfare i requisiti della AWS PCS sottorete. Per ulteriori informazioni, consulta [AWS PCS VPCe requisiti e considerazioni sulla sottorete](#). Ti consigliamo di selezionare una sottorete privata per evitare di esporre gli endpoint dello scheduler alla rete Internet pubblica.
 - Gruppi di sicurezza: specifica i gruppi di sicurezza che desideri associare AWS PCS alle interfacce di rete che crea per il tuo cluster. È necessario selezionare almeno un gruppo di sicurezza che consenta la comunicazione tra il cluster e i relativi nodi di elaborazione. Per ulteriori informazioni, consulta [Requisiti e considerazioni sui gruppi di sicurezza](#).
4. (Facoltativo) In Crittografia, puoi definire una chiave personalizzata per crittografare i dati del controller impostando questi campi:
- KMSID chiave: lascia che usi aws/pcs la KMS chiave che PCS crea. Seleziona un alias di KMS chiave esistente per utilizzare una KMS chiave personalizzata. Tieni presente che l'account utilizzato per creare il cluster deve disporre kms : Decrypt dei privilegi sulla chiave personalizzata KMS.
5. (Facoltativo) Nella sezione di configurazione Slurm, è possibile specificare le opzioni di configurazione Slurm che sostituiscono i valori predefiniti impostati da: AWS PCS
- Ridimensiona il tempo di inattività: controlla per quanto tempo i nodi di elaborazione con provisioning dinamico rimangono attivi dopo il completamento o la fine dei lavori su di essi assegnati. L'impostazione di questo valore su un valore più lungo può aumentare la probabilità che un processo successivo possa essere eseguito sul nodo, ma può

comportare un aumento dei costi. Un valore più breve ridurrà i costi, ma potrebbe aumentare la percentuale di tempo che il HPC sistema impiega per il provisioning dei nodi anziché per l'esecuzione dei job su di essi.

- Prolog: si tratta di un percorso completo per accedere a una directory di script Prolog sulle istanze del gruppo di nodi di calcolo. [Corrisponde all'impostazione Prolog in Slurm](#). Nota che questa deve essere una directory, non un percorso verso un eseguibile specifico.
 - Epilog: si tratta di un percorso completo verso una directory di script di epilog sulle istanze del gruppo di nodi di calcolo. [Corrisponde all'impostazione Epilog in Slurm](#). Nota che questa deve essere una directory, non il percorso di un eseguibile specifico.
 - Seleziona i parametri del tipo: questo aiuta a controllare l'algoritmo di selezione delle risorse utilizzato da Slurm. L'impostazione di questo valore su CR_CPU_Memory attiverà la pianificazione basata sulla memoria, mentre impostandolo su attiverà la sola pianificazione. CR_CPU CPU Questo parametro corrisponde all'[SelectTypeParameters](#) impostazione in Slurm dove è impostata su by. SelectType select/cons_tres AWS PCS
6. (Facoltativo) In Tag, aggiungi qualsiasi tag al tuo AWS PCS cluster.
 7. Scegli Create cluster (Crea cluster). Il campo Status viene visualizzato Creating durante la AWS PCS creazione del cluster. Questo processo può richiedere alcuni minuti.

Important

Può esserci solo 1 cluster in uno Creating stato Regione AWS per ogni stato Account AWS. AWS PCS restituisce un errore se c'è già un cluster in uno Creating stato quando si tenta di creare un cluster.


AWS CLI

Come creare un cluster

1. Crea un cluster con il comando seguente. Prima di eseguire il comando, apporta le modifiche seguenti:
 - Replace (Sostituisci) *region* con l'ID del Regione AWS quale vuoi creare il cluster, ad esempio `us-east-1`.
 - Replace (Sostituisci) *my-cluster* con un nome per il tuo cluster. Il nome può contenere solo caratteri alfanumerici (con distinzione tra lettere maiuscole e minuscole) e trattini.

Deve iniziare con un carattere alfabetico e non può essere più lungo di 40 caratteri. Il nome deve essere univoco all'interno Regione AWS e nel Account AWS luogo in cui si sta creando il cluster.

- Replace (Sostituisci) **23.11** con qualsiasi versione supportata di Slurm.

 Note

AWS PCS attualmente supporta Slurm 23.11.

- Replace (Sostituisci) **SMALL** con qualsiasi dimensione di cluster supportata. Ciò determina il numero di processi e nodi di calcolo simultanei che il AWS PCS cluster può gestire. Può essere impostato solo al momento della creazione del cluster. Per ulteriori informazioni sul dimensionamento, vedere [Scelta della dimensione del AWS PCS cluster](#).
- Sostituisci il valore di `subnetIds` con il tuo. Ti consigliamo di selezionare una sottorete privata per evitare di esporre gli endpoint dello scheduler alla rete Internet pubblica.
- Specificate `securityGroupIds` quella che desiderate associare AWS PCS alle interfacce di rete che crea per il cluster. I gruppi di sicurezza devono trovarsi nello VPC stesso gruppo del cluster. È necessario selezionare almeno un gruppo di sicurezza che consenta la comunicazione tra il cluster e i relativi nodi di elaborazione. Per ulteriori informazioni, consulta [Requisiti e considerazioni sui gruppi di sicurezza](#).
- Facoltativamente, puoi ottimizzare il comportamento di Slurm aggiungendo un'opzione. `--slurm-configuration` Ad esempio, è possibile impostare il tempo di inattività per la riduzione della scala su 60 minuti (3600 secondi) con. `--slurm configuration scaleDownIdleTime=3600`
- Facoltativamente, puoi fornire una KMS chiave personalizzata per crittografare i dati del controller utilizzando. `--kms-key-id` **kms-key** Sostituiscilo **kms-key** con un ID chiave o un alias esistente KMSARN. Tieni presente che l'account utilizzato per creare il cluster deve disporre `kms:Decrypt` dei privilegi sulla chiave personalizzata KMS.

```
aws pcs create-cluster --region region \  
  --cluster-name my-cluster \  
  --scheduler type=SLURM,version=23.11 \  
  --size SMALL \  
  --networking subnetIds=subnet-ExampleId1,securityGroupIds=sg-ExampleId1
```

2. Il provisioning del cluster può richiedere diversi minuti. È possibile eseguire query sullo stato del cluster con il comando seguente. Non procedere alla creazione di code o gruppi di nodi di calcolo finché non viene visualizzato il campo di stato del cluster. ACTIVE

```
aws pcs get-cluster --region region --cluster-identifier my-cluster
```

Important

Può esserci solo 1 cluster in uno `Creating` stato per ogni stato. Regione AWS Account AWS PCS restituisce un errore se c'è già un cluster in uno `Creating` stato quando si tenta di creare un cluster.

Passaggi successivi consigliati per il cluster

- Aggiungi gruppi di nodi di calcolo.
- Aggiungi code.
- Attivare la registrazione nel log.

Eliminazione di un cluster in AWS PCS

Questo argomento fornisce una panoramica su come eliminare un AWS PCS cluster.

Considerazioni sull'eliminazione di un cluster AWS PCS

- Tutte le code associate al cluster devono essere eliminate prima che il cluster possa essere eliminato. Per ulteriori informazioni, consulta [Eliminazione di una coda in AWS PCS](#).
- Tutti i gruppi di nodi di calcolo associati al cluster devono essere eliminati prima che il cluster possa essere eliminato. Per ulteriori informazioni, consulta [Eliminazione di un gruppo di nodi di calcolo in AWS PCS](#).

Eliminare il cluster

È possibile utilizzare AWS Management Console o AWS CLI per eliminare un cluster.

AWS Management Console

Per eliminare un cluster

1. Aprire la [AWS PCS console](#).
2. Seleziona il cluster da eliminare.
3. Scegli Elimina.
4. Viene visualizzato il campo Stato del cluster `Deleting`. Per il completamento possono essere necessari alcuni minuti.

AWS CLI

Per eliminare un cluster

1. Utilizzate il seguente comando per eliminare un cluster, con queste sostituzioni:
 - Replace (Sostituisci) *region-code* con Regione AWS il cluster in uso.
 - Replace (Sostituisci) *my-cluster* con il nome o l'ID del tuo cluster.

```
aws pcs delete-cluster --region region-code --cluster-identifier my-cluster
```

2. L'eliminazione del cluster può richiedere diversi minuti. Puoi controllare lo stato del tuo cluster con il seguente comando.

```
aws pcs get-cluster --region region-code --cluster-identifier my-cluster
```

Scelta della dimensione del AWS PCS cluster

AWS PCS fornisce cluster ad alta disponibilità e sicuri, automatizzando al contempo attività chiave come l'applicazione di patch, il provisioning dei nodi e gli aggiornamenti.

Quando si crea un cluster, si seleziona una dimensione in base a due fattori:

- Il numero di nodi di elaborazione che gestirà
- Il numero di processi attivi e in coda che si prevede di eseguire nel cluster

Dimensione del cluster Slurm	Numero di istanze gestite	Numero di lavori attivi e in coda
Small	Fino a 32	Fino a 256
Media	Fino a 512	Fino a 8192
Large	Fino al 2048	Fino a 16384

Esempi

- Se il tuo cluster avrà fino a 24 istanze gestite ed eseguirà fino a 100 job, scegli Small.
- Se il cluster avrà fino a 24 istanze gestite e gestirà fino a 1000 job, scegli Medium.
- Se il tuo cluster avrà fino a 1000 istanze gestite e gestirà fino a 100 job, scegli Large.
- Se il tuo cluster avrà fino a 1000 istanze gestite e gestirà fino a 10.000 job, scegli Large.

Utilizzo dei segreti del cluster in AWS PCS

Come parte della creazione di un cluster, AWS PCS crea un cluster secret necessario per connettersi al job scheduler del cluster. È inoltre possibile creare gruppi di nodi di AWS PCS calcolo, che definiscono set di istanze da avviare in risposta a eventi di scalabilità. AWS PCS configura le istanze lanciate da tali gruppi di nodi di calcolo con il cluster secret in modo che possano connettersi al job scheduler. Ci sono casi in cui potresti voler configurare i client Slurm manualmente. Gli esempi includono la creazione di un nodo di accesso persistente o la configurazione di un gestore del flusso di lavoro con funzionalità di gestione dei lavori.

AWS PCS memorizza il segreto del cluster come [segreto gestito](#) con il prefisso pcs ! in AWS Secrets Manager. Il costo del segreto è incluso nel costo di utilizzo AWS PCS.

Warning

Non modificare il segreto del cluster. AWS PCS non sarà in grado di comunicare con il cluster se si modifica il segreto del cluster. AWS PCS non supporta la rotazione del segreto del cluster. È necessario creare un nuovo cluster se è necessario modificare il segreto del cluster.

Indice

- [Trova il segreto del cluster Slurm](#)
 - [Usa AWS Secrets Manager per trovare il segreto del cluster](#)
 - [Usa AWS PCS per trovare il segreto del cluster](#)
- [Ottieni il segreto del cluster Slurm](#)

Trova il segreto del cluster Slurm

Puoi trovare i segreti AWS PCS gestiti utilizzando la AWS Secrets Manager console oAPI, direttamente da AWS PCS, o utilizzando i tag.

Usa AWS Secrets Manager per trovare il segreto del cluster

AWS Management Console

1. Vai alla [console Secrets Manager](#).
2. Scegli Segreti, quindi cerca il pcs ! prefisso.

Note

Un segreto AWS PCS del cluster ha un nome nel formato in `pcs!slurm-secret-cluster-id` cui *cluster-id* è l'ID del AWS PCS cluster.

AWS CLI

Ogni segreto AWS PCS del cluster è inoltre contrassegnato con `aws:pcs:cluster-id`. È possibile ottenere l'ID segreto di un cluster con il comando seguente. Effettua queste sostituzioni prima di eseguire il comando:

- Sostituisci *region* con il Regione AWS per creare il cluster, ad esempio. `us-east-1`
- Sostituiscilo *cluster-id* con l'ID del AWS PCS cluster per trovare il segreto del cluster.

```
aws secretsmanager list-secrets \  
  --region region \  
  --filters Key=tag-key,Values=aws:pcs:cluster-id \  
  --
```

```
Key=tag-value,Values=cluster-id
```

Usa AWS PCS per trovare il segreto del cluster

È possibile utilizzare il AWS CLI per trovare il ARN segreto di un AWS PCS cluster. Immettete il comando seguente, effettuando le seguenti sostituzioni:

- Sostituisci *region* con il Regione AWS per creare il tuo cluster, ad esempio. `us-east-1`
- Sostituiscilo *my-cluster* con il nome o l'identificatore del cluster.

```
aws pcs get-cluster --region region --cluster-identifier my-cluster
```

L'output di esempio seguente proviene dal `get-cluster` comando. Potete usare `secretArn` e `secretVersion` insieme per ottenere il segreto.

```
{
  "cluster": {
    "name": "pcsdemo",
    "id": "s3431v9rx2",
    "arn": "arn:aws:pcs:us-east-1:012345678901:cluster/s3431v9rx2",
    "status": "ACTIVE",
    "createdAt": "2024-07-12T15:32:27.225136+00:00",
    "modifiedAt": "2024-07-12T15:32:27.225136+00:00",
    "scheduler": {
      "type": "SLURM",
      "version": "23.11"
    },
    "size": "SMALL",
    "networking": {
      "subnetIds": [
        "subnet-0123456789abcdef"
      ],
      "securityGroupIds": [
        "sg-0123456789abcde"
      ]
    },
    "endpoints": [
      {
        "type": "SLURMCTLD",
        "privateIpAddress": "127.0.0.1",

```

```

        "port": "6817"
      }
    ],
    "secretArn": "arn:aws:secretsmanager:us-east-1:012345678901:secret:pcs!slurm-
secret-s3431v9rx2-FN7tJF",
    "secretVersion": "ff58d1fd-070e-4bbc-98a0-64ef967cebcc"
  }
}

```

Ottieni il segreto del cluster Slurm

È possibile utilizzare Secrets Manager per ottenere la versione corrente con codifica base64 di un segreto del cluster Slurm. L'esempio seguente utilizza il AWS CLI Effettua le seguenti sostituzioni prima di eseguire il comando.

- Sostituisci *region* con il Regione AWS per creare il cluster, ad esempio. `us-east-1`
- Sostituisci *secret-arn* con il `secretArn` da un AWS PCS cluster.

```

aws secretsmanager get-secret-value \
  --region region \
  --secret-id 'secret-arn' \
  --version-stage AWSCURRENT \
  --query 'SecretString' \
  --output text

```

Per informazioni su come utilizzare il segreto del cluster Slurm, vedere. [Utilizzo di istanze autonome come AWS PCS nodi di accesso](#)

Autorizzazioni

Si utilizza un IAM principal per ottenere il segreto del cluster Slurm. Il IAM preside deve avere il permesso di leggere il segreto. Per ulteriori informazioni, consulta [i termini e i concetti relativi ai ruoli](#) nella Guida AWS Identity and Access Management per l'utente.

La seguente IAM policy di esempio consente l'accesso a un cluster secret di esempio.

```

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

```

```
{
  "Sid": "AllowSecretValueRetrievalAndVersionListing",
  "Effect": "Allow",
  "Action": [
    "secretsmanager:GetSecretValue",
    "secretsmanager:ListSecretVersionIds"
  ],
  "Resource": "arn:aws:secretsmanager:us-east-1:012345678901:secret:pcs!
slurm-secret-s3431v9rx2-FN7tJF"
}
```

AWS PCSgruppi di nodi di calcolo

Un gruppo di nodi di AWS PCS calcolo è una raccolta logica di nodi (EC2istanze Amazon). Questi nodi possono essere utilizzati per eseguire lavori di elaborazione e per fornire un accesso interattivo basato su shell a un sistema. HPC Un gruppo di nodi di calcolo è costituito da regole per la creazione di nodi, tra cui quali tipi di EC2 istanze Amazon utilizzare, quante istanze eseguire, se utilizzare istanze Spot o istanze On-demand, quali sottoreti e gruppi di sicurezza utilizzare e come configurare ogni istanza all'avvio. Quando tali regole vengono aggiornate, aggiorna le risorse associate al gruppo di nodi di calcolo in modo che AWS PCS corrispondano.

Argomenti

- [Creazione di un gruppo di nodi di calcolo in AWS PCS](#)
- [Aggiornamento di un gruppo di nodi di AWS PCS calcolo](#)
- [Eliminazione di un gruppo di nodi di calcolo in AWS PCS](#)
- [Ricerca di istanze di gruppi di nodi di calcolo in AWS PCS](#)

Creazione di un gruppo di nodi di calcolo in AWS PCS

Questo argomento fornisce una panoramica delle opzioni disponibili e descrive gli elementi da considerare quando si crea un gruppo di nodi di calcolo in AWS Parallel Computing Service (AWS PCS). Se è la prima volta che crei un gruppo di nodi di calcolo in AWS PCS, ti consigliamo di seguire il tutorial in [Guida introduttiva con AWS PCS](#). Il tutorial può aiutarti a creare un HPC sistema funzionante senza approfondire tutte le opzioni disponibili e le architetture di sistema possibili.

Prerequisiti

- Quote di servizio sufficienti per avviare il numero desiderato di EC2 istanze nel tuo. Regione AWS Puoi utilizzarle [AWS Management Console](#) per controllare e richiedere aumenti delle quote di servizio.
- Una o più sottoreti esistenti VPC che soddisfano i requisiti AWS PCS di rete. Si consiglia di comprendere a fondo questi requisiti prima di implementare un cluster per l'uso in produzione. Per ulteriori informazioni, consulta [AWS PCS VPC e requisiti e considerazioni sulla sottorete](#). È inoltre possibile utilizzare un CloudFormation modello per creare sottoreti VPC e. AWS fornisce una HPC ricetta per il modello. CloudFormation Per ulteriori informazioni, vedere [aws-hpc-recipes](#) on GitHub.
- Un profilo di IAM istanza con le autorizzazioni per avviare l' AWS PCS RegisterComputeNodeGroupInstance API azione e l'accesso a tutte AWS le altre risorse necessarie per le istanze del gruppo di nodi. Per ulteriori informazioni, consulta [IAM profili di istanza per AWS Parallel Computing Service](#).
- Un modello di avvio per le istanze del gruppo di nodi. Per ulteriori informazioni, consulta [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#).
- Per creare un gruppo di nodi di calcolo che utilizzi istanze Amazon EC2 Spot, devi avere il ruolo AWSServiceRoleForEC2Spot collegato al servizio nel tuo. Account AWS Per ulteriori informazioni, consulta [Ruolo di Amazon EC2 Spot per AWS PCS](#).

Crea un gruppo di nodi di calcolo in AWS PCS


È possibile creare un gruppo di nodi di calcolo utilizzando AWS Management Console o il. AWS CLI

AWS Management Console

Per creare il gruppo di nodi di calcolo utilizzando la console

1. Apri la [AWS PCS console](#).
2. Seleziona il cluster in cui desideri creare un gruppo di nodi di calcolo. Passa ai gruppi di nodi di calcolo e scegli Crea.
3. Nella sezione Configurazione del gruppo di nodi di calcolo, fornisci un nome per il tuo gruppo di nodi. Il nome può contenere solo caratteri alfanumerici e trattini con distinzione tra maiuscole e minuscole. Deve iniziare con un carattere alfabetico e non può superare i 25 caratteri. Il nome deve essere univoco all'interno del cluster.
4. In Computing configuration, inserisci o seleziona questi valori:

- a. EC2modello di avvio: seleziona un modello di avvio personalizzato da utilizzare per questo gruppo di nodi. I modelli di avvio possono essere utilizzati per personalizzare le impostazioni di rete come sottorete e gruppi di sicurezza, configurazione di monitoraggio e archiviazione a livello di istanza. Se non hai preparato un modello di lancio, scopri come [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#) crearne uno.

 Important

AWS PCS crea un modello di avvio gestito per ogni gruppo di nodi di calcolo. Questi sono `pcs-identifier-do-not-delete` denominati. Non selezionarli quando crei o aggiorni un gruppo di nodi di calcolo, altrimenti il gruppo di nodi non funzionerà correttamente.

- b. EC2versione del modello di avvio: seleziona una versione del modello di avvio personalizzato. Puoi scegliere una versione specifica, che può migliorare la riproducibilità. Se modifichi la versione in un secondo momento, devi aggiornare il gruppo di nodi di calcolo per rilevare le modifiche nel modello di avvio. Per ulteriori informazioni, consulta [Aggiornamento di un gruppo di nodi di AWS PCS calcolo](#).
- c. AMIID: se il modello di lancio non include un AMI ID o se desideri sovrascrivere il valore nel modello di lancio, fornisci un AMI ID qui. Nota che quello AMI usato per il gruppo di nodi deve essere compatibile con AWS PCS. Puoi anche selezionare un campione AMI fornito da AWS. Per ulteriori informazioni su questo argomento, vedere [Amazon Machine Images \(AMIs\) per AWS PCS](#).
- d. IAMprofilo di istanza: scegli un profilo di istanza per il gruppo di nodi. Un profilo di istanza concede all'istanza le autorizzazioni per accedere a AWS risorse e servizi in modo sicuro. Se non ne hai uno pronto, scopri come [IAMprofili di istanza per AWS Parallel Computing Service](#) crearne uno.
- e. Sottoreti: scegli una o più sottoreti nel luogo in VPC cui è distribuito il AWS PCS cluster. Se si selezionano più sottoreti, EFA le comunicazioni non saranno disponibili tra i nodi e la comunicazione tra nodi in sottoreti diverse potrebbe avere una latenza maggiore. Assicurati che le sottoreti che specifichi qui corrispondano a quelle definite nel modello di avvio. EC2
- f. Istanze: scegli uno o più tipi di istanze per soddisfare le richieste di scalabilità nel gruppo di nodi. Tutti i tipi di istanza devono avere la stessa architettura del processore (x864_64

- o arm64) e lo stesso numero di vCPUs. Se le istanze lo sono GPU, tutti i tipi di istanza devono avere lo stesso numero di GPU.
- g. Configurazione di scalabilità: specifica il numero minimo e massimo di istanze per il gruppo di nodi. È possibile definire una configurazione statica, in cui è in esecuzione un numero fisso di nodi, o una configurazione dinamica, in cui è possibile eseguire fino al numero massimo di nodi. Per una configurazione statica, imposta minimo e massimo sullo stesso numero, maggiore di zero. Per una configurazione dinamica, imposta il numero minimo di istanze su zero e il numero massimo di istanze su un numero maggiore di zero. AWS PCS non supporta gruppi di nodi di calcolo con un mix di istanze statiche e dinamiche.
5. (Facoltativo) In Impostazioni aggiuntive, specifica quanto segue:
 - a. Opzione di acquisto: seleziona tra istanze Spot e On-demand.
 - b. Strategia di allocazione: se hai selezionato l'opzione di acquisto Spot, puoi specificare come vengono scelti i pool di capacità Spot al momento del lancio delle istanze nel gruppo di nodi. Per ulteriori informazioni, consulta [Strategie di allocazione per istanze Spot nella Guida](#) per l'utente di Amazon Elastic Compute Cloud. Questa opzione non ha effetto se hai selezionato l'opzione di acquisto On-demand.
 6. (Facoltativo) Nella sezione delle impostazioni Slurm personalizzate, fornisci questi valori:
 - a. Peso: questo valore imposta la priorità dei nodi del gruppo ai fini della pianificazione. I nodi con pesi inferiori hanno una priorità più alta e le unità sono arbitrarie. Per ulteriori informazioni, consulta [Weight](#) nella Slurm documentazione.
 - b. Memoria reale: questo valore imposta la dimensione (in GB) della memoria reale sui nodi del gruppo di nodi. È pensato per essere utilizzato insieme all'`CR_CPU_Memory` opzione nella Slurm configurazione del cluster in AWS PCS. Per ulteriori informazioni, [RealMemory](#) consulta la Slurm documentazione.
 7. (Facoltativo) In Tag, aggiungi qualsiasi tag al gruppo di nodi di calcolo.
 8. Scegli Crea gruppo di nodi di calcolo. Il campo Status mostra Creating mentre effettua il provisioning del gruppo di nodi. Questo processo può richiedere diversi minuti.

Fase successiva consigliata

- Aggiungi il tuo gruppo di nodi a una coda AWS PCS per consentirgli di elaborare i lavori.

AWS CLI

Per creare il tuo gruppo di nodi di calcolo utilizzando AWS CLI

Crea la tua coda con il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:

1. Replace (Sostituisci) *region* con l'ID di in Regione AWS cui creare il cluster, ad esempio `us-east-1`.
2. Replace (Sostituisci) *my-cluster* con il nome o `clusterId` del cluster.
3. Replace (Sostituisci) *my-node-group* con il nome del gruppo di nodi di calcolo. Il nome può contenere solo caratteri alfanumerici (con distinzione tra lettere maiuscole e minuscole) e trattini. Deve iniziare con un carattere alfabetico e non può essere più lungo di 25 caratteri. Il nome deve essere univoco all'interno del cluster.
4. Replace (Sostituisci) *subnet-ExampleID1* con una o più sottoreti IDs del cluster. VPC
5. Replace (Sostituisci) *lt-ExampleID1* con l'ID per il tuo modello di lancio personalizzato. Se non ne hai uno già pronto, scopri [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#) come crearne uno.

Important

AWS PCS crea un modello di avvio gestito per ogni gruppo di nodi di calcolo. Questi sono `pcs-identifier-do-not-delete` denominati. Non selezionarli quando crei o aggiorni un gruppo di nodi di calcolo, altrimenti il gruppo di nodi non funzionerà correttamente.

6. Replace (Sostituisci) *launch-template-version* con una versione specifica del modello di lancio se desideri associare il tuo gruppo di nodi a una versione specifica.
7. Replace (Sostituisci) *arn:InstanceProfile* con il profilo ARN della tua IAM istanza. Se non ne hai uno pronto, consulta la sezione [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#) per maggiori informazioni.
8. Replace (Sostituisci) *min-instances* e *max-instances* con valori interi. È possibile definire una configurazione statica, in cui è in esecuzione un numero fisso di nodi, o una configurazione dinamica, in cui è possibile eseguire fino al numero massimo di nodi. Per una configurazione statica, imposta minimo e massimo sullo stesso numero, maggiore di zero. Per una configurazione dinamica, imposta il numero minimo di istanze su zero e il numero

massimo di istanze su un numero maggiore di zero. AWS PCS non supporta gruppi di nodi di calcolo con un mix di istanze statiche e dinamiche.

9. Replace (Sostituisci) *t3.large* con un altro tipo di istanza. È possibile aggiungere altri tipi di istanze specificando un elenco di `instanceType` impostazioni. Ad esempio, `--instance-configs instanceType=c6i.16xlarge,instanceType=c6a.16xlarge`. Tutti i tipi di istanza devono avere la stessa architettura del processore (x86_64 o arm64) e lo stesso numero di vCPUs. Se le istanze lo sono GPUs, tutti i tipi di istanza devono avere lo stesso numero di GPUs.

```
aws pcs create-compute-node-group --region region \
  --cluster-identifier my-cluster \
  --compute-node-group-name my-node-group \
  --subnet-ids subnet-ExampleID1 \
  --custom-launch-template id=lt-ExampleID1,version='launch-template-version' \
  --iam-instance-profile arn=arn:InstanceProfile \
  --scaling-config minInstanceCount=min-instances,maxInstanceCount=max-instance \
  --instance-configs instanceType=t3.large
```

Esistono diverse impostazioni di configurazione opzionali che è possibile aggiungere al `create-compute-node-group` comando.

- È possibile specificare `--amiId` se il modello di avvio personalizzato non include un riferimento a un AMI valore o se si desidera sovrascrivere tale valore. Nota che quello AMI usato per il gruppo di nodi deve essere compatibile con AWS PCS. Puoi anche selezionare un campione AMI fornito da AWS. Per ulteriori informazioni su questo argomento, vedere [Amazon Machine Images \(AMIs\) per AWS PCS](#).
- È possibile selezionare tra istanze on-demand (ONDEMAND) e Spot (SPOT) utilizzando `--purchase-option`. L'impostazione predefinita è On-demand. Se scegli le istanze Spot, puoi anche utilizzarle `--allocation-strategy` per definire in che modo AWS PCS sceglie i pool di capacità Spot quando avvia le istanze nel gruppo di nodi. Per ulteriori informazioni, consulta [Strategie di allocazione per istanze Spot nella Guida](#) per l'utente di Amazon Elastic Compute Cloud.
- È possibile fornire opzioni di Slurm configurazione per i nodi del gruppo di nodi utilizzando `--slurm-configuration`. È possibile impostare il peso (priorità di pianificazione) e la memoria reale. I nodi con pesi inferiori hanno una priorità più alta e le unità sono arbitrarie. Per ulteriori informazioni, consulta [Weight](#) nella Slurm documentazione. La memoria reale imposta la dimensione (in GB) della memoria reale sui nodi del gruppo di nodi. È pensata per essere

utilizzata insieme all'`CR_CPU_Memory` opzione per il cluster presente AWS PCS nella Slurm configurazione. Per ulteriori informazioni, [RealMemory](#) consulta la Slurm documentazione.

⚠ Important

La creazione del gruppo di nodi di calcolo può richiedere diversi minuti.

Puoi interrogare lo stato del tuo gruppo di nodi con il seguente comando. Non sarai in grado di associare il gruppo di nodi a una coda finché non ne raggiungerà ACTIVE lo stato.

```
aws pcs get-compute-node-group --region region \  
  --cluster-identifier my-cluster \  
  --compute-node-group-identifier my-node-group
```

Aggiornamento di un gruppo di nodi di AWS PCS calcolo

Questo argomento fornisce una panoramica delle opzioni disponibili e descrive gli elementi da considerare quando si aggiorna un gruppo di nodi di AWS PCS calcolo.

Opzioni per l'aggiornamento di un gruppo di AWS PCS nodi di calcolo

L'aggiornamento di un gruppo di nodi di AWS PCS calcolo consente di modificare le proprietà delle istanze avviate da AWSPCS, nonché le regole relative al modo in cui tali istanze vengono avviate. Ad esempio, puoi sostituire le istanze del gruppo di nodi AMI for con un'altra su cui è installato un software diverso. In alternativa, puoi aggiornare i gruppi di sicurezza per modificare la connettività di rete in entrata o in uscita. Puoi anche modificare la configurazione di scalabilità o persino modificare l'opzione di acquisto preferita da o verso le istanze Spot.

Le seguenti impostazioni dei gruppi di nodi non possono essere modificate dopo la creazione:

- Nome
- Istanze

Considerazioni sull'aggiornamento di un gruppo di nodi di AWS PCS calcolo

I gruppi di nodi di calcolo definiscono EC2 le istanze utilizzate per elaborare i lavori, fornire l'accesso interattivo alla shell e altre attività. Sono spesso associati a una o più AWS PCS code. Quando

aggiorni il gruppo di nodi di calcolo per modificarne il comportamento (o quello dei nodi), considera quanto segue:

- Le modifiche alle proprietà del gruppo di nodi di calcolo diventano effettive quando lo stato del gruppo di nodi di calcolo passa da Aggiornamento ad Attivo. Le nuove istanze vengono avviate con le proprietà aggiornate.
- Gli aggiornamenti che non influiscono sulla configurazione di nodi specifici non influiscono sui nodi in esecuzione. Ad esempio, l'aggiunta di una sottorete e la modifica della strategia di allocazione.
- Se si aggiorna il modello di avvio per un gruppo di nodi di calcolo, è necessario aggiornare il gruppo di nodi di calcolo per utilizzare la nuova versione.
- Per aggiungere o rimuovere un gruppo di sicurezza dai nodi di un gruppo di nodi di calcolo, modifica il relativo modello di avvio e aggiorna il gruppo di nodi di calcolo. Le nuove istanze vengono lanciate con il set aggiornato di gruppi di sicurezza.
- Se modifichi direttamente un gruppo di sicurezza utilizzato da un gruppo di nodi di calcolo, ciò ha effetto immediato sulle istanze in esecuzione e future.
- Se aggiungi o rimuovi le autorizzazioni dal profilo dell'IAM istanza utilizzato da un gruppo di nodi di calcolo, ciò ha effetto immediato sulle istanze in esecuzione e future.
- Per modificare le istanze AMI utilizzate da un gruppo di nodi di calcolo, aggiorna il gruppo di nodi di calcolo (o il relativo modello di avvio) per utilizzare il nuovo gruppo di nodi di calcolo AMI e attendi che le istanze vengano sostituite. AWS PCS
- AWS PCS sostituisce le istanze esistenti nel gruppo di nodi dopo un'operazione di aggiornamento del gruppo di nodi. Se ci sono processi in esecuzione su un nodo, tali processi possono essere completati prima di AWS PCS sostituire il nodo. I processi utente interattivi (ad esempio sulle istanze del nodo di accesso) vengono terminati. Lo stato del gruppo di nodi torna a Active quando AWS PCS contrassegna le istanze da sostituire, ma la sostituzione effettiva avviene quando le istanze sono inattive.
- Se riduci il numero massimo di istanze consentite in un gruppo di nodi di calcolo, AWS PCS rimuove i nodi da Slurm per raggiungere il nuovo numero massimo. AWS PCS termina le istanze in esecuzione associate ai nodi Slurm rimossi. I job in esecuzione sui nodi rimossi hanno esito negativo e ritornano nelle rispettive code.
- AWS PCS crea un modello di avvio gestito per ogni gruppo di nodi di calcolo. Sono `pcs-identifier-do-not-delete` denominati. Non selezionarli quando crei o aggiorni un gruppo di nodi di calcolo, altrimenti il gruppo di nodi non funzionerà correttamente.

- Se aggiorni un gruppo di nodi di calcolo per utilizzare Spot come opzione di acquisto, devi avere il ruolo `AWSServiceRoleForEC2Spot` collegato al servizio nel tuo account. Per ulteriori informazioni, consulta [Ruolo di Amazon EC2 Spot per AWS PCS](#).

Per aggiornare un gruppo di nodi di calcolo AWS PCS

È possibile aggiornare un gruppo di nodi utilizzando la Console AWS di gestione o il AWSCLI.

AWS Management Console

Per aggiornare un gruppo di nodi di calcolo

1. Apri la AWS PCS console all'indirizzo `https://console.aws.amazon.com/pcs/home#/clusters`
2. Seleziona il cluster in cui desideri aggiornare un gruppo di nodi di calcolo.
3. Passa ai gruppi di nodi di calcolo, vai al gruppo di nodi che desideri aggiornare, quindi seleziona Modifica.
4. Nelle sezioni Configurazione informatica, Impostazioni aggiuntive e Impostazioni di Slurmpersonalizzazione, aggiorna tutti i valori tranne:
 - Istanze: non è possibile modificare le istanze in un gruppo di nodi di calcolo.
5. Scegli Aggiorna. Il campo Stato mostrerà Aggiornamento durante l'applicazione delle modifiche.

Important

Gli aggiornamenti dei gruppi di nodi di calcolo possono richiedere diversi minuti.

AWS CLI

Per aggiornare un gruppo di nodi di calcolo

1. Aggiorna il tuo gruppo di nodi di calcolo con il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:
 - a. Replace (Sostituisci) *region-code* con la AWS regione in cui vuoi creare il cluster.

- b. Replace (Sostituisci) *my-node-group* con il nome o `computeNodeId` per il gruppo di nodi di calcolo.
- c. Replace (Sostituisci) *my-cluster* con il nome o `clusterId` del tuo cluster.

```
aws pcs update-compute-node-group --region region-code \  
  --cluster-identifier my-cluster \  
  --compute-node-group-identifier my-node-group
```

2. Aggiorna tutti i parametri del gruppo di nodi ad eccezione di `--instance-configs`. Ad esempio, per impostare un nuovo AMI ID, passa `--amiId my-custom-ami-id` dove *my-custom-ami-id* viene sostituito dal tuo AMI preferito.

Important

L'aggiornamento del gruppo di nodi di calcolo può richiedere diversi minuti.

Puoi interrogare lo stato del tuo gruppo di nodi con il seguente comando.

```
aws pcs get-compute-node-group --region region-code \  
  --cluster-identifier my-cluster \  
  --compute-node-group-identifier my-node-group
```

Eliminazione di un gruppo di nodi di calcolo in AWS PCS

Questo argomento fornisce una panoramica delle opzioni disponibili e descrive gli elementi da considerare quando si elimina un gruppo di nodi di calcolo in AWS PCS.

