Guía del usuario

AWS PCS

AWS PCS: Guía del usuario

Table of Contents

¿Qué es el Servicio de Computación AWS Paralela?

AWS Parallel Computing Service (AWS PCS) es un servicio gestionado que facilita la ejecución y el escalado de las cargas de trabajo informáticas de alto rendimiento (HPC) y la creación de modelos científicos y de ingeniería basados en el AWS uso de Slurm. Úselo AWS PCS para crear clústeres de procesamiento que integren la mejor AWS computación, almacenamiento, redes y visualización de su clase. Ejecute simulaciones o cree modelos científicos y de ingeniería. Optimice y simplifique las operaciones de sus clústeres mediante las funciones integradas de administración y observabilidad. Permita que sus usuarios se centren en la investigación y la innovación al permitirles ejecutar sus aplicaciones y trabajos en un entorno familiar.

Conceptos clave

Un clúster AWS PCS tiene una o más colas asociadas a al menos un grupo de nodos de cómputo. Los trabajos se envían a colas y se ejecutan en EC2 instancias definidas por grupos de nodos de procesamiento. Puede utilizar estas bases para implementar HPC arquitecturas sofisticadas.

Clúster

Un clúster es un recurso para administrar recursos y ejecutar cargas de trabajo. Un clúster es un AWS PCS recurso que define un conjunto de configuraciones de procesamiento, redes, almacenamiento, identidad y programador de tareas. Para crear un clúster, especifique qué programador de tareas desea usar (actualmente Slurm), qué configuración de programador desea, qué controlador de servicios desea administrar el clúster y en qué VPC desea que se lancen los recursos del clúster. El programador acepta y programa los trabajos, y también lanza los nodos de cómputo (EC2instancias) que procesan esos trabajos.

Grupo de nodos de cómputo

Un grupo de nodos de procesamiento es un conjunto de nodos de procesamiento que se AWS PCS utiliza para ejecutar trabajos o proporcionar acceso interactivo a un clúster. Al definir un grupo de nodos de cómputo, se especifican características comunes, como los tipos de EC2 instancias de Amazon, el número mínimo y máximo de instancias, VPC las subredes de destino, Amazon Machine Image (AMI), la opción de compra y la configuración de lanzamiento personalizada. AWS PCSusa esta configuración para lanzar, administrar y terminar de manera eficiente los nodos de cómputo de un grupo de nodos de cómputo.

Queue

Cuando quieres ejecutar un trabajo en un clúster específico, lo envías a una cola determinada (también denominada partición). El trabajo permanece en la cola hasta que se AWS PCS programe su ejecución en un grupo de nodos de procesamiento. Asocia uno o más grupos de nodos de cómputo a cada cola. Se necesita una cola para programar y ejecutar los trabajos en los recursos del grupo de nodos de cómputo subyacentes mediante diversas políticas de programación ofrecidas por el programador de trabajos. Los usuarios no envían los trabajos directamente a un nodo de cómputo o a un grupo de nodos de cómputo.

Administrador de sistemas

Un administrador del sistema implementa, mantiene y opera un clúster. Pueden acceder a AWS PCS través de AWS Management Console AWS PCSAPI, y AWS SDK. Tienen acceso a clústeres específicos a través de SSH o AWS Systems Manager, donde pueden ejecutar tareas administrativas, ejecutar trabajos, administrar datos y realizar otras actividades basadas en el shell. Para obtener más información, consulte la Documentación de AWS Systems Manager.

Usuario final

El usuario final no tiene day-to-day la responsabilidad de implementar u operar un clúster. Utilizan una interfaz de terminal (por ejemploSSH) para acceder a los recursos del clúster, ejecutar tareas, administrar datos y realizar otras actividades basadas en el shell.

Configuración del servicio de computación AWS paralela

Realice las siguientes tareas para configurar AWS Parallel Computing Service (AWS PCS).

Temas

Inscríbase en una Cuenta de AWS

Si no tiene uno Cuenta de AWS, complete los siguientes pasos para crearlo.