Considerazioni sull'eliminazione di un gruppo di nodi di calcolo

I gruppi di nodi di calcolo definiscono EC2 le istanze utilizzate per elaborare i lavori, fornire l'accesso interattivo alla shell e altre attività. Sono spesso associati a una o più AWS PCS code. Prima di eliminare un gruppo di nodi di calcolo, considerate quanto segue:

- Tutte EC2 le istanze avviate dal gruppo di nodi di calcolo verranno terminate. Ciò annullerà i processi in esecuzione su queste istanze e interromperà l'esecuzione dei processi interattivi.

- È necessario dissociare il gruppo di nodi di calcolo da tutte le code prima di poterlo eliminare. Per ulteriori informazioni, consulta [Aggiornamento di una AWS PCS coda](#).

Eliminare il gruppo di nodi di calcolo

È possibile utilizzare AWS Management Console o AWS CLI per eliminare un gruppo di nodi di calcolo.

AWS Management Console

Per eliminare un gruppo di nodi di calcolo

1. Apri la [AWS PCSconsole](#).
2. Seleziona il cluster del gruppo di nodi di calcolo.
3. Passa ai gruppi di nodi di calcolo e seleziona il gruppo di nodi di calcolo da eliminare.
4. Scegli Elimina.
5. Viene visualizzato il campo Stato. Deleting Per il completamento possono essere necessari alcuni minuti.

Note

È possibile utilizzare i comandi nativi dello scheduler per confermare l'eliminazione del gruppo di nodi di calcolo. Ad esempio, usa `sinfo` o `squeue` per Slurm.

AWS CLI


Per eliminare un gruppo di nodi di calcolo

- Usa il comando seguente per eliminare un gruppo di nodi di calcolo, con queste sostituzioni:
 - Replace (Sostituisci) *region-code* con Regione AWS il cluster in uso.
 - Replace (Sostituisci) *my-node-group* con il nome o l'ID del gruppo di nodi di calcolo.
 - Replace (Sostituisci) *my-cluster* con il nome o l'ID del tuo cluster.

```
aws pcs delete-compute-node-group --region region-code \
```

```
--compute-node-group-identifier my-node-group \  
--cluster-identifier my-cluster
```

L'eliminazione del gruppo di nodi di calcolo può richiedere diversi minuti.

 Note

È possibile utilizzare i comandi nativi dello scheduler per confermare l'eliminazione del gruppo di nodi di calcolo. Ad esempio, usa `sinfo` o `squeue` per Slurm.

Ricerca di istanze di gruppi di nodi di calcolo in AWS PCS

Ogni gruppo di nodi di AWS PCS calcolo può avviare EC2 istanze con configurazioni condivise. Puoi utilizzare i EC2 tag per trovare istanze in un gruppo di nodi di calcolo in o con. AWS Management Console AWS CLI

AWS Management Console

Per trovare le istanze del tuo gruppo di nodi di calcolo

1. [Apri la console.AWS PCS](#)
2. Seleziona il cluster .
3. Scegli i gruppi di nodi Compute.
4. Trova l'ID per il gruppo di nodi di accesso che hai creato.
5. Vai alla [EC2console](#) e scegli Istanze.
6. Cerca le istanze con il tag seguente. Replace (Sostituisci) *node-group-id* con l'ID (non il nome) del gruppo di nodi di calcolo.

```
aws:pcs:compute-node-group-id=node-group-id
```

7. (Facoltativo) Puoi modificare il valore dello stato dell'istanza nel campo di ricerca per trovare le istanze che sono in fase di configurazione o che sono state terminate di recente.
8. Trova l'ID e l'indirizzo IP dell'istanza per ogni istanza nell'elenco delle istanze con tag.

AWS CLI

Per trovare le istanze del tuo gruppo di nodi, usa i comandi che seguono. Prima di eseguire i comandi, effettua le seguenti sostituzioni:

- Sostituisci *region-code* con il Regione AWS del tuo cluster. Esempio: us-east-1
- Sostituisci *node-group-id* con l'ID (non il nome) del tuo gruppo di nodi di calcolo.
- Sostituisci *running* con altri stati di istanza come *pending* o *terminated* per trovare EC2 istanze in altri stati.

```
aws ec2 describe-instances \
  --region region-code --filters \
  "Name=tag:aws:pcs:compute-node-group-id,Values=node-group-id" \
  "Name=instance-state-name,Values=running" \
  --query 'Reservations[*].Instances[*].
{InstanceID:InstanceId,State:State.Name,PublicIP:PublicIpAddress,PrivateIP:PrivateIpAddress}'
```

Il comando restituisce un output simile al seguente: Il valore di `PublicIP` è `null` se l'istanza si trova in una sottorete privata.

```
[
  [
    {
      "InstanceID": "i-0123456789abcdefa",
      "State": "running",
      "PublicIP": "18.189.32.188",
      "PrivateIP": "10.0.0.1"
    }
  ]
]
```

Note

Se prevedi `describe-instances` di restituire un numero elevato di istanze, devi utilizzare le opzioni per più pagine. Per ulteriori informazioni, consulta [DescribeInstances](#) Amazon Elastic Compute Cloud API Reference.

Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS

In AmazonEC2, un modello di avvio può memorizzare una serie di preferenze in modo da non doverle specificare singolarmente all'avvio delle istanze. AWS PCS incorpora modelli di lancio come modo flessibile per configurare i gruppi di nodi di calcolo. Quando crei un gruppo di nodi, fornisci un modello di lancio. AWS PCS ne crea un modello di lancio derivato che include trasformazioni per garantire che funzioni con il servizio.

Capire quali sono le opzioni e le considerazioni da prendere in considerazione quando si scrive un modello di lancio personalizzato può aiutarvi a scriverne uno da utilizzare. AWS PCS Per ulteriori informazioni sui modelli di avvio, consulta [Launching an Instance from a Launch an instance from a launch template](#) nella Amazon EC2 User Guide.

Argomenti

- [Panoramica](#)
- [Creare un modello di avvio di base](#)
- [Lavorare con i dati EC2 degli utenti Amazon](#)
- [Prenotazioni di capacità in AWS PCS](#)
- [Parametri utili del modello di lancio](#)

Panoramica

Sono [disponibili oltre 30 parametri](#) che puoi includere in un modello di EC2 lancio, che controllano molti aspetti della configurazione delle istanze. La maggior parte sono completamente compatibili con AWS PCS, ma ci sono alcune eccezioni.

I seguenti parametri del modello EC2 Launch verranno ignorati AWS PCS poiché queste proprietà devono essere gestite direttamente dal servizio:

- Tipo di istanza/Specificare gli attributi del tipo di istanza (`InstanceRequirements`): non AWS PCS supporta la selezione dell'istanza basata sugli attributi.
- Tipo di istanza (`InstanceType`) — Specificate i tipi di istanza quando create un gruppo di nodi.
- Dettagli avanzati/ profilo di IAM istanza (`IamInstanceProfile`): fornite questo parametro quando create o aggiornate il gruppo di nodi.
- Dettagli avanzati/disable API termination (`DisableApiTermination`): AWS PCS deve controllare il ciclo di vita delle istanze del gruppo di nodi che avvia.

- **Dettagli avanzati/disable API stop (DisableApiStop):** AWS PCS deve controllare il ciclo di vita delle istanze del gruppo di nodi che avvia.
- **Advanced Details/stop — Hibernate behavior () —** non supporta l'ibernazione delle istanze. `HibernationOptions` AWS PCS
- **Advanced Details/Elastic GPU (ElasticGpuSpecifications)** — Amazon Elastic Graphics ha raggiunto la fine del ciclo di vita l'8 gennaio 2024.
- **Dettagli avanzati/Elastic inference (ElasticInferenceAccelerators):** Amazon Elastic Inference non è più disponibile per i nuovi clienti.
- **AdvancedDetails/specifica CPU opzioni/thread per core (ThreadsPerCore):** AWS PCS imposta il numero di thread per core su 1.

Questi parametri hanno requisiti speciali che supportano la compatibilità con: AWS PCS

- **Dati utente (UserData):** devono essere codificati in più parti. Per informazioni, consulta [Lavorare con i dati EC2 degli utenti Amazon](#).
- **Immagini dell'applicazione e del sistema operativo (ImageId):** puoi includerle. Tuttavia, se si specifica un AMI ID quando si crea o si aggiorna il gruppo di nodi, questo sostituirà il valore nel modello di avvio. Il AMI file fornito deve essere compatibile con AWS PCS. Per ulteriori informazioni, vedere "[Amazon Machine Images \(AMIs\) per AWS PCS](#)".
- **Impostazioni di rete/Firewall (security groups) (SecurityGroups):** non è possibile impostare un elenco di nomi di gruppi di sicurezza in un AWS PCS modello di avvio. È possibile impostare un elenco di gruppi di sicurezza IDs (`SecurityGroupIds`), a meno che non si definiscano interfacce di rete nel modello di avvio. Quindi, è necessario specificare il gruppo di sicurezza IDs per ogni interfaccia. Per ulteriori informazioni, consulta [Gruppi di sicurezza in AWS PCS](#).
- **Impostazioni di rete/Configurazione di rete avanzata (NetworkInterfaces):** se utilizzi EC2 istanze con una singola scheda di rete e non richiedi alcuna configurazione di rete specializzata, AWS PCS puoi configurare il networking delle istanze automaticamente. Per configurare più schede di rete o abilitare Elastic Fabric Adapter sulle tue istanze, usa `NetworkInterfaces`. Ogni interfaccia di rete deve avere un elenco di gruppi di sicurezza IDs in `Groups`. Per ulteriori informazioni, consulta [Interfacce di rete multiple in AWS PCS](#).
- **Dettagli avanzati/Prenotazione della capacità (CapacityReservationSpecification):** può essere impostato, ma non può fare riferimento a uno specifico `CapacityReservationId` quando si lavora con AWS PCS. Tuttavia, è possibile fare riferimento a un gruppo di prenotazione di capacità, laddove tale gruppo contenga una o più prenotazioni di capacità. Per ulteriori informazioni, consulta [Prenotazioni di capacità in AWS PCS](#).

Creare un modello di avvio di base

È possibile creare un modello di lancio utilizzando AWS Management Console o il AWS CLI.

AWS Management Console

Per creare un modello di avvio

1. Apri la [EC2console Amazon](#) e seleziona Launch templates.
2. Scegli Crea modello di avvio.
3. In Nome e descrizione del modello Launch, inserisci un nome univoco e distintivo per il nome del modello Launch
4. In Key pair (login) in Key pair name, seleziona la coppia di SSH chiavi che verrà utilizzata per accedere alle EC2 istanze gestite da AWS PCS. Questo passaggio è facoltativo, ma è consigliato.
5. In Impostazioni di rete, quindi Firewall (gruppi di sicurezza), scegli i gruppi di sicurezza da collegare all'interfaccia di rete. Tutti i gruppi di sicurezza nel modello di avvio devono provenire dal tuo AWS PCS clusterVPC. Come minimo, scegli:
 - Un gruppo di sicurezza che consente la comunicazione con il AWS PCS cluster
 - Un gruppo di sicurezza che consente la comunicazione tra EC2 istanze lanciate da AWS PCS
 - (Facoltativo) Un gruppo di sicurezza che consente l'SSHaccesso in entrata a istanze interattive
 - (Facoltativo) Un gruppo di sicurezza che consente ai nodi di elaborazione di effettuare connessioni in uscita a Internet
 - (Facoltativo) Gruppi di sicurezza che consentono l'accesso a risorse di rete come file system condivisi o un server di database.
6. Il tuo nuovo ID del modello di lancio sarà accessibile nella EC2 console Amazon alla voce Launch templates. L'ID del modello di lancio avrà il modulo `lt-0123456789abcdef01`.

Fase successiva consigliata

- Usa il nuovo modello di lancio per creare o aggiornare un gruppo di nodi di AWS PCS calcolo.

AWS CLI

Per creare un modello di avvio

Crea il tuo modello di lancio con il comando che segue.

- Prima di eseguire il comando, apporta le modifiche seguenti:
 - a. Replace (Sostituisci) *region-code* con il Regione AWS quale stai lavorando AWS PCS
 - b. Replace (Sostituisci) *my-launch-template-name* con un nome per il tuo modello. Deve essere unico per l' Account AWS oggetto Regione AWS che stai utilizzando.
 - c. Replace (Sostituisci) *my-ssh-key-name* con il nome della tua SSH chiave preferita.
 - d. Replace (Sostituisci) *sg-ExampleID1* e *sg-ExampleID2* con un gruppo di sicurezza IDs che consente la comunicazione tra le EC2 istanze e lo scheduler e la comunicazione tra le EC2 istanze. Se disponi di un solo gruppo di sicurezza che abilita tutto questo traffico, puoi rimuovere *sg-ExampleID2* anche la virgola che lo precede. Puoi anche aggiungere altri gruppi IDs di sicurezza. Tutti i gruppi di sicurezza che includi nel modello di lancio devono provenire dal tuo AWS PCS clusterVPC.

```
aws ec2 create-launch-template --region region-code \  
  --launch-template-name my-template-name \  
  --launch-template-data '{"KeyName":"my-ssh-key-name","SecurityGroupIds":  
  ["sg-ExampleID1","sg-ExampleID2"]}'
```

AWS CLI Verrà emesso un testo simile al seguente. L'ID del modello di avvio si trova in `LaunchTemplateId`.

```
{  
  "LaunchTemplate": {  
    "LatestVersionNumber": 1,  
    "LaunchTemplateId": "lt-0123456789abcdef01",  
    "LaunchTemplateName": "my-launch-template-name",  
    "DefaultVersionNumber": 1,  
    "CreatedBy": "arn:aws:iam::123456789012:user/Bob",  
    "CreateTime": "2019-04-30T18:16:06.000Z"  
  }  
}
```


Fase successiva consigliata

- Usa il nuovo modello di lancio per creare o aggiornare un gruppo di nodi di AWS PCS calcolo.

Lavorare con i dati EC2 degli utenti Amazon

Puoi fornire i dati EC2 utente nel modello di lancio che `cloud-init` viene eseguito all'avvio delle istanze. I blocchi di dati utente con il tipo di contenuto `cloud-config` vengono eseguiti prima che l'istanza si registri con il AWS PCS API, mentre i blocchi di dati utente con il tipo di contenuto `text/x-shellscript` vengono eseguiti dopo il completamento della registrazione, ma prima dell'avvio del demone Slurm. Per ulteriori informazioni sui tipi di contenuto, consultare la documentazione di [cloud-init](#).

I nostri dati utente possono eseguire scenari di configurazione comuni, tra cui, a titolo esemplificativo ma non esaustivo, i seguenti:

- [Inclusi utenti o gruppi](#)
- [Installazione di pacchetti](#)
- [Creazione di partizioni e file system](#)
- Montaggio di file system di rete

I dati utente nei modelli di avvio devono essere in formato di [archivio MIME multipart](#). Questo perché i dati utente vengono uniti ad altri dati AWS PCS utente necessari per configurare i nodi nel gruppo di nodi. È possibile combinare più blocchi di dati utente in un unico file composto da MIME più parti.

Un file composto da MIME più parti è composto dai seguenti componenti:

- Il tipo di contenuto e la dichiarazione di delimitazione della parte: `Content-Type: multipart/mixed; boundary="==BOUNDARY=="`
- La dichiarazione di MIME versione: `MIME-Version: 1.0`
- Uno o più blocchi di dati utente che contengono i seguenti componenti:
 - Il limite di apertura che segnala l'inizio di un blocco di dati utente: `--==BOUNDARY==`. È necessario mantenere vuota la linea prima di questo limite.
 - La dichiarazione del tipo di contenuto per il blocco: `Content-Type: text/cloud-config; charset="us-ascii"` o `Content-Type: text/x-shellscript; charset="us-ascii"`. È necessario lasciare vuota la riga dopo la dichiarazione del tipo di contenuto.

- Il contenuto dei dati utente, ad esempio un elenco di comandi o `cloud-config` direttive di shell.
- Il limite di chiusura che segnala la fine del file composto da MIME più parti: `--==BOUNDARY==--`. È necessario mantenere vuota la linea prima del limite di chiusura.

Note

Se aggiungi dati utente a un modello di lancio nella EC2 console Amazon, puoi incollarli come testo normale. Oppure puoi caricarli da un file. Se usi AWS CLI o an AWS SDK, devi prima codificare in base64 i dati utente e inviare quella stringa come valore del `UserData` parametro quando chiami [CreateLaunchTemplate](#), come mostrato in questo file. JSON

```
{
  "LaunchTemplateName": "base64-user-data",
  "LaunchTemplateData": {
    "UserData":
"ewogICAgIkxhdW5jaFR1bXBsYXR1TmFtZSI6ICJpbmNyZWZzZS1jb250YWluZXItZm9sdW..."
  }
}
```

Esempi

- [Esempio: installare il software da un archivio di pacchetti](#)
- [Esempio: esegui script da un bucket S3](#)
- [Esempio: imposta le variabili di ambiente globali](#)
- [Utilizzo di file system di rete con AWS PCS](#)
- [Esempio: utilizzare un EFS file system come home directory condivisa](#)

Esempio: installare il software AWS PCS da un archivio di pacchetti

Fornisci questo script come valore di `"userData"` nel tuo modello di lancio. Per ulteriori informazioni, consulta [Lavorare con i dati EC2 degli utenti Amazon](#).

Questo script utilizza `cloud-config` per installare pacchetti software su istanze di gruppi di nodi al momento del lancio. Per ulteriori informazioni, consulta i [formati dei dati utente nella documentazione](#) di `cloud-init`. Questo esempio installa `and. curl llvm`

Note

Le istanze devono essere in grado di connettersi agli archivi di pacchetti configurati.

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary==="MYBOUNDARY==="

--===MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

packages:
- python3-devel
- rust
- golang

--===MYBOUNDARY===--
```

Esempio: esegui script aggiuntivi AWS PCS da un bucket S3

Fornisci questo script come valore di "userData" nel tuo modello di lancio. Per ulteriori informazioni, consulta [Lavorare con i dati EC2 degli utenti Amazon](#).

Questo script utilizza cloud-config per importare uno script da un bucket S3 ed eseguirlo su istanze di gruppi di nodi all'avvio. Per ulteriori informazioni, consulta i formati [dei dati utente nella documentazione di cloud-init](#).

Sostituisci i seguenti valori in questo script con i tuoi dati:

- *my-bucket-name* — Il nome di un bucket S3 da cui il tuo account può leggere.
- *path* — Il percorso relativo alla radice del bucket S3.
- *shell* — La shell Linux da usare per eseguire lo script, ad esempio. bash

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary==="MYBOUNDARY==="

--===MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

runcmd:
```

```
- aws s3 cp s3://my-bucket-name/path /tmp/script.sh
- /usr/bin/shell /tmp/script.sh

--==MYBOUNDARY==--
```

Il profilo di IAM istanza per il gruppo di nodi deve avere accesso al bucket. La seguente IAM politica è un esempio del bucket nello script di dati utente riportato sopra.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3::my-bucket-name",
        "arn:aws:s3::my-bucket-name/path/*"
      ]
    }
  ]
}
```

Esempio: imposta le variabili di ambiente globali per AWS PCS

Fornisci questo script come valore di "userData" nel tuo modello di lancio. Per ulteriori informazioni, consulta [Lavorare con i dati EC2 degli utenti Amazon](#).

L'esempio seguente utilizza /etc/profile.d per impostare variabili globali su istanze di gruppi di nodi.

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/x-shellscript; charset="us-ascii"

#!/bin/bash
touch /etc/profile.d/awspcs-userdata-vars.sh
echo MY_GLOBAL_VAR1=100 >> /etc/profile.d/awspcs-userdata-vars.sh
```

```
echo MY_GLOBAL_VAR2=abc >> /etc/profile.d/awspcs-userdata-vars.sh

---MYBOUNDARY---
```

Esempio: utilizzare un EFS file system come home directory condivisa per AWS PCS

Fornisci questo script come valore di "userData" nel tuo modello di lancio. Per ulteriori informazioni, consulta [Lavorare con i dati EC2 degli utenti Amazon](#).

Questo esempio estende l'esempio EFS mount in [Utilizzo di file system di rete con AWS PCS](#) per implementare una home directory condivisa. Viene eseguito il backup del contenuto di /home prima del montaggio del EFS file system. I contenuti vengono quindi copiati rapidamente nella memoria condivisa dopo il completamento del montaggio.

Sostituisci i seguenti valori in questo script con i tuoi dati:

- */mount-point-directory* — Il percorso su un'istanza in cui si desidera montare il EFS file system.
- *filesystem-id* — L'ID del file system per il EFS file system.

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="===MYBOUNDARY==="

---MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

packages:
  - amazon-efs-utils

runcmd:
  - mkdir -p /tmp/home
  - rsync -a /home/ /tmp/home
  - echo "filesystem-id:/ /mount-point-directory efs tls,_netdev" >> /etc/fstab
  - mount -a -t efs defaults
  - rsync -a --ignore-existing /tmp/home/ /home
  - rm -rf /tmp/home/

---MYBOUNDARY---
```

Attivazione della modalità senza password SSH

È possibile basarsi sull'esempio della home directory condivisa per implementare SSH connessioni tra istanze del cluster utilizzando le chiavi. SSH Per ogni utente che utilizza il file system home condiviso, esegui uno script simile al seguente:

```
#!/bin/bash

mkdir -p $HOME/.ssh && chmod 700 $HOME/.ssh
touch $HOME/.ssh/authorized_keys
chmod 600 $HOME/.ssh/authorized_keys

if [ ! -f "$HOME/.ssh/id_rsa" ]; then
    ssh-keygen -t rsa -b 4096 -f $HOME/.ssh/id_rsa -N ""
    cat ~/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_keys
fi
```

Note

Le istanze devono utilizzare un gruppo di sicurezza che consenta SSH le connessioni tra i nodi del cluster.

Prenotazioni di capacità in AWS PCS

Puoi prenotare la EC2 capacità di Amazon in una zona di disponibilità specifica e per una durata specifica utilizzando le prenotazioni di capacità su richiesta o i blocchi di EC2 capacità per assicurarti di avere la capacità di elaborazione necessaria quando ne hai bisogno.

Note

AWS PCS supporta On-Demand Capacity Reservations (ODCR) ma attualmente non supporta Capacity Blocks for ML.

Utilizzo con ODCRs AWS PCS

Puoi scegliere in che modo AWS PCS consumare le tue istanze riservate. Se crei un'istanza aperta ODCR, tutte le istanze corrispondenti avviate da AWS PCS o altri processi presenti nel tuo

account vengono conteggiate nella prenotazione. Con un targeting ODCR, solo le istanze avviate con lo specifico ID di prenotazione vengono conteggiate ai fini della prenotazione. Per i carichi di lavoro urgenti, i target ODCRs sono più comuni.

Puoi configurare un gruppo di nodi di AWS PCS calcolo per utilizzare un target ODCR aggiungendolo a un modello di avvio. Ecco i passaggi per farlo:

1. Crea una prenotazione di capacità mirata su richiesta (ODCR).
2. Aggiungi il ODCR a un gruppo di prenotazione della capacità.
3. Associa il gruppo Capacity Reservation a un modello di lancio.
4. Crea o aggiorna un gruppo di nodi di AWS PCS calcolo per utilizzare il modello di lancio.

Esempio: prenota e utilizza istanze hpc6a.48xlarge con un target ODCR

Questo comando di esempio crea un target per 32 istanze hpc6a.48xlarge. ODCR Per avviare le istanze riservate in un gruppo di posizionamento, aggiungetele al comando. `--placement-group-arn` È possibile definire una data di fine con `--end-date` e `--end-date-type`, in caso contrario, la prenotazione continuerà fino a quando non verrà terminata manualmente.

```
aws ec2 create-capacity-reservation \  
  --instance-type hpc6a.48xlarge \  
  --instance-platform Linux/UNIX \  
  --availability-zone us-east-2a \  
  --instance-count 32 \  
  --instance-match-criteria targeted
```

Il risultato di questo comando sarà un ARN per il nuovo ODCR. Per utilizzarlo AWS PCS, è necessario aggiungerlo a un gruppo di prenotazione della capacità. ODCR Questo perché AWS PCS non supporta individui ODCRs. Per ulteriori informazioni, consulta [Capacity Reservation groups](#) nella Amazon Elastic Compute Cloud User Guide.

Ecco come aggiungerli ODCR a un gruppo di prenotazione di capacità denominato `EXAMPLE-CR-GROUP`.

```
aws resource-groups group-resources --group EXAMPLE-CR-GROUP \  
  --resource-arns arn:aws:ec2:sa-east-1:123456789012:capacity-reservation/  
  cr-1234567890abcdef1
```

Dopo averlo ODCR creato e aggiunto a un gruppo di prenotazione della capacità, ora può essere collegato a un gruppo di nodi di AWS PCS calcolo aggiungendolo a un modello di avvio. Ecco un esempio di modello di lancio che fa riferimento al gruppo Capacity Reservation.

```
{
  "CapacityReservationSpecification": {
    "CapacityReservationResourceGroupArn": "arn:aws:resource-groups:us-
east-2:123456789012:group/EXAMPLE-CR-GROUP"
  }
}
```

Infine, crea o aggiorna un gruppo di nodi di AWS PCS calcolo per utilizzare le istanze `hpc6a.48xlarge` e utilizza il modello di avvio che fa riferimento a tali istanze nel relativo gruppo di prenotazione della capacità. ODCR Per un gruppo di nodi statico, imposta il numero minimo e massimo di istanze in base alla dimensione della prenotazione (32). Per un gruppo di nodi dinamico, imposta il numero minimo di istanze su 0 e il massimo fino alla dimensione della prenotazione.

Questo esempio è una semplice implementazione di un singolo ODCR che ha fornito il provisioning per un gruppo di nodi di calcolo. Tuttavia, AWS PCS supporta molti altri design. Ad esempio, puoi suddividere un gruppo di grandi dimensioni ODCR o Capacity Reservation tra più gruppi di nodi di elaborazione. In alternativa, puoi utilizzare il ODCRs fatto che un altro AWS account abbia creato e condiviso con il tuo. Il vincolo principale è che deve ODCRs sempre essere contenuto in un gruppo di Capacity Reservation.

Per ulteriori informazioni, consulta [On-Demand Capacity Reservations e Capacity Blocks for ML](#) nella Amazon Elastic Compute Cloud User Guide.

Parametri utili del modello di lancio

Questa sezione descrive alcuni parametri del modello di lancio che possono essere ampiamente utili.

AWS PCS

Attiva il monitoraggio dettagliato CloudWatch

Puoi abilitare la raccolta di CloudWatch metriche a intervalli più brevi utilizzando un parametro del modello di avvio.

AWS Management Console

Nelle pagine della console per la creazione o la modifica dei modelli di avvio, questa opzione si trova nella sezione Dettagli avanzati. Imposta CloudWatch il monitoraggio dettagliato su Abilita.

YAML

```
Monitoring:
  Enabled: True
```

JSON

```
{"Monitoring": {"Enabled": "True"}}
```

Per ulteriori informazioni, consulta [Attivare o disattivare il monitoraggio dettagliato per le istanze](#) nella Amazon Elastic Compute Cloud User Guide for Linux Instances.

Instance Metadata Service versione 2 (v2) IMDS

L'utilizzo della IMDS versione v2 con EC2 le istanze offre significativi miglioramenti della sicurezza e aiuta a mitigare i potenziali rischi associati all'accesso ai metadati delle istanze negli ambienti. AWS

AWS Management Console

Nelle pagine della console per la creazione o la modifica dei modelli di avvio, questa opzione si trova nella sezione Dettagli avanzati. Imposta Metadati accessibili su Enabled, la versione Metadata solo su V2 (token richiesto) e il limite dell'hop di risposta dei metadati su 4.

YAML

```
MetadataOptions:
  HttpEndpoint: enabled
  HttpTokens: required
  HttpPutResponseHopLimit: 4
```

JSON

```
{
  "MetadataOptions": {
    "HttpEndpoint": "enabled",
    "HttpPutResponseHopLimit": 4,
    "HttpTokens": "required"
  }
}
```

AWS PCScode

Una AWS PCS coda è un'astrazione leggera rispetto all'implementazione nativa di una coda di lavoro da parte dello scheduler. Nel caso di Slurm, una AWS PCS coda è equivalente a una partizione Slurm.

Gli utenti inviano i lavori a una coda in cui risiedono fino a quando non è possibile programmarne l'esecuzione sui nodi forniti da uno o più gruppi di nodi di calcolo. Un AWS PCS cluster può avere più code di lavoro. Ad esempio, puoi creare una coda che utilizza Amazon EC2 On-demand Instances per lavori ad alta priorità e un'altra coda che utilizza Amazon EC2 Spot Instances per lavori a bassa priorità.

Argomenti

- [Creazione di una coda in AWS PCS](#)
- [Aggiornamento di una AWS PCS coda](#)
- [Eliminazione di una coda in AWS PCS](#)

Creazione di una coda in AWS PCS

Questo argomento fornisce una panoramica delle opzioni disponibili e descrive gli elementi da considerare quando si crea una coda in AWS PCS

Prerequisiti

- Un AWS PCS cluster: le code possono essere create solo in associazione con un cluster specifico PCS.
- Uno o più gruppi di nodi di AWS PCS calcolo: una coda deve essere associata ad almeno un gruppo di nodi di PCS calcolo.

Per creare una coda in AWS PCS

È possibile creare una coda utilizzando AWS Management Console o il AWS CLI

AWS Management Console

Per creare una coda utilizzando la console

1. Apri la AWS PCS console all'indirizzo `https://console.aws.amazon.com/pcs/home#/clusters`
2. Seleziona il cluster in cui desideri creare una coda. Vai a Queues e scegli Crea coda.
3. Nella sezione Configurazione della coda, fornisci i seguenti valori:
 - a. Nome della coda: un nome per la coda. Il nome può contenere solo caratteri alfanumerici (con distinzione tra lettere maiuscole e minuscole) e trattini. Deve iniziare con un carattere alfabetico e non può superare i 25 caratteri. Il nome deve essere univoco all'interno del cluster.
 - b. Gruppi di nodi di calcolo: seleziona uno o più gruppi di nodi di calcolo per servire questa coda. Un gruppo di nodi di calcolo può essere associato a più di una coda.
4. (Facoltativo) In Tag, aggiungi qualsiasi tag alla coda AWS PCS
5. Scegliere Crea coda. Il campo Stato mostrerà Creazione durante la configurazione della coda. La creazione della coda può richiedere diversi minuti.

Passaggio successivo consigliato

- Invia un lavoro alla tua nuova coda

AWS CLI

Per creare una coda utilizzando AWS CLI

Crea la tua coda con il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:

1. Replace (Sostituisci) *region-code* con la AWS regione in cui vuoi creare il cluster.
2. Replace (Sostituisci) *my-queue* con il nome della coda. Il nome può contenere solo caratteri alfanumerici (con distinzione tra lettere maiuscole e minuscole) e trattini. Deve iniziare con un carattere alfabetico e non può superare i 25 caratteri. Il nome deve essere univoco all'interno del cluster.
3. Replace (Sostituisci) *my-cluster* con il nome o clusterId del tuo cluster.

4. Sostituisci il valore di `computeNodeGroupId` con il tuo identificatore di gruppo di nodi di calcolo. Nota che non puoi specificare i nomi dei gruppi di nodi di calcolo durante la creazione di una coda.

```
aws pcs create-queue --region region-code \  
  --queue-name my-queue \  
  --cluster-identifier my-cluster \  
  --compute-node-group-configurations \  
  computeNodeGroupId=computeNodeGroupExampleID1
```

La creazione della coda può richiedere diversi minuti. È possibile interrogare lo stato della coda con il seguente comando. Non potrai inviare lavori alla coda finché non verrà raggiunto lo stato corrispondente. ACTIVE

```
aws pcs get-queue --region region-code \  
  --cluster-identifier my-cluster \  
  --queue-identifier my-queue
```

Passaggio successivo consigliato

- Invia un lavoro alla tua nuova coda

Aggiornamento di una AWS PCS coda

Questo argomento fornisce una panoramica delle opzioni disponibili e descrive gli elementi da considerare quando si aggiorna una AWS PCS coda.

Considerazioni sull'aggiornamento di una coda AWS PCS

Gli aggiornamenti della coda non influiranno sui job in esecuzione, ma il cluster potrebbe non essere in grado di accettare nuovi lavori durante l'aggiornamento della coda.

Per aggiornare un gruppo di nodi di AWS PCS calcolo

È possibile aggiornare un gruppo di nodi utilizzando la Console AWS di gestione o il AWSCLI.

AWS Management Console

Per aggiornare una coda

1. Apri la AWS PCS console all'indirizzo `https://console.aws.amazon.com/pcs/home#/clusters`
2. Seleziona il cluster in cui desideri aggiornare una coda.
3. Vai a Code, vai alla coda che desideri aggiornare, quindi seleziona Modifica.
4. Nella sezione di configurazione della coda, aggiorna uno dei seguenti valori:
 - Gruppi di nodi: aggiungi o rimuovi i gruppi di nodi di calcolo dall'associazione alla coda.
 - Tag: aggiungi o rimuovi tag per la coda.
5. Scegli Aggiorna. Il campo Stato mostrerà Aggiornamento durante l'applicazione delle modifiche.

Important

Gli aggiornamenti delle code possono richiedere diversi minuti.

AWS CLI

Per aggiornare una coda

1. Aggiorna la coda con il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:
 - a. Replace (Sostituisci) *region-code* con il Regione AWS quale vuoi creare il tuo cluster.
 - b. Replace (Sostituisci) *my-queue* con il nome o `computeNodeGroupId` per la tua coda.
 - c. Replace (Sostituisci) *my-cluster* con il nome o `clusterId` del tuo cluster.
 - d. Per modificare le associazioni dei gruppi di nodi di calcolo, fornisci un elenco aggiornato per `--compute-node-group-configurations`.
 - Ad esempio, per aggiungere un secondo gruppo di nodi di calcolo:
`computeNodeGroupExampleID2`

```
--compute-node-group-configurations  
computeNodeId=computeNodeGroupExampleID1, computeNodeId=computeNodeGro
```

```
aws pcs update-queue --region region-code \  
  --queue-identifier my-queue \  
  --cluster-identifier my-cluster \  
  --compute-node-group-configurations \  
  computeNodeId=computeNodeGroupExampleID1
```

2. L'aggiornamento della coda può richiedere diversi minuti. È possibile interrogare lo stato della coda con il seguente comando. Non potrai inviare lavori alla coda finché non verrà raggiunto lo stato corrispondente. ACTIVE

```
aws pcs get-queue --region region-code \  
  --cluster-identifier my-cluster \  
  --queue-identifier my-queue
```

Passaggi successivi consigliati

- Invia un lavoro alla tua coda aggiornata.

Eliminazione di una coda in AWS PCS

Questo argomento fornisce una panoramica su come eliminare una coda in AWS PCS

Considerazioni sull'eliminazione di una coda

- Se ci sono lavori in esecuzione nella coda, questi verranno terminati dallo scheduler quando la coda viene eliminata. I lavori in sospeso in coda verranno annullati. Valuta la possibilità di attendere il completamento dei lavori in coda o di interromperli o annullarli manualmente utilizzando i comandi nativi dello scheduler (come per Slurm). `scancel`

Elimina la coda

È possibile utilizzare AWS Management Console o AWS CLI per eliminare una coda.

AWS Management Console

Per eliminare una coda

1. Apri la [AWS PCSconsole](#).
2. Seleziona il cluster della coda.
3. Vai a Code e seleziona la coda da eliminare.
4. Scegli Elimina.
5. Viene visualizzato il campo Stato. Deleting Per il completamento possono essere necessari alcuni minuti.

Note

È possibile utilizzare i comandi nativi dello scheduler per confermare l'eliminazione della coda. Ad esempio, usa `sinfo` o `squeue` per Slurm.

AWS CLI

Per eliminare una coda

- Utilizzate il seguente comando per eliminare una coda, con queste sostituzioni:
 - Replace (Sostituisci) *region-code* con Regione AWS il cluster in uso.
 - Replace (Sostituisci) *my-queue* con il nome o l'ID della coda.
 - Replace (Sostituisci) *my-cluster* con il nome o l'ID del tuo cluster.

```
aws pcs delete-queue --region region-code \  
  --queue-identifier my-queue \  
  --cluster-identifier my-cluster
```

L'eliminazione della coda può richiedere diversi minuti.

Note

È possibile utilizzare i comandi nativi dello scheduler per confermare l'eliminazione della coda. Ad esempio, usa `sinfo` o `squeue` per Slurm.

AWS PCS nodi di accesso

Un AWS PCS cluster in genere necessita di almeno 1 nodo di accesso per supportare l'accesso interattivo e la gestione dei lavori. Un modo per eseguire questa operazione è utilizzare un gruppo di nodi di AWS PCS calcolo statici configurato per la funzionalità del nodo di accesso. Puoi anche configurare un'EC2istanza autonoma che funga da nodo di accesso.

Argomenti

- [Utilizzo di un gruppo di nodi di AWS PCS calcolo per fornire nodi di accesso](#)
- [Utilizzo di istanze autonome come AWS PCS nodi di accesso](#)

Utilizzo di un gruppo di nodi di AWS PCS calcolo per fornire nodi di accesso

Questo argomento fornisce una panoramica delle opzioni di configurazione suggerite e descrive gli aspetti da considerare quando si utilizza un gruppo di nodi di AWS PCS calcolo per fornire un accesso persistente e interattivo al cluster.

Creazione di un gruppo di nodi di AWS PCS calcolo per i nodi di accesso

Dal punto di vista operativo, questo non è molto diverso dalla creazione di un normale gruppo di nodi di calcolo. Tuttavia, ci sono alcune scelte di configurazione chiave:

- Imposta una configurazione di scalabilità statica di almeno un'EC2istanza nel gruppo di nodi di calcolo.
- Scegli l'opzione di acquisto su richiesta per evitare che le tue istanze vengano recuperate.
- Scegli un nome informativo per il gruppo di nodi di calcolo, ad esempio `login`.
- Se desideri che le istanze del nodo di accesso siano accessibili al di fuori delle tueVPC, prendi in considerazione l'utilizzo di una sottorete pubblica.
- Se intendi consentire SSH l'accesso, il modello di avvio dovrà disporre di un gruppo di sicurezza che esponga la SSH porta agli indirizzi IP da te scelti.

- Il profilo dell'IAM istanza deve avere solo le AWS autorizzazioni che desideri siano concesse agli utenti finali. Per informazioni dettagliate, vedi [IAM profili di istanza per AWS Parallel Computing Service](#).
- Valuta la possibilità AWS di consentire a Systems Manager Session Manager di gestire le istanze di accesso.
- Prendi in considerazione la possibilità di limitare l'accesso alle AWS credenziali dell'istanza solo agli utenti amministrativi
- Seleziona tipi di istanze meno costosi rispetto ai normali gruppi di nodi di calcolo, poiché i nodi di accesso funzioneranno continuamente.
- Usa lo stesso (o un derivato) AMI degli altri gruppi di nodi di calcolo per assicurarti che su tutte le istanze sia installato lo stesso software. Per ulteriori informazioni sulla personalizzazione, consulta AMIs [Amazon Machine Images \(AMIs\) per AWS PCS](#)
- Configura lo stesso file system di rete (AmazonEFS, Amazon FSx for Lustre, ecc.) montato sui nodi di accesso e sulle istanze di calcolo. Per ulteriori informazioni, consulta [Utilizzo di file system di rete con AWS PCS](#).

Accedi ai tuoi nodi di accesso

Una volta raggiunto ACTIVE lo stato del nuovo gruppo di nodi di calcolo, puoi trovare le EC2 istanze che ha creato e accedervi. Per ulteriori informazioni, consulta [Ricerca di istanze di gruppi di nodi di calcolo in AWS PCS](#).

Aggiornamento di un gruppo di nodi di AWS PCS calcolo per i nodi di accesso

È possibile aggiornare un gruppo di nodi di accesso utilizzando UpdateComputeNodeGroup. Come parte del processo di aggiornamento del gruppo di nodi, le istanze in esecuzione verranno sostituite. Tieni presente che ciò interromperà tutte le sessioni o i processi utente attivi sull'istanza. I job Slurm in esecuzione o in coda non subiranno alcuna modifica. Per ulteriori informazioni, consulta [Aggiornamento di un gruppo di nodi di AWS PCS calcolo](#).

Puoi anche modificare il modello di avvio utilizzato dal tuo gruppo di nodi di calcolo. È necessario utilizzare UpdateComputeNodeGroup per applicare il modello di avvio aggiornato al gruppo di nodi di calcolo. Le nuove EC2 istanze lanciate nel gruppo di nodi di calcolo utilizzano il modello di avvio aggiornato. Per ulteriori informazioni, consulta [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#).

Eliminazione di un gruppo di nodi di AWS PCS calcolo per i nodi di accesso

È possibile aggiornare un gruppo di nodi di accesso utilizzando il meccanismo di eliminazione del gruppo di nodi di calcolo in AWS PCS. Le istanze in esecuzione verranno terminate come parte dell'eliminazione del gruppo di nodi. Tieni presente che ciò interromperà tutte le sessioni o i processi utente attivi sull'istanza. I job Slurm in esecuzione o in coda non subiranno alcuna modifica. Per ulteriori informazioni, consulta [Eliminazione di un gruppo di nodi di calcolo in AWS PCS](#).

Utilizzo di istanze autonome come AWS PCS nodi di accesso

Puoi configurare EC2 istanze indipendenti per interagire con lo scheduler Slurm di un AWS PCS cluster. Ciò è utile per creare nodi di accesso, workstation o host dedicati alla gestione del flusso di lavoro che funzionano con AWS PCS i cluster ma operano al di fuori della gestione. AWS PCS. A tale scopo, ogni istanza autonoma deve:

1. Avere installata una versione del software Slurm compatibile.
2. Essere in grado di connettersi all'endpoint AWS PCS Slurmctld del cluster.
3. Configurare correttamente Slurm Auth e Cred Kiosk Daemon () con l'endpoint e il segreto del cluster. sackd AWS PCS [Per ulteriori informazioni, consulta sackd nella documentazione di Slurm.](#)

Questo tutorial ti aiuta a configurare un'istanza indipendente che si connette a un cluster. AWS PCS

Indice

- [Passaggio 1: recuperare l'indirizzo e il segreto per il cluster di destinazione AWS PCS](#)
- [Passaggio 2: avviare un'EC2istanza](#)
- [Passaggio 3: installa Slurm sull'istanza](#)
- [Fase 4 — Recuperare e archiviare il segreto del cluster](#)
- [Fase 5 — Configurare la connessione al cluster AWS PCS](#)
- [Fase 6 — \(Facoltativo\) Verifica della connessione](#)

Passaggio 1: recuperare l'indirizzo e il segreto per il cluster di destinazione AWS PCS

Recupera i dettagli sul AWS PCS cluster di destinazione utilizzando AWS CLI il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:

- Replace (Sostituisci) *region-code* con il Regione AWS luogo in cui è in esecuzione il cluster di destinazione.
- Replace (Sostituisci) *cluster-ident* con il nome o l'identificatore del cluster di destinazione

```
aws pcs get-cluster --region region-code --cluster-identifier cluster-ident
```

Il comando restituirà un output simile a questo esempio.

```
{
  "cluster": {
    "name": "independent-instance-demo",
    "id": "s3431v9rx2",
    "arn": "arn:aws:pcs:us-east-1:012345678901:cluster/s3431v9rx2",
    "status": "ACTIVE",
    "createdAt": "2024-07-12T15:32:27.225136+00:00",
    "modifiedAt": "2024-07-12T15:32:27.225136+00:00",
    "scheduler": {
      "type": "SLURM",
      "version": "23.11"
    },
    "size": "SMALL",
    "networking": {
      "subnetIds": [
        "subnet-0123456789abcdef"
      ],
      "securityGroupIds": [
        "sg-0123456789abcdef"
      ]
    },
    "endpoints": [
      {
        "type": "SLURMCTLD",
        "privateIpAddress": "10.3.149.220",
        "port": "6817"
      }
    ],
    "authKey": {
      "secretArn": "arn:aws:secretsmanager:us-east-1:123456789012:secret:pcs!slurm-secret-s3431v9rx2-FN7tJFf",
      "secretVersion": "ff58d1fd-070e-4bbc-98a0-64ef967cebcc"
    }
  }
}
```

```
}  
}
```

In questo esempio, l'endpoint del controller Slurm del cluster ha un indirizzo IP di 10.3.149.220 ed è in esecuzione sulla porta. 6817 `secretArn` verrà utilizzato nei passaggi successivi per recuperare il segreto del cluster. L'indirizzo IP e la porta verranno utilizzati nei passaggi successivi per configurare il `sackd` servizio.

Passaggio 2: avviare un'EC2istanza

Per avviare un'istanza EC2

1. Apri la [EC2console Amazon](#).
2. Nel pannello di navigazione, scegli Instances (Istanze) e quindi scegli Launch instances (Avvia istanze) per aprire la nuova procedura guidata di avvio dell'istanza.
3. (Facoltativo) Nella sezione Nome e tag, fornisci un nome per l'istanza, ad esempio PCS-LoginNode. Il nome viene assegnato all'istanza come tag di risorsa (Name=PCS-LoginNode).
4. Nella sezione Immagini dell'applicazione e del sistema operativo, selezionate uno AMI per uno dei sistemi operativi supportati da AWS PCS. Per ulteriori informazioni, consulta [Sistemi operativi supportati](#).
5. Nella sezione Tipo di istanza, seleziona un tipo di istanza supportato. Per ulteriori informazioni, consulta [Tipi di istanze supportati](#).
6. Nella sezione Key pair, seleziona la coppia di SSH chiavi da usare per l'istanza.
7. Nella sezione Impostazioni di rete:
 - Scegli Modifica.
 - i. Seleziona il VPC tipo del tuo AWS PCS cluster.
 - ii. Per Firewall (gruppi di sicurezza), scegli Seleziona gruppo di sicurezza esistente.
 - A. Seleziona un gruppo di sicurezza che consenta il traffico tra l'istanza e il controller AWS PCS Slurm del cluster di destinazione. Per ulteriori informazioni, consulta [Requisiti e considerazioni sui gruppi di sicurezza](#).
 - B. (Facoltativo) Seleziona un gruppo di sicurezza che consenta l'SSHaccesso in entrata alla tua istanza.
8. Nella sezione Archiviazione, configura i volumi di archiviazione in base alle esigenze. Assicurati di configurare uno spazio sufficiente per installare applicazioni e librerie adatte al tuo caso d'uso.

- In Avanzato, scegli un IAM ruolo che consenta l'accesso al segreto del cluster. Per ulteriori informazioni, consulta [Ottieni il segreto del cluster Slurm](#).
- Nel riquadro Riepilogo, scegli Launch instance.

Passaggio 3: installa Slurm sull'istanza

Quando l'istanza è stata lanciata e diventa attiva, connettiti ad essa utilizzando il tuo meccanismo preferito. Utilizza il programma di installazione Slurm fornito da AWS per installare Slurm sull'istanza. Per ulteriori informazioni, consulta [Programma di installazione Slurm](#).

Scarica il programma di installazione di Slurm, decomprimilo e usa lo script per installare Slurm. `installer.sh` Per ulteriori informazioni, consulta [Passaggio 3: installa Slurm](#).

Fase 4 — Recuperare e archiviare il segreto del cluster

Queste istruzioni richiedono il. AWS CLI Per ulteriori informazioni, vedere [Installazione o aggiornamento alla versione più recente di AWS CLI nella Guida per l'AWS Command Line Interface utente della versione 2](#).

Memorizza il segreto del cluster con i seguenti comandi.

- Crea la directory di configurazione per Slurm.

```
sudo mkdir -p /etc/slurm
```

- Recupera, decodifica e archivia il segreto del cluster. Prima di eseguire questo comando, sostituisci *region-code* con la regione in cui è in esecuzione il cluster di destinazione e sostituisci *secret-arn* con il valore `secretArn` recuperato nel [passaggio 1](#).

```
sudo aws secretsmanager get-secret-value \  
  --region region-code \  
  --secret-id 'secret-arn' \  
  --version-stage AWSCURRENT \  
  --query 'SecretString' \  
  --output text | base64 -d > /etc/slurm/slurm.key
```

Warning

In un ambiente multiutente, qualsiasi utente con accesso all'istanza potrebbe essere in grado di recuperare il segreto del cluster se può accedere al servizio di metadati

dell'istanza (). IMDS Questo, a sua volta, potrebbe consentire loro di impersonare altri utenti. Valuta la possibilità di limitare l'accesso solo IMDS agli utenti root o amministrativi. In alternativa, prendi in considerazione l'utilizzo di un meccanismo diverso che non si basi sul profilo dell'istanza per recuperare e configurare il segreto.

- Imposta proprietà e autorizzazioni sul file chiave Slurm.

```
sudo chmod 0600 /etc/slurm/slurm.key
sudo chown slurm:slurm /etc/slurm/slurm.key
```

Note

La chiave Slurm deve essere di proprietà dell'utente e del gruppo con cui viene eseguito il servizio. sackd

Fase 5 — Configurare la connessione al cluster AWS PCS

Per stabilire una connessione al AWS PCS cluster, sackd avvialo come servizio di sistema seguendo questi passaggi.

1. Imposta il file di ambiente per il sackd servizio con il comando che segue. Prima di eseguire il comando, sostituisci *ip-address* e *port* con i valori recuperati dagli endpoint nel [passaggio 1](#).

```
sudo echo "SACKD_OPTIONS='--conf-server=ip-address:port'" > /etc/sysconfig/sackd
```

2. Crea un file systemd di servizio per la gestione del sackd processo.

```
sudo cat << EOF > /etc/systemd/system/sackd.service
[Unit]
Description=Slurm auth and cred kiosk daemon
After=network-online.target remote-fs.target
Wants=network-online.target
ConditionPathExists=/etc/sysconfig/sackd

[Service]
Type=notify
EnvironmentFile=/etc/sysconfig/sackd
User=slurm
Group=slurm
```

```
RuntimeDirectory=slurm
RuntimeDirectoryMode=0755
ExecStart=/opt/aws/pcs/scheduler/slurm-23.11/sbin/sackd --systemd \${SACKD_OPTIONS}
ExecReload=/bin/kill -HUP \${MAINPID}
KillMode=process
LimitNOFILE=131072
LimitMEMLOCK=infinity
LimitSTACK=infinity

[Install]
WantedBy=multi-user.target
EOF
```

3. Imposta la proprietà del file sackd di servizio.

```
sudo chown root:root /etc/systemd/system/sackd.service && \
sudo chmod 0644 /etc/systemd/system/sackd.service
```

4. Abilita il sackd servizio.

```
sudo systemctl daemon-reload && sudo systemctl enable sackd
```

5. Avviare il servizio sackd.

```
sudo systemctl start sackd
```

Fase 6 — (Facoltativo) Verifica della connessione

Verificare che il sackd servizio sia in esecuzione. Di seguito è riportato un output di esempio. Se ci sono errori, di solito vengono visualizzati qui.

```
[root@ip-10-3-27-112 ~]# systemctl status sackd
[x] sackd.service - Slurm auth and cred kiosk daemon
   Loaded: loaded (/etc/systemd/system/sackd.service; enabled; vendor preset: disabled)
   Active: active (running) since Tue 2024-07-16 16:34:55 UTC; 8s ago
 Main PID: 9985 (sackd)
   CGroup: /system.slice/sackd.service
           ##9985 /opt/aws/pcs/scheduler/slurm-23.11/sbin/sackd --systemd --conf-
server=10.3.149.220:6817

Jul 16 16:34:55 ip-10-3-27-112.ec2.internal systemd[1]: Starting Slurm auth and cred
kiosk daemon...
```

```
Jul 16 16:34:55 ip-10-3-27-112.ec2.internal systemd[1]: Started Slurm auth and cred
kiosk daemon.
Jul 16 16:34:55 ip-10-3-27-112.ec2.internal sackd[9985]: sackd: running
```

Conferma che le connessioni al cluster funzionino utilizzando i comandi del client Slurm come `esinfo`. `squeue` Ecco un esempio di output da `sinfo`

```
[root@ip-10-3-27-112 ~]# /opt/aws/pcs/scheduler/slurm-23.11/bin/sinfo
PARTITION AVAIL TIMELIMIT NODES STATE NODELIST
all up infinite 4 idle~ compute-[1-4]
```

Dovresti anche essere in grado di inviare offerte di lavoro. Ad esempio, un comando simile a questo esempio avvierebbe un processo interattivo su 1 nodo del cluster.

```
/opt/aws/pcs/scheduler/slurm-23.11/bin/srun --nodes=1 -p all --pty bash -i
```

AWS PCS Rete

Il tuo AWS PCS cluster viene creato in un AmazonVPC. Questo capitolo include i seguenti argomenti sul networking per lo scheduler e i nodi del cluster.

Ad eccezione della scelta di una sottorete in cui avviare le istanze, è necessario utilizzare i modelli di EC2 avvio per configurare la rete per i gruppi di nodi di AWS PCS calcolo. Per ulteriori informazioni sui modelli di avvio, consulta [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#)

Argomenti

- [AWS PCSVPCe requisiti e considerazioni sulla sottorete](#)
- [Creare un file VPC per il tuo AWS PCS cluster](#)
- [Gruppi di sicurezza in AWS PCS](#)
- [Interfacce di rete multiple in AWS PCS](#)
- [Gruppi di collocamento per EC2 istanze in AWS PCS](#)
- [Utilizzo di Elastic Fabric Adapter \(EFA\) con AWS PCS](#)

AWS PCSVPCe requisiti e considerazioni sulla sottorete

Quando si crea un AWS PCS cluster, si specifica VPC una sottorete al suo interno. VPC Questo argomento fornisce una panoramica dei requisiti e delle considerazioni AWS PCS specifici per le

VPC sottoreti e le sottoreti utilizzate con il cluster. Se non ne hai uno con VPC cui utilizzarlo AWS PCS, puoi crearne uno utilizzando un modello fornito AWS. AWS CloudFormation Per ulteriori informazioni VPCs, consulta [Virtual private cloud \(VPC\)](#) nella Amazon VPC User Guide.

VPC requisiti e considerazioni

Quando si crea un cluster, VPC ciò che si specifica deve soddisfare i requisiti e le considerazioni seguenti:

- VPC È necessario disporre di un numero sufficiente di indirizzi IP disponibili per il cluster, tutti i nodi e le altre risorse del cluster che si desidera creare. Per ulteriori informazioni, consulta la sezione [Indirizzamento IP per le tue sottoreti VPCs e subnet](#) nella Amazon VPC User Guide.
- VPC Deve avere un DNS nome host e DNS un supporto per la risoluzione. In caso contrario, i nodi non possono registrare il cluster di clienti. Per ulteriori informazioni, consulta [DNS gli attributi per te VPC](#) nella Amazon VPC User Guide.
- VPC Potrebbe essere necessario utilizzare gli VPC endpoint AWS PrivateLink per poter contattare il AWS PCS API. Per ulteriori informazioni, consulta [Connect your VPC to services using AWS PrivateLink](#) nella Amazon VPC User Guide.

Considerazioni e requisiti relativi alle sottoreti

Quando crei un cluster Slurm, AWS PCS crea un'[interfaccia di rete elastica \(ENI\)](#) nella sottorete specificata. Questa interfaccia di rete consente la comunicazione tra il controller dello scheduler e il cliente. VPC L'interfaccia di rete consente inoltre a Slurm di comunicare con i componenti distribuiti nell'account del cliente. È possibile specificare la sottorete per un cluster solo al momento della creazione.

Requisiti relativi alla sottorete per i cluster

La [sottorete](#) specificata quando si crea un cluster deve soddisfare i seguenti requisiti:

- La sottorete deve avere almeno un indirizzo IP da utilizzare da. AWS PCS
- La sottorete non può risiedere in AWS Outposts AWS Wavelength, o in una AWS zona locale.
- La sottorete può essere pubblica o privata. Si consiglia di specificare una sottorete privata, se possibile. Una sottorete pubblica è una sottorete con una tabella di routing che include un percorso verso un [gateway Internet](#); una sottorete privata è una sottorete con una tabella di routing che non include un percorso verso un gateway Internet.

Requisiti relativi alla sottorete per i nodi

È possibile distribuire nodi e altre risorse del cluster nella sottorete specificata al momento della creazione del AWS PCS cluster e in altre sottoreti dello stesso VPC.

Qualsiasi sottorete in cui vengono distribuiti nodi e risorse del cluster deve soddisfare i seguenti requisiti:

- È necessario assicurarsi che la sottorete disponga di un numero sufficiente di indirizzi IP disponibili per distribuire tutti i nodi e le risorse del cluster.
- Se si prevede di distribuire nodi in una sottorete pubblica, tale sottorete deve assegnare automaticamente IPv4 gli indirizzi pubblici.
- Se la sottorete in cui distribuisce i nodi è una sottorete privata e la relativa tabella di routing non include un percorso verso un [dispositivo di traduzione degli indirizzi di rete \(NAT\)](#), aggiungi gli endpoint utilizzando il client IPv4. VPC AWS PrivateLink VPC VPC. Gli endpoint sono necessari per tutti i AWS servizi contattati dai nodi. L'unico endpoint richiesto è quello di consentire AWS PCS al nodo di richiamare l'azione `registerNodeGroupInstancesAPI`.
- Lo stato della sottorete pubblica o privata non ha alcun impatto AWS PCS; gli endpoint richiesti devono essere raggiungibili.

Creare un file VPC per il tuo AWS PCS cluster

Puoi creare un Amazon Virtual Private Cloud (AmazonVPC) per i tuoi cluster all'interno di AWS Parallel Computing Service (AWS PCS).

Usa Amazon VPC per lanciare VPC risorse in una rete virtuale che hai definito. Questa rete virtuale è simile a una rete tradizionale da gestire all'interno del proprio data center, ma con i vantaggi dell'infrastruttura scalabile di Amazon Web Services. Ti consigliamo di avere una conoscenza approfondita del VPC servizio Amazon prima di distribuire i VPC cluster di produzione. Per ulteriori informazioni, consulta [What is AmazonVPC?](#) in modalità visuale d'autore. Guida per VPC l'utente di Amazon.

Un PCS cluster, nodi e risorse di supporto (come file system e servizi di directory) vengono distribuiti all'interno di AmazonVPC. Se desideri utilizzare un Amazon esistente VPC con PCS, questo deve soddisfare i requisiti descritti in [AWS PCS VPC e requisiti e considerazioni sulla sottorete](#). Questo argomento descrive come creare un file VPC che soddisfi PCS i requisiti utilizzando un AWS CloudFormation modello AWS fornito. Dopo l'implementazione di un modello, puoi visualizzare le

risorse create dal modello per sapere esattamente quali risorse ha creato e la configurazione di tali risorse.

Prerequisiti

Per creare un Amazon VPC per PCS, devi disporre delle IAM autorizzazioni necessarie per creare VPC risorse Amazon. Queste risorse sono sottoreti VPCs, gruppi di sicurezza, tabelle e percorsi di routing, Internet e gateway. NAT Per ulteriori informazioni, consulta [Create a VPC with a public subnet](#) nella Amazon VPC User Guide. Per rivedere l'elenco completo di Amazon EC2, consulta [Azioni, risorse e chiavi di condizione per Amazon EC2](#) nel Service Authorization Reference.

Crea un Amazon VPC

Crea un file VPC copiando e incollando quello appropriato URL per il Regione AWS luogo in cui lo PCS utilizzerai. Puoi anche scaricare il AWS CloudFormation modello e caricarlo tu stesso sulla [AWS CloudFormation console](#).

- US East (N. Virginia) (Stati Uniti orientali (Virginia settentrionale)) (us-east-1)

```
https://console.aws.amazon.com/cloudformation/home?region=us-east-1#/stacks/create/review?stackName=hpc-networking&templateURL=https://aws-hpc-recipes.s3.us-east-1.amazonaws.com/main/recipes/net/hpc_large_scale/assets/main.yaml
```

- US East (Ohio) (Stati Uniti orientali (Ohio)) (us-east-2)

```
https://console.aws.amazon.com/cloudformation/home?region=us-east-2#/stacks/create/review?stackName=hpc-networking&templateURL=https://aws-hpc-recipes.s3.us-east-1.amazonaws.com/main/recipes/net/hpc_large_scale/assets/main.yaml
```

- US West (Oregon) (Stati Uniti occidentali (Oregon)) (us-west-2)

```
https://console.aws.amazon.com/cloudformation/home?region=us-west-2#/stacks/create/review?stackName=hpc-networking&templateURL=https://aws-hpc-recipes.s3.us-east-1.amazonaws.com/main/recipes/net/hpc_large_scale/assets/main.yaml
```

- Solo modello

```
https://aws-hpc-recipes.s3.us-east-1.amazonaws.com/main/recipes/net/hpc_large_scale/assets/main.yaml
```

Per creare un Amazon VPC per PCS

1. Apri il modello nella [AWS CloudFormation console](#).

Note

Questi sono precompilati nel modello in modo che tu possa semplicemente lasciarli come valori predefiniti.

2. In Fornisci un nome per lo stack, quindi per nome dello stack, inserisci `hpc-networking`
3. In Parametri, inserisci i seguenti dettagli:
 - a. Sotto VPC, quindi CidrBlock, inserisci `10.3.0.0/16`
 - b. In Sottoreti A:
 - i. Quindi A, CidrPublicSubnet inserisci `10.3.0.0/20`
 - ii. Poi CidrPrivateSubnetA, entra `10.3.128.0/20`
 - c. In Sottoreti B:
 - i. Quindi CidrPublicSubnetB, inserisci `10.3.16.0/20`
 - ii. Poi CidrPrivateSubnetA, entra `10.3.144.0/20`
 - d. In Sottoreti C:
 - i. Per ProvisionSubnetsC, seleziona `True`

Note

Se ne stai creando una VPC in una regione con meno di tre zone di disponibilità, questa opzione verrà ignorata se impostata su `True`

- ii. Quindi CidrPublicSubnetB, inserisci `10.3.32.0/20`
 - iii. Poi CidrPrivateSubnetA, entra `10.3.160.0/20`
4. In Capacità, seleziona la casella Riconosco che AWS CloudFormation potrebbe creare IAM risorse.

Monitora lo stato dello AWS CloudFormation stack. Una volta raggiunta la `CREATE_COMPLETE` soglia, le VPC risorse sono pronte per l'uso.

Note

Per visualizzare tutte le risorse create dal AWS CloudFormation modello, apri la [AWS CloudFormation console](#). Scegli lo stack hpc-networking, quindi la scheda Resources (Risorse).

Gruppi di sicurezza in AWS PCS

I gruppi di sicurezza in Amazon EC2 agiscono come firewall virtuali per controllare il traffico in entrata e in uscita verso le istanze. Utilizza un modello di avvio per un gruppo di nodi di AWS PCS calcolo per aggiungere o rimuovere gruppi di sicurezza alle relative istanze. Se il modello di lancio non contiene interfacce di rete, utilizzalo SecurityGroupIds per fornire un elenco di gruppi di sicurezza. Se il modello di lancio definisce le interfacce di rete, è necessario utilizzare il Groups parametro per assegnare gruppi di sicurezza a ciascuna interfaccia di rete. Per ulteriori informazioni sui modelli di avvio, consulta. [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#)

Note

Le modifiche alla configurazione del gruppo di sicurezza nel modello di lancio influiscono solo sulle nuove istanze lanciate dopo l'aggiornamento del gruppo di nodi di calcolo.

Requisiti e considerazioni sui gruppi di sicurezza

AWS PCS crea un'[interfaccia di rete elastica per più account \(ENI\)](#) nella sottorete specificata durante la creazione di un cluster. Ciò fornisce allo HPC scheduler, che è in esecuzione in un account gestito da AWS, un percorso per comunicare con le EC2 istanze avviate da AWS PCS. A tale scopo, è necessario fornire un gruppo di sicurezza ENI che consenta la comunicazione bidirezionale tra lo scheduler ENI e le istanze del cluster. EC2

Un modo semplice per farlo consiste nel creare un gruppo di sicurezza permissivo e autoreferenziale che consenta il traffico TCP /IP su tutte le porte tra tutti i membri del gruppo. È possibile collegarlo sia al cluster che alle istanze del gruppo di nodi. EC2

Esempio di configurazione permissiva del gruppo di sicurezza

Tipo di regola	Protocolli	Porte	Origine	Destinazione
In entrata	Tutti	Tutti	Personale	
In uscita	Tutti	Tutti		0.0.0.0/0
In uscita	Tutti	Tutti		Personale

[Queste regole consentono a tutto il traffico di fluire liberamente tra il controller Slurm e i nodi, consentono tutto il traffico in uscita verso qualsiasi destinazione e abilitano il traffico. EFA](#)

Esempio di configurazione restrittiva del gruppo di sicurezza

È inoltre possibile limitare le porte aperte tra il cluster e i relativi nodi di elaborazione. Per lo scheduler Slurm, il gruppo di sicurezza collegato al cluster deve consentire le seguenti porte:

- 6817: abilita le connessioni in entrata da e verso le istanze `slurmctld` EC2
- 6818: abilita le connessioni in uscita da e l'esecuzione su istanze `slurmctld` `slurmd` EC2

Il gruppo di sicurezza collegato ai nodi di elaborazione deve consentire le seguenti porte:

- 6817: abilita le connessioni in uscita da istanze a `slurmctld` partire da istanze. EC2
- 6818: abilita le connessioni in entrata e in uscita `slurmd` da e verso le istanze del gruppo di nodi `slurmctld` `slurmd`
- 60001—63000: connessioni in entrata e in uscita tra istanze di gruppi di nodi da supportare `srunk`
- EFA traffico tra istanze del gruppo di nodi. Per ulteriori informazioni, consulta [Preparare un gruppo EFA di sicurezza abilitato](#) nella Guida utente per le istanze Linux
- Qualsiasi altro traffico internodale richiesto dal carico di lavoro

Interfacce di rete multiple in AWS PCS

Alcune EC2 istanze dispongono di più schede di rete. Ciò consente loro di fornire prestazioni di rete più elevate, comprese capacità di larghezza di banda superiori a 100 Gbps e una migliore gestione dei pacchetti. Per ulteriori informazioni sulle istanze con più schede di rete, consulta le [interfacce di rete elastiche](#) nella Amazon Elastic Compute Cloud User Guide.

Configura schede di rete aggiuntive per le istanze in un gruppo di nodi di AWS PCS calcolo aggiungendo interfacce di rete al relativo modello di lancio. EC2 Di seguito è riportato un esempio di modello di avvio che abilita due schede di rete, ad esempio quelle disponibili su un'istanza. `hpc7a.96xlarge` Nota i seguenti dettagli:

- La sottorete per ogni interfaccia di rete deve essere la stessa scelta durante la configurazione del gruppo di nodi di AWS PCS calcolo che utilizzerà il modello di avvio.
- Il dispositivo di rete principale, su cui avverranno le comunicazioni di rete di routine, ad SSH esempio HTTPS il traffico, viene stabilito impostando un di. `DeviceIndex 0` Le altre interfacce di rete dispongono `DeviceIndex` di. `1` Può esserci una sola interfaccia di rete principale, tutte le altre interfacce sono secondarie.
- Tutte le interfacce di rete devono avere un'interfaccia univoca. `NetworkCardIndex` Una pratica consigliata consiste nel numerarle in sequenza così come sono definite nel modello di avvio.
- I gruppi di sicurezza per ogni interfaccia di rete vengono impostati utilizzando `Groups`. In questo esempio, un gruppo di SSH sicurezza in entrata (`sg-SshSecurityGroupId`) viene aggiunto all'interfaccia di rete principale, oltre al gruppo di sicurezza che abilita le comunicazioni all'interno del cluster (`sg-ClusterSecurityGroupId`). Infine, un gruppo di sicurezza che consente le connessioni in uscita a Internet (`sg-InternetOutboundSecurityGroupId`) viene aggiunto alle interfacce primarie e secondarie.

```
{
  "NetworkInterfaces": [
    {
      "DeviceIndex": 0,
      "NetworkCardIndex": 0,
      "SubnetId": "subnet-SubnetId",
      "Groups": [
        "sg-SshSecurityGroupId",
        "sg-ClusterSecurityGroupId",
        "sg-InternetOutboundSecurityGroupId"
      ]
    },
    {
      "DeviceIndex": 1,
      "NetworkCardIndex": 1,
      "SubnetId": "subnet-SubnetId",
      "Groups": ["sg-InternetOutboundSecurityGroupId"]
    }
  ]
}
```

```
}
```

Gruppi di collocamento per EC2 istanze in AWS PCS

È possibile utilizzare un gruppo di collocamento per influenzare il posizionamento delle EC2 istanze in base alle esigenze del carico di lavoro che viene eseguito su di esse.

Tipi di gruppi di posizionamento

- **Cluster:** raggruppa le istanze in una zona di disponibilità per ottimizzare le comunicazioni a bassa latenza.
- **Partizione:** distribuisce le istanze su partizioni logiche per massimizzare la resilienza.
- **Spread:** impone rigorosamente l'avvio di un numero limitato di istanze su hardware distinto, il che può anche favorire la resilienza.

Per ulteriori informazioni, consulta [i gruppi di posizionamento per le tue EC2 istanze Amazon](#) nella Amazon Elastic Compute Cloud User Guide.

Ti consigliamo di includere un gruppo di posizionamento del cluster quando configuri un gruppo di nodi di AWS PCS calcolo per utilizzare Elastic Fabric Adapter (EFA).

Per creare un gruppo di posizionamento del cluster che funzioni con EFA:

1. Crea un gruppo di posizionamento con il tipo cluster per il gruppo di nodi di calcolo.

- Utilizzate il seguente AWS CLI comando:

```
aws ec2 create-placement-group --strategy cluster --group-name PLACEMENT-GROUP-NAME
```

- Potete anche utilizzare un CloudFormation modello per creare un gruppo di posizionamento. Per ulteriori informazioni, consulta [Lavorare con CloudFormation i modelli](#) nella Guida AWS CloudFormation per l'utente. Scarica il modello dal seguente link URL e caricalo nella [CloudFormation console](#).

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/enable_efa/assets/efa-placement-group.yaml
```

2. Includi il gruppo di posizionamento nel modello di EC2 lancio per il gruppo di nodi di AWS PCS calcolo.

Utilizzo di Elastic Fabric Adapter (EFA) con AWS PCS

Elastic Fabric Adapter (EFA) è un'interconnessione di rete avanzata ad alte prestazioni AWS che puoi collegare all'EC2istanza per accelerare le applicazioni di High Performance Computing (HPC) e machine learning. L'abilitazione delle applicazioni in esecuzione su un AWS PCS cluster EFA implica la configurazione delle istanze del gruppo di nodi di AWS PCS calcolo da utilizzare come segue. EFA

Indice

- [Installa EFA su un dispositivo -compatibile AWS PCS AMI](#)
- [Identifica le EFA istanze abilitate EC2](#)
- [Determina quante interfacce di rete sono disponibili](#)
- [Crea un gruppo di sicurezza per supportare EFA le comunicazioni](#)
- [\(Facoltativo\) Create un gruppo di collocamento](#)
- [Crea o aggiorna un modello di EC2 lancio](#)
- [Crea o aggiorna un gruppo di nodi di calcolo](#)
- [\(Facoltativo\) Test EFA](#)
- [\(Facoltativo\) Utilizzate un CloudFormation modello per creare un modello di EFA avvio abilitato](#)

Installa EFA su un dispositivo -compatibile AWS PCS AMI

Il EFA driver AMI utilizzato nel gruppo di nodi di AWS PCS calcolo deve essere installato e caricato. Per informazioni su come creare un software personalizzato AMI con EFA software installato, vedere [Immagine di macchine Amazon personalizzate \(AMIs\) per AWS PCS](#).

Identifica le EFA istanze abilitate EC2

Per poter essere utilizzati EFA, tutti i tipi di istanze consentiti per un gruppo di AWS PCS calcolo devono supportare EFA e avere lo stesso numero di istanze vCPUs (e GPUs se appropriato). Per un elenco di istanze EFA abilitate, consulta [Elastic Fabric Adapter per HPC carichi di lavoro ML su Amazon EC2 nella Amazon Elastic Compute Cloud User Guide](#). Puoi anche usare il AWS CLI per visualizzare un elenco di tipi di istanze che supportano. EFA Replace (Sostituisci) *region-code* con il Regione AWS luogo in cui si utilizza AWS PCS, ad esempio `us-east-1`.

```
aws ec2 describe-instance-types \  
  --region region-code \  
  --output text
```

```
--filters Name=network-info.efa-supported,Values=true \  
--query "InstanceTypes[*].[InstanceType]" \  
--output text | sort
```

Determina quante interfacce di rete sono disponibili

Alcune EC2 istanze dispongono di più schede di rete. Ciò consente loro di averne più EFAs di una. Per ulteriori informazioni, consulta [Interfacce di rete multiple in AWS PCS](#).