Para suscribirte a una Cuenta de AWS

1. Abrir https://portal.aws.amazon.com/billing/registro.
2. Siga las instrucciones que se le indiquen.

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

   Cuando te registras en un Cuenta de AWS, Usuario raíz de la cuenta de AWSse crea un. El usuario raíz tendrá acceso a todos los Servicios de AWS y recursos de esa cuenta. Como práctica recomendada de seguridad, asigne acceso administrativo a un usuario y utilice únicamente el usuario raíz para realizar tareas que requieren acceso de usuario raíz.

AWS te envía un correo electrónico de confirmación una vez finalizado el proceso de registro. En cualquier momento, puede ver la actividad de su cuenta actual y administrarla accediendo a https://aws.amazon.com/y seleccionando Mi cuenta.

Creación de un usuario con acceso administrativo

Después de crear un usuario administrativo Cuenta de AWS, asegúrelo Usuario raíz de la cuenta de AWS AWS IAM Identity Center, habilite y cree un usuario administrativo para no usar el usuario root en las tareas diarias.

Proteja su Usuario raíz de la cuenta de AWS

1. Inicie sesión AWS Management Consolecomo propietario de la cuenta seleccionando el usuario root e introduciendo su dirección de Cuenta de AWS correo electrónico. En la siguiente página, escriba su contraseña.

   Para obtener ayuda para iniciar sesión con el usuario raíz, consulte Iniciar sesión como usuario raíz en la Guía del usuario de AWS Sign-In .
2. Activa la autenticación multifactorial (MFA) para tu usuario root.

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

Creación de un usuario con acceso administrativo

1. Habilite IAM Identity Center.

   Consulte las instrucciones en Activar AWS IAM Identity Center en la Guía del usuario de AWS IAM Identity Center .
2. En IAM Identity Center, conceda acceso administrativo a un usuario.

   Para ver un tutorial sobre cómo usar el Directorio de IAM Identity Center como fuente de identidad, consulte Configurar el acceso de los usuarios con la configuración predeterminada Directorio de IAM Identity Center en la Guía del AWS IAM Identity Center usuario.

Iniciar sesión como usuario con acceso de administrador

- Para iniciar sesión con su usuario de IAM Identity Center, utilice el inicio de sesión URL que se envió a su dirección de correo electrónico cuando creó el usuario de IAM Identity Center.

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

Concesión de acceso a usuarios adicionales

1. En IAM Identity Center, cree un conjunto de permisos que siga la práctica recomendada de aplicar permisos con privilegios mínimos.

Para conocer las instrucciones, consulte Create a permission set en la Guía del usuario de AWS IAM Identity Center .

2. Asigne usuarios a un grupo y, a continuación, asigne el acceso de inicio de sesión único al grupo.

   Para conocer las instrucciones, consulte Add groups en la Guía del usuario de AWS IAM Identity Center .

Instale el AWS CLI

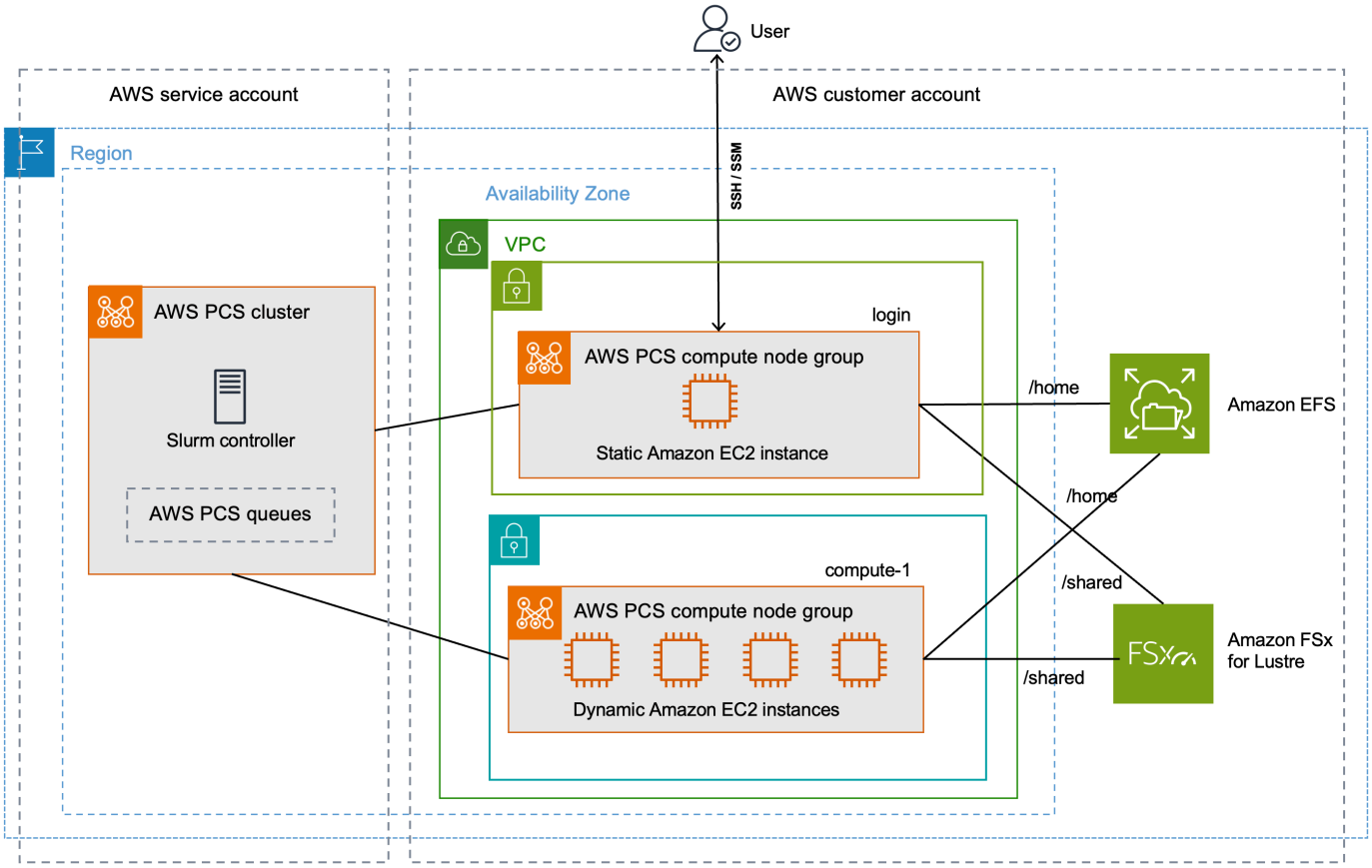
Debe utilizar la última versión de AWS CLI. Para obtener más información, consulte Instalar o actualizar a la última versión de AWS CLI en la Guía del AWS Command Line Interface usuario de la versión 2.

Introduzca el siguiente comando en una línea de comandos para comprobar su AWS CLI estado; debería mostrar información de ayuda.

```
aws pcs help
```

Empezar con AWS PCS

Este es un tutorial para crear un clúster sencillo que puedes usar para probar AWS PCS. La siguiente figura muestra el diseño del clúster.



El tutorial sobre el diseño de clústeres incluye los siguientes componentes clave:

- A VPC y subredes que cumplen los requisitos AWS PCS de red.
- Un sistema de EFS archivos de Amazon, que se utilizará como directorio principal compartido.
- Un sistema de archivos Amazon FSx for Lustre, que proporciona un directorio compartido de alto rendimiento.
- Un AWS PCS clúster que proporciona un controlador Slurm.
- 2 grupos de nodos de cómputo.
  - El grupo de `login` nodos, que proporciona acceso interactivo al sistema basado en una consola.

  - El grupo de `compute-1` nodos proporciona instancias que se escalan elásticamente para ejecutar trabajos.
- 1 cola que envía los trabajos a las EC2 instancias del grupo de nodos. `compute-1`

El clúster requiere AWS recursos adicionales, como grupos de seguridad, IAM funciones y plantillas de EC2 lanzamiento, que no se muestran en el diagrama.

Temas

Requisitos previos para empezar con AWS PCS

Antes de comenzar este tutorial, instale y configure las siguientes herramientas y recursos que necesita para crear y administrar un AWS PCS clúster.

- AWS CLI— Una herramienta de línea de comandos para trabajar con AWS servicios, que incluye AWS PCS. Para obtener más información, consulte Instalar o actualizar a la última versión de AWS CLI en la Guía del AWS Command Line Interface usuario de la versión 2. Tras instalarlo AWS CLI, le recomendamos que también lo configure. Para obtener más información, consulte Configurar el AWS CLI en la Guía del AWS Command Line Interface usuario de la versión 2.