Crea un gruppo di sicurezza per supportare EFA le comunicazioni

AWS CLI

È possibile utilizzare il seguente AWS CLI comando per creare un gruppo di sicurezza che supporti EFA. Il comando restituisce un ID del gruppo di sicurezza. Effettua le seguenti sostituzioni:

- *region-code*— Specificare il Regione AWS luogo in cui si utilizza AWS PCS, ad esempio `east-1`.
- *vpc-id*— Specificate l'ID per VPC cui utilizzate AWS PCS.
- *efa-group-name*— Fornisci il nome scelto per il gruppo di sicurezza.

```
aws ec2 create-security-group \  
  --group-name efa-group-name \  
  --description "Security group to enable EFA traffic" \  
  --vpc-id vpc-id \  
  --region region-code
```

Utilizzate i seguenti comandi per allegare le regole del gruppo di sicurezza in entrata e in uscita. Effettua la seguente sostituzione:

- *efa-secgroup-id*— Fornisci l'ID del gruppo di EFA sicurezza che hai appena creato.

```
aws ec2 authorize-security-group-ingress \  
  --group-id efa-secgroup-id \  
  --protocol -1 \  
  --source-group efa-secgroup-id
```

```
aws ec2 authorize-security-group-egress \  
  --group-id efa-secgroup-id \  
  --protocol -1 \  
  --source-group efa-secgroup-id
```

CloudFormation template

È possibile utilizzare un CloudFormation modello per creare un gruppo di sicurezza che supporti EFA. Scarica il modello dal seguente link URL, quindi caricalo nella [AWS CloudFormation console](#).

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/enable_efa/assets/efa-sg.yaml
```

Con il modello aperto nella AWS CloudFormation console, inserisci le seguenti opzioni.

- In Fornisci un nome per lo stack
 - In Nome dello stack, inserisci un nome come. `efa-sg-stack`
- In Parametri
 - In SecurityGroupName, inserisci un nome come `efa-sg`.
 - Sotto VPC, seleziona il VPC luogo in cui utilizzerai AWS PCS.

Completa la creazione dello CloudFormation stack e monitorane lo stato. Quando raggiunge CREATE_COMPLETE il gruppo EFA di sicurezza è pronto per l'uso.

(Facoltativo) Create un gruppo di collocamento

Si consiglia di avviare tutte le istanze da utilizzare EFA in un gruppo di posizionamento del cluster per ridurre al minimo la distanza fisica tra di esse. Ti consigliamo di creare un gruppo di posizionamento per ogni gruppo di nodi di calcolo che utilizzerai. EFA Vedi [Gruppi di collocamento per EC2 istanze in AWS PCS](#) per creare un gruppo di posizionamento per il tuo gruppo di nodi di calcolo.

Crea o aggiorna un modello di EC2 lancio

EFA le interfacce di rete sono configurate nel modello di EC2 avvio per un gruppo di nodi di AWS PCS calcolo. Se sono presenti più schede di rete, è EFA possibile configurarne più di una. Il gruppo EFA di sicurezza e il gruppo di posizionamento opzionale sono inclusi anche nel modello di lancio.

Ecco un esempio di modello di avvio per istanze con due schede di rete, come hpc7a.96xlarge. Le istanze verranno avviate nel gruppo di collocamento del cluster. subnet - *SubnetID1* pg - *PlacementGroupId1*

I gruppi di sicurezza devono essere aggiunti in modo specifico a ciascuna EFA interfaccia. Tutti EFA hanno bisogno del gruppo di sicurezza che abilita EFA il traffico (sg - *EfaSecGroupId*). Gli altri gruppi di sicurezza, in particolare quelli che gestiscono traffico regolare come SSH oHTTPS, devono essere collegati solo all'interfaccia di rete principale (indicata da un DeviceIndex of0). I modelli di avvio in cui sono definite le interfacce di rete non supportano l'impostazione dei gruppi di sicurezza utilizzando il SecurityGroupIds parametro: è necessario impostare un valore per Groups ogni interfaccia di rete configurata.

```
{
  "Placement": {
    "GroupId": "pg-PlacementGroupId1"
  },
  "NetworkInterfaces": [
    {
      "DeviceIndex": 0,
      "InterfaceType": "efa",
      "NetworkCardIndex": 0,
      "SubnetId": "subnet-SubnetId1",
      "Groups": [
        "sg-SecurityGroupId1",
        "sg-EfaSecGroupId"
      ]
    },
    {
      "DeviceIndex": 1,
      "InterfaceType": "efa",
      "NetworkCardIndex": 1,
      "SubnetId": "subnet-SubnetId1"
      "Groups": ["sg-EfaSecGroupId"]
    }
  ]
}
```

Crea o aggiorna un gruppo di nodi di calcolo

Crea o aggiorna un gruppo di nodi di AWS PCS calcolo con istanze con lo stesso numero di istanzevCPUs, la stessa architettura di processori e tutte supportate. EFA Configura il gruppo di nodi

di calcolo per utilizzarlo AMI con il EFA software installato su di esso e per utilizzare il modello di avvio che configura le interfacce di rete EFA abilitate.

(Facoltativo) Test EFA

È possibile dimostrare la comunicazione EFA abilitata tra due nodi in un gruppo di nodi di calcolo eseguendo il `fi_pingpong` programma, incluso nell'installazione del EFA software. Se questo test ha esito positivo, è probabile che EFA sia configurato correttamente.

Per iniziare, sono necessarie due istanze in esecuzione nel gruppo di nodi di calcolo. Se il gruppo di nodi di calcolo utilizza una capacità statica, dovrebbero esserci già delle istanze disponibili. Per un gruppo di nodi di calcolo che utilizza capacità dinamica, puoi avviare due nodi utilizzando il comando `salloc`. Ecco un esempio tratto da un cluster con un gruppo di nodi dinamico denominato `hpc7g` associato a una coda denominata `all`

```
% salloc --nodes 2 -p all
salloc: Granted job allocation 6
salloc: Waiting for resource configuration
... a few minutes pass ...
salloc: Nodes hpc7g-[1-2] are ready for job
```

Scopri l'indirizzo IP per i due nodi allocati utilizzando `scontrol`. Nell'esempio che segue, gli indirizzi sono `10.3.140.69` for `hpc7g-1` e `10.3.132.211` for `hpc7g-2`.

```
% scontrol show nodes hpc7g-[1-2]
NodeName=hpc7g-1 Arch=aarch64 CoresPerSocket=1
  CPUAlloc=0 CPUEfctv=64 CPUTot=64 CPUload=0.00
  AvailableFeatures=hpc7g
  ActiveFeatures=hpc7g
  Gres=(null)
  NodeAddr=10.3.140.69 NodeHostName=ip-10-3-140-69 Version=23.11.8
  OS=Linux 5.10.218-208.862.amzn2.aarch64 #1 SMP Tue Jun 4 16:52:10 UTC 2024
  RealMemory=124518 AllocMem=0 FreeMem=110763 Sockets=64 Boards=1
  State=IDLE+CLOUD ThreadsPerCore=1 TmpDisk=0 Weight=1 Owner=N/A MCS_label=N/A
  Partitions=efa
  BootTime=2024-07-02T19:00:09 SlurmdStartTime=2024-07-08T19:33:25
  LastBusyTime=2024-07-08T19:33:25 ResumeAfterTime=None
  CfgTRES=cpu=64,mem=124518M,billing=64
  AllocTRES=
  CapWatts=n/a
  CurrentWatts=0 AveWatts=0
```

```

ExtSensorsJoules=n/a ExtSensorsWatts=0 ExtSensorsTemp=n/a
Reason=Maintain Minimum Number Of Instances [root@2024-07-02T18:59:00]
InstanceId=i-04927897a9ce3c143 InstanceType=hpc7g.16xlarge

NodeName=hpc7g-2 Arch=aarch64 CoresPerSocket=1
CPUAlloc=0 CPUEfctv=64 CPUTot=64 CPULoad=0.00
AvailableFeatures=hpc7g
ActiveFeatures=hpc7g
Gres=(null)
NodeAddr=10.3.132.211 NodeHostName=ip-10-3-132-211 Version=23.11.8
OS=Linux 5.10.218-208.862.amzn2.aarch64 #1 SMP Tue Jun 4 16:52:10 UTC 2024
RealMemory=124518 AllocMem=0 FreeMem=110759 Sockets=64 Boards=1
State=IDLE+CLOUD ThreadsPerCore=1 TmpDisk=0 Weight=1 Owner=N/A MCS_label=N/A
Partitions=efa
BootTime=2024-07-02T19:00:09 SlurmdStartTime=2024-07-08T19:33:25
LastBusyTime=2024-07-08T19:33:25 ResumeAfterTime=None
CfgTRES=cpu=64,mem=124518M,billing=64
AllocTRES=
CapWatts=n/a
CurrentWatts=0 AveWatts=0
ExtSensorsJoules=n/a ExtSensorsWatts=0 ExtSensorsTemp=n/a
Reason=Maintain Minimum Number Of Instances [root@2024-07-02T18:59:00]
InstanceId=i-0a2c82623cb1393a7 InstanceType=hpc7g.16xlarge

```

Connect a uno dei nodi (in questo caso, hpc7g-1) utilizzando SSH (oSSM). Tieni presente che si tratta di un indirizzo IP interno, quindi potresti dover connetterti da uno dei tuoi nodi di accesso se lo utilizzi SSH. Tieni inoltre presente che l'istanza deve essere configurata con una SSH chiave tramite il modello di avvio del gruppo di nodi di calcolo.

```
% ssh ec2-user@10.3.140.69
```

Ora, avvia `fi_pingpong` in modalità server.

```
/opt/amazon/efa/bin/fi_pingpong -p efa
```

Connect alla seconda istanza (hpc7g-2).

```
% ssh ec2-user@10.3.132.211
```

Esegui `fi_pingpong` in modalità client, con connessione al server attiva hpc7g-1. L'output dovrebbe essere simile a quello dell'esempio seguente.

```
% /opt/amazon/efa/bin/fi_pingpong -p efa 10.3.140.69

bytes  #sent  #ack  total  time  MB/sec  usec/xfer  Mxfers/sec
64     10    =10   1.2k   0.00s  3.08    20.75     0.05
256    10    =10   5k     0.00s  21.24   12.05     0.08
1k     10    =10   20k    0.00s  82.91   12.35     0.08
4k     10    =10   80k    0.00s  311.48  13.15     0.08
[error] util/pingpong.c:1876: fi_close (-22) fid 0
```

(Facoltativo) Utilizzate un CloudFormation modello per creare un modello di EFA avvio abilitato

Poiché la configurazione dipende da diverse dipendenze EFA, è stato fornito un CloudFormation modello che è possibile utilizzare per configurare un gruppo di nodi di calcolo. Supporta istanze con un massimo di quattro schede di rete. Per ulteriori informazioni sulle istanze con più schede di rete, consulta le [interfacce di rete elastiche](#) nella Amazon Elastic Compute Cloud User Guide.

Scarica il CloudFormation modello da quanto segue URL, quindi caricalo sulla CloudFormation console Regione AWS dove lo usi. AWS PCS

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/enable_efa/assets/pcs-1t-efa.yaml
```

Con il modello aperto nella AWS CloudFormation console, inserisci i seguenti valori. Tieni presente che il modello fornirà alcuni valori di parametro predefiniti: puoi lasciarli come valori predefiniti.

- In Fornisci un nome per lo stack
 - In Nome dello stack, inserisci un nome descrittivo. Ti consigliamo di incorporare il nome che sceglierai per il tuo gruppo di nodi di AWS PCS calcolo, ad esempio. ***NODEGROUPNAME***-efa-1t
- In Parametri
 - In NumberOfNetworkCards, scegli il numero di schede di rete nelle istanze che faranno parte del tuo gruppo di nodi.
 - In VpcId, scegli VPC dove viene distribuito il tuo AWS PCS cluster.
 - In NodeGroupSubnetId, scegli la sottorete del cluster in VPC cui verranno lanciate le istanze EFA abilitate.
 - Sotto PlacementGroupName, lascia il campo vuoto per creare un nuovo gruppo di posizionamento del cluster per il gruppo di nodi. Se disponi di un gruppo di collocamento esistente che desideri utilizzare, inseriscine il nome qui.

- In `ClusterSecurityGroupId`, scegli il gruppo di sicurezza che stai utilizzando per consentire l'accesso ad altre istanze del cluster e al AWS PCS API. Molti clienti scelgono il gruppo di sicurezza predefinito dal proprio cluster VPC.
- In `SshSecurityGroupId`, fornisci l'ID di un gruppo di sicurezza che stai utilizzando per consentire SSH l'accesso in entrata ai nodi del cluster.
- Per `SshKeyName`, seleziona la SSH coppia di chiavi per l'accesso ai nodi del cluster.
- Per `LaunchTemplateName`, inserisci un nome descrittivo per il modello di lancio, ad esempio. `NODEGROUPNAME-efa-1t` Il nome deve essere univoco per il luogo Account AWS in Regione AWS cui lo AWS PCS utilizzerai.
- In Capacità
 - Seleziona la casella Riconosco che AWS CloudFormation potrebbe creare IAM risorse.

Monitora lo stato dello CloudFormation stack. Quando raggiunge `CREATE_COMPLETE` il modello di lancio è pronto per essere utilizzato. Usalo con un gruppo di nodi di AWS PCS calcolo, come descritto sopra in [Crea o aggiorna un gruppo di nodi di calcolo](#).

Utilizzo di file system di rete con AWS PCS

È possibile collegare volumi di storage di rete ai nodi avviati in un gruppo di nodi di calcolo AWS Parallel Computing Service (AWS PCS) per fornire una posizione persistente in cui è possibile scrivere e accedere a dati e file. È possibile utilizzare i volumi forniti dai AWS servizi. I volumi includono [Amazon Elastic File System](#) (AmazonEFS), [Amazon FSx for NetApp ONTAP](#), [Amazon FSx for Open ZFS](#), [Amazon FSx for Lustre](#) e [Amazon File Cache](#). Puoi anche utilizzare volumi autogestiti, come NFS i server.

In questo argomento vengono fornite considerazioni ed esempi sull'utilizzo dei file system in rete con AWS PCS

Considerazioni sull'utilizzo dei file system di rete

I dettagli di implementazione per i vari file system sono diversi, ma ci sono alcune considerazioni comuni.

- Il software del file system pertinente deve essere installato sull'istanza. Ad esempio, per utilizzare Amazon FSx for Lustre, deve essere presente il Lustre pacchetto appropriato. Ciò può essere

ottenuto includendolo nel gruppo di nodi di calcolo AMI o utilizzando uno script che viene eseguito all'avvio dell'istanza.

- Deve esserci un percorso di rete tra il volume di storage condiviso e le istanze del gruppo di nodi di calcolo.
- Le regole del gruppo di sicurezza relative sia al volume di archiviazione condiviso che alle istanze del gruppo di nodi di calcolo devono consentire le connessioni alle porte pertinenti.
- È necessario mantenere uno spazio dei nomi POSIX utente e di gruppo coerente tra le risorse che accedono ai file system. In caso contrario, i job e i processi interattivi eseguiti sul PCS cluster potrebbero riscontrare errori di autorizzazione.
- I montaggi del file system vengono eseguiti utilizzando modelli di EC2 avvio. Errori o timeout nel montaggio di un file system di rete possono impedire la disponibilità di istanze per l'esecuzione dei job. Ciò, a sua volta, può comportare costi imprevisti. Per ulteriori informazioni sul debug dei modelli di avvio, consulta [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#)

Esempi di supporti di rete

Puoi creare file system utilizzando AmazonEFS, Amazon FSx for Lustre, Amazon FSx for Open ZFS e Amazon File Cache. Espandi la sezione pertinente di seguito per vedere un esempio di ogni montaggio di rete.

Amazon EFS

Configurazione del file system

Crea un EFS file system Amazon. Assicurati che abbia un target di montaggio in ogni zona di disponibilità in cui avviare le istanze del gruppo di nodi di PCS calcolo. Assicurati inoltre che ogni target di montaggio sia associato a un gruppo di sicurezza che consenta l'accesso in entrata e in uscita dalle istanze del gruppo di nodi di PCS calcolo. Per ulteriori informazioni, consulta [Mount targets and security group](#) nella Amazon Elastic File System User Guide.

Modello di lancio

Aggiungi i gruppi di sicurezza dalla configurazione del file system al modello di lancio che utilizzerai per il gruppo di nodi di calcolo.

Includi i dati utente che utilizzano un `cloud-config` meccanismo per montare il EFS file system Amazon. Sostituisci i seguenti valori in questo script con i tuoi dati:

- *mount-point-directory*— Il percorso su ogni istanza in cui monterai Amazon EFS

- *filesystem-id*— L'ID del file system per il EFS file system

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

packages:
  - amazon-efs-utils

runcmd:
  - mkdir -p /mount-point-directory
  - echo "filesystem-id:/ /mount-point-directory efs tls,_netdev" >> /etc/fstab
  - mount -a -t efs defaults

--==MYBOUNDARY==--
```

Amazon FSx per Lustre

Configurazione del file system

Crea un file system FSx for Lustre nel luogo in VPC cui lo AWS PCS utilizzerai. Per ridurre al minimo i trasferimenti tra zone, esegui la distribuzione in una sottorete nella stessa zona di disponibilità, dove lancerai la maggior parte delle istanze del gruppo di nodi di calcolo. PCS Assicurati che il file system sia associato a un gruppo di sicurezza che consenta l'accesso in entrata e in uscita dalle istanze del gruppo di nodi di calcolo. PCS Per ulteriori informazioni sui gruppi di sicurezza, consulta [il controllo degli accessi al file system con Amazon VPC](#) nella Guida per l'utente di Amazon FSx for Lustre.

Modello di lancio

Includi i dati utente utilizzati c`loud-config` per montare il file system FSx for Lustre. Sostituisci i seguenti valori in questo script con i tuoi dati:

- *mount-point-directory*— Il percorso su un'istanza in cui si desidera montare FSx Lustre
- *filesystem-id*— L'ID del file system per il file system FSx for Lustre
- *mount-name*— Il nome di montaggio per il file FSx system for Lustre
- *region-code*— Il Regione AWS luogo in cui è distribuito il file system FSx for Lustre (deve essere lo stesso del sistema in uso) AWS PCS
- (Facoltativo)*latest*: qualsiasi versione Lustre supportata da for Lustre FSx

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

runcmd:
- amazon-linux-extras install -y lustre=latest
- mkdir -p /mount-point-directory
- mount -t lustre filesystem-id.fsx.region-code.amazonaws.com@tcp:/mount-name /mount-point-directory

--==MYBOUNDARY==
```

Amazon FSx per Open ZFS

Configurazione del file system

Crea un ZFS file system FSx for Open nel VPC luogo in cui lo utilizzerai AWS PCS. Per ridurre al minimo i trasferimenti tra zone, esegui la distribuzione in una sottorete nella stessa zona di disponibilità, dove lancerai la maggior parte delle istanze del gruppo di nodi di AWS PCS calcolo. Assicurati che il file system sia associato a un gruppo di sicurezza che consenta l'accesso in entrata e in uscita dalle istanze del gruppo di nodi di calcolo. AWS PCS Per ulteriori informazioni sui gruppi di sicurezza, consulta [Managing file system access with Amazon VPC](#) nella FSxfor Open ZFS User Guide.

Modello di lancio

Includi i dati utente utilizzati `cloud-config` per montare il volume root di un ZFS file system FSx for Open. Sostituisci i seguenti valori in questo script con i tuoi dati:

- *mount-point-directory*— Il percorso su un'istanza in cui desideri montare il tuo FSx for Open ZFS share
- *filesystem-id*— L'ID del file system FSx per il ZFS file system for Open
- *region-code*— Il Regione AWS luogo in cui è distribuito ZFS il file system FSx for Open (deve essere lo stesso del AWS PCS sistema in uso)

```
MIME-Version: 1.0
```

```
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

runcmd:
- mkdir -p /mount-point-directory
- mount -t nfs -o noatime,nfsvers=4.2,sync,rsize=1048576,wsiz=1048576 filesystem-id.fsx.region-code.amazonaws.com:/fsx/ /mount-point-directory

--==MYBOUNDARY==
```

Cache di file Amazon

Configurazione del file system

Crea un [Amazon File Cache](#) nel VPC luogo in cui lo utilizzerai AWS PCS. Per ridurre al minimo i trasferimenti tra zone, scegli una sottorete nella stessa zona di disponibilità in cui lancerai la maggior parte delle istanze del gruppo di nodi di PCS calcolo. Assicurati che la File Cache sia associata a un gruppo di sicurezza che consenta il traffico in entrata e in uscita sulla porta 988 tra le tue istanze e la File Cache. PCS Per ulteriori informazioni sui gruppi di sicurezza, consulta [la sezione Controllo dell'accesso alla cache con Amazon VPC](#) nella Amazon File Cache User Guide.

Modello di lancio

Aggiungi i gruppi di sicurezza dalla configurazione del file system al modello di avvio che utilizzerai per il gruppo di nodi di calcolo.

Includi i dati utente utilizzati `cloud-config` per montare Amazon File Cache. Sostituisci i seguenti valori in questo script con i tuoi dati:

- *mount-point-directory*— Il percorso su un'istanza in cui si desidera montare FSx Lustre
- *cache-dns-name*— Il nome Domain Name System (DNS) per la File Cache
- *mount-name*— Il nome di montaggio per la File Cache

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary="==MYBOUNDARY=="

--==MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"
```

```
runcmd:  
- amazon-linux-extras install -y lustre=2.12  
- mkdir -p /mount-point-directory  
- mount -t lustre -o relatime,flock cache-dns-name@tcp:/mount-name /mount-point-  
directory  
  
--==MYBOUNDARY==
```

Amazon Machine Images (AMIs) per AWS PCS

AWS PCS funziona con AMIs ciò che offri, offrendo una grande flessibilità nel software e nella configurazione presenti sui nodi del cluster. Se stai provando AWS PCS, puoi usare un esempio AMI fornito e gestito da AWS. Se lo utilizzi AWS PCS in produzione, ti consigliamo di crearne uno tuo AMI. Questo argomento spiega come scoprire e utilizzare l'esempio AMI, nonché come crearne e utilizzarne uno personalizzato AMI.

Argomenti

- [Utilizzo di Amazon Machine Images \(AMIs\) di esempio con AWS PCS](#)
- [Immagini di macchine Amazon personalizzate \(AMIs\) per AWS PCS](#)
- [Programmi di installazione software per cui creare software personalizzati AMI AWS PCS](#)

Utilizzo di Amazon Machine Images (AMIs) di esempio con AWS PCS

AWS fornisce [esempi AMI](#) che è possibile utilizzare come punto di partenza per lavorare AWS PCS.

Important

AMIs Gli esempi sono a scopo dimostrativo e non sono consigliati per carichi di lavoro di produzione.

Trova il campione attuale AWS PCS AMI

AWS Management Console

AWSPCS AMI campioni hanno la seguente convenzione di denominazione:

```
aws-pcs-sample_ami-OS-architecture-schdeulder-scheduler-major-version
```

Valori accettati

- *OS* – amzn2
- *architecture* — oppure x86_64 arm64
- *scheduler* – slurm
- *scheduler-major-version* – 23.11

Per trovare AWS PCS un campione AMIs

1. Apri la [EC2console Amazon](#).
2. Vai a AMIs.
3. Scegliere Immagini pubbliche.
4. In Trova AMI per attributo o tag, cerca un nome AMI utilizzando il nome del modello.

Esempi

- Slurm 23.11 che supporta Graviton AMI

```
aws-pcs-sample_ami-amzn2-arm64-slurm-23.11
```

- Esempio per istanze x86 AMI

```
aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11
```

Note

Se ce ne sono più AMIs, usa il timestamp AMI con il timestamp più recente.

5. Usa l'AMIID quando crei o aggiorni un gruppo di nodi di calcolo.

AWS CLI

Puoi trovare l' AWS PCS esempio più recente AMI con i comandi seguenti. Replace (Sostituisci) *region-code* con il Regione AWS luogo in cui si utilizza AWS PCS, ad esempio `us-east-1`.

- x86_64

```
aws ec2 describe-images --region region-code --owners amazon 533267220047
654654292779 654654317195 975050324343 \
--filters 'Name=name,Values=aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11*' \
          'Name=state,Values=available' \
--query 'sort_by(Images, &CreationDate)[-1].[Name,ImageId]' --output text
```

- Arm 64

```
aws ec2 describe-images --region region-code --owners amazon 533267220047
654654292779 654654317195 975050324343 \
--filters 'Name=name,Values=aws-pcs-sample_ami-amzn2-arm64-slurm-23.11*' \
          'Name=state,Values=available' \
--query 'sort_by(Images, &CreationDate)[-1].[Name,ImageId]' --output text
```

Usa l'AMIID quando crei o aggiorni un gruppo di nodi di calcolo.

Scopri di più sul campione AWS PCS AMIs

Per visualizzare i contenuti e i dettagli di configurazione per le versioni correnti e precedenti dell'AWS PCS esempio AMIs, vedere [Note di rilascio per AWS PCS un esempio AMIs](#).

Creare un tuo AMIs compatibile con AWS PCS

Per imparare a crearne AMIs di tuoi compatibili AWS PCS, vedi [Immagini di macchine Amazon personalizzate \(AMIs\) per AWS PCS](#).

Immagini di macchine Amazon personalizzate (AMIs) per AWS PCS

AWS PCS è progettato per funzionare con Amazon Machine Images (AMI) che porti al servizio. Questi AMIs possono avere software e configurazioni arbitrari installati su di essi, purché abbiano l'AWS PCS agente e una versione compatibile di Slurm installati e configurati correttamente. È necessario utilizzare i programmi AWS di installazione forniti per installare il software sul dispositivo personalizzato. AWS PCS AMI Ti consigliamo di utilizzare i programmi AWS di installazione forniti per installare Slurm sul tuo dispositivo personalizzato, AMI ma puoi installare Slurm da solo se preferisci (scelta non consigliata).

Note

Se vuoi provare AWS PCS senza crearne uno personalizzato AMI, puoi usare un esempio fornito da. AMI AWS Per ulteriori informazioni, consulta [Utilizzo di Amazon Machine Images \(AMIs\) di esempio con AWS PCS](#).

Questo tutorial ti aiuta a crearne uno AMI che possa essere utilizzato con gruppi di nodi di PCS calcolo per alimentare i tuoi carichi di lavoro HPC e di AI/ML.

Argomenti

- [Passaggio 1: avvia un'istanza temporanea](#)
- [Fase 2 — Installare l' AWS PCS agente](#)
- [Passaggio 3: installa Slurm](#)
- [Fase 4 — \(Facoltativo\) Installare driver, librerie e software applicativi aggiuntivi](#)
- [Passaggio 5: crea una versione compatibile con AMI AWS PCS](#)
- [Passaggio 6: utilizza il custom AMI con un gruppo di nodi di AWS PCS calcolo](#)
- [Passaggio 7: terminare l'istanza temporanea](#)

Passaggio 1: avvia un'istanza temporanea

Avvia un'istanza temporanea che puoi utilizzare per installare e configurare il AWS PCS software e lo scheduler Slurm. Questa istanza viene utilizzata per creare un'AMI istanza compatibile con. AWS PCS

Per avviare un'istanza temporanea

1. Apri la [EC2 console Amazon](#).
2. Nel riquadro di navigazione, scegli Istanze, quindi scegli Avvia istanze per aprire la nuova procedura guidata di avvio dell'istanza.
3. (Facoltativo) Nella sezione Nome e tag, fornisci un nome per l'istanza, ad esempio. PCS-AMI-instance Il nome viene assegnato all'istanza come tag di risorsa (Name=PCS-AMI-instance).
4. Nella sezione Immagini dell'applicazione e del sistema operativo, selezionate uno AMI per uno dei [sistemi operativi supportati](#).

5. Nella sezione Instance type (Tipo di istanza), seleziona un [tipo di istanza supportato](#).
6. Nella sezione Key pair (Coppia di chiavi), seleziona la coppia di chiavi da utilizzare per l'istanza.
7. Nella sezione Impostazioni di rete:
 - Per Firewall (gruppi di sicurezza), scegli Seleziona gruppo di sicurezza esistente, quindi seleziona un gruppo di sicurezza che consenta SSH l'accesso in entrata all'istanza.
8. Nella sezione Storage (Archiviazione), configura i volumi secondo necessità. Assicurati di configurare uno spazio sufficiente per installare le tue applicazioni e librerie.
9. Nel pannello Summary (Riepilogo), scegliere Launch instance (Avvia istanza).

Fase 2 — Installare l' AWS PCSagente

Installa l'agente che configura le istanze lanciate da AWS PCS per l'uso con Slurm.

Per installare l'agente di AWS PCS

1. Connettersi all'istanza avviata. Per ulteriori informazioni, consulta [Connect to your Linux instance](#).
2. (Facoltativo) Per assicurarti che tutti i pacchetti software siano aggiornati, esegui un rapido aggiornamento del software sull'istanza. Questo processo può richiedere alcuni minuti.
 - Amazon Linux 2, RHEL 9, Rocky Linux 9

```
sudo yum update -y
```


- Ubuntu 22.04

```
sudo apt-get update && sudo apt-get upgrade -y
```

3. Riavviare l'istanza e riconnettersi a essa.
4. Scarica i file di installazione dell' AWS PCSagente. I file di installazione sono impacchettati in un file tarball () `.tar.gz` compresso. Per scaricare l'ultima versione stabile, utilizzare il comando seguente. Sostituisci *region* con il Regione AWS punto in cui hai lanciato l'istanza temporanea, ad esempio. `us-east-1`

```
curl https://aws-pcs-repo-region.s3.amazonaws.com/aws-pcs-agent/aws-pcs-agent-v1.0.0-1.tar.gz -o aws-pcs-agent-v1.0.0-1.tar.gz
```

È inoltre possibile ottenere la versione più recente sostituendo il numero di versione con quello indicato `latest` nel comando precedente (ad esempio: `aws-pcs-agent-v1-latest.tar.gz`).

 Note

Questo potrebbe cambiare nelle future versioni del software dell' AWS PCSagente.

5. (Facoltativo) Verificate l'autenticità e l'integrità del tarball del AWS PCS software. È consigliabile eseguire questa operazione per verificare l'identità dell'autore del software e che il file non sia stato alterato o danneggiato dopo la pubblicazione.
 - a. Scarica la GPG chiave pubblica AWS PCS e importala nel tuo portachiavi. Sostituisci *region* con il Regione AWS punto in cui hai lanciato l'istanza temporanea. Il comando dovrebbe restituire un valore di chiave. Registra il valore della chiave; lo utilizzerai nel passaggio successivo.


```
wget https://aws-pcs-repo-public-keys-region.s3.amazonaws.com/aws-pcs-public-key.pub && \  
  gpg --import aws-pcs-public-key.pub
```

- b. Eseguite il comando seguente per verificare l'impronta digitale della GPG chiave.

```
gpg --fingerprint 7EEF030EDDF5C21C
```

Il comando dovrebbe restituire un'impronta digitale identica alla seguente:

```
1C24 32C1 862F 64D1 F90A 239A 7EEF 030E DDF5 C21C
```

 Important

Non eseguire lo script di installazione dell' AWS PCSagente se l'impronta digitale non corrisponde. Contatta il [supporto AWS](#).


- c. Scaricate il file della firma e verificate la firma del file tarball del AWS PCS software. Replace (Sostituisci) *region* con il Regione AWS punto in cui hai lanciato l'istanza temporanea, ad esempio. `us-east-1`

```
wget https://aws-pcs-repo-region.s3.amazonaws.com/aws-pcs-agent/aws-pcs-agent-  
v1.0.0-1.tar.gz.sig && \  
gpg --verify ./aws-pcs-agent-v1.0.0-1.tar.gz.sig
```

L'output visualizzato dovrebbe essere simile al seguente:

```
gpg: assuming signed data in './aws-pcs-agent-v1.0.0-1.tar.gz'  
gpg: Signature made Thu Aug 8 18:50:19 2024 CEST  
gpg: using RSA key 4BAA531875430EB0739E6D961BA7F0AF6E34C496  
gpg: Good signature from "AWS PCS Packages (AWS PCS Packages)" [unknown]  
gpg: WARNING: This key is not certified with a trusted signature!  
gpg: There is no indication that the signature belongs to the owner.  
Primary key fingerprint: 1C24 32C1 862F 64D1 F90A 239A 7EEF 030E DDF5 C21C  
Subkey fingerprint: 4BAA 5318 7543 0EB0 739E 6D96 1BA7 F0AF 6E34 C496
```

Se il risultato include `Good signature` e l'impronta digitale corrisponde all'impronta digitale restituita nel passaggio precedente, procedi al passaggio successivo.

 **Important**

Non eseguite lo script di installazione del AWS PCS software se l'impronta digitale non corrisponde. Contatta il [supporto AWS](#).

6. Estrai i file dal file compresso `.tar.gz` e vai alla directory estratta.

```
tar -xf aws-pcs-agent-v1.0.0-1.tar.gz && \  
cd aws-pcs-agent
```

7. Installare il software AWS PCS.

```
sudo ./installer.sh
```

8. Controlla il file della versione del AWS PCS software per confermare l'avvenuta installazione.

```
cat /opt/aws/pcs/version
```

L'output visualizzato dovrebbe essere simile al seguente:

```
AGENT_INSTALL_DATE='Mon Aug 12 12:28:43 UTC 2024'
```

```
AGENT_VERSION='1.0.0'  
AGENT_RELEASE='1'
```

Passaggio 3: installa Slurm

Installa una versione di Slurm compatibile con. AWS PCS

Per installare Slurm

1. Connect alla stessa istanza temporanea in cui è stato installato il AWS PCS software.
2. Scarica il software di installazione Slurm. Il programma di installazione di Slurm è impacchettato in un file tarball () compresso. `.tar.gz` Per scaricare l'ultima versione stabile, utilizzare il comando seguente. Sostituisci *region* con la Regione AWS tua istanza temporanea, ad esempio. `us-east-1`

```
curl https://aws-pcs-repo-region.s3.amazonaws.com/aws-pcs-slurm/aws-pcs-  
slurm-23.11-installer-23.11.9-1.tar.gz \  
-o aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz
```

È inoltre possibile ottenere la versione più recente sostituendo il numero di versione con `latest` il comando precedente (ad esempio: `aws-pcs-slurm-23.11-installer-latest.tar.gz`).

Note

Questo potrebbe cambiare nelle future versioni del software di installazione Slurm.

3. (Facoltativo) Verifica l'autenticità e l'integrità del tarball del programma di installazione di Slurm. È consigliabile eseguire questa operazione per verificare l'identità dell'autore del software e che il file non sia stato alterato o danneggiato dopo la pubblicazione.
 - a. Scarica la GPG chiave pubblica AWS PCS e importala nel tuo portachiavi. Sostituisci *region* con il Regione AWS punto in cui hai lanciato l'istanza temporanea. Il comando dovrebbe restituire un valore di chiave. Registra il valore della chiave; lo utilizzerai nel passaggio successivo.

```
wget https://aws-pcs-repo-public-keys-region.s3.amazonaws.com/aws-pcs-public-  
key.pub && \  
gpg --import aws-pcs-public-key.pub
```

- b. Eseguite il comando seguente per verificare l'impronta digitale della GPG chiave.

```
gpg --fingerprint 7EEF030EDDF5C21C
```

Il comando dovrebbe restituire un'impronta digitale identica alla seguente:

```
1C24 32C1 862F 64D1 F90A 239A 7EEF 030E DDF5 C21C
```

⚠ Important

Non eseguire lo script di installazione di Slurm se l'impronta digitale non corrisponde. Contatta il [supporto AWS](#).

- c. Scarica il file della firma e verifica la firma del file tarball del programma di installazione di Slurm. Replace (Sostituisci) *region* con il Regione AWS punto in cui hai lanciato l'istanza temporanea, ad esempio. us-east-1

```
wget https://aws-pcs-repo-region.s3.amazonaws.com/aws-pcs-slurm/aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz.sig && \  
gpg --verify ./aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz.sig
```

L'output visualizzato dovrebbe essere simile al seguente:

```
gpg: assuming signed data in './aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz'  
gpg: Signature made Thu Aug 8 14:23:38 2024 CEST  
gpg: using RSA key 4BAA531875430EB0739E6D961BA7F0AF6E34C496  
gpg: Good signature from "AWS PCS Packages (AWS PCS Packages)" [unknown]  
gpg: WARNING: This key is not certified with a trusted signature!  
gpg: There is no indication that the signature belongs to the owner.  
Primary key fingerprint: 1C24 32C1 862F 64D1 F90A 239A 7EEF 030E DDF5 C21C  
Subkey fingerprint: 4BAA 5318 7543 0EB0 739E 6D96 1BA7 F0AF 6E34 C496
```

Se il risultato include `Good signature` e l'impronta digitale corrisponde all'impronta digitale restituita nel passaggio precedente, procedi al passaggio successivo.

⚠ Important

Non eseguire lo script di installazione di Slurm se l'impronta digitale non corrisponde. Contatta il [supporto AWS](#).

4. Estrarre i file dal file `.tar.gz` compresso e andare alla directory estratta.

```
tar -xf aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz && \  
cd aws-pcs-slurm-23.11-installer
```

5. Installa Slurm. Il programma di installazione scarica, compila e installa Slurm e le sue dipendenze. L'operazione richiede alcuni minuti, a seconda delle specifiche dell'istanza temporanea selezionata.

```
sudo ./installer.sh -y
```

6. Controlla il file della versione dello scheduler per confermare l'installazione.

```
cat /opt/aws/pcs/scheduler/slurm-23.11/version
```

L'output visualizzato dovrebbe essere simile al seguente:

```
SLURM_INSTALL_DATE='Mon Aug 12 12:38:56 UTC 2024'  
SLURM_VERSION='23.11.9'  
PCS_SLURM_RELEASE='1'
```

Fase 4 — (Facoltativo) Installare driver, librerie e software applicativi aggiuntivi

Installa driver, librerie e software applicativi aggiuntivi sull'istanza temporanea. Le procedure di installazione variano a seconda delle applicazioni e delle librerie specifiche. Se non ne hai mai creato uno personalizzato AMI per AWS PCS prima, ti consigliamo di crearne e testarne uno AMI con solo il AWS PCS software e Slurm installati, quindi di aggiungere in modo incrementale il tuo software e le tue configurazioni una volta confermato il successo iniziale.

Esempi

- Software Elastic Fabric Adapter (). EFA Per ulteriori informazioni, consulta Guida [introduttiva a EFA e MPI per i HPC carichi di lavoro su Amazon EC2 nella Amazon](#) Elastic Compute Cloud User Guide.
- Client Amazon Elastic File System (AmazonEFS). Per ulteriori informazioni, consulta [Installazione manuale del EFS client Amazon](#) nella Amazon Elastic File System User Guide.
- Client Lustre, per utilizzare Amazon FSx for Lustre e Amazon File Cache. Per ulteriori informazioni, consulta [Installazione del client Lustre](#) nella guida FSxper l'utente di for Lustre.
- CloudWatch Agente Amazon, per utilizzare CloudWatch Logs and Metrics. Per ulteriori informazioni, consulta [Installa l' CloudWatch agente](#) nella Amazon CloudWatch User Guide.
- AWS Neuron, per usare i tipi di istanza trn* e inf*. [Per ulteriori informazioni, consultate la documentazione di Neuron.AWS](#)
- NVIDIA Driver e CUDADCGM, per utilizzare i tipi di istanza p* o g*.

Passaggio 5: crea una versione compatibile con AMI AWS PCS

Dopo aver installato i componenti software richiesti, ne crei uno AMI che puoi riutilizzare per avviare istanze in gruppi di nodi di AWS PCS calcolo.

Per creare un file AMI dalla tua istanza temporanea

1. Apri la [EC2console Amazon](#).
2. Nel pannello di navigazione, seleziona Instances (Istanze).
3. Seleziona l'istanza temporanea che hai creato. Scegli Azioni, Immagine, Crea immagine.
4. Per Create image (Crea immagine), effettua le seguenti operazioni:
 - a. Per Nome immagine, inserisci un nome descrittivo per. AMI
 - b. (Facoltativo) Per Descrizione dell'immagine, immettere una breve descrizione dello scopo di. AMI
 - c. Scegliere Create Image (Crea immagine).
5. Nel riquadro di navigazione, scegli AMIs.
6. Individua AMI quello che hai creato nell'elenco. Attendi che il suo stato cambi da In sospeso a Disponibile, quindi usalo con un gruppo di nodi di AWS PCS calcolo.

Passaggio 6: utilizza il custom AMI con un gruppo di nodi di AWS PCS calcolo

Puoi usare il tuo personalizzato AMI con un gruppo di nodi di AWS PCS calcolo nuovo o esistente.

New compute node group

Per utilizzare il personalizzato AMI

1. Apri la [AWS PCSconsole](#).
2. Nel pannello di navigazione scegliere Clusters (Cluster).
3. Scegli il cluster in cui utilizzerai quello personalizzatoAMI, quindi seleziona Gruppi di nodi di calcolo.
4. Crea un nuovo gruppo di nodi di calcolo. Per ulteriori informazioni, consulta [Creazione di un gruppo di nodi di calcolo in AWS PCS](#). In AMIID, cerca il nome o l'ID del personalizzato AMI che desideri utilizzare. Completa la configurazione del gruppo di nodi di calcolo, quindi scegli Crea gruppo di nodi di calcolo.
5. (Facoltativo) Conferma che AMI supporta l'avvio dell'istanza. Avvia un'istanza nel gruppo di nodi di calcolo. Puoi farlo configurando il gruppo di nodi di calcolo in modo che abbia una singola istanza statica oppure puoi inviare un lavoro a una coda che utilizza il gruppo di nodi di calcolo.
 - a. Controlla la EC2 console Amazon finché un'istanza non appare etichettata con il nuovo ID del gruppo di nodi di calcolo. Per ulteriori informazioni su questo argomento, consulta.. [Ricerca di istanze di gruppi di nodi di calcolo in AWS PCS](#)
 - b. Quando vedi un'istanza avviarsi e completare il processo di bootstrap, verifica che stia utilizzando il valore previstoAMI. Per fare ciò, seleziona l'istanza, quindi controlla l'AMIID in Dettagli. Dovrebbe corrispondere a AMI quello configurato nelle impostazioni del gruppo di nodi di calcolo.
 - c. (Facoltativo) Aggiorna la configurazione di ridimensionamento dei gruppi di nodi di calcolo ai tuoi valori preferiti.

Existing compute node group

Per utilizzare la versione personalizzata AMI

1. Apri la [AWS PCSconsole](#).
2. Nel pannello di navigazione scegliere Clusters (Cluster).

3. Scegli il cluster in cui utilizzerai quello personalizzatoAMI, quindi seleziona Gruppi di nodi di calcolo.
4. Seleziona il gruppo di nodi che desideri configurare e scegli Modifica. In AMIID, cerca il nome o l'ID del personalizzato che AMI desideri utilizzare. Completa la configurazione del gruppo di nodi di calcolo, quindi scegli Aggiorna. Le nuove istanze lanciate nel gruppo di nodi di calcolo utilizzeranno l'ID aggiornato. AMI Le istanze esistenti continueranno a utilizzare le vecchie AMI fino AWS PCS a quando non verranno sostituite. Per ulteriori informazioni, consulta [Aggiornamento di un gruppo di nodi di AWS PCS calcolo](#).
5. (Facoltativo) Conferma che AMI supporta l'avvio dell'istanza. Avvia un'istanza nel gruppo di nodi di calcolo. Puoi farlo configurando il gruppo di nodi di calcolo in modo che abbia una singola istanza statica oppure puoi inviare un lavoro a una coda che utilizza il gruppo di nodi di calcolo.
 - a. Controlla la EC2 console Amazon finché un'istanza non appare etichettata con il nuovo ID del gruppo di nodi di calcolo. Per ulteriori informazioni su questo argomento, consulta.. [Ricerca di istanze di gruppi di nodi di calcolo in AWS PCS](#)
 - b. Quando vedi un'istanza avviarsi e completare il processo di bootstrap, verifica che stia utilizzando il valore previstoAMI. Per fare ciò, seleziona l'istanza, quindi controlla l'AMIID in Dettagli. Dovrebbe corrispondere a AMI quello configurato nelle impostazioni del gruppo di nodi di calcolo.
 - c. (Facoltativo) Aggiorna la configurazione di ridimensionamento dei gruppi di nodi di calcolo ai tuoi valori preferiti.

Passaggio 7: terminare l'istanza temporanea

Dopo aver confermato che l'istanza AMI funziona come previsto AWS PCS, puoi chiudere l'istanza temporanea per evitare di incorrere in addebiti.

Per terminare l'istanza temporanea

1. Apri la [EC2console Amazon](#).
2. Nel pannello di navigazione, seleziona Instances (Istanze).
3. Seleziona l'istanza temporanea che hai creato e scegli Azioni, Stato dell'istanza, Termina istanza.
4. Quando ti viene richiesto di confermare, scegli Termina.

Programmi di installazione software per cui creare software personalizzati AMIs AWS PCS

AWS fornisce un file scaricabile che consente di installare il AWS PCS software su un'istanza. AWS fornisce inoltre software in grado di scaricare, compilare e installare le versioni pertinenti di Slurm e delle sue dipendenze. È possibile utilizzare queste istruzioni per crearne di personalizzate AMIs da utilizzare con AWS PCS oppure è possibile utilizzare metodi personalizzati.

Indice

- [AWS PCS programma di installazione del software](#)
- [Programma di installazione Slurm](#)
- [Sistemi operativi supportati](#)
- [Tipi di istanze supportati](#)
- [Versioni Slurm supportate](#)
- [Verifica gli installatori utilizzando un checksum](#)

AWS PCS programma di installazione del software

Il programma di installazione del AWS PCS software configura un'istanza su cui lavorare AWS PCS durante il processo di avvio dell'istanza. È necessario utilizzare i programmi AWS di installazione forniti per installare il software sul AWS PCS dispositivo personalizzato. AMI

Programma di installazione Slurm

Il programma di installazione di Slurm scarica, compila e installa le versioni pertinenti di Slurm e delle sue dipendenze. Puoi usare il programma di installazione di Slurm per creare file personalizzati per. AMIs AWS PCS È inoltre possibile utilizzare i propri meccanismi se sono coerenti con la configurazione software fornita dal programma di installazione di Slurm.

Il software AWS fornito installa quanto segue:

- [Slurm alla versione principale e di manutenzione richiesta \(attualmente versione 23.11.8\) - Licenza 2 GPL](#)
 - Slurm è costruito con `set to --sysconfdir /etc/slurm`
 - Slurm è costruito con l'opzione `e --enable-pam --without-munge`

- Slurm è costruito con l'opzione `--sharedstatedir=/run/slurm/`
- Slurm è costruito con e supporta PMIX JWT
- Slurm è installato su `/opt/aws/pcs/schedulers/slurm-23.11`
- [Open PMIX \(versione 4.2.6\) — Licenza](#)
 - Open PMIX è installato come sottodirectory di `/opt/aws/pcs/scheduler/`
- [libjwt \(versione 1.15.3\) — Licenza -2.0 MPL](#)
 - libjwt è installato come sottodirectory di `/opt/aws/pcs/scheduler/`

Il software AWS fornito modifica la configurazione del sistema come segue:

- Il `systemd` file Slurm creato dalla build viene copiato con il nome del file. `/etc/systemd/system/slurmd-23.11.service`
- Se non esistono, vengono creati un utente e un gruppo Slurm (`slurm:slurm`) con/of. UID/GID 401
- Su Amazon Linux 2 e Rocky Linux 9 l'installazione aggiunge il EPEL repository per installare il software richiesto per creare Slurm o le sue dipendenze.
- Durante RHEL9 l'installazione abiliterà `codeready-builder-for-rhel-9-rhui-rpms` e `epel-release-latest-9` installerà il software richiesto `fedoraproject` per creare Slurm o le sue dipendenze.

Sistemi operativi supportati

Il software AWS PCS e i programmi di installazione di Slurm supportano i seguenti sistemi operativi:

- Amazon Linux 2
- RedHat Enterprise Linux 9
- Rocky Linux 9
- Ubuntu 22.04

Note

AWS Deep Learning AMIs (DLAMI) le versioni basate su Amazon Linux 2 e Ubuntu 22.04 devono essere compatibili con il software AWS PCS e i programmi di installazione Slurm.

Per ulteriori informazioni, consulta [Choosing Your DLAMI nella Developer Guide](#).AWS Deep Learning AMIs

Tipi di istanze supportati

AWS PCS il software e i programmi di installazione Slurm supportano qualsiasi tipo di istanza x86_64 o arm64 che possa eseguire uno dei sistemi operativi supportati.

Versioni Slurm supportate

Sono supportate le seguenti versioni principali di Slurm:

- Slurm 23.11

Verifica gli installatori utilizzando un checksum

È possibile utilizzare i SHA256 checksum per verificare i file tarball (.tar.gz) del programma di installazione. È consigliabile eseguire questa operazione per verificare l'identità dell'autore del software e che l'applicazione non sia stata alterata o danneggiata dopo la pubblicazione.

Per verificare un tarball

Utilizzate l'utilità sha256sum per il SHA256 checksum e specificate il nome del file tarball. È necessario eseguire il comando dalla directory in cui è stato salvato il file tarball.

- SHA256

```
$ sha256sum tarball_filename.tar.gz
```

Il comando deve restituire un valore di checksum nel formato seguente.

```
checksum_value tarball_filename.tar.gz
```

Confrontate il valore di checksum restituito dal comando con il valore di checksum fornito nella tabella seguente. Se i checksum corrispondono, è sicuro eseguire lo script di installazione.

⚠ Important

Se i checksum non corrispondono, non eseguite lo script di installazione. Contattare [AWS Support](#).

Ad esempio, il comando seguente genera il SHA256 checksum per il tarball Slurm 23.11.9.

```
$ sha256sum aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz
```

Output di esempio:

```
1de7d919c8632fe8e2806611bed4fde1005a4fadc795412456e935c7bba2a9b8 aws-pcs-slurm-23.11-
installer-23.11.9-1.tar.gz
```

La tabella seguente elenca i checksum per le versioni recenti dei programmi di installazione. Replace (Sostituisci) *us-east-1* con il Regione AWS luogo in cui si utilizza. AWS PCS

Installer (Programma di installazione)	Scarica URL	SHA256checksum
Slurm 23.11.9	<code>https://aws-pcs-repo- <i>us-east-1</i> .s3.amazonaws.com/aws-pcs-slurm/aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz</code>	<code>1de7d919c8632fe8e2806611bed4fde1005a4fadc795412456e935c7bba2a9b8</code>
AWS PCSagente 1.0.0	<code>https://aws-pcs-repo- <i>us-east-1</i> .s3.amazonaws.com/aws-pcs-agent/aws-pcs-agent-v1.0.0-1.tar.gz</code>	<code>d2d3d68d00c685435c38af471d7e2492dde5ce9eb222d7b6ef0042144b134ce0</code>

Versioni Slurm in AWS PCS

SchedMD migliora continuamente Slurm con nuove funzionalità, ottimizzazioni e patch di sicurezza. SchedMD rilascia una nuova versione principale a [intervalli regolari](#) e prevede di supportare fino a 3 versioni alla volta. AWS PCS supporta inizialmente Slurm 23.11. Puoi aggiornare la tua versione principale di Slurm dopo il rilascio di una nuova versione. AWS PCS è progettato per aggiornare automaticamente il controller Slurm con versioni patch.

Quando SchedMD termina [il supporto](#) per una particolare versione principale, termina AWS PCS anche il supporto per quella versione principale. AWS PCS invia un avviso anticipato se una versione principale di Slurm è prossima alla fine del ciclo di vita, per aiutare i clienti a sapere quando aggiornare i propri cluster a una versione più recente supportata.

Ti consigliamo di utilizzare l'ultima versione supportata di Slurm per distribuire il tuo cluster, per accedere ai progressi e ai miglioramenti più recenti.

Domande frequenti sulle versioni di Slurm

Per quanto tempo AWS PCS supporta una versione di Slurm?

AWS PCS segue i cicli di supporto di SchedMD per le versioni principali. AWS PCS supporta fino a 3 versioni principali in qualsiasi momento. Dopo il rilascio di una nuova versione principale, SchedMD AWS PCS ritira la versione più vecchia supportata. AWS PCS rilascia una nuova versione principale di Slurm il prima possibile, ma potrebbe esserci un ritardo tra la versione di SchedMD e la sua disponibilità in AWS PCS.

Quando mi AWS PCS comunica le versioni di End of Support Life (EOSL) for Slurm?

AWS PCS ti avvisa più volte, con una cadenza predeterminata, prima della data EOSL.

Cosa devo fare quando si avvicina una versione di Slurm? EOSL

È necessario aggiornare prima le versioni di Slurm EOSL per contribuire a mantenere un ambiente sicuro e supportato.

Come posso aggiornare i miei cluster per utilizzare una nuova versione principale di Slurm?

Per aggiornare la versione di Slurm, è necessario creare un nuovo cluster. È inoltre necessario eseguire l'aggiornamento al AWS PCS software equivalente in uso AMI e utilizzarlo per creare i gruppi di nodi di calcolo per il nuovo cluster.

In che modo i miei cluster riceveranno nuove versioni di patch Slurm?

AWS PCS è progettato per applicare automaticamente le patch per risolvere le vulnerabilità e le esposizioni comuni di Slurm (. CVEs AWS PCS applica le patch ai controller del cluster eseguiti in account interni di proprietà dei servizi. È necessario utilizzare le AWS PCS API azioni AWS Management Console o per installare le patch sulle istanze di EC2 Account AWS

Cosa succede se non aggiorno Slurm entro la data? EOSL

AWS PCS è progettato per bloccare i cluster che hanno una versione Slurm non supportata. È necessario aggiornare la versione principale Slurm del controller del cluster e il AWS PCS software installato sui gruppi di nodi di calcolo.

Quante versioni di Slurm supporta? AWS PCS

AWS PCS supporta fino a 3 versioni principali di Slurm in qualsiasi momento, incluse la versione principale attuale e le 2 precedenti.

Quali aggiornamenti della versione di Slurm devo applicare?

Ti consigliamo vivamente di utilizzare la stessa versione principale per tutti i componenti del cluster e di installare le patch più recenti non appena vengono rilasciate. I gruppi di nodi AMIs per il calcolo devono utilizzare una versione del software Slurm compatibile con la versione Slurm del controller del cluster. La versione principale di Slurm in uso AMIs deve essere compresa tra 2 versioni della versione principale di Slurm sul controller del cluster. La versione Slurm installata nelle AMI e sulle EC2 istanze in esecuzione nel cluster non può essere più recente della versione Slurm sul controller del cluster. Per mantenere il supporto per il cluster, è AMIs necessario utilizzare una versione software supportata. AWS PCS

Cosa succede se aggiorno la versione principale di Slurm ma utilizzo il software Slurm precedente nei miei gruppi di nodi AMI for compute?

È necessario aggiornare il AWS PCS software alla stessa versione per utilizzare la nuova funzionalità Slurm. Per un AWS PCS supporto completo, tutti i componenti Slurm devono utilizzare versioni supportate. In sintesi:

- Siamo in grado di fornire un supporto completo quando il controller del cluster e tutti i componenti (AWS PCS pacchetti) di Account AWS entrambi utilizzano le versioni supportate.
- AWS PCS è progettato per arrestare un cluster se la versione Slurm del relativo controller arriva. EOSL

- Se la versione Slurm dei componenti è alla tua Account AWS portataEOSL, il tuo cluster non sarà supportato.

In quale ordine devo aggiornare i componenti del mio Cluster?

È necessario aggiornare la versione Slurm del controller del cluster prima di utilizzarla e AMI con una versione Slurm più recente. Si aggiorna un gruppo di nodi di calcolo per utilizzare il. AMI AWS PCSutilizza il AMI per avviare nuove EC2 istanze nel gruppo di nodi di calcolo. AWS PCSnon aggiorna EC2 le istanze esistenti con processi in esecuzione; AWS PCS è progettato per terminare tali istanze dopo il completamento dei processi.

AWS PCSOffre un supporto esteso per le versioni di Slurm?

No. Comunicheremo informazioni dettagliate sulle opzioni di supporto esteso, inclusi eventuali costi aggiuntivi e la copertura di supporto specifica fornita.

Servizio di sicurezza nel settore del calcolo AWS parallelo

La sicurezza del cloud AWS è la massima priorità. In qualità di AWS cliente, puoi beneficiare di data center e architetture di rete progettati per soddisfare i requisiti delle organizzazioni più sensibili alla sicurezza.

La sicurezza è una responsabilità condivisa tra te e te. AWS Il [modello di responsabilità condivisa](#) descrive questo aspetto come sicurezza del cloud e sicurezza nel cloud:

- Sicurezza del cloud: AWS è responsabile della protezione dell'infrastruttura che gestisce AWS i servizi in Cloud AWS. AWS fornisce inoltre servizi che è possibile utilizzare in modo sicuro. I revisori esterni testano e verificano regolarmente l'efficacia della nostra sicurezza nell'ambito dei [AWS Programmi di AWS conformità dei Programmi di conformità](#) dei di . Per ulteriori informazioni sui programmi di conformità che si applicano al servizio di elaborazione AWS parallela, vedere [AWS Servizi nell'ambito del programma di conformitàAWS](#) .
- Sicurezza nel cloud: la tua responsabilità è determinata dal AWS servizio che utilizzi. Sei anche responsabile di altri fattori, tra cui la riservatezza dei dati, i requisiti della tua azienda e le leggi e normative vigenti.

Questa documentazione ti aiuta a capire come applicare il modello di responsabilità condivisa durante l'utilizzo AWS PCS. Negli argomenti seguenti viene illustrato come eseguire la configurazione AWS PCS per soddisfare gli obiettivi di sicurezza e conformità. Imparerai anche a utilizzare altri AWS servizi che ti aiutano a monitorare e proteggere AWS PCS le tue risorse.

Argomenti

- [Protezione dei dati in AWS Parallel Computing Service](#)
- [Accedi al servizio AWS Parallel Computing utilizzando un'interfaccia endpoint \(AWS PrivateLink\)](#)
- [Servizio di Identity and Access Management per AWS Parallel Computing](#)
- [Convalida della conformità per Parallel Computing Service AWS](#)
- [Resilienza nel servizio di elaborazione AWS parallela](#)
- [Servizio di sicurezza dell'infrastruttura nel servizio di elaborazione AWS parallela](#)
- [Analisi e gestione delle vulnerabilità in Parallel Computing Service AWS](#)
- [Prevenzione del confused deputy tra servizi](#)
- [Best practice di sicurezza per AWS Parallel Computing Service](#)

Protezione dei dati in AWS Parallel Computing Service

Il modello di [responsabilità AWS condivisa modello](#) si applica alla protezione dei dati in AWS Parallel Computing Service. Come descritto in questo modello, AWS è responsabile della protezione dell'infrastruttura globale che gestisce tutti i Cloud AWS. L'utente è responsabile del controllo dei contenuti ospitati su questa infrastruttura. L'utente è inoltre responsabile della configurazione della protezione e delle attività di gestione per i Servizi AWS utilizzati. Per ulteriori informazioni sulla privacy dei dati, consulta la sezione [Privacy dei dati FAQ](#). Per informazioni sulla protezione dei dati in Europa, consulta il [Modello di responsabilitàAWS condivisa e GDPR](#) il post sul blog sulla AWS sicurezza.

Ai fini della protezione dei dati, ti consigliamo di proteggere Account AWS le credenziali e di configurare i singoli utenti con AWS IAM Identity Center o AWS Identity and Access Management (IAM). In tal modo, a ogni utente verranno assegnate solo le autorizzazioni necessarie per svolgere i suoi compiti. Ti suggeriamo, inoltre, di proteggere i dati nei seguenti modi:

- Utilizza l'autenticazione a più fattori (MFA) con ogni account.
- Usa SSL/TLS per comunicare con AWS le risorse. Richiediamo TLS 1.2 e consigliamo TLS 1.3.
- Configurazione API e registrazione delle attività degli utenti con AWS CloudTrail.
- Utilizza soluzioni di AWS crittografia, insieme a tutti i controlli di sicurezza predefiniti all'interno Servizi AWS.
- Utilizza i servizi di sicurezza gestiti avanzati, come Amazon Macie, che aiutano a individuare e proteggere i dati sensibili archiviati in Amazon S3.
- Se hai bisogno di FIPS 140-3 moduli crittografici convalidati per accedere AWS tramite un'interfaccia a riga di comando o un'API, usa un endpoint. FIPS Per ulteriori informazioni sugli FIPS endpoint disponibili, vedere [Federal Information Processing Standard \(\) 140-3. FIPS](#)

Ti consigliamo vivamente di non inserire mai informazioni riservate o sensibili, ad esempio gli indirizzi e-mail dei clienti, nei tag o nei campi di testo in formato libero, ad esempio nel campo Nome. Ciò include quando lavori AWS PCS o Servizi AWS utilizzi in altro modo la console, API AWS CLI, o. AWS SDKs I dati inseriti nei tag o nei campi di testo in formato libero utilizzati per i nomi possono essere utilizzati per la fatturazione o i log di diagnostica. Se fornisci un URL a un server esterno, ti consigliamo vivamente di non includere le informazioni sulle credenziali URL per convalidare la tua richiesta a quel server.

Crittografia a riposo

La crittografia è abilitata per impostazione predefinita per i dati inattivi quando si crea un cluster AWS Parallel Computing Service (AWS PCS) con AWS Management Console, AWS CLI AWS PCSAPI, o AWS SDKs AWS PCSutilizza una KMSchiave AWS proprietaria per crittografare i dati inattivi. Per ulteriori informazioni, consulta [Customer keys and AWS keys](#) nella AWS KMS Developer Guide. Il segreto del cluster viene archiviato AWS Secrets Manager e crittografato con la KMS chiave gestita di Secrets Manager. Per ulteriori informazioni, consulta [Utilizzo dei segreti del cluster in AWS PCS](#).

In un AWS PCS cluster, i seguenti dati sono inattivi:

- Stato dell'utilità di pianificazione: include i dati sui processi in esecuzione e sui nodi a cui è stato assegnato il provisioning nel cluster. Questi sono i dati in cui Slurm persiste nei dati definiti nel tuo `StateSaveLocation` `slurm.conf` Per ulteriori informazioni, consulta la descrizione contenuta [StateSaveLocation](#) nella documentazione di Slurm. AWS PCSelimina i dati del lavoro dopo il completamento di un lavoro.
- Scheduler auth secret: lo AWS PCS utilizza per autenticare tutte le comunicazioni dello scheduler nel cluster.

Per le informazioni sullo stato dello scheduler, crittografa AWS PCS automaticamente i dati e i metadati prima di scriverli nel file system. Il file system crittografato utilizza l'algoritmo di crittografia AES -256 standard del settore per i dati inattivi.

Crittografia in transito

Le tue connessioni AWS PCS API utilizzano la TLS crittografia con il processo di firma Signature Version 4, indipendentemente dal fatto che utilizzi () o AWS Command Line Interface AWS CLI AWS SDKs Per ulteriori informazioni, consulta [Firmare AWS API le richieste](#) nella Guida AWS Identity and Access Management per l'utente. AWS gestisce il controllo degli accessi tramite API le IAM politiche relative alle credenziali di sicurezza utilizzate per la connessione.

AWS PCSutilizza TLS per connettersi ad altri AWS servizi.

All'interno di un cluster Slurm, lo scheduler è configurato con il plug-in di autenticazione che fornisce l'`auth/slurmutenticazione` per tutte le comunicazioni dello scheduler. Slurm non fornisce la crittografia a livello di applicazione per le sue comunicazioni, tutti i dati che fluiscono tra le istanze del cluster rimangono locali EC2 VPC e quindi sono soggetti a VPC crittografia se tali istanze supportano la crittografia in transito. Per ulteriori informazioni, consulta [Encryption in transit](#) nella Amazon Elastic

Compute Cloud User Guide. La comunicazione è crittografata tra il controller (fornito in un account di servizio) e i nodi del cluster del tuo account.

Gestione delle chiavi

AWS PCS utilizza una KMS chiave AWS proprietaria per crittografare i dati. Per ulteriori informazioni, consulta [Customer keys and AWS keys](#) nella AWS KMS Developer Guide. Il segreto del cluster viene archiviato AWS Secrets Manager e crittografato con la KMS chiave gestita di Secrets Manager. Per ulteriori informazioni, consulta [Utilizzo dei segreti del cluster in AWS PCS](#).

Riservatezza del traffico Internet

AWS PCS le risorse di calcolo per un cluster risiedono entro 1 unità VPC nell'account del cliente. Pertanto, tutto il traffico AWS PCS di servizio interno all'interno di un cluster rimane all'interno della AWS rete e non viaggia su Internet. La comunicazione tra l'utente e AWS PCS i nodi può viaggiare su Internet e consigliamo di utilizzare SSH il nostro Systems Manager per connettersi ai nodi. Per ulteriori informazioni, vedi [Cos'è AWS Systems Manager?](#) nella Guida AWS Systems Manager per l'utente.

Puoi anche utilizzare le seguenti offerte per connettere la tua rete locale a: AWS

- AWS Site-to-Site VPN. Per ulteriori informazioni, vedi [Cos'è AWS Site-to-Site VPN?](#) nella Guida AWS Site-to-Site VPN per l'utente.
- Un AWS Direct Connect. Per ulteriori informazioni, vedi [Cos'è AWS Direct Connect?](#) nella Guida AWS Direct Connect per l'utente.

È possibile accedere AWS PCS API a per eseguire attività amministrative relative al servizio. Tu e i tuoi utenti accedete alle porte degli endpoint Slurm per interagire direttamente con lo scheduler.

Crittografia del traffico API

Per accedere a AWS PCS API, i client devono supportare Transport Layer Security (TLS) 1.2 o versione successiva. Richiediamo TLS 1.2 e consigliamo TLS 1.3. I client devono inoltre supportare suite di crittografia con Perfect Forward Secrecy (PFS), come Ephemeral Diffie-Hellman () o Elliptic Curve Diffie-Hellman Ephemeral (). DHE 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 Puoi anche utilizzare AWS Security Token Service (AWS STS) per generare credenziali di sicurezza temporanee per firmare le richieste.

Crittografia del traffico dati

La crittografia dei dati in transito è abilitata dalle EC2 istanze supportate che accedono all'endpoint dello scheduler e tra le ComputeNodeGroup istanze dall'interno di. Cloud AWS Per ulteriori informazioni, consulta [Crittografia in transito](#).

Accedi al servizio AWS Parallel Computing utilizzando un'interfaccia endpoint ()AWS PrivateLink

È possibile utilizzare AWS PrivateLink per creare una connessione privata tra l'utente VPC e AWS Parallel Computing Service (AWS PCS). Puoi accedere AWS PCS come se fossi nel tuo computerVPC, senza l'uso di un gateway, NAT dispositivo, VPN connessione o AWS Direct Connect connessione Internet. Le istanze del tuo account VPC non hanno bisogno di indirizzi IP pubblici per accedervi AWS PCS.

Stabilisci questa connessione privata creando un endpoint di interfaccia attivato da AWS PrivateLink. In ciascuna sottorete viene creata un'interfaccia di rete endpoint da abilitare per l'endpoint di interfaccia. Si tratta di interfacce di rete gestite dai richiedenti che fungono da punto di ingresso per il traffico destinato a. AWS PCS

Per ulteriori informazioni, consulta [Access Servizi AWS through AWS PrivateLink](#) nella Guida.AWS PrivateLink

Considerazioni per AWS PCS

Prima di configurare un endpoint di interfaccia per AWS PCS, consulta [Accedere a un AWS servizio utilizzando un VPC endpoint di interfaccia nella Guida](#).AWS PrivateLink

AWS PCSsupporta l'effettuazione di chiamate a tutte le sue API azioni tramite l'endpoint dell'interfaccia.

Se VPC non disponi di un accesso diretto a Internet, devi configurare un VPC endpoint per consentire alle istanze del gruppo di nodi di calcolo di avviare l'azione. AWS PCS [RegisterComputeNodeGroupInstanceAPI](#)

Crea un endpoint di interfaccia per AWS PCS

Puoi creare un endpoint di interfaccia per AWS PCS utilizzare la VPC console Amazon o AWS Command Line Interface (AWS CLI). Per ulteriori informazioni, consulta la sezione [Creazione di un endpoint di interfaccia](#) nella Guida per l'utente di AWS PrivateLink .

Crea un endpoint di interfaccia per AWS PCS utilizzare il seguente nome di servizio:

```
com.amazonaws.region.pcs
```

Replace (Sostituisci) *region* con l'ID di Regione AWS in cui creare l'endpoint, ad esempio. us-east-1

Se abiliti private DNS per l'endpoint dell'interfaccia, puoi effettuare API richieste AWS PCS utilizzando il nome regionale DNS predefinito. Ad esempio pcs.us-east-1.amazonaws.com.

Creazione di una policy dell' endpoint per l'endpoint dell'interfaccia

Una policy per gli endpoint è una IAM risorsa che è possibile collegare a un endpoint di interfaccia. La policy predefinita per gli endpoint consente l'accesso completo AWS PCS tramite l'endpoint dell'interfaccia. Per controllare l'accesso consentito AWS PCS dal tuo dispositivoVPC, allega una policy personalizzata per l'endpoint all'endpoint di interfaccia.

Una policy di endpoint specifica le informazioni riportate di seguito:

- I principali che possono eseguire azioni (IAM utenti Account AWS e IAM ruoli).
- Le azioni che possono essere eseguite.
- Le risorse in cui è possibile eseguire le operazioni.

Per ulteriori informazioni, consulta la sezione [Controllo dell'accesso ai servizi con policy di endpoint](#) nella Guida di AWS PrivateLink .

Esempio: politica VPC degli endpoint per le azioni AWS PCS

Di seguito è riportato l'esempio di una policy dell'endpoint personalizzata. Quando si allega questa policy all'endpoint di interfaccia, si concede l'accesso alle AWS PCS azioni elencate per tutti i principali attori del cluster con i valori specificati *cluster-id*. Sostituisci *region* con l'ID Regione AWS del cluster, ad esempio us-east-1. Replace (Sostituisci) *account-id* con il Account AWS numero del cluster.