- IAMPermisos necesarios: el responsable de IAM seguridad que utilice debe tener permisos para trabajar con AWS PCS IAM funciones, funciones vinculadas a servicios AWS CloudFormation VPC, a y recursos relacionados. Para obtener más informaciónServicio de Gestión de Identidad y Acceso para Computación AWS Paralela, consulte Crear un rol vinculado a un servicio en la Guía del AWS Identity and Access Management usuario. Debe completar todos los pasos de esta guía como el mismo usuario. Ejecute el siguiente comando para comprobar el usuario actual:

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

- Le recomendamos que complete los pasos de la línea de comandos de este tema en un shell de Bash. Si no está utilizando un intérprete de comandos Bash, algunos comandos de script, como los caracteres de continuación de línea y la forma en que se establecen y utilizan las variables, requieren ajustes para su intérprete de comandos. Además, las reglas de entrecomillado y escape de su intérprete de comandos pueden ser diferentes. Para obtener más información, consulte Comillas y literales con cadenas AWS CLI en la Guía del AWS Command Line Interface usuario de la versión 2.

Cree subredes VPC a y para AWS PCS

Puede crear subredes a VPC y con una CloudFormation plantilla. Utilice lo siguiente URL para descargar la CloudFormation plantilla y, a continuación, cárguela en la AWS CloudFormation consola para crear una CloudFormation pila nueva. Para obtener más información, consulte Uso de la AWS CloudFormation consola en la Guía del AWS CloudFormation usuario.

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

Con la plantilla abierta en la AWS CloudFormation consola, introduzca las siguientes opciones. Puede utilizar los valores predeterminados proporcionados en la plantilla.

- En Proporcione un nombre de pila:
  - En Nombre de pila, ingresa:

    ```
    hpc-networking
    ```

- En Parámetros:
  - En VPC:

    - En CidrBlock, introduzca:

      ```
      10.3.0.0/16
      ```

  - En las subredes A:
    - En CidrPublicSubnetA, introduzca:

      ```
      10.3.0.0/20
      ```

    - En CidrPrivateSubnetA, introduzca:

      ```
      10.3.128.0/20
      ```

  - En las subredes B:
    - En CidrPublicSubnetB, introduzca:

      ```
      10.3.16.0/20
      ```

    - En CidrPrivateSubnetB, introduzca:

      ```
      10.3.144.0/20
      ```

  - En Subredes C:
    - Para ProvisionSubnetsC, seleccione True
    - En CidrPublicSubnetC, introduzca:

      ```
      10.3.32.0/20
      ```

    - En CidrPrivateSubnetC, introduzca:

      ```
      10.3.160.0/20
      ```

- En Capacidades:
  - Marque la casilla de Reconozco que AWS CloudFormation podría crear IAM recursos.

Supervisa el estado de la CloudFormation pila. Cuando llegue`CREATE_COMPLETE`, busque el ID del grupo de seguridad predeterminado en el nuevoVPC. Utilizará el ID más adelante en el tutorial.

Busque el grupo de seguridad predeterminado para el clúster VPC

Para encontrar el ID del grupo de seguridad predeterminado en el nuevoVPC, siga este procedimiento:

- Ve a la VPCconsola de Amazon.
- En el VPCpanel de control, selecciona Filtrar por VPC.
  - Elige el VPC lugar por el que empieza el nombrehpc-networking.
  - En Seguridad, selecciona Grupos de seguridad.
- Busque el ID del grupo de seguridad del grupo denominadodefault. Tiene la descripcióndefault VPC security group. El ID se utiliza más adelante para configurar las plantillas de EC2 lanzamiento.

Crear grupos de seguridad para AWS PCS

AWS PCSdepende de los grupos de seguridad para administrar el tráfico de red que entra y sale de un clúster y sus grupos de nodos de cómputo. Para obtener información detallada sobre este tema, consulteRequisitos y consideraciones sobre los grupos de seguridad.

En este paso, utilizará una CloudFormation plantilla para dos grupos de seguridad.