```
{
  "Statement": [
    {
      "Action": [
        "pcs:CreateCluster",
        "pcs:ListClusters",
        "pcs>DeleteCluster",
        "pcs:GetCluster",
      ],
      "Effect": "Allow",
      "Principal": "*",
      "Resource": [
        "arn:aws:pcs:region:account-id:cluster/cluster-id*"
      ]
    }
  ]
}
```

Servizio di Identity and Access Management per AWS Parallel Computing

AWS Identity and Access Management (IAM) è un dispositivo Servizio AWS che aiuta un amministratore a controllare in modo sicuro l'accesso alle AWS risorse. IAM gli amministratori controllano chi può essere autenticato (effettuato l'accesso) e autorizzato (disporre delle autorizzazioni) a utilizzare le risorse. AWS PCS IAM è un dispositivo Servizio AWS che puoi utilizzare senza costi aggiuntivi.

Argomenti

- [Destinatari](#)
- [Autenticazione con identità](#)
- [Gestione dell'accesso con policy](#)
- [Come funziona AWS Parallel Computing Service con IAM](#)
- [Esempi di policy basate sull'identità per Parallel Computing Service AWS](#)
- [AWS politiche gestite per AWS Parallel Computing Service](#)
- [Ruoli collegati ai servizi per l' AWS PCS](#)
- [Ruolo di Amazon EC2 Spot per AWS PCS](#)

- [Autorizzazioni minime per AWS PCS](#)
- [IAM profili di istanza per AWS Parallel Computing Service](#)
- [Risoluzione dei problemi di identità e accesso al AWS Parallel Computing Service](#)

Destinatari

Il modo in cui usi AWS Identity and Access Management (IAM) varia a seconda del lavoro che svolgi. AWS PCS

Utente del servizio: se utilizzi il AWS PCS servizio per svolgere il tuo lavoro, l'amministratore ti fornisce le credenziali e le autorizzazioni necessarie. Man mano che utilizzi più AWS PCS funzionalità per svolgere il tuo lavoro, potresti aver bisogno di autorizzazioni aggiuntive. La comprensione della gestione dell'accesso ti consente di richiedere le autorizzazioni corrette all'amministratore. Se non riesci ad accedere a una funzionalità in AWS PCS, consulta [Risoluzione dei problemi di identità e accesso al AWS Parallel Computing Service](#).

Amministratore del servizio: se sei responsabile delle AWS PCS risorse della tua azienda, probabilmente hai pieno accesso a AWS PCS. È tuo compito determinare a quali AWS PCS funzionalità e risorse devono accedere gli utenti del servizio. È quindi necessario inviare richieste all'IAM amministratore per modificare le autorizzazioni degli utenti del servizio. Consulta le informazioni contenute in questa pagina per comprendere i concetti di base di IAM. Per ulteriori informazioni su come la tua azienda può utilizzare IAM con AWS PCS, consulta [Come funziona AWS Parallel Computing Service con IAM](#).

IAM amministratore: se sei un IAM amministratore, potresti voler conoscere i dettagli su come scrivere politiche a cui gestire l'accesso AWS PCS. Per visualizzare esempi di policy AWS PCS basate sull'identità che puoi utilizzare in IAM, consulta [Esempi di policy basate sull'identità per Parallel Computing Service AWS](#)

Autenticazione con identità

L'autenticazione è il modo in cui accedi AWS utilizzando le tue credenziali di identità. È necessario autenticarsi (accedere a AWS) come Utente root dell'account AWS, come IAM utente o assumendo un ruolo. IAM

È possibile accedere AWS come identità federata utilizzando le credenziali fornite tramite una fonte di identità. AWS IAM Identity Center Gli utenti (IAM Identity Center), l'autenticazione Single Sign-On della tua azienda e le tue credenziali di Google o Facebook sono esempi di identità federate. Quando

accedi come identità federata, l'amministratore aveva precedentemente configurato la federazione delle identità utilizzando i ruoli. IAM Quando si accede AWS utilizzando la federazione, si assume indirettamente un ruolo.

A seconda del tipo di utente, puoi accedere al AWS Management Console o al portale di AWS accesso. Per ulteriori informazioni sull'accesso a AWS, vedi [Come accedere al tuo Account AWS nella Guida per l'Accedi ad AWS utente](#).

Se accedi a AWS livello di codice, AWS fornisce un kit di sviluppo software (SDK) e un'interfaccia a riga di comando () per firmare crittograficamente le tue richieste utilizzando le tue credenziali. CLI Se non utilizzi AWS strumenti, devi firmare tu stesso le richieste. Per ulteriori informazioni sull'utilizzo del metodo consigliato per firmare autonomamente le richieste, consulta [Firmare AWS API le richieste nella Guida per l'IAMutente](#).

A prescindere dal metodo di autenticazione utilizzato, potrebbe essere necessario specificare ulteriori informazioni sulla sicurezza. Ad esempio, ti AWS consiglia di utilizzare l'autenticazione a più fattori (MFA) per aumentare la sicurezza del tuo account. Per ulteriori informazioni, consulta [Autenticazione a più fattori nella Guida per l'AWS IAM Identity Center utente](#) e [Utilizzo dell'autenticazione a più fattori \(MFA\) AWS nella Guida per l'IAMutente](#).

Account AWS utente root

Quando si crea un account Account AWS, si inizia con un'identità di accesso che ha accesso completo a tutte Servizi AWS le risorse dell'account. Questa identità è denominata utente Account AWS root ed è accessibile effettuando l'accesso 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 l'elenco completo delle attività che richiedono l'accesso come utente root, consulta [Attività che richiedono le credenziali dell'utente root](#) nella Guida per l'IAMutente.

Identità federata

Come procedura consigliata, richiedi agli utenti umani, compresi gli utenti che richiedono l'accesso come amministratore, di utilizzare la federazione con un provider di identità per accedere Servizi AWS utilizzando credenziali temporanee.

Un'identità federata è un utente dell'elenco utenti aziendale, di un provider di identità Web AWS Directory Service, della directory Identity Center o di qualsiasi utente che accede utilizzando le Servizi AWS credenziali fornite tramite un'origine di identità. Quando le identità federate accedono Account AWS, assumono ruoli e i ruoli forniscono credenziali temporanee.

Per la gestione centralizzata degli accessi, consigliamo di utilizzare AWS IAM Identity Center. Puoi creare utenti e gruppi in IAM Identity Center oppure puoi connetterti e sincronizzarti con un set di utenti e gruppi nella tua fonte di identità per utilizzarli su tutte le tue applicazioni. Account AWS Per informazioni su IAM Identity Center, vedi [Cos'è IAM Identity Center?](#) nella Guida AWS IAM Identity Center per l'utente.

IAM users and groups

Un [IAMutente](#) è un'identità interna all'utente Account AWS che dispone di autorizzazioni specifiche per una singola persona o applicazione. Laddove possibile, consigliamo di fare affidamento su credenziali temporanee anziché creare IAM utenti con credenziali a lungo termine come password e chiavi di accesso. Tuttavia, se hai casi d'uso specifici che richiedono credenziali a lungo termine con IAM gli utenti, ti consigliamo di ruotare le chiavi di accesso. Per ulteriori informazioni, consulta [Ruotare regolarmente le chiavi di accesso per i casi d'uso che richiedono credenziali a lungo termine](#) nella Guida per l'utente. IAM

Un [IAMgruppo](#) 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 assegnare un nome a un gruppo IAMAdminse concedere a tale gruppo le autorizzazioni per IAM amministrare le risorse.

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 IAM utente \(anziché un ruolo\)](#) nella Guida per l'IAMutente.

IAMruoli

Un [IAMruolo](#) è un'identità interna all'utente Account AWS che dispone di autorizzazioni specifiche. È simile a un IAM utente, ma non è associato a una persona specifica. È possibile assumere temporaneamente un IAM ruolo in AWS Management Console [cambiando ruolo](#). È possibile assumere un ruolo chiamando un' AWS APIoperazione AWS CLI or o utilizzando un'operazione personalizzataURL. Per ulteriori informazioni sui metodi di utilizzo dei ruoli, vedere [Utilizzo IAM dei ruoli](#) nella Guida per l'IAMutente.

IAMi ruoli 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 informazioni sui ruoli per la federazione, vedere [Creazione di un ruolo per un provider di identità di terze parti](#) nella Guida per l'IAMutente. Se utilizzi IAM Identity Center, configuri un set di autorizzazioni. Per controllare a cosa possono accedere le identità dopo l'autenticazione, IAM Identity Center correla il set di autorizzazioni a un ruolo in IAM. Per informazioni sui set di autorizzazioni, consulta [Set di autorizzazioni](#) nella Guida per l'utente di AWS IAM Identity Center .
- **Autorizzazioni IAM utente temporanee:** un IAM utente o un ruolo può assumere il IAM ruolo di assumere temporaneamente autorizzazioni diverse per un'attività specifica.
- **Accesso su più account:** puoi utilizzare un IAM ruolo per consentire a qualcuno (un responsabile fidato) di un altro account di accedere alle risorse del tuo account. I ruoli sono lo strumento principale per concedere l'accesso multi-account. Tuttavia, con alcuni Servizi AWS, è possibile allegare una policy direttamente a una risorsa (anziché utilizzare un ruolo come proxy). Per conoscere la differenza tra ruoli e politiche basate sulle risorse per l'accesso tra account diversi, consulta la [sezione Accesso alle risorse su più account IAM nella Guida per l'utente](#). IAM
- **Accesso tra servizi:** alcuni Servizi AWS utilizzano funzionalità in altri. Servizi AWS Ad esempio, quando effettui una chiamata in un servizio, è normale che quel 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.
- **Sessioni di accesso inoltrato (FAS):** quando utilizzi un IAM utente o un ruolo per eseguire azioni AWS, sei considerato un principale. Quando si utilizzano alcuni servizi, è possibile eseguire un'operazione che attiva un'altra operazione in un servizio diverso. FASutilizza le autorizzazioni del principale che chiama un Servizio AWS, in combinazione con la richiesta di effettuare richieste Servizio AWS ai servizi downstream. FASle richieste vengono effettuate solo quando un servizio riceve una richiesta che richiede interazioni con altri Servizi AWS o risorse per essere completata. In questo caso è necessario disporre delle autorizzazioni per eseguire entrambe le azioni. Per i dettagli FAS delle politiche relative alle richieste, consulta [Forward access sessions](#).
- **Ruolo di servizio:** un ruolo di servizio è un [IAMruolo](#) che un servizio assume per eseguire azioni per conto dell'utente. Un IAM amministratore può creare, modificare ed eliminare un ruolo di servizio dall'internoIAM. Per ulteriori informazioni, vedere [Creazione di un ruolo per delegare le autorizzazioni a un utente Servizio AWS nella Guida per l'IAMutente](#).
- **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 al servizio vengono visualizzati nel tuo account Account AWS e sono di proprietà del servizio. Un IAM amministratore può visualizzare, ma non modificare le autorizzazioni per i ruoli collegati al servizio.

- Applicazioni in esecuzione su Amazon EC2: puoi utilizzare un IAM ruolo per gestire le credenziali temporanee per le applicazioni in esecuzione su un'EC2istanza e che effettuano AWS CLI o effettuano AWS API richieste. Ciò è preferibile alla memorizzazione delle chiavi di accesso all'interno dell'EC2istanza. Per assegnare un AWS ruolo a un'EC2istanza e renderlo disponibile per tutte le sue applicazioni, create un profilo di istanza collegato all'istanza. Un profilo di istanza contiene il ruolo e consente ai programmi in esecuzione sull'EC2istanza di ottenere credenziali temporanee. Per ulteriori informazioni, consulta [Usare un IAM ruolo per concedere le autorizzazioni alle applicazioni in esecuzione su EC2 istanze Amazon nella Guida](#) per l'IAMutente.

Per sapere se utilizzare IAM ruoli o IAM utenti, consulta [Quando creare un IAM ruolo \(anziché un utente\)](#) nella Guida per l'IAMutente.

Gestione dell'accesso con policy

Puoi controllare l'accesso AWS creando policy e associandole a AWS identità o risorse. Una policy è un oggetto AWS che, se associato a un'identità o a una risorsa, ne definisce le autorizzazioni. AWS valuta queste politiche quando un principale (utente, utente root o sessione di ruolo) effettua una richiesta. Le autorizzazioni nelle policy determinano l'approvazione o il rifiuto della richiesta. La maggior parte delle politiche viene archiviata AWS come JSON documenti. Per ulteriori informazioni sulla struttura e il contenuto dei documenti relativi alle JSON politiche, vedere [Panoramica delle JSON politiche](#) nella Guida per l'IAMutente.

Gli amministratori possono utilizzare AWS JSON le politiche per specificare chi ha accesso a cosa. 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 azioni sulle risorse di cui hanno bisogno, un IAM amministratore può creare IAM politiche. L'amministratore può quindi aggiungere le IAM politiche ai ruoli e gli utenti possono assumerli.

IAMle politiche definiscono le autorizzazioni per un'azione indipendentemente dal metodo utilizzato per eseguire l'operazione. Ad esempio, supponiamo di disporre di una policy che consente l'operazione `iam:GetRole`. Un utente con tale criterio può ottenere informazioni sul ruolo da AWS Management Console, da o da. AWS CLI AWS API

Policy basate su identità

I criteri basati sull'identità sono documenti relativi alle politiche di JSON autorizzazione che è possibile allegare a un'identità, ad esempio un IAM utente, un gruppo di utenti o un ruolo. Tali policy definiscono le azioni che utenti e ruoli possono eseguire, su quali risorse e in quali condizioni. [Per informazioni su come creare una politica basata sull'identità, consulta Creazione di politiche nella Guida per l'utente. IAM IAM](#)

Le policy basate su identità possono essere ulteriormente classificate come policy inline o policy gestite. Le policy inline sono integrate direttamente in un singolo utente, gruppo o ruolo. Le politiche gestite sono politiche autonome che puoi allegare a più utenti, gruppi e ruoli all'interno del tuo Account AWS. Le politiche gestite includono politiche AWS gestite e politiche gestite dai clienti. Per informazioni su come scegliere tra una politica gestita o una politica in linea, consulta [Scelta tra politiche gestite e politiche in linea nella Guida](#) per l'IAM utente.

Policy basate su risorse

Le politiche basate sulle risorse sono documenti di JSON policy allegati a una risorsa. Esempi di politiche basate sulle risorse sono le policy di trust dei IAM ruoli e le policy dei bucket di 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 collegata 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 AWS gestite contenute IAM in una policy basata sulle risorse.

Elenchi di controllo degli accessi () ACLs

Le liste di controllo degli accessi (ACLs) controllano quali principali (membri dell'account, utenti o ruoli) dispongono delle autorizzazioni per accedere a una risorsa. ACLs sono simili alle politiche basate sulle risorse, sebbene non utilizzino il formato del documento di policy. JSON

Amazon S3 e Amazon VPC sono esempi di servizi che supportano. AWS WAF ACLs Per ulteriori informazioni ACLs, consulta la [panoramica di Access control list \(ACL\)](#) nella Amazon Simple Storage Service Developer Guide.

Altri tipi di policy

AWS supporta tipi di policy aggiuntivi e 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 di autorizzazioni è una funzionalità avanzata in cui si impostano le autorizzazioni massime che una politica basata sull'identità può concedere a un'entità (utente o ruolo). IAM 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 nella Guida per l'utente. IAM IAM](#)
- **Politiche di controllo del servizio (SCPs):** SCPs sono JSON politiche che specificano le autorizzazioni massime per un'organizzazione o un'unità organizzativa (OU) in. AWS Organizations AWS Organizations è un servizio per il raggruppamento e la gestione centralizzata di più Account AWS di proprietà dell'azienda. Se abiliti tutte le funzionalità di un'organizzazione, puoi applicare le politiche di controllo del servizio (SCPs) a uno o tutti i tuoi account. SCP Limita le autorizzazioni per le entità negli account dei membri, inclusa ciascuna Utente root dell'account AWS. Per ulteriori informazioni su Organizations and SCPs, consulta [le politiche di controllo dei servizi](#) nella Guida AWS Organizations per l'utente.
- **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 [le politiche di sessione](#) nella Guida IAM per l'utente.

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 determinare se consentire una richiesta quando sono coinvolti più tipi di policy, consulta [Logica di valutazione delle politiche](#) nella Guida per l'IAM utente.

Come funziona AWS Parallel Computing Service con IAM

Prima di IAM utilizzarlo per gestire l'accesso a AWS PCS, scopri con quali IAM funzionalità è disponibile l'uso AWS PCS.

IAM funzionalità che è possibile utilizzare con AWS Parallel Computing Service

IAM caratteristica	AWS PCS supporto
Policy basate su identità	Sì
Policy basate su risorse	No
Azioni di policy	Sì
Risorse relative alle policy	Sì
Chiavi di condizione della policy (specifica del servizio)	Sì
ACLs	No
ABAC(tag nelle politiche)	Sì
Credenziali temporanee	Sì
Autorizzazioni del principale	Sì
Ruoli di servizio	No
Ruoli collegati al servizio	Sì

Per avere una panoramica generale del funzionamento AWS PCS e degli altri AWS servizi con la maggior parte delle IAM funzionalità, consulta [AWS i servizi che funzionano con](#) la maggior parte delle funzionalità IAM nella Guida per l'IAM utente.

Politiche basate sull'identità per AWS PCS

Supporta le policy basate su identità: sì

Le politiche basate sull'identità sono documenti relativi alle politiche di JSON autorizzazione che è possibile allegare a un'identità, ad esempio un IAM utente, un gruppo di utenti o un ruolo. Tali policy definiscono le azioni che utenti e ruoli possono eseguire, su quali risorse e in quali condizioni. [Per informazioni su come creare una politica basata sull'identità, consulta Creazione di politiche nella Guida per l'utente. IAM IAM](#)

Con le politiche IAM basate sull'identità, puoi specificare azioni e risorse consentite o negate, nonché le condizioni in base alle quali le azioni sono consentite o negate. 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 ulteriori informazioni su tutti gli elementi che è possibile utilizzare in una JSON politica, vedere il [riferimento agli elementi IAM JSON della politica](#) nella Guida per l'IAM utente.

Esempi di policy basate sull'identità per AWS PCS

Per visualizzare esempi di politiche basate sull' AWS PCS identità, vedere. [Esempi di policy basate sull'identità per Parallel Computing Service AWS](#)

Politiche basate sulle risorse all'interno AWS PCS

Supporta le policy basate su risorse: no

Le politiche basate sulle risorse sono documenti di JSON policy allegati a una risorsa. Esempi di politiche basate sulle risorse sono le policy di trust dei IAM ruoli e le policy dei bucket di 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 collegata 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 abilitare l'accesso tra più account, puoi specificare un intero account o IAM entità in un altro account come principale in una politica 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 il principale e la risorsa sono diversi Account AWS, un IAM amministratore dell'account fidato deve inoltre concedere all'entità principale (utente o ruolo) l'autorizzazione ad 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 la sezione [Cross Account Resource Access IAM nella Guida IAM per l'utente](#).

Azioni politiche per AWS PCS

Supporta le operazioni di policy: sì

Gli amministratori possono utilizzare AWS JSON le policy per specificare chi ha accesso a cosa. Cioè, quale principale può eseguire operazioni su quali risorse, e in quali condizioni.

L'Actionelemento di una JSON policy descrive le azioni che è possibile utilizzare per consentire o negare l'accesso a una policy. Le azioni politiche in genere hanno lo stesso nome dell' AWS APIoperazione associata. Esistono alcune eccezioni, come le azioni basate solo sulle autorizzazioni che non hanno un'operazione corrispondente. API Esistono anche alcune operazioni che richiedono più operazioni in una policy. Queste operazioni aggiuntive sono denominate operazioni dipendenti.

Includi le operazioni in una policy per concedere le autorizzazioni a eseguire l'operazione associata.

Per visualizzare un elenco di AWS PCS azioni, vedere Azioni [definite da AWS Parallel Computing Service nel Service Authorization Reference](#).

Le azioni politiche in AWS PCS uso utilizzano il seguente prefisso prima dell'azione:

```
pcs
```

Per specificare più operazioni in una sola istruzione, occorre separarle con la virgola.

```
"Action": [  
  "pcs:action1",  
  "pcs:action2"  
]
```

Per visualizzare esempi di politiche AWS PCS basate sull'identità, vedere. [Esempi di policy basate sull'identità per Parallel Computing Service AWS](#)

Risorse politiche per AWS PCS

Supporta le risorse di policy: sì

Gli amministratori possono utilizzare AWS JSON le policy per specificare chi ha accesso a cosa. Cioè, quale principale può eseguire operazioni su quali risorse, e in quali condizioni.

L'elemento Resource JSON policy specifica l'oggetto o gli oggetti a cui si applica l'azione. Le istruzioni devono includere un elemento Resource o un elemento NotResource. Come best practice, specifica una risorsa utilizzando il relativo [Amazon Resource Name \(ARN\)](#). 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 dei tipi di AWS PCS risorse e relativi ARNs, consulta [Resources Defined by AWS Parallel Computing Service nel Service Authorization Reference](#). Per sapere con quali azioni è possibile specificare il tipo ARN di ciascuna risorsa, vedere [Azioni definite dal servizio AWS Parallel Computing](#).

Per visualizzare esempi di politiche AWS PCS basate sull'identità, vedere [Esempi di policy basate sull'identità per Parallel Computing Service AWS](#)

Chiavi relative alle condizioni delle politiche per AWS PCS

Supporta le chiavi di condizione delle policy specifiche del servizio: sì

Gli amministratori possono utilizzare AWS JSON le politiche per specificare chi ha accesso a cosa. 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 si specificano più valori per una singola chiave di condizione, AWS valuta la condizione utilizzando un'operazione logica OR. 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, è possibile concedere a un IAM utente l'autorizzazione ad accedere a una risorsa solo se è contrassegnata con il

suo nome IAM utente. Per ulteriori informazioni, consulta [gli elementi IAM della politica: variabili e tag](#) nella Guida IAM per l'utente.

AWS supporta chiavi di condizione globali e chiavi di condizione specifiche del servizio. Per visualizzare tutte le chiavi di condizione AWS globali, consulta le chiavi di [contesto delle condizioni AWS globali nella Guida](#) per l'IAMutente.

Per visualizzare un elenco di chiavi di AWS PCS condizione, vedere [Condition Keys for AWS Parallel Computing Service nel Service](#) Authorization Reference. Per sapere con quali azioni e risorse è possibile utilizzare una chiave di condizione, consulta [Actions Defined by AWS Parallel Computing Service](#).

Per visualizzare esempi di politiche AWS PCS basate sull'identità, vedere. [Esempi di policy basate sull'identità per Parallel Computing Service AWS](#)

ACLsin AWS PCS

SupportiACLs: no

Le liste di controllo degli accessi (ACLs) controllano quali principali (membri dell'account, utenti o ruoli) dispongono delle autorizzazioni per accedere a una risorsa. ACLs sono simili alle politiche basate sulle risorse, sebbene non utilizzino il formato del documento di policy. JSON

ABACcon AWS PCS

Supporti ABAC (tag nelle politiche): Sì

Il controllo degli accessi basato sugli attributi (ABAC) è una strategia di autorizzazione che definisce le autorizzazioni in base agli attributi. In AWS, questi attributi sono chiamati tag. È possibile allegare tag a IAM entità (utenti o ruoli) e a molte AWS risorse. L'etichettatura di entità e risorse è il primo passo di ABAC. Quindi si progettano ABAC politiche per consentire le operazioni quando il tag del principale corrisponde al tag sulla risorsa a cui sta tentando di accedere.

ABAC è utile in ambienti in rapida crescita e aiuta in situazioni in cui la gestione delle politiche diventa complicata.

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 Parziale.

Per ulteriori informazioni su ABAC, vedere [Cos'è? ABAC](#) nella Guida IAM per l'utente. Per visualizzare un tutorial con i passaggi per la configurazione ABAC, consulta [Utilizzare il controllo di accesso basato sugli attributi \(ABAC\)](#) nella Guida per l'IAM utente.

Utilizzo di credenziali temporanee con AWS PCS

Supporta le credenziali temporanee: sì

Alcuni Servizi AWS non funzionano quando si accede utilizzando credenziali temporanee. Per ulteriori informazioni, incluse quelle che Servizi AWS funzionano con credenziali temporanee, consulta la sezione [Servizi AWS relativa alla funzionalità IAM nella Guida](#) per l'IAM utente.

Si utilizzano credenziali temporanee se si accede AWS Management Console utilizzando qualsiasi metodo tranne il nome utente e la password. Ad esempio, quando accedete AWS utilizzando il link Single Sign-on (SSO) della vostra azienda, tale processo crea automaticamente credenziali temporanee. Le credenziali temporanee vengono create in automatico anche quando accedi alla console come utente e poi cambi ruolo. Per ulteriori informazioni sul cambio di ruolo, consulta [Passare a un ruolo \(console\)](#) nella Guida per l'IAM utente.

È possibile creare manualmente credenziali temporanee utilizzando AWS CLI o AWS API. È quindi possibile utilizzare tali credenziali temporanee per accedere. AWS consiglia di generare dinamicamente credenziali temporanee anziché utilizzare chiavi di accesso a lungo termine. Per ulteriori informazioni, vedere [Credenziali di sicurezza temporanee](#) in IAM.

Autorizzazioni principali per più servizi per AWS PCS

Supporta sessioni di accesso diretto (FAS): Sì

Quando utilizzi un IAM utente o un ruolo per eseguire azioni AWS, sei considerato un principale. Quando si utilizzano alcuni servizi, è possibile eseguire un'operazione che attiva un'altra operazione in un servizio diverso. FAS utilizza le autorizzazioni del principale che chiama un Servizio AWS, in combinazione con la richiesta Servizio AWS per effettuare richieste ai servizi downstream. FAS le richieste vengono effettuate solo quando un servizio riceve una richiesta che richiede interazioni con altri Servizi AWS o risorse per essere completata. In questo caso è necessario disporre delle autorizzazioni per eseguire entrambe le azioni. Per i dettagli FAS delle politiche relative alle richieste, consulta [Forward access sessions](#).

Ruoli di servizio per AWS PCS

Supporta i ruoli di servizio: No

Un ruolo di servizio è un [IAMruolo](#) che un servizio assume per eseguire azioni per conto dell'utente. Un IAM amministratore può creare, modificare ed eliminare un ruolo di servizio dall'interno IAM. Per ulteriori informazioni, vedere [Creazione di un ruolo per delegare le autorizzazioni a un utente Servizio AWS nella Guida per l'IAMutente](#).

Warning

La modifica delle autorizzazioni per un ruolo di servizio potrebbe compromettere la funzionalità. AWS PCS Modifica i ruoli di servizio solo quando viene AWS PCS fornita una guida in tal senso.

Ruoli collegati ai servizi per AWS PCS

Supporta ruoli collegati ai servizi: Sì

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 al servizio vengono visualizzati nel tuo account Account AWS e sono di proprietà del servizio. Un IAM amministratore può visualizzare, ma non modificare le autorizzazioni per i ruoli collegati al servizio.

[Per informazioni dettagliate sulla creazione o la gestione di ruoli collegati ai servizi, consulta AWS Servizi compatibili con. 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.

Esempi di policy basate sull'identità per Parallel Computing Service AWS

Per impostazione predefinita, gli utenti e i ruoli non dispongono dell'autorizzazione per creare o modificare risorse. AWS PCS Inoltre, non possono eseguire attività utilizzando AWS Management Console, AWS Command Line Interface (AWS CLI) o AWS API. Per concedere agli utenti il permesso di eseguire azioni sulle risorse di cui hanno bisogno, un IAM amministratore può creare IAM policy. L'amministratore può quindi aggiungere le IAM politiche ai ruoli e gli utenti possono assumerli.

Per informazioni su come creare una politica IAM basata sull'identità utilizzando questi documenti di esempio JSON, consulta [Creazione di IAM politiche](#) nella Guida per l'IAM utente.

Per informazioni dettagliate sulle azioni e sui tipi di risorse definiti da AWS PCS, incluso il formato di ARNs per ogni tipo di risorsa, vedere [Actions, Resources and Condition Keys for AWS Parallel Computing Service nel Service Authorization](#) Reference.

Argomenti

- [Best practice per le policy](#)
- [Utilizzo della AWS PCS console](#)
- [Consentire agli utenti di visualizzare le loro autorizzazioni](#)

Best practice per le policy

Le politiche basate sull'identità determinano se qualcuno può creare, accedere o eliminare AWS PCS risorse nel tuo account. Queste azioni possono comportare costi aggiuntivi per l'Account AWS. Quando crei o modifichi policy basate su identità, segui queste linee guida e raccomandazioni:

- Inizia con le policy AWS gestite e passa alle autorizzazioni con privilegi minimi: per iniziare a concedere autorizzazioni a utenti e carichi di lavoro, utilizza le politiche gestite che concedono le autorizzazioni per molti casi d'uso comuni. AWS Sono disponibili nel tuo Account AWS. Ti consigliamo di ridurre ulteriormente le autorizzazioni definendo politiche gestite dai AWS clienti specifiche per i tuoi casi d'uso. Per ulteriori informazioni, consulta [le politiche AWS gestite o le politiche AWS gestite per le funzioni lavorative](#) nella Guida per l'IAM utente.
- Applica le autorizzazioni con privilegi minimi: quando imposti le autorizzazioni con le IAM politiche, concedi solo le autorizzazioni necessarie per eseguire un'attività. Puoi farlo definendo le azioni che possono essere intraprese su risorse specifiche in condizioni specifiche, note anche come autorizzazioni con privilegi minimi. Per ulteriori informazioni sull'utilizzo per applicare le autorizzazioni, consulta [Politiche](#) e autorizzazioni nella Guida IAM per l'utente. IAM IAM
- Utilizza le condizioni nelle IAM politiche per limitare ulteriormente l'accesso: puoi aggiungere una condizione alle tue politiche per limitare l'accesso ad azioni e risorse. Ad esempio, puoi scrivere una condizione di policy per specificare che tutte le richieste devono essere inviate utilizzando SSL. È inoltre possibile utilizzare condizioni per concedere l'accesso alle azioni di servizio se vengono utilizzate tramite uno specifico Servizio AWS, ad esempio AWS CloudFormation. Per ulteriori informazioni, consulta [Elementi IAM JSON della politica: Condizione](#) nella Guida IAM per l'utente.

- Usa IAM Access Analyzer per convalidare IAM le tue policy e garantire autorizzazioni sicure e funzionali: IAM Access Analyzer convalida le policy nuove ed esistenti in modo che aderiscano al linguaggio delle IAM policy () e alle best practice. JSON IAM IAMAccess Analyzer fornisce più di 100 controlli delle politiche e consigli pratici per aiutarti a creare policy sicure e funzionali. Per ulteriori informazioni, vedere [Convalida delle policy di IAM Access Analyzer nella Guida per l'utente](#). IAM
- Richiedi l'autenticazione a più fattori (MFA): se hai uno scenario che richiede l'utilizzo di IAM utenti o di un utente root Account AWS, attiva questa opzione MFA per una maggiore sicurezza. Per richiedere MFA quando vengono richiamate API le operazioni, aggiungi MFA delle condizioni alle tue politiche. Per ulteriori informazioni, consulta [Configurazione dell'APIaccesso MFA protetto nella Guida](#) per l'IAMutente.

Per ulteriori informazioni sulle procedure consigliate inIAM, consulta la sezione [Procedure consigliate in materia di sicurezza IAM nella Guida](#) per l'IAMutente.

Utilizzo della AWS PCS console

Per accedere alla console di AWS Parallel Computing Service, è necessario disporre di un set minimo di autorizzazioni. Queste autorizzazioni devono consentirti di elencare e visualizzare i dettagli sulle AWS PCS risorse del tuo. Account AWS Se crei una policy basata sull'identità più restrittiva rispetto alle autorizzazioni minime richieste, la console non funzionerà nel modo previsto per le entità (utenti o ruoli) associate a tale policy.

Non è necessario concedere autorizzazioni minime per la console agli utenti che effettuano chiamate solo verso il AWS CLI o il. AWS API Consenti invece l'accesso solo alle azioni che corrispondono all'APIoperazione che stanno cercando di eseguire.

Per ulteriori informazioni sulle autorizzazioni minime richieste per utilizzare la AWS PCS console, consulta [Autorizzazioni minime per AWS PCS](#).

Consentire agli utenti di visualizzare le loro autorizzazioni

Questo esempio mostra come è possibile creare una politica che consenta IAM agli utenti di visualizzare le politiche in linea e gestite allegate alla loro identità utente. Questa politica include le autorizzazioni per completare questa azione sulla console o utilizzando o a livello di codice. AWS CLI
AWS API

```
{
```

```
"Version": "2012-10-17",
"Statement": [
  {
    "Sid": "ViewOwnUserInfo",
    "Effect": "Allow",
    "Action": [
      "iam:GetUserPolicy",
      "iam:ListGroupsWithUser",
      "iam:ListAttachedUserPolicies",
      "iam:ListUserPolicies",
      "iam:GetUser"
    ],
    "Resource": ["arn:aws:iam::*:user/${aws:username}"]
  },
  {
    "Sid": "NavigateInConsole",
    "Effect": "Allow",
    "Action": [
      "iam:GetGroupPolicy",
      "iam:GetPolicyVersion",
      "iam:GetPolicy",
      "iam:ListAttachedGroupPolicies",
      "iam:ListGroupPolicies",
      "iam:ListPolicyVersions",
      "iam:ListPolicies",
      "iam:ListUsers"
    ],
    "Resource": "*"
  }
]
```


AWS politiche gestite per AWS Parallel Computing Service

Una politica AWS gestita è una politica autonoma creata e amministrata da AWS. AWS le politiche gestite sono progettate per fornire autorizzazioni per molti casi d'uso comuni, in modo da poter iniziare ad assegnare autorizzazioni a utenti, gruppi e ruoli.

Tieni presente che le policy AWS gestite potrebbero non concedere le autorizzazioni con il privilegio minimo per i tuoi casi d'uso specifici, poiché sono disponibili per tutti i clienti. AWS Ti consigliamo pertanto di ridurre ulteriormente le autorizzazioni definendo [policy gestite dal cliente](#) specifiche per i tuoi casi d'uso.

Non è possibile modificare le autorizzazioni definite nelle politiche gestite. AWS Se AWS aggiorna le autorizzazioni definite in una politica AWS gestita, l'aggiornamento ha effetto su tutte le identità principali (utenti, gruppi e ruoli) a cui è associata la politica. AWS è più probabile che aggiorni una policy AWS gestita quando ne Servizio AWS viene lanciata una nuova o quando diventano disponibili nuove API operazioni per i servizi esistenti.

Per ulteriori informazioni, consulta [le politiche AWS gestite](#) nella Guida IAM per l'utente.

AWS politica gestita: AWSPCSServiceRolePolicy

Non puoi collegarti AWSPCSServiceRolePolicy alle tue IAM entità. Questa policy è associata a un ruolo collegato al servizio che consente di eseguire azioni AWS PCS per conto dell'utente. Per ulteriori informazioni, consulta [Ruoli collegati ai servizi per l' AWS PCS](#).

Dettagli dell'autorizzazione

Questa policy include le seguenti autorizzazioni:

- `ec2`— Consente AWS PCS di creare e gestire EC2 risorse Amazon.
- `iam`— Consente di AWS PCS creare un ruolo collegato ai servizi per la EC2 flotta Amazon e di passare il ruolo ad Amazon. EC2
- `cloudwatch`— Consente di AWS PCS pubblicare le metriche del servizio su Amazon CloudWatch.
- `secretsmanager`— Consente di AWS PCS gestire i segreti per le risorse AWS PCS del cluster.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "PermissionsToCreatePCSNetworkInterfaces",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": "arn:aws:ec2:*:*:network-interface/*",
      "Condition": {
        "Null": {
          "aws:RequestTag/AWSPCSManaged": "false"
        }
      }
    },
    {
      "Sid": "PermissionsToCreatePCSNetworkInterfacesInSubnet",
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:*:*:subnet/*",
        "arn:aws:ec2:*:*:security-group*"
      ]
    },
    {
      "Sid": "PermissionsToManagePCSNetworkInterfaces",
      "Effect": "Allow",
      "Action": [
        "ec2>DeleteNetworkInterface",
        "ec2:CreateNetworkInterfacePermission"
      ],
      "Resource": "arn:aws:ec2:*:*:network-interface/*",
      "Condition": {
        "Null": {
          "aws:ResourceTag/AWSPCSManaged": "false"
        }
      }
    },
    {
      "Sid": "PermissionsToDescribePCSResources",
```

```

    "Effect": "Allow",
    "Action": [
      "ec2:DescribeSubnets",
      "ec2:DescribeVpcs",
      "ec2:DescribeNetworkInterfaces",
      "ec2:DescribeLaunchTemplates",
      "ec2:DescribeLaunchTemplateVersions",
      "ec2:DescribeInstances",
      "ec2:DescribeInstanceTypes",
      "ec2:DescribeInstanceStatus",
      "ec2:DescribeInstanceAttribute",
      "ec2:DescribeSecurityGroups",
      "ec2:DescribeKeyPairs",
      "ec2:DescribeImages",
      "ec2:DescribeImageAttribute"
    ],
    "Resource": "*"
  },
  {
    "Sid": "PermissionsToCreatePCSLaunchTemplates",
    "Effect": "Allow",
    "Action": [
      "ec2:CreateLaunchTemplate"
    ],
    "Resource": "arn:aws:ec2:*:*:launch-template/*",
    "Condition": {
      "Null": {
        "aws:RequestTag/AWSPCSManaged": "false"
      }
    }
  },
  {
    "Sid": "PermissionsToManagePCSLaunchTemplates",
    "Effect": "Allow",
    "Action": [
      "ec2>DeleteLaunchTemplate",
      "ec2>DeleteLaunchTemplateVersions",
      "ec2>CreateLaunchTemplateVersion"
    ],
    "Resource": "arn:aws:ec2:*:*:launch-template/*",
    "Condition": {
      "Null": {
        "aws:ResourceTag/AWSPCSManaged": "false"
      }
    }
  }
}

```

```

    }
  },
  {
    "Sid": "PermissionsToTerminatePCSMangedInstances",
    "Effect": "Allow",
    "Action": [
      "ec2:TerminateInstances"
    ],
    "Resource": "arn:aws:ec2:*:*:instance/*",
    "Condition": {
      "Null": {
        "aws:ResourceTag/AWSPCSManaged": "false"
      }
    }
  },
  {
    "Sid": "PermissionsToPassRoleToEC2",
    "Effect": "Allow",
    "Action": "iam:PassRole",
    "Resource": [
      "arn:aws:iam:*:*:role/*/AWSPCS*",
      "arn:aws:iam:*:*:role/AWSPCS*",
      "arn:aws:iam:*:*:role/aws-pcs/*",
      "arn:aws:iam:*:*:role/*/aws-pcs*"
    ],
    "Condition": {
      "StringEquals": {
        "iam:PassedToService": [
          "ec2.amazonaws.com"
        ]
      }
    }
  },
  {
    "Sid": "PermissionsToControlClusterInstanceAttributes",
    "Effect": "Allow",
    "Action": [
      "ec2:RunInstances",
      "ec2:CreateFleet"
    ],
    "Resource": [
      "arn:aws:ec2:*:*:image/*",
      "arn:aws:ec2:*:*:snapshot/*",
      "arn:aws:ec2:*:*:subnet*"
    ]
  }
}

```

```

        "arn:aws:ec2:*:*:network-interface/*",
        "arn:aws:ec2:*:*:security-group/*",
        "arn:aws:ec2:*:*:volume/*",
        "arn:aws:ec2:*:*:key-pair/*",
        "arn:aws:ec2:*:*:launch-template/*",
        "arn:aws:ec2:*:*:placement-group/*",
        "arn:aws:ec2:*:*:capacity-reservation/*",
        "arn:aws:resource-groups:*:*:group/*",
        "arn:aws:ec2:*:*:fleet/*"
    ]
},
{
    "Sid": "PermissionsToProvisionClusterInstances",
    "Effect": "Allow",
    "Action": [
        "ec2:RunInstances",
        "ec2:CreateFleet"
    ],
    "Resource": [
        "arn:aws:ec2:*:*:instance/*"
    ],
    "Condition": {
        "Null": {
            "aws:RequestTag/AWSPCSManaged": "false"
        }
    }
},
{
    "Sid": "PermissionsToTagPCSResources",
    "Effect": "Allow",
    "Action": [
        "ec2:CreateTags"
    ],
    "Resource": [
        "*"
    ],
    "Condition": {
        "StringEquals": {
            "ec2:CreateAction": [
                "RunInstances",
                "CreateLaunchTemplate",
                "CreateFleet",
                "CreateNetworkInterface"
            ]
        }
    }
}

```

```

    }
  },
  {
    "Sid": "PermissionsToPublishMetrics",
    "Effect": "Allow",
    "Action": "cloudwatch:PutMetricData",
    "Resource": "*",
    "Condition": {
      "StringEquals": {
        "cloudwatch:namespace": "AWS/PCS"
      }
    }
  },
  {
    "Sid": "PermissionsToManageSecret",
    "Effect": "Allow",
    "Action": [
      "secretsmanager:DescribeSecret",
      "secretsmanager:GetSecretValue",
      "secretsmanager:PutSecretValue",
      "secretsmanager:UpdateSecretVersionStage",
      "secretsmanager>DeleteSecret"
    ],
    "Resource": "arn:aws:secretsmanager:*:*:secret:pcs!*",
    "Condition": {
      "StringEquals": {
        "secretsmanager:ResourceTag/aws:secretsmanager:owningService":
"pcs",
        "aws:ResourceAccount": "${aws:PrincipalAccount}"
      }
    }
  }
]
}

```

AWS PCS aggiornamenti alle politiche AWS gestite

Visualizza i dettagli sugli aggiornamenti delle politiche AWS gestite AWS PCS da quando questo servizio ha iniziato a tenere traccia di queste modifiche. Per ricevere avvisi automatici sulle modifiche a questa pagina, iscriviti al RSS feed nella pagina della cronologia dei AWS PCS documenti.