- Un grupo de seguridad de clúster, que permite las comunicaciones entre el AWS PCS controlador, los nodos de cómputo y los nodos de inicio de sesión.
- Un grupo de SSH seguridad entrante que, si lo desea, puede añadir a sus nodos de inicio de sesión para facilitar SSH el acceso

Cree los grupos de seguridad para AWS PCS

Puede crear subredes VPC y subredes con esta CloudFormation plantilla. Utilice lo siguiente URL para descargar la CloudFormation plantilla y, a continuación, cárguela en la AWS CloudFormation consola para crear una CloudFormation pila nueva. Para obtener más información, consulte Uso de la AWS CloudFormation consola en la Guía del AWS CloudFormation usuario.

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

Con la plantilla abierta en la AWS CloudFormation consola, introduzca las siguientes opciones. Tenga en cuenta que algunas opciones se rellenarán automáticamente en la plantilla; simplemente puede dejarlas como valores predeterminados.

- En Indique un nombre para la pila
  - En Nombre de pila, ingresa:

    ```
    getstarted-sg
    ```
- En Parámetros
  - En VpcId, elija el VPC lugar por el que empieza el nombre`hpc-networking`.
  - (Opcional) En ClientIpCidr, introduzca un rango de IP más restrictivo para el grupo de SSH seguridad entrante. Le recomendamos que lo restrinja con su propia IP/subred (x.x.x.x/32 para su propia IP o x.x.x.x/24 para el rango). PUBLICSustituya x.x.x.x por su propia IP. Puede obtener su IP pública mediante herramientas como https://ifconfig.co/)

Supervisa el estado de la CloudFormation pila. Cuando llegue al grupo `CREATE_COMPLETE` de seguridad, los recursos estarán listos.

Se han creado dos grupos de seguridad, con los nombres:

- `cluster-getstarted-sg`— este es el grupo de seguridad del clúster
- `inbound-ssh-getstarted-sg`— se trata de un grupo de seguridad que permite el acceso entrante SSH

Crear un clúster en AWS PCS

En AWS PCS, un clúster es un recurso persistente para administrar recursos y ejecutar cargas de trabajo. Se crea un clúster para un programador específico (AWS PCSactualmente compatible con Slurm) en una subred de un programador nuevo o existente. VPC El clúster acepta y programa trabajos, y también lanza los nodos de cómputo (EC2instancias) que procesan esos trabajos.

Cómo crear el clúster

1. Abre la AWS PCSconsola y selecciona Crear clúster.
2. En la sección Configuración del clúster, introduce los siguientes campos:

- Nombre del clúster: introduzca `get-started`
- Tamaño de la controladora: seleccione Pequeña

3. En la sección Redes, seleccione valores para los siguientes campos:
   - VPC— Elija el VPC nombre `hpc-networking:Large-Scale-HPC`
   - Subred: seleccione la subred por la que comienza el nombre `hpc-networking:PrivateSubnetA`
   - Grupos de seguridad: seleccione el nombre del grupo de seguridad del clúster `cluster-getstarted-sg`
4. Elija Create cluster.

> **Note**
>
> El campo Estado muestra la opción Crear mientras se aprovisiona el clúster. La creación del clúster puede tardar varios minutos.

Cree almacenamiento compartido para AWS PCS Amazon Elastic File System

Amazon Elastic File System (AmazonEFS) es un AWS servicio que proporciona almacenamiento de archivos totalmente elástico y sin servidor para que pueda compartir datos de archivos sin aprovisionar ni administrar la capacidad de almacenamiento y el rendimiento. Para obtener más información, consulte ¿Qué es Amazon Elastic File System? en la Guía del usuario de Amazon Elastic File System.

El clúster de AWS PCS demostración utiliza un sistema de EFS archivos para proporcionar un directorio principal compartido entre los nodos del clúster. Cree un sistema de EFS archivos similar VPC al del clúster.

Para crear tu sistema de EFS archivos de Amazon

1. Ve a la EFSconsola de Amazon.
2. Asegúrate de que esté configurada de la misma manera en la Región de AWS que la probarás AWS PCS.

3. Seleccione Crear sistema de archivos.
4. En la página Crear sistema de archivos, defina los siguientes parámetros:
   - En Nombre, introduzca `getstarted-efs`.
   - En Virtual Private Cloud (VPC), elija el VPC nombre `hpc-networking:Large-Scale-HPC`
   - Seleccione Crear. De este modo, volverá a la página de sistemas de archivos.
5. Anote el ID del sistema de archivos del sistema de `getstarted-efs` archivos. Usará esta información más tarde.

Crea almacenamiento compartido AWS PCS en Amazon FSx for Lustre

Amazon FSx for Lustre hace que sea fácil y rentable lanzar y ejecutar el popular sistema de archivos Lustre de alto rendimiento. Utiliza Lustre para cargas de trabajo en las que la velocidad es importante, como el aprendizaje automático, la informática de alto rendimiento (HPC), el procesamiento de vídeo y la elaboración de modelos financieros. Para obtener más información, consulta ¿Qué es Amazon FSx for Lustre? en la Guía del usuario FSx de Amazon for Lustre.

El clúster de AWS PCS demostración puede utilizar un sistema de archivos FSx for Lustre para proporcionar un directorio compartido de alto rendimiento entre los nodos del clúster. Cree un sistema de archivos FSx para Lustre en el mismo lugar VPC que su clúster.

Para crear su sistema de FSx archivos para Lustre

1. Ve a la FSxconsola de Amazon.
2. Asegúrese de que la consola esté configurada para usar lo Región de AWS mismo que su clúster.
3. Seleccione Crear sistema de archivos.
   - En Seleccione el tipo de sistema de archivos, elija Amazon FSx for Lustre y, a continuación, elija Siguiente.
4. En la página Especificar los detalles del sistema de archivos, defina los siguientes parámetros:
   - En Detalles del sistema de archivos
     - En Nombre, introduzca `getstarted-fsx`.
     - Para el tipo de implementación y almacenamiento, selecciona Persistente, SSD

    - Para obtener un rendimiento por unidad de almacenamiento, elija 125 MB/s/TiB
    - Para Capacidad de almacenamiento, introduzca 1,2 TiB
    - Para la configuración de metadatos, elija Automática
    - Para el tipo de compresión de datos, elija LZ4
  - En Red y seguridad
    - Para Virtual Private Cloud (VPC), elija el VPC nombre `hpc-networking:Large-Scale-HPC`
    - En el VPCcaso de los grupos de seguridad, deje el nombre del grupo de seguridad `default`
    - En Subred, elija la subred en la que comience el nombre `hpc-networking:PrivateSubnetA`
  - Deje las demás opciones con sus valores predeterminados.
  - Elija Next (Siguiente).
5. En la página Revisar y crear, elija Crear sistema de archivos. De este modo, volverá a la página Sistemas de archivos.
6. Navegue a la página de detalles del sistema de archivos FSx para Lustre que creó.
7. Anote el ID del sistema de archivos y el nombre del montaje. Usará esta información más tarde.

> **Note**
>
> El campo Estado muestra la opción Crear mientras se aprovisiona el sistema de archivos. La creación del sistema de archivos puede tardar varios minutos. Espere a que se complete antes de continuar con el resto del tutorial.

Cree grupos de nodos de cómputo en AWS PCS

Un grupo de nodos de cómputo es un conjunto virtual de nodos de cómputo (EC2instancias) que AWS PCS se lanza y administra. Al definir un grupo de nodos de cómputo, se especifican características comunes, como los tipos de EC2 instancias, el número mínimo y máximo de instancias, VPC las subredes de destino, la opción de compra preferida y la configuración de lanzamiento personalizada. AWS PCSLanza, administra y termina de manera eficiente los nodos de cómputo de un grupo de nodos de cómputo, de acuerdo con estos ajustes. El clúster de demostración utiliza un grupo de nodos de cálculo para proporcionar nodos de inicio de sesión para

el acceso de los usuarios y un grupo de nodos de cálculo independiente para procesar los trabajos. En los temas siguientes se describen los procedimientos para configurar estos grupos de nodos de procesamiento en el clúster.

Temas

Cree un perfil de instancia para AWS PCS

Los grupos de nodos de cómputo requieren un perfil de instancia cuando se crean. Si utilizas el AWS Management Console para crear un rol para AmazonEC2, la consola crea automáticamente un perfil de instancia y le asigna el mismo nombre que el rol. Para obtener más información, consulte Uso de perfiles de instancia en la Guía del AWS Identity and Access Management usuario.