Modifica	Descrizione	Data
AWS PCS ha iniziato il rilevamento delle modifiche	AWS PCS ha iniziato a tenere traccia delle modifiche per le sue politiche AWS gestite.	28 agosto 2024

Ruoli collegati ai servizi per l' AWS PCS

AWS Parallel Computing Service utilizza AWS Identity and Access Management (IAM) ruoli collegati [ai servizi](#). Un ruolo collegato al servizio è un tipo unico di IAM ruolo a cui è collegato direttamente. AWS PCS I ruoli collegati ai servizi sono predefiniti AWS PCS e includono tutte le autorizzazioni richieste dal servizio per chiamare altri servizi per conto dell'utente. AWS

Un ruolo collegato al servizio semplifica la configurazione AWS PCS perché non è necessario aggiungere manualmente le autorizzazioni necessarie. AWS PCS definisce le autorizzazioni dei ruoli collegati ai servizi e, se non diversamente definito, solo può assumerne i ruoli. AWS PCS Le autorizzazioni definite includono la politica di attendibilità e la politica di autorizzazione e tale politica di autorizzazione non può essere associata a nessun'altra entità. IAM

È possibile eliminare un ruolo collegato ai servizi solo dopo aver eliminato le risorse correlate. In questo modo proteggi AWS PCS le tue risorse perché non puoi rimuovere accidentalmente l'autorizzazione ad accedere alle risorse.

Per informazioni su altri servizi che supportano i ruoli collegati ai servizi, consulta [AWS i servizi che funzionano con IAM](#) e cerca i servizi con Sì nella colonna Ruoli collegati ai servizi. Scegli Sì in corrispondenza di un link per visualizzare la documentazione relativa al ruolo collegato ai servizi per tale servizio.

Autorizzazioni di ruolo collegate ai servizi per AWS PCS

AWS PCS utilizza il ruolo collegato al servizio denominato AWSServiceRoleForPCS: Consenti di AWS PCS gestire le risorse AmazonEC2.

Il ruolo AWSServiceRoleForPCS collegato al servizio prevede che i seguenti servizi assumano il ruolo:

- `pcs.amazonaws.com`

La politica di autorizzazione dei ruoli denominata [AWSPCSServiceRolePolicy](#) consente di AWS PCS completare azioni su risorse specifiche.

Per consentire a utenti, gruppi o ruoli di creare, modificare o eliminare un ruolo orientato ai servizi, devi configurare le autorizzazioni. Per ulteriori informazioni, consulta le [autorizzazioni dei ruoli collegati al servizio](#) nella Guida per l'utente. IAM

Creazione di un ruolo collegato al servizio per AWS PCS

Non è necessario creare manualmente un ruolo collegato al servizio. AWS PCS crea automaticamente un ruolo collegato ai servizi quando crei un cluster.

Modifica di un ruolo collegato al servizio per AWS PCS

AWS PCS non consente di modificare il ruolo collegato al `AWSServiceRoleForPCS` servizio. Dopo aver creato un ruolo collegato al servizio, non potrai modificarne il nome perché varie entità potrebbero farvi riferimento. Tuttavia, è possibile modificare la descrizione del ruolo utilizzando IAM. Per ulteriori informazioni, consulta [Modifica di un ruolo collegato al servizio nella Guida](#) per l'IAM utente.

Eliminazione di un ruolo collegato al servizio per AWS PCS

Se non è più necessario utilizzare una funzionalità o un servizio che richiede un ruolo collegato al servizio, 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 al servizio prima di poterlo eliminare manualmente.

Note

Se il AWS PCS servizio utilizza il ruolo quando si tenta di eliminare le risorse, l'eliminazione potrebbe non riuscire. In questo caso, attendi alcuni minuti e quindi ripeti l'operazione.

Per rimuovere AWS PCS le risorse utilizzate da `AWSServiceRoleForPCS`

È necessario eliminare tutti i cluster per eliminare il ruolo collegato al `AWSServiceRoleForPCS` servizio. Per ulteriori informazioni, consulta [Eliminare](#) un cluster.

Per eliminare manualmente il ruolo collegato al servizio utilizzando IAM

Usa la IAM console AWS CLI, o il AWS API per eliminare il ruolo collegato al AWSServiceRoleForPCS servizio. Per ulteriori informazioni, vedere [Eliminazione di un ruolo collegato al servizio nella Guida per l'utente](#). IAM

Regioni supportate per i ruoli collegati ai servizi AWS PCS

AWS PCSsupporta l'utilizzo di ruoli collegati al servizio in tutte le regioni in cui il servizio è disponibile. Per ulteriori informazioni, consulta [AWS Regioni ed endpoint](#).

Ruolo di Amazon EC2 Spot per AWS PCS

Se desideri creare un gruppo di nodi di AWS PCS calcolo che utilizzi Spot come opzione di acquisto, devi avere anche il ruolo AWSServiceRoleForEC2Spotcollegato al servizio. Account AWSÈ possibile utilizzare il seguente AWS CLI comando per creare il ruolo. Per ulteriori informazioni, consulta [Creare un ruolo collegato al servizio e Creare un ruolo per delegare le autorizzazioni a un AWS servizio nella Guida per l'utente](#).AWS Identity and Access Management

```
aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

Note

Se hai Account AWS già un ruolo, ricevi il seguente errore. AWSServiceRoleForEC2Spot IAM

```
An error occurred (InvalidInput) when calling the CreateServiceLinkedRole operation: Service role name AWSServiceRoleForEC2Spot has been taken in this account, please try a different suffix.
```

Autorizzazioni minime per AWS PCS

Questa sezione descrive le IAM autorizzazioni minime richieste per un'IAMidentità (utente, gruppo o ruolo) per utilizzare il servizio.

Indice

- [Autorizzazioni minime per utilizzare le azioni API](#)

- [Autorizzazioni minime richieste per utilizzare i tag](#)
- [Autorizzazioni minime richieste per supportare i log](#)
- [Autorizzazioni minime per un amministratore del servizio](#)

Autorizzazioni minime per utilizzare le azioni API

API azione	Autorizzazioni minime	Autorizzazioni aggiuntive per la console
CreateCluster	<pre>ec2:CreateNetworkInterface, ec2:DescribeVpcs, ec2:DescribeSubnets, ec2:DescribeSecurityGroups, ec2:GetSecurityGroupsForVpc, iam:CreateServiceLinkedRole, secretsmanager:CreateSecret, secretsmanager:TagResource, pcs:CreateCluster</pre>	
ListClusters	<pre>pcs:ListClusters</pre>	
GetCluster	<pre>pcs:GetCluster</pre>	<pre>ec2:DescribeSubnets</pre>
DeleteCluster	<pre>pcs>DeleteCluster</pre>	
CreateComputeNodeGroup	<pre>ec2:DescribeVpcs, ec2:DescribeSubnets, ec2:DescribeSecurityGroups, ec2:DescribeLaunchTemplates,</pre>	<pre>iam:ListInstanceProfiles, ec2:DescribeImages, pcs:GetCluster</pre>

API/Azione	Autorizzazioni minime	Autorizzazioni aggiuntive per la console
	ec2:DescribeLaunchTemplateVersions, ec2:DescribeInstanceTypes, ec2:RunInstances, ec2:CreateFleet, ec2:CreateTags, iam:PassRole, iam:GetInstanceProfile, pcs:CreateComputeNodeGroup	
ListComputerNodeGroups	pcs:ListComputeNodeGroups	pcs:GetCluster
GetComputeNodeGroup	pcs:GetComputeNodeGroup	ec2:DescribeSubnets
UpdateComputeNodeGroup	ec2:DescribeVpcs, ec2:DescribeSubnets, ec2:DescribeSecurityGroups, ec2:DescribeLaunchTemplates, ec2:DescribeLaunchTemplateVersions, ec2:DescribeInstanceTypes, ec2:RunInstances, ec2:CreateFleet, ec2:CreateTags, iam:PassRole, iam:GetInstanceProfile, pcs:UpdateComputeNodeGroup	pcs:GetComputeNodeGroup, iam:ListInstanceProfiles, ec2:DescribeImages, pcs:GetCluster

APIazione	Autorizzazioni minime	Autorizzazioni aggiuntive per la console
DeleteComputeNodeGroup	<code>pcs:DeleteComputeNodeGroup</code>	
CreateQueue	<code>pcs:CreateQueue</code>	<code>pcs:ListComputeNodeGroups</code> , <code>pcs:GetCluster</code>
ListQueues	<code>pcs:ListQueues</code>	<code>pcs:GetCluster</code>
GetQueue	<code>pcs:GetQueue</code>	
UpdateQueue	<code>pcs:UpdateQueue</code>	<code>pcs:ListComputeNodeGroups</code> , <code>pcs:GetQueue</code>
DeleteQueue	<code>pcs>DeleteQueue</code>	

Autorizzazioni minime richieste per utilizzare i tag

Le seguenti autorizzazioni sono necessarie per utilizzare i tag con le risorse in AWS PCS

```
pcs:ListTagsForResource
pcs:TagResource
pcs:UntagResource
```

Autorizzazioni minime richieste per supportare i log

AWS PCS invia i dati di registro ad Amazon CloudWatch Logs (CloudWatch Logs). Devi assicurarti che la tua identità disponga delle autorizzazioni minime per utilizzare Logs. CloudWatch Per ulteriori informazioni, consulta [Panoramica della gestione delle autorizzazioni di accesso alle risorse CloudWatch Logs](#) nella Amazon CloudWatch Logs User Guide.

Per informazioni sulle autorizzazioni richieste a un servizio per inviare log a CloudWatch Logs, consulta [Enabling logging from services AWS nella Amazon CloudWatch Logs User Guide](#).

Autorizzazioni minime per un amministratore del servizio

La seguente IAM politica specifica le autorizzazioni minime richieste per un'IAM identità (utente, gruppo o ruolo) per configurare e gestire il servizio. AWS PCS

Note

Gli utenti che non configurano e gestiscono il servizio non richiedono queste autorizzazioni. Gli utenti che eseguono solo processi utilizzano secure shell (SSH) per connettersi al cluster. AWS Identity and Access Management (IAM) non gestisce l'autenticazione o l'autorizzazione per SSH.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface",
        "ec2:DescribeImages",
        "ec2:DescribeInstanceTypes",
        "ec2:DescribeLaunchTemplates",
        "ec2:DescribeLaunchTemplateVersions",
        "ec2:DescribeSecurityGroups",
        "ec2:DescribeSubnets",
        "ec2:DescribeVpcs",
        "ec2:GetSecurityGroupsForVpc",
        "firehose:*",
        "iam:GetInstanceProfile",
        "iam:ListInstanceProfiles",
        "iam:PassRole",
        "kms:*",
        "logs:*",
        "pcs:*",
        "s3:*"
      ],
      "Resource": "*"
    }
  ]
}
```

```
    }  
  ]  
}
```

È possibile escludere le seguenti autorizzazioni dalla politica e utilizzare invece la politica gestita corrispondente in IAM:

- "firehose:*"

AmazonKinesisFirehoseFullAccess

- "kms:*"

AWSKeyManagementServicePowerUser

- "logs:*"

CloudWatchLogsFullAccess

- "s3:*"

AmazonS3FullAccess

IAM profili di istanza per AWS Parallel Computing Service

Le applicazioni eseguite su un'EC2 istanza devono includere AWS credenziali in tutte AWS API le richieste effettuate. Ti consigliamo di utilizzare un IAM ruolo per gestire le credenziali temporanee sull'EC2 istanza. A tale scopo, puoi definire un profilo di istanza e collegarlo alle tue istanze. Per ulteriori informazioni, consulta [IAMi ruoli per Amazon EC2 nella Amazon](#) Elastic Compute Cloud User Guide.

Note

Quando usi per AWS Management Console creare un IAM ruolo per AmazonEC2, la console crea automaticamente un profilo di istanza e gli assegna lo stesso nome del IAM ruolo. Se utilizzi le AWS CLI AWS API azioni o an AWS SDK per creare il IAM ruolo, crei il profilo dell'istanza come azione separata. Per ulteriori informazioni, consulta [Profili di istanza](#) nella Guida per l'utente di Amazon Elastic Compute Cloud.

È necessario specificare il profilo ARN di un'istanza quando si creano gruppi di nodi di calcolo. È possibile scegliere diversi profili di istanza per alcuni o tutti i gruppi di nodi di calcolo.

Requisiti del profilo di istanza

Nome del profilo dell'istanza

Il profilo dell'istanza ARN deve iniziare con `AWSPCS` o contenere `/aws-pcs/` nel suo percorso.

Example

- `arn:aws:iam::*:instance-profile/AWSPCS-example-role-1` e
- `arn:aws:iam::*:instance-profile/aws-pcs/example-role-2`.

Autorizzazioni

Come minimo, il profilo dell'istanza di AWS PCS deve includere la seguente politica. Consente ai nodi di elaborazione di notificare il AWS PCS servizio quando diventano operativi.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "pcs:RegisterComputeNodeGroupInstance"
      ],
      "Resource": "*",
      "Effect": "Allow"
    }
  ]
}
```

Politiche aggiuntive

Potresti prendere in considerazione l'aggiunta di politiche gestite al profilo dell'istanza. Per esempio:

- [AmazonS3 ReadOnlyAccess](#) fornisce accesso in sola lettura a tutti i bucket S3.
- [AmazonSSMManaged InstanceCore](#) abilita le funzionalità principali del servizio AWS Systems Manager, come l'accesso remoto direttamente dalla console di gestione Amazon.
- [CloudWatchAgentServerPolicy](#) contiene le autorizzazioni necessarie per l'uso AmazonCloudWatchAgent sui server.

Puoi anche includere IAM politiche personalizzate che supportano il tuo caso d'uso specifico.

Creazione di un profilo dell'istanza

Puoi creare un profilo di istanza direttamente dalla EC2 console Amazon. Per ulteriori informazioni, consulta [Usare i profili di istanza](#) nella Guida AWS Identity and Access Management per l'utente.

Risoluzione dei problemi di identità e accesso al AWS Parallel Computing Service

Utilizza le seguenti informazioni per aiutarti a diagnosticare e risolvere i problemi più comuni che potresti riscontrare quando lavori con AWS PCS e IAM.

Argomenti

- [Non sono autorizzato a eseguire alcuna azione in AWS PCS](#)
- [Non sono autorizzato a eseguire iam: PassRole](#)
- [Voglio consentire a persone esterne a me di accedere Account AWS alle mie AWS PCS risorse](#)

Non sono autorizzato a eseguire alcuna azione in AWS PCS

Se ricevi un errore che indica che non disponi dell'autorizzazione per eseguire un'operazione, le tue policy devono essere aggiornate in modo che ti sei consentito eseguire tale operazione.

L'errore di esempio seguente si verifica quando l'utente `mateojacksonIAMutente` tenta di utilizzare la console per visualizzare i dettagli su una `my-example-widget` risorsa fittizia ma non dispone delle autorizzazioni fittizie `pcs:GetWidget`.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
pcs:GetWidget on resource: my-example-widget
```

In questo caso, la policy per l'utente `mateojackson` deve essere aggiornata per consentire l'accesso alla risorsa `my-example-widget` utilizzando l'azione `pcs:GetWidget`.

Se hai bisogno di assistenza, contatta l'amministratore. AWS L'amministratore è la persona che ti ha fornito le credenziali di accesso.

Non sono autorizzato a eseguire iam: PassRole

Se ricevi un messaggio di errore indicante che non sei autorizzato a eseguire l'iam:PassRoleazione, le tue politiche devono essere aggiornate per consentirti di assegnare un ruolo a AWS PCS.

Alcuni Servizi AWS consentono di trasferire un ruolo esistente a quel servizio invece di creare un nuovo ruolo di servizio o un ruolo collegato al servizio. Per eseguire questa operazione, è necessario disporre delle autorizzazioni per trasmettere il ruolo al servizio.

L'errore di esempio seguente si verifica quando un IAM utente denominato marymajor tenta di utilizzare la console per eseguire un'azione in. AWS PCS 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.

Se hai bisogno di assistenza, contatta AWS l'amministratore. L'amministratore è la persona che ti ha fornito le credenziali di accesso.

Voglio consentire a persone esterne a me di accedere Account AWS alle mie AWS PCS risorse

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 i servizi che supportano politiche basate sulle risorse o liste di controllo degli accessi (ACLs), puoi utilizzare tali politiche per concedere alle persone l'accesso alle tue risorse.

Per ulteriori informazioni, consulta gli argomenti seguenti:

- Per sapere se AWS PCS supporta queste funzionalità, consulta. [Come funziona AWS Parallel Computing Service con IAM](#)
- Per informazioni su Account AWS come fornire l'accesso alle risorse di tua proprietà, consulta [Fornire l'accesso a un IAM utente di un altro Account AWS utente di tua proprietà](#) nella Guida per l'IAMutente.

- Per scoprire come fornire l'accesso alle tue risorse a terze parti Account AWS, consulta [Fornire l'accesso a persone Account AWS di proprietà di terzi](#) nella Guida per l'IAMutente.
- Per informazioni su come fornire l'accesso tramite la federazione delle identità, consulta [Fornire l'accesso agli utenti autenticati esternamente \(federazione delle identità\)](#) nella Guida per l'IAMutente.
- Per conoscere la differenza tra l'utilizzo di ruoli e politiche basate sulle risorse per l'accesso tra account diversi, consulta la sezione Accesso alle [risorse tra account nella Guida per l'utente](#). IAM IAM

Convalida della conformità per Parallel Computing Service AWS

Per sapere se un Servizio AWS programma rientra nell'ambito di specifici programmi di conformità, consulta Servizi AWS la sezione [Scope by Compliance Program Servizi AWS](#) e scegli il programma di conformità che ti interessa. Per informazioni generali, consulta Programmi di [AWS conformità Programmi](#) di di .

È possibile scaricare report di audit di terze parti utilizzando AWS Artifact. Per ulteriori informazioni, consulta [Scaricamento dei report in AWS Artifact](#) .

La vostra responsabilità di conformità durante l'utilizzo Servizi AWS è determinata dalla sensibilità dei dati, dagli obiettivi di conformità dell'azienda e dalle leggi e dai regolamenti applicabili. AWS fornisce le seguenti risorse per contribuire alla conformità:

- [Guide introduttive su sicurezza e conformità](#): queste guide all'implementazione illustrano considerazioni sull'architettura e forniscono passaggi per implementare ambienti di base incentrati sulla AWS sicurezza e la conformità.
- [Architettura per la HIPAA sicurezza e la conformità su Amazon Web Services](#): questo white paper descrive in che modo le aziende possono utilizzare AWS per creare applicazioni idonee. HIPAA

Note

Non tutte sono idonee. Servizi AWS HIPAA Per ulteriori informazioni, consulta la [Guida ai servizi HIPAA idonei](#).

- [AWS Risorse per la conformità](#): questa raccolta di cartelle di lavoro e guide potrebbe riguardare il settore e la località in cui operi.

- [AWS Guide alla conformità dei clienti](#): comprendi il modello di responsabilità condivisa attraverso la lente della conformità. Le guide riassumono le migliori pratiche per la protezione Servizi AWS e mappano le linee guida per i controlli di sicurezza su più framework (tra cui il National Institute of Standards and Technology (NIST), il Payment Card Industry Security Standards Council (PCI) e l'International Organization for Standardization ()). ISO
- [Evaluating Resources with Rules](#) nella Guida per gli AWS Config sviluppatori: il AWS Config servizio valuta la conformità delle configurazioni delle risorse alle pratiche interne, alle linee guida del settore e alle normative.
- [AWS Security Hub](#)— Ciò Servizio AWS fornisce una visione completa dello stato di sicurezza interno. AWS La Centrale di sicurezza utilizza i controlli di sicurezza per valutare le risorse AWS e verificare la conformità agli standard e alle best practice del settore della sicurezza. Per un elenco dei servizi e dei controlli supportati, consulta la pagina [Documentazione di riferimento sui controlli della Centrale di sicurezza](#).
- [Amazon GuardDuty](#): Servizio AWS rileva potenziali minacce ai tuoi carichi di lavoro Account AWS, ai contenitori e ai dati monitorando l'ambiente alla ricerca di attività sospette e dannose. GuardDuty può aiutarti a soddisfare vari requisiti di conformità, ad esempio PCI DSS soddisfacendo i requisiti di rilevamento delle intrusioni imposti da determinati framework di conformità.
- [AWS Audit Manager](#)— Ciò Servizio AWS consente di verificare continuamente AWS l'utilizzo per semplificare la gestione del rischio e la conformità alle normative e agli standard di settore.

Resilienza nel servizio di elaborazione AWS parallela

L'infrastruttura AWS globale è costruita attorno a Regioni AWS zone di disponibilità. Regioni AWS forniscono più zone di disponibilità fisicamente separate e isolate, collegate con reti a bassa latenza, ad alto throughput e altamente ridondanti. Con le zone di disponibilità, puoi progettare e gestire applicazioni e database che eseguono automaticamente il failover 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 tradizionali.

[Per ulteriori informazioni sulle zone di disponibilità, vedere Global Regioni AWS Infrastructure.AWS](#)

Servizio di sicurezza dell'infrastruttura nel servizio di elaborazione AWS parallela

In quanto servizio gestito, AWS Parallel Computing Service è protetto dalla sicurezza di rete AWS globale. Per informazioni sui servizi AWS di sicurezza e su come AWS protegge l'infrastruttura, consulta [AWS Cloud Security](#). Per progettare il tuo AWS ambiente utilizzando le migliori pratiche per la sicurezza dell'infrastruttura, vedi [Infrastructure Protection](#) in Security Pillar AWS Well-Architected Framework.

Si utilizzano API chiamate AWS pubblicate per accedere tramite AWS PCS la rete. I client devono supportare quanto segue:

- Transport Layer Security (TLS). Richiediamo TLS 1.2 e consigliamo TLS 1.3.
- Suite di cifratura con Perfect Forward Secrecy (PFS) come (Ephemeral Diffie-Hellman) o DHE (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 O puoi utilizzare [AWS Security Token Service](#) (AWS STS) per generare credenziali di sicurezza temporanee per sottoscrivere le richieste.

Quando AWS PCS crea un cluster, il servizio avvia il controller Slurm in un account di proprietà del servizio, separato dai nodi di elaborazione dell'account. Per collegare la comunicazione tra il controller e i nodi di elaborazione, AWS PCS crea un'interfaccia di rete elastica tra account () nel tuo ENI VPC Il controller Slurm lo utilizza ENI per gestire e comunicare con i nodi di calcolo tra diversi nodi Account AWS, mantenendo la sicurezza e l'isolamento delle risorse e facilitando al contempo le operazioni AI/ML efficienti. HPC

Analisi e gestione delle vulnerabilità in Parallel Computing Service AWS

La configurazione e i controlli IT sono una responsabilità condivisa tra te AWS e te. Per ulteriori informazioni, consulta il [modello di responsabilitàAWS condivisa](#). AWS gestisce le attività di sicurezza di base per l'infrastruttura sottostante nell'account di servizio, come l'applicazione di patch al sistema operativo sulle istanze del controller, la configurazione del firewall e il ripristino di emergenza AWS

dell'infrastruttura. Queste procedure sono state riviste e certificate dalle terze parti appropriate. Per ulteriori dettagli, consulta [Best practice per la sicurezza, l'identità e la conformità](#).

Sei responsabile della sicurezza dell'infrastruttura sottostante nei tuoi Account AWS:

- Mantieni il codice, inclusi aggiornamenti e patch di sicurezza.
- Patch e aggiorna il sistema operativo sulle istanze del gruppo di nodi.
- Aggiorna lo scheduler per mantenerlo all'interno delle versioni supportate.
- Autentica e crittografa le comunicazioni tra i client utente e i nodi a cui si connettono.

Prevenzione del confused deputy tra servizi

Con "confused deputy" si intende un problema di sicurezza in cui un'entità che non dispone dell'autorizzazione per eseguire una certa operazione può costringere un'entità con più privilegi a eseguire tale operazione. Nel frattempo AWS, l'impersonificazione tra servizi può portare al confuso problema del vicesceriffo. La rappresentazione tra servizi può verificarsi quando un servizio (il servizio chiamante) effettua una chiamata a un altro servizio (il servizio chiamato). Il servizio chiamante può essere manipolato per utilizzare le proprie autorizzazioni e agire sulle risorse di un altro cliente, a cui normalmente non avrebbe accesso. Per evitare ciò, AWS fornisce strumenti per poterti a proteggere i tuoi dati per tutti i servizi con entità di servizio a cui è stato concesso l'accesso alle risorse del tuo account.

Si consiglia di utilizzare le chiavi di contesto [aws:SourceArn](#) [aws:SourceAccount](#) global condition nelle politiche delle risorse per limitare le autorizzazioni che AWS Parallel Computing Service (AWS PCS) concede a un altro servizio alla risorsa. Utilizza `aws:SourceArn` se desideri consentire l'associazione di una sola risorsa all'accesso tra servizi. Utilizza `aws:SourceAccount` se desideri consentire l'associazione di qualsiasi risorsa in tale account all'uso tra servizi.

Il modo più efficace per proteggersi dal confuso problema del vice consiste nell'utilizzare la chiave di contesto ARN della condizione `aws:SourceArn` globale con l'intera risorsa. Se non conosci la dimensione completa ARN della risorsa o se stai specificando più risorse, usa la chiave `aws:SourceArn` global context condition con caratteri jolly (*) per le parti sconosciute di ARN. Ad esempio `arn:aws:service:*:123456789012:*`.

Se il `aws:SourceArn` valore non contiene l'ID dell'account, ad esempio un bucket Amazon S3ARN, devi utilizzare entrambe le chiavi di contesto della condizione globale per limitare le autorizzazioni.

Il valore di `aws:SourceArn` deve essere un cluster. ARN

L'esempio seguente mostra come utilizzare le chiavi di contesto `aws:SourceArn` e `aws:SourceAccount` global condition AWS PCS per evitare il confuso problema del vice.

```
{
  "Version": "2012-10-17",
  "Statement": {
    "Sid": "ConfusedDeputyPreventionExamplePolicy",
    "Effect": "Allow",
    "Principal": {
      "Service": "pcs.amazonaws.com"
    },
    "Action": "sts:AssumeRole",
    "Condition": {
      "ArnLike": {
        "aws:SourceArn": [
          "arn:aws:pcs:us-east-1:123456789012:cluster/*"
        ]
      },
      "StringEquals": {
        "aws:SourceAccount": "123456789012"
      }
    }
  }
}
```

IAM ruolo per EC2 le istanze Amazon fornite come parte di un gruppo di nodi di calcolo

AWS PCS orchestra automaticamente la EC2 capacità di Amazon per ciascuno dei gruppi di nodi di calcolo configurati in un cluster. Quando creano un gruppo di nodi di calcolo, gli utenti devono fornire un profilo di IAM istanza tramite il campo `iamInstanceProfileArn`. Il profilo dell'istanza specifica le autorizzazioni associate alle istanze assegnate. EC2 AWS PCS accetta qualsiasi ruolo che abbia `AWSPCS` come prefisso il nome del ruolo o `/aws-pcs/` come parte del percorso del ruolo. L'`iam:PassRole` autorizzazione è richiesta sull'IAM identità (utente o ruolo) che crea o aggiorna un gruppo di nodi di calcolo. Quando un utente chiama le `UpdateComputeNodeGroup` API azioni `CreateComputeNodeGroup` o, AWS PCS verifica se l'utente è autorizzato a eseguire l'`iam:PassRole` azione.

La politica di esempio seguente concede le autorizzazioni per l'assegnazione solo IAM dei ruoli il cui nome inizia con `AWSPCS`

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": "iam:PassRole",
      "Resource": "arn:aws:iam::123456789012:role/AWSPCS*",
      "Condition": {
        "StringEquals": {
          "iam:PassedToService": [
            "ec2.amazonaws.com"
          ]
        }
      }
    }
  ]
}
```

Best practice di sicurezza per AWS Parallel Computing Service

Questa sezione descrive le migliori pratiche di sicurezza specifiche di AWS Parallel Computing Service (AWS PCS). Per ulteriori informazioni sulle best practice di sicurezza in AWS, consulta [Best practice for Security, Identity and Compliance](#).

AMI-sicurezza correlata

- Non utilizzare AWS PCS sample AMIs per carichi di lavoro di produzione. I campioni non AMIs sono supportati e sono destinati esclusivamente ai test.
- Aggiorna regolarmente il sistema operativo e il software delle AWS PCS istanze per mitigare le vulnerabilità.
- Utilizzatelo AWS Systems Manager per automatizzare l'applicazione delle patch e mantenere la conformità con le vostre politiche di sicurezza.
- Utilizza solo AWS PCS pacchetti ufficiali autenticati scaricati da fonti ufficiali. AWS
- Aggiorna regolarmente AWS PCS i pacchetti sui nodi di calcolo per ricevere patch e miglioramenti di sicurezza. Valuta la possibilità di automatizzare questo processo per ridurre al minimo le vulnerabilità.

Sicurezza di Slurm Workload Manager

- Implementa i controlli di accesso e le restrizioni di rete per proteggere i nodi di controllo e calcolo di Slurm. Consenti solo a utenti e sistemi affidabili di inviare lavori e accedere ai comandi di gestione Slurm.
- Utilizza le funzionalità di sicurezza integrate di Slurm, come l'autenticazione Slurm, per garantire che gli invii di lavori e le comunicazioni siano autenticati.
- Aggiorna le versioni di Slurm per mantenere operazioni fluide e supporto per i cluster.

Important

Qualsiasi cluster che utilizza una versione di Slurm che ha raggiunto la fine del periodo di supporto (EOSL) viene interrotto immediatamente. Usa il link nella parte superiore delle pagine della guida per l'utente per iscriverti al RSS feed di AWS PCS documentazione per ricevere notifiche quando si avvicina una versione di Slurm. EOSL

Monitoraggio e registrazione

- Usa Amazon CloudWatch Logs e AWS CloudTrail per monitorare e registrare le azioni nei tuoi cluster e. Account AWS Usa i dati per la risoluzione dei problemi e il controllo.

Sicurezza di rete

- Implementa AWS PCS i cluster in un ambiente separato VPC per isolare l'HPCAmbiente dal resto del traffico di rete.
- Utilizza i gruppi di sicurezza e gli elenchi di controllo degli accessi alla rete (ACLs) per controllare il traffico in entrata e in uscita verso istanze e sottoreti. AWS PCS
- Usa i AWS PrivateLink nostri VPC endpoint per mantenere il traffico di rete tra i cluster e gli altri servizi all'interno della rete. AWS AWS

Registrazione e monitoraggio per AWS PCS

Il monitoraggio è un elemento importante per mantenere l'affidabilità, la disponibilità e le prestazioni delle AWS PCS e altre AWS risorse a disposizione. AWS fornisce i seguenti strumenti di monitoraggio per osservare AWS PCS, segnalare quando qualcosa non va e intraprendere azioni automatiche quando necessario:

- Amazon CloudWatch monitora AWS le tue risorse e le applicazioni su cui esegui AWS in tempo reale. Puoi raccogliere i parametri e tenerne traccia, creare pannelli di controllo personalizzati e impostare allarmi per inviare una notifica o intraprendere azioni quando un parametro specificato raggiunge una determinata soglia. Ad esempio, puoi tenere CloudWatch traccia CPU dell'utilizzo o di altri parametri delle tue EC2 istanze Amazon e avviare automaticamente nuove istanze quando necessario. Per ulteriori informazioni, consulta la [Amazon CloudWatch User Guide](#).
- Amazon CloudWatch Logs ti consente di monitorare, archiviare e accedere ai tuoi file di registro da EC2 istanze Amazon e altre fonti. CloudTrail CloudWatch Logs possono monitorare le informazioni contenute nei file di registro e avvisarti quando vengono raggiunte determinate soglie. Puoi inoltre archiviare i dati del log in storage estremamente durevole. Per ulteriori informazioni, consulta la [Amazon CloudWatch Logs User Guide](#).
- AWS CloudTrail acquisisce le API chiamate e gli eventi correlati effettuati da o per conto del tuo AWS account e invia i file di registro a un bucket Amazon S3 da te specificato. Puoi identificare quali utenti e account hanno chiamato AWS, l'indirizzo IP di origine da cui sono state effettuate le chiamate e quando sono avvenute le chiamate. Per ulteriori informazioni, consulta la [Guida per l'utente AWS CloudTrail](#).

AWS PCS registri dello scheduler

Puoi AWS PCS configurare l'invio di dati di registrazione dettagliati dal tuo programma di pianificazione del cluster ad Amazon CloudWatch Logs, Amazon Simple Storage Service (Amazon S3) e Amazon Data Firehose. Questo può aiutare nel monitoraggio e nella risoluzione dei problemi. È possibile configurare i registri dello AWS PCS scheduler utilizzando la AWS PCS console, nonché a livello di codice utilizzando o. AWS CLI SDK

Indice

- [Prerequisiti](#)
- [Configurazione dei registri dello scheduler utilizzando la console AWS PCS](#)

- [Configurazione dei registri dello scheduler utilizzando il AWS CLI](#)
 - [Crea una destinazione di consegna](#)
 - [Abilita il AWS PCS cluster come fonte di consegna](#)
 - [Connect l'origine di consegna del cluster alla destinazione di consegna](#)
- [Scheduler log stream, percorsi e nomi](#)
- [Esempio di record AWS PCS di log dello scheduler](#)

Prerequisiti

Il IAM principale utilizzato per gestire il cluster deve consentire. AWS PCS

`pcs:AllowVendedLogDeliveryForResource` Ecco un esempio di AWS IAM policy che lo abilita.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "PcsAllowVendedLogsDelivery",
      "Effect": "Allow",
      "Action": ["pcs:AllowVendedLogDeliveryForResource"],
      "Resource": [
        "arn:aws:pcs:::cluster/*"
      ]
    }
  ]
}
```

Configurazione dei registri dello scheduler utilizzando la console AWS PCS

Per configurare i log dello AWS PCS scheduler nella console, segui questi passaggi:

1. [Apri la console AWS PCS](#)
2. Scegli Clusters e vai alla pagina dei dettagli del AWS PCS cluster in cui abiliterai la registrazione.
3. Scegliere Logs (Log).
4. In Consegne dei log — Scheduler Logs — opzionale
 - a. Aggiungi fino a tre destinazioni di consegna dei log. Le scelte includono CloudWatch Logs, Amazon S3 o Firehose.
 - b. Scegli Aggiorna le consegne dei log.

Puoi riconfigurare, aggiungere o rimuovere le consegne di log rivisitando questa pagina.

Configurazione dei registri dello scheduler utilizzando il AWS CLI

A tale scopo, sono necessarie almeno una destinazione di consegna, un'origine di consegna (il PCS cluster) e una consegna, ovvero una relazione che collega un'origine a una destinazione.

Crea una destinazione di consegna

È necessaria almeno una destinazione di consegna per ricevere i log dello scheduler da un AWS PCS cluster. Puoi saperne di più su questo argomento nella PutDeliveryDestination sezione della Guida per l' CloudWatch APIutente.

Per creare una destinazione di consegna utilizzando il AWS CLI

- Crea una destinazione con il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:
 - Replace (Sostituisci) *region-code* con Regione AWS dove creerai la tua destinazione. In genere si tratta della stessa regione in cui viene distribuito il AWS PCS cluster.
 - Replace (Sostituisci) *pcs-logs-destination* con il tuo nome preferito. Deve essere unico per tutte le destinazioni di consegna del tuo account.
 - Replace (Sostituisci) *resource-arn* con il ARN per un gruppo di log esistente in CloudWatch Logs, un bucket S3 o un flusso di distribuzione in Firehose. Esempi includono:
 - CloudWatch Gruppo di log

```
arn:aws:logs:region-code:account-id:log-group:/log-group-name:*
```

- Bucket S3

```
arn:aws:s3:::bucket-name
```

- Flusso di distribuzione Firehose

```
arn:aws:firehose:region-code:account-id:deliverystream/stream-name
```

```
aws logs put-delivery-destination --region region-code \  
--name pcs-logs-destination \  
--delivery-destination-configuration destinationResourceArn=resource-arn
```

Prendi nota del campo ARN relativo alla nuova destinazione di consegna, poiché ti servirà per configurare le consegne.

Abilita il AWS PCS cluster come fonte di consegna

Per raccogliere i log dello scheduler AWSPCS, configura il cluster come fonte di consegna. Per ulteriori informazioni, [PutDeliverySource](#) consulta Amazon CloudWatch Logs API Reference.

Per configurare un cluster come fonte di distribuzione utilizzando il AWS CLI

- Abilita la consegna dei log dal tuo cluster con il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:
 - Replace (Sostituisci) *region-code* con il Regione AWS luogo in cui è distribuito il cluster.
 - Replace (Sostituisci) *cluster-logs-source-name* con un nome per questa fonte. Deve essere unico per tutte le fonti di consegna nel tuo Account AWS. Valuta la possibilità di incorporare il nome o l'ID del AWS PCS cluster.
 - Replace (Sostituisci) *cluster-arn* con il ARN per il tuo cluster AWS PCS

```
aws logs put-delivery-source \  
  --region region-code \  
  --name cluster-logs-source-name \  
  --resource-arn cluster-arn \  
  --log-type PCS_SCHEDULER_LOGS
```

Connect l'origine di consegna del cluster alla destinazione di consegna

Affinché i dati di log dello scheduler fluiscono dal cluster alla destinazione, è necessario configurare una consegna che li colleghi. Per ulteriori informazioni, [CreateDelivery](#) consulta Amazon CloudWatch Logs API Reference.

Per creare una consegna utilizzando il AWS CLI

- Crea una consegna utilizzando il comando che segue. Prima di eseguire il comando, apporta le modifiche seguenti:
 - Replace (Sostituisci) *region-code* con il Regione AWS luogo in cui esistono la fonte e la destinazione.
 - Replace (Sostituisci) *cluster-logs-source-name* con il nome della fonte di consegna indicato dall'alto.

- Replace (Sostituisci) *destination-arn* ARN da una destinazione di consegna in cui desideri che i registri vengano consegnati.

```
aws logs create-delivery \  
  --region region-code \  
  --delivery-source-name cluster-logs-source \  
  --delivery-destination-arn destination-arn
```

Scheduler log stream, percorsi e nomi

Il percorso e il nome dei log dello AWS PCS scheduler dipendono dal tipo di destinazione.

- CloudWatch Registri
 - Uno stream CloudWatch Logs segue questa convenzione di denominazione.

```
AWSLogs/PCS/${cluster_id}/${log_name}_${scheduler_major_version}.log
```

Example

```
AWSLogs/PCS/abcdef0123/slurmctld_24.05.log
```

- Bucket S3
 - Un percorso di output del bucket S3 segue questa convenzione di denominazione:

```
AWSLogs/${account-id}/PCS/${region}/${cluster_id}/${log_name}/  
${scheduler_major_version}/yyyy/MM/dd/HH/
```

Example

```
AWSLogs/111111111111/PCS/us-east-2/abcdef0123/slurmctld/24.05/2024/09/01/00.
```

- Il nome di un oggetto S3 segue questa convenzione:

```
PCS_${log_name}_${scheduler_major_version}_#{expr date 'event_timestamp', format:  
"yyyy-MM-dd-HH"}_${cluster_id}_${hash}.log
```

Example

```
PCS_slurmctld_24.05_2024-09-01-00_abcdef0123_0123abcdef.log
```

Esempio di record AWS PCS di log dello scheduler

AWSPCSi log dello scheduler sono strutturati. Includono campi come l'identificatore del cluster, il tipo di scheduler, le versioni principali e di patch, oltre al messaggio di registro emesso dal processo del controller Slurm. Ecco un esempio.

```
{
  "resource_id": "s3431v9rx2",
  "resource_type": "PCS_CLUSTER",
  "event_timestamp": 1721230979,
  "log_level": "info",
  "log_name": "slurmctld",
  "scheduler_type": "slurm",
  "scheduler_major_version": "23.11",
  "scheduler_patch_version": "8",
  "node_type": "controller_primary",
  "message": "[2024-07-17T15:42:58.614+00:00] Running as primary controller\n"
}
```

Servizio di monitoraggio del calcolo AWS parallelo con Amazon CloudWatch

Amazon CloudWatch fornisce il monitoraggio dello stato e delle prestazioni del cluster AWS Parallel Computing Service (AWS PCS) raccogliendo parametri dal cluster a intervalli regolari. Queste metriche vengono mantenute, consentendoti di accedere ai dati storici e ottenere informazioni dettagliate sulle prestazioni del cluster nel tempo.

CloudWatch consente inoltre di monitorare le EC2 istanze lanciate da AWS PCS per soddisfare i requisiti di scalabilità. Sebbene sia possibile controllare i log sulle istanze in esecuzione, le CloudWatch metriche e i dati di registrazione vengono in genere eliminati una volta terminate le istanze. Tuttavia, è possibile configurare l' CloudWatch agente sulle istanze utilizzando un modello di EC2 avvio per mantenere le metriche e i log anche dopo la chiusura dell'istanza, abilitando il monitoraggio e l'analisi a lungo termine.

Esplora gli argomenti di questa sezione per ulteriori informazioni sul monitoraggio e l'utilizzo. AWS PCS CloudWatch

Argomenti

- [Monitoraggio delle AWS PCS metriche tramite CloudWatch](#)
- [Monitoraggio delle AWS PCS istanze tramite Amazon CloudWatch](#)

Monitoraggio delle AWS PCS metriche tramite CloudWatch

Puoi monitorare lo stato AWS PCS del cluster utilizzando Amazon CloudWatch, che raccoglie i dati dal cluster e li trasforma in metriche quasi in tempo reale. Queste statistiche vengono conservate per un periodo di 15 mesi, in modo da poter accedere alle informazioni storiche e avere una prospettiva migliore sulle prestazioni del cluster. Le metriche del cluster vengono inviate a CloudWatch intervalli di 1 minuto. Per ulteriori informazioni su CloudWatch, consulta [What Is Amazon CloudWatch?](#) nella Amazon CloudWatch User Guide.

AWS PCS pubblica le seguenti metriche nello spazio dei PCS nomi AWS/in. CloudWatch Hanno un'unica dimensione, `ClusterId`

Nome	Descrizione	unità
ActualCapacity	IdleCapacity + UtilizedCapacity	Conteggio
CapacityUtilization	UtilizedCapacity / ActualCapacity	Conteggio
DesiredCapacity	ActualCapacity + PendingCapacity	Conteggio
IdleCapacity	Numero di istanze in esecuzione ma non assegnate ai job	Conteggio
UtilizedCapacity	Numero di istanze in esecuzione e assegnate ai job	Conteggio

Monitoraggio delle AWS PCS istanze tramite Amazon CloudWatch

AWSPCS avvia EC2 le istanze Amazon in base alle esigenze per soddisfare i requisiti di scalabilità definiti nei gruppi di nodi di PCS calcolo. Puoi monitorare queste istanze mentre sono in esecuzione utilizzando Amazon CloudWatch. Puoi controllare i log delle istanze in esecuzione accedendovi e utilizzando strumenti interattivi da riga di comando. Tuttavia, per impostazione predefinita, i dati CloudWatch delle metriche vengono conservati solo per un periodo limitato dopo la chiusura di un'istanza e i log delle istanze vengono generalmente eliminati insieme ai volumi che supportano l'istanza. EBS Per conservare le metriche o i dati di registrazione delle istanze avviate PCS dopo la loro chiusura, puoi configurare l' CloudWatch agente sulle istanze con un modello di avvio. EC2 Questo argomento fornisce una panoramica del monitoraggio delle istanze in esecuzione e fornisce esempi su come configurare i parametri e i log delle istanze persistenti.

Monitoraggio delle istanze in esecuzione

Ricerca di istanze AWS PCS

Per monitorare le istanze avviate da PCS, trova le istanze in esecuzione associate a un cluster o a un gruppo di nodi di calcolo. Quindi, nella EC2 console di una determinata istanza, controlla le sezioni Stato e allarmi e Monitoraggio. Se l'accesso di accesso è configurato per tali istanze, puoi connetterti ad esse e controllare i vari file di registro sulle istanze. Per ulteriori informazioni sull'identificazione delle istanze gestite da, consulta. PCS [Ricerca di istanze di gruppi di nodi di calcolo in AWS PCS](#)

Abilitazione di metriche dettagliate

Per impostazione predefinita, le metriche delle istanze vengono raccolte a intervalli di 5 minuti. Per raccogliere le metriche a intervalli di un minuto, abilita il CloudWatch monitoraggio dettagliato nel modello di lancio del gruppo di nodi di calcolo. Per ulteriori informazioni, consulta [Attiva il monitoraggio dettagliato CloudWatch](#).

Configurazione di metriche e log persistenti delle istanze

Puoi conservare i parametri e i log delle tue istanze installando e configurando l'agente CloudWatch Amazon su di esse. Si compone di tre passaggi principali:

1. Creare una configurazione CloudWatch dell'agente.
2. Archivia la configurazione dove può essere recuperata dalle PCS istanze.
3. Scrivi un modello di EC2 avvio che installi il software dell' CloudWatch agente, recuperi la configurazione e avvii l' CloudWatch agente utilizzando la configurazione.

Per ulteriori informazioni, consulta [Raccogli metriche, log e tracce con l' CloudWatch agente](#) nella Amazon CloudWatch User Guide e. [Utilizzo dei modelli di EC2 lancio di Amazon con AWS PCS](#)

Crea una configurazione dell'agente CloudWatch

Prima di distribuire l' CloudWatch agente sulle istanze, è necessario generare un file di JSON configurazione che specifichi le metriche, i log e le tracce da raccogliere. I file di configurazione possono essere creati utilizzando una procedura guidata o manualmente, utilizzando un editor di testo. Il file di configurazione verrà creato manualmente per questa dimostrazione.

Sul computer in cui è AWS CLI installato, crea un file di CloudWatch configurazione denominato `config.json` con il contenuto che segue. È inoltre possibile utilizzare quanto segue URL per scaricare una copia del file.

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/cloudwatch/assets/config.json
```

Note

- I percorsi di log nel file di esempio sono per Amazon Linux 2. Se le tue istanze utilizzeranno un sistema operativo di base diverso, modifica i percorsi in modo appropriato.
- Per acquisire altri registri, aggiungi altre voci in `collect_list`
- I valori in `{brackets}` sono variabili basate su modelli. Per l'elenco completo delle variabili supportate, consulta [Creare o modificare manualmente il file di configurazione dell' CloudWatch agente](#) nella Amazon CloudWatch User Guide.
- Puoi scegliere di omettere `logs` o `metrics` di non raccogliere questi tipi di informazioni.

```
{
  "agent": {
    "metrics_collection_interval": 60
  },
  "logs": {
    "logs_collected": {
      "files": {
        "collect_list": [
          {
            "file_path": "/var/log/cloud-init.log",
            "log_group_class": "STANDARD",
            "log_group_name": "/PCSLogs/instances",
            "log_stream_name": "{instance_id}.cloud-init.log",
```

```

        "retention_in_days": 30
    },
    {
        "file_path": "/var/log/cloud-init-output.log",
        "log_group_class": "STANDARD",
        "log_stream_name": "{instance_id}.cloud-init-output.log",
        "log_group_name": "/PCSLogs/instances",
        "retention_in_days": 30
    },
    {
        "file_path": "/var/log/amazon/pcs/bootstrap.log",
        "log_group_class": "STANDARD",
        "log_stream_name": "{instance_id}.bootstrap.log",
        "log_group_name": "/PCSLogs/instances",
        "retention_in_days": 30
    },
    {
        "file_path": "/var/log/slurmd.log",
        "log_group_class": "STANDARD",
        "log_stream_name": "{instance_id}.slurmd.log",
        "log_group_name": "/PCSLogs/instances",
        "retention_in_days": 30
    },
    {
        "file_path": "/var/log/messages",
        "log_group_class": "STANDARD",
        "log_stream_name": "{instance_id}.messages",
        "log_group_name": "/PCSLogs/instances",
        "retention_in_days": 30
    },
    {
        "file_path": "/var/log/secure",
        "log_group_class": "STANDARD",
        "log_stream_name": "{instance_id}.secure",
        "log_group_name": "/PCSLogs/instances",
        "retention_in_days": 30
    }
]
}
},
"metrics": {
    "aggregation_dimensions": [
        [

```

```
        "InstanceId"
      ]
    ],
    "append_dimensions": {
      "AutoScalingGroupName": "${aws:AutoScalingGroupName}",
      "ImageId": "${aws:ImageId}",
      "InstanceId": "${aws:InstanceId}",
      "InstanceType": "${aws:InstanceType}"
    },
    "metrics_collected": {
      "cpu": {
        "measurement": [
          "cpu_usage_idle",
          "cpu_usage_iowait",
          "cpu_usage_user",
          "cpu_usage_system"
        ],
        "metrics_collection_interval": 60,
        "resources": [
          "*"
        ],
        "totalcpu": false
      },
      "disk": {
        "measurement": [
          "used_percent",
          "inodes_free"
        ],
        "metrics_collection_interval": 60,
        "resources": [
          "*"
        ]
      },
      "diskio": {
        "measurement": [
          "io_time"
        ],
        "metrics_collection_interval": 60,
        "resources": [
          "*"
        ]
      },
      "mem": {
        "measurement": [
```

```

        "mem_used_percent"
    ],
    "metrics_collection_interval": 60
},
"swap": {
    "measurement": [
        "swap_used_percent"
    ],
    "metrics_collection_interval": 60
}
}
}
}
}

```

Questo file indica all' CloudWatch agente di monitorare diversi file che possono essere utili per diagnosticare errori relativi, ad esempio, al bootstrap, all'autenticazione e all'accesso e ad altri domini di risoluzione dei problemi. Ciò include:

- `/var/log/cloud-init.log`— Output dalla fase iniziale della configurazione dell'istanza
- `/var/log/cloud-init-output.log`— Output dei comandi eseguiti durante la configurazione dell'istanza
- `/var/log/amazon/pcs/bootstrap.log`— Output da operazioni PCS specifiche eseguite durante la configurazione dell'istanza
- `/var/log/slurmd.log`— Output dal demone slurmd del gestore del carico di lavoro Slurm
- `/var/log/messages`— Messaggi di sistema dal kernel, dai servizi di sistema e dalle applicazioni
- `/var/log/secure`— Registri relativi ai tentativi di autenticazioneSSH, come sudo e altri eventi di sicurezza

I file di registro vengono inviati a un gruppo di CloudWatch log denominato `/PCSLogs/instances`. I flussi di registro sono una combinazione dell'ID dell'istanza e del nome di base del file di registro. Il gruppo di log ha un tempo di conservazione di 30 giorni.

Inoltre, il file indica all' CloudWatch agente di raccogliere diverse metriche comuni, aggregandole per ID di istanza.

Memorizza la configurazione

Il file di configurazione dell' CloudWatch agente deve essere archiviato dove possono accedervi le istanze del nodo di PCS calcolo. Esistono due modi comuni per eseguire questa operazione. Puoi

caricarlo in un bucket Amazon S3 a cui le tue istanze del gruppo di nodi di calcolo avranno accesso tramite il loro profilo di istanza. In alternativa, puoi archivarlo come parametro in SSM Amazon Systems Manager Parameter Store.

Carica in un bucket S3

Per archiviare il file in S3, usa i AWS CLI comandi che seguono. Prima di eseguire il comando, effettua le seguenti sostituzioni:

- Replace (Sostituisci) *DOC-EXAMPLE-BUCKET* con il tuo nome di bucket S3

Innanzitutto, (questo è facoltativo se hai un bucket esistente), crea un bucket per contenere i tuoi file di configurazione.

```
aws s3 mb s3://DOC-EXAMPLE-BUCKET
```

Quindi, carica il file nel bucket.

```
aws s3 cp ./config.json s3://DOC-EXAMPLE-BUCKET/
```

Memorizza come parametro SSM

Per memorizzare il file come SSM parametro, utilizzate il comando che segue. Prima di eseguire il comando, effettuate le seguenti sostituzioni:

- Replace (Sostituisci) *region-code* con la AWS regione in cui lavori. AWS PCS
- (Facoltativo) Sostituisci *AmazonCloudWatch-PCS* con il tuo nome per il parametro. Tieni presente che se modifichi il prefisso del nome da, AmazonCloudWatch- dovrai aggiungere specificamente l'accesso in lettura al SSM parametro nel profilo dell'istanza del gruppo di nodi.

```
aws ssm put-parameter \  
  --region region-code \  
  --name "AmazonCloudWatch-PCS" \  
  --type String \  
  --value file://config.json
```

Scrivi un modello di EC2 lancio

I dettagli specifici per il modello di lancio dipendono dal fatto che il file di configurazione sia archiviato in S3 o SSM.

Usa una configurazione archiviata in S3

Questo script installa CloudWatch l'agente, importa un file di configurazione da un bucket S3 e avvia l'agente con esso. CloudWatch Sostituisci i seguenti valori in questo script con i tuoi dati:

- *DOC-EXAMPLE-BUCKET* — Il nome di un bucket S3 da cui il tuo account può leggere
- */config.json* — Percorso relativo alla radice del bucket S3 in cui è archiviata la configurazione

```
MIME-Version: 1.0
Content-Type: multipart/mixed; boundary==="MYBOUNDARY==="

--===MYBOUNDARY==
Content-Type: text/cloud-config; charset="us-ascii"

packages:
- amazon-cloudwatch-agent

runcmd:
- aws s3 cp s3://DOC-EXAMPLE-BUCKET/config.json /etc/s3-cw-config.json
- /opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent-ctl -a fetch-config -m
  ec2 -s -c file:///etc/s3-cw-config.json

--===MYBOUNDARY===--
```

Il profilo di IAM istanza per il gruppo di nodi deve avere accesso al bucket. Ecco un esempio di IAM policy per il bucket nello script di dati utente riportato sopra.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:ListBucket"
      ],
    }
  ],
}
```

```

        "Resource": [
            "arn:aws:s3:::DOC-EXAMPLE-BUCKET",
            "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*"
        ]
    }
]
}

```

Tieni inoltre presente che le istanze devono consentire il traffico in uscita verso S3 e gli endpoint. CloudWatch Ciò può essere ottenuto utilizzando gruppi o VPC endpoint di sicurezza, a seconda dell'architettura del cluster.

Utilizza una configurazione memorizzata in SSM

Questo script installa CloudWatch l'agente, importa un file di configurazione da un SSM parametro e avvia l' CloudWatch agente con esso. Sostituisci i seguenti valori in questo script con i tuoi dati:

- (Facoltativo) Sostituisci *AmazonCloudWatch-PCS* con il tuo nome per il parametro.

```

MIME-Version: 1.0
Content-Type: multipart/mixed; boundary=="MYBOUNDARY=="

--MYBOUNDARY--
Content-Type: text/cloud-config; charset="us-ascii"

packages:
- amazon-cloudwatch-agent

runcmd:
- /opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent-ctl -a fetch-config -m
  ec2 -s -c ssm:AmazonCloudWatch-PCS

--MYBOUNDARY--

```

La politica di IAM istanza per il gruppo di nodi deve avere il codice CloudWatchAgentServerPolicyallegato.

Se il nome del parametro non inizia con, sarà AmazonCloudWatch- necessario aggiungere specificamente l'accesso in lettura al SSM parametro nel profilo dell'istanza del gruppo di nodi. Ecco un esempio di IAM policy che lo illustra per prefix *DOC-EXAMPLE-PREFIX*.

```
{
  "Version" : "2012-10-17",
  "Statement" : [
    {
      "Sid" : "CustomCwSsmMParamReadOnly",
      "Effect" : "Allow",
      "Action" : [
        "ssm:GetParameter"
      ],
      "Resource" : "arn:aws:ssm:*:*:parameter/DOC-EXAMPLE-PREFIX*"
    }
  ]
}
```

Tieni inoltre presente che le istanze devono consentire il traffico in uscita verso gli endpoint e. SSM CloudWatch. Ciò può essere ottenuto utilizzando gruppi o VPC endpoint di sicurezza, a seconda dell'architettura del cluster.

Registrazione delle API chiamate al AWS Parallel Computing Service utilizzando AWS CloudTrail

AWS PCS è integrato con AWS CloudTrail, un servizio che fornisce una registrazione delle azioni intraprese da un utente, ruolo o AWS servizio in AWS PCS. CloudTrail acquisisce tutte le API chiamate AWS PCS come eventi. Le chiamate acquisite includono chiamate dalla AWS PCS console e chiamate in codice alle AWS PCS API operazioni. Se crei un trail, puoi abilitare la distribuzione continua di CloudTrail eventi a un bucket Amazon S3, inclusi gli eventi per. AWS PCS. Se non configuri un percorso, puoi comunque visualizzare gli eventi più recenti nella CloudTrail console nella cronologia degli eventi. Utilizzando le informazioni raccolte da CloudTrail, puoi determinare a quale richiesta è stata inviata AWS PCS, l'indirizzo IP da cui è stata effettuata, chi ha effettuato la richiesta, quando è stata effettuata e dettagli aggiuntivi.

Per ulteriori informazioni CloudTrail, consulta la [Guida AWS CloudTrail per l'utente](#).

AWS PCS informazioni in CloudTrail

CloudTrail è abilitato sul tuo account al Account AWS momento della creazione dell'account. Quando si verifica un'attività in AWS PCS, tale attività viene registrata in un CloudTrail evento insieme ad altri eventi AWS di servizio nella cronologia degli eventi. Puoi visualizzare, cercare e scaricare

eventi recenti in Account AWS. Per ulteriori informazioni, consulta [Visualizzazione degli eventi con la cronologia degli CloudTrail eventi](#).

Per una registrazione continua degli eventi del tuo Account AWS, inclusi gli eventi di AWS PCS, crea un percorso. Un trail consente di CloudTrail inviare file di log a un bucket Amazon S3. Per impostazione predefinita, quando si crea un percorso nella console, questo sarà valido in tutte le Regioni AWS. Il trail registra gli eventi di tutte le regioni della AWS partizione e consegna i file di log al bucket Amazon S3 specificato. Inoltre, puoi configurare altri AWS servizi per analizzare ulteriormente e agire in base ai dati sugli eventi raccolti nei log. CloudTrail Per ulteriori informazioni, consulta gli argomenti seguenti:

- [Panoramica della creazione di un percorso](#)
- [CloudTrail servizi e integrazioni supportati](#)
- [Configurazione delle SNS notifiche Amazon per CloudTrail](#)
- [Ricezione di file di CloudTrail registro da più regioni](#) e [ricezione di file di CloudTrail registro da più account](#)

Tutte le AWS PCS azioni vengono registrate CloudTrail e documentate nel [AWS Parallel Computing Service API Reference](#). Ad esempio, le chiamate a `CreateComputeNodeGroupUpdateQueue`, e `DeleteCluster` le azioni generano voci nei file di CloudTrail registro.

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 AWS servizio.

Per ulteriori informazioni, vedete l'[CloudTrail userIdentityelemento](#).

Informazioni sulle voci dei file di CloudTrail registro da AWS PCS

Un trail è una configurazione che consente la consegna di eventi come file di registro a un bucket S3 specificato dall'utente. CloudTrail i file di registro contengono una o più voci di registro. Un evento

rappresenta una singola richiesta proveniente da qualsiasi fonte e include informazioni sull'azione richiesta, la data e l'ora dell'azione, i parametri della richiesta e così via. CloudTrail i file di registro non sono una traccia stack ordinata delle API chiamate pubbliche, quindi non vengono visualizzati in un ordine specifico.

L'esempio seguente mostra una voce di CloudTrail registro per un>CreateQueueazione.

```
{
  "eventVersion": "1.09",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "AIDACKCEVSQ6C2EXAMPLE:admin",
    "arn": "arn:aws:sts::012345678910:assumed-role/Admin/admin",
    "accountId": "012345678910",
    "accessKeyId": "ASIAY36PTPIEXAMPLE",
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "AROAY36PTPIEXAMPLE",
        "arn": "arn:aws:iam::012345678910:role/Admin",
        "accountId": "012345678910",
        "userName": "Admin"
      },
      "attributes": {
        "creationDate": "2024-07-16T17:05:51Z",
        "mfaAuthenticated": "false"
      }
    }
  },
  "eventTime": "2024-07-16T17:13:09Z",
  "eventSource": "pcs.amazonaws.com",
  "eventName": "CreateQueue",
  "awsRegion": "us-east-1",
  "sourceIPAddress": "127.0.0.1",
  "userAgent": "Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/126.0.0.0 Safari/537.36",
  "requestParameters": {
    "clientToken": "c13b7baf-2894-42e8-acec-example",
    "clusterIdentifier": "abcdef0123",
    "computeNodeGroupConfigurations": [
      {
        "computeNodeId": "abcdef0123"
      }
    ]
  }
}
```

```
    ],
    "queueName": "all"
  },
  "responseElements": {
    "queue": {
      "arn": "arn:aws:pcs:us-east-1:609783872011:cluster/abcdef0123/queue/
abcdef0123",
      "clusterId": "abcdef0123",
      "computeNodeGroupConfigurations": [
        {
          "computeNodeId": "abcdef0123"
        }
      ],
      "createdAt": "2024-07-16T17:13:09.276069393Z",
      "id": "abcdef0123",
      "modifiedAt": "2024-07-16T17:13:09.276069393Z",
      "name": "all",
      "status": "CREATING"
    }
  },
  "requestID": "a9df46d7-3f6d-43a0-9e3f-example",
  "eventID": "7ab18f88-0040-47f5-8388-example",
  "readOnly": false,
  "eventType": "AwsApiCall",
  "managementEvent": true,
  "recipientAccountId": "012345678910",
  "eventCategory": "Management",
  "tlsDetails": {
    "tlsVersion": "TLSv1.3",
    "cipherSuite": "TLS_AES_128_GCM_SHA256",
    "clientProvidedHostHeader": "pcs.us-east-1.amazonaws.com"
  },
  "sessionCredentialFromConsole": "true"
}
```

Endpoint e quote di servizio per AWS PCS

Le sezioni seguenti descrivono gli endpoint e le quote di servizio per AWS Parallel Computing Service (AWS PCS). Le quote di servizio, precedentemente denominate limiti, rappresentano il numero massimo di risorse o operazioni di servizio per l'utente. Account AWS

Hai Account AWS delle quote predefinite per ogni servizio. AWS Salvo diversa indicazione, ogni quota si applica a una regione specifica. Se per alcune quote è possibile richiedere aumenti, altre quote non possono essere modificate.

Per ulteriori informazioni, consulta [AWS Service Quotas](#) in Riferimenti generali di AWS .

Indice

- [Endpoint del servizio](#)
- [Quote del servizio](#)
 - [Quote interne](#)
 - [Quote pertinenti per altri servizi AWS](#)

Endpoint del servizio

Nome Regione	Regione	Endpoint	Protocollo
US East (N. Virginia)	us-east-1	pcs.us-east-1.amazonaws.com	HTTPS
Stati Uniti orientali (Ohio)	us-east-2	pcs.us-east-2.amazonaws.com	HTTPS
US West (Oregon)	us-west-2	pcs.us-west-2.amazonaws.com	HTTPS
Asia Pacific (Singapore)	ap-southeast-1	pcs.ap-southeast-1.amazonaws.com	HTTPS
Asia Pacific (Sydney)	ap-southeast-2	pcs.ap-southeast-2.amazonaws.com	HTTPS

Nome Regione	Regione	Endpoint	Protocollo
Asia Pacifico (Tokyo)	ap-northeast-1	pcs.ap-northeast-1 .amazonaws.com	HTTPS
Europe (Frankfurt)	eu-central-1	pcs.eu-central-1.a mazonaws.com	HTTPS
Europa (Irlanda)	eu-west-1	pcs.eu-west-1.amaz onaws.com	HTTPS
Europa (Stoccolma)	eu-north-1	pcs.eu-north-1.ama zonaws.com	HTTPS

Quote del servizio

Nome	Impostazione predefinita	Regolabile	Descrizione
Cluster	5	Sì	Il numero massimo di cluster per Regione AWS

Note

I valori predefiniti sono le quote iniziali impostate da AWS. Questi valori predefiniti sono separati dai valori effettivi delle quote applicate e dalle quote massime possibili del servizio. Per ulteriori informazioni, consulta [Terminologia di Service Quotas](#) nella Guida per l'utente di Service Quotas.

Queste quote di servizio sono elencate in AWS Parallel Computing Service (PCS) nel [AWS Management Console](#). Per richiedere un aumento della quota per i valori indicati come regolabili, vedere [Requesting a Quote Increase](#) nella Service Quotas User Guide.

⚠ Important

Ricordati di controllare l' Regione AWS impostazione corrente in. AWS Management Console

Quote interne

Le seguenti quote sono interne e non regolabili.

Nome	Impostazione predefinita	Regolabile	Descrizione
Creazione simultanea di cluster	1	No	Il numero massimo di cluster nello Creating stato per. Regione AWS

Quote pertinenti per altri servizi AWS

AWS PCS utilizza altri AWS servizi. Le quote di servizio per tali servizi influiscono sull'utilizzo di AWS PCS.

Quote EC2 di servizi Amazon che influiscono AWS PCS

- Richieste di istanze Spot
- Esecuzione di istanze su richiesta
- Modelli di avvio
- Versioni del modello di avvio
- EC2API Richieste Amazon

Per ulteriori informazioni, consulta le [quote dei EC2 servizi Amazon](#) nella Amazon Elastic Compute Cloud User Guide.

Note di rilascio per AWS PCS un esempio AMIs

AWS PCSAMIs gli esempi hanno una cadenza di rilascio notturna per le patch di sicurezza. Queste patch di sicurezza incrementali non sono incluse nelle note di rilascio ufficiali.

Important

AMIs Gli esempi sono a scopo dimostrativo e non sono consigliati per carichi di lavoro di produzione.

Indice

- [AWS PCS esempio x86_64 AMI per Slurm 23.11 \(Amazon Linux 2\)](#)
- [AWS PCS esempio Arm64 AMI per Slurm 23.11 \(Amazon Linux 2\)](#)

AWS PCS esempio x86_64 AMI per Slurm 23.11 (Amazon Linux 2)

Questo documento descrive le modifiche, le aggiunte, i problemi noti e le correzioni più recenti per AWS PCS Sample x86_64 AMI (Amazon Linux 2).

- Data di creazione: 15 luglio 2024
- Data di rilascio: 22 agosto 2024
- Ultimo aggiornamento: 22 agosto 2024

AMI nome

- `aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11`

EC2 Istanze supportate

- Tutte le istanze con processore x86 a 64 bit. Per trovare istanze compatibili, accedi alla [EC2 console Amazon](#). Scegli Tipi di istanza, quindi cerca `Architectures=x86_64`.

AMI contenuti

- AWS Servizio supportato: AWS PCS

- Sistema operativo: Amazon Linux 2
- Architettura di calcolo: x86_64
- Kernel Linux: 5.10.220-209.867.amzn2.x86_64
- EBStipo di volume: gp2
- AWS PCSprogramma di installazione Slurm 23.11:23.11.9-1
- AWS PCSprogramma di installazione del software: 1.0.0-1
- EFAinstallatore: 1.33.0
- GDRCopy: 2.4
- NVIDIAAutista: 535.154.05
- NVIDIACUDA: 12.2.2_535.104.05

Note

- Nessuno

Data di rilascio: 2024-08-22

Aggiornato

- Nessuna. Prima versione.

Aggiunto

- Nessuna. Prima versione.

Rimosso

- Nessuna. Prima versione.

AWS PCSesempio Arm64 AMI per Slurm 23.11 (Amazon Linux 2)

Questo documento descrive le ultime modifiche, aggiunte, problemi noti e correzioni per AWS PCS Sample Arm64 (AMI Amazon Linux 2).

- Data di creazione: 15 luglio 2024

- Data di rilascio: 22 agosto 2024
- Ultimo aggiornamento: 22 agosto 2024

AMInome

- `aws-pcs-sample_ami-amzn2-arm64-slurm-23.11`

EC2Istanze supportate

- Tutte le istanze con processore Arm a 64 bit. Per trovare istanze compatibili, accedi alla [EC2console Amazon](#). Scegli Tipi di istanza, quindi cerca `Architectures=arm64`.

AMIcontenuti

- AWS Servizio supportato: AWS PCS
- Sistema operativo: Amazon Linux 2
- Architettura di calcolo: arm64
- Kernel Linux: 5.10.220-209.867.amzn2.aarch64
- EBStipo di volume: gp2
- AWS PCSprogramma di installazione Slurm 23.11:23.11.9-1
- AWS PCSprogramma di installazione del software: 1.0.0-1
- EFAinstallatore: 1.33.0
- GDRCopy: 2.4
- NVIDIAAutista: 535.154.05
- NVIDIAACUDA: 12.2.2_535.104.05

Note

- Nessuno

Data di rilascio: 2024-08-22

Aggiornato

- Nessuna. Prima versione.

Aggiunto

- Nessuna. Prima versione.

Rimosso

- Nessuna. Prima versione.

Cronologia dei documenti per la Guida per l'utente di AWS PCS

La tabella seguente descrive le versioni della documentazione per AWS PCS.

Data	Modifica	Aggiornamenti della documentazione	APIversioni aggiornate
28 agosto 2024	È stata aggiunta la pagina delle politiche gestite	Per ulteriori informazioni, consulta AWS politiche gestite per AWS Parallel Computing Service .	N/D
28 agosto 2024	AWS PCSrilascio	Versione iniziale della guida AWS PCS per l'utente.	AWS SDK: 2024-08-28

AWS Glossario

Per la AWS terminologia più recente, consultate il [AWS glossario](#) nella sezione Reference. Glossario AWS

Le traduzioni sono generate tramite traduzione automatica. In caso di conflitto tra il contenuto di una traduzione e la versione originale in Inglese, quest'ultima prevarrà.