En el siguiente procedimiento, se utiliza AWS Management Console para crear un rol para AmazonEC2, que también crea el perfil de instancia para los grupos de nodos de cómputo.

Para crear el perfil de rol y de instancia

- Vaya a la consola de IAM.
- En Access management (Administración de acceso), seleccione Policies (Políticas).
  - Elija Create Policy.
  - En Especificar permisos, en Editor de políticas, elija JSON.
  - Sustituya el contenido del editor de texto por lo siguiente:

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

```
    ]
}
```

- Elija Next (Siguiente).
- En Revisar y crear, escriba el nombre de la políticaAWSPCS-getstarted-policy.
- Elija Crear política.

- En Access management (Administración de acceso), elija Roles (Roles).
- Elija Crear rol.
- En Seleccione una entidad de confianza:
  - En el tipo de entidad de confianza, seleccione AWS servicio
  - En Caso de uso, selecciona EC2.
    - A continuación, en Elegir un caso de uso para el servicio especificado, selecciona EC2.
  - Elija Next (Siguiente).
- En Añadir permisos:
  - En Políticas de permisos, busque AWSPCS-getstarted-policy.
  - Marque la casilla situada junto a AWSPCS-getstarted-policy para añadirla al rol.
  - En Políticas de permisos, busque A. mazonSSMManaged InstanceCore
  - Marque la casilla situada junto mazonSSMManaged InstanceCore a A para añadirla al rol.
  - Elija Next (Siguiente).
- En Nombre, revisa y crea:
  - En Detalles del rol:
    - En Role name (Nombre del rol), introduzca AWSPCS-getstarted-role.
  - Elija Create role (Crear rol).

Cree plantillas de lanzamiento para AWS PCS

Al crear un grupo de nodos de cómputo, se proporciona una plantilla de EC2 lanzamiento que se AWS PCS utiliza para configurar EC2 las instancias que lanza. Esto incluye ajustes como los grupos de seguridad y los scripts que se ejecutan cuando se lanza la instancia.

En este paso, se utilizará una CloudFormation plantilla para crear dos plantillas de EC2 lanzamiento. Una plantilla se usará para crear nodos de inicio de sesión y la otra se usará para crear nodos de cómputo. La diferencia clave entre ellos es que los nodos de inicio de sesión se pueden configurar para permitir el SSH acceso entrante.

Acceda a la plantilla CloudFormation

Utilice lo siguiente URL para descargar la CloudFormation plantilla y, a continuación, cárguela en la AWS CloudFormation consola para crear una CloudFormation pila nueva. Para obtener más información, consulte Uso de la AWS CloudFormation consola en la Guía del AWS CloudFormation usuario.

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/getting_started/assets/pcs-lt-efs-fsxl.yaml
```

Utilice la CloudFormation plantilla para crear plantillas de EC2 lanzamiento

Utilice el siguiente procedimiento para completar la CloudFormation plantilla en la AWS CloudFormation consola

- En Proporcione un nombre de pila:
  - En Nombre de pila, ingresa`getstarted-lt`.
- En Parámetros:
  - Bajo seguridad
    - Para VpcSecurityGroupId, seleccione el grupo de seguridad nombrado `default` en su clústerVPC.
    - Para ClusterSecurityGroupId, seleccione el grupo denominado `cluster-getstarted-sg`
    - Para SshSecurityGroupId, seleccione el grupo denominado `inbound-ssh-getstarted-sg`
    - Para SshKeyName, selecciona tu SSH key pair preferido.
  - En Sistemas de archivos
    - Para EfsFilesystemId, introduzca el ID del sistema de EFS archivos del sistema de archivos que creó anteriormente en el tutorial.
    - Para FSxLustreFilesystemId, introduzca el ID del sistema de archivos del sistema FSx de archivos de Lustre que creó anteriormente en el tutorial.
    - Para FSxLustreFilesystemMountName, introduzca el nombre de montaje correspondiente al mismo sistema FSx de archivos Lustre.
- Seleccione Siguiente y, a continuación, vuelva a seleccionar Siguiente.
- Elija Enviar.

Supervisa el estado de la CloudFormation pila. Cuando llegue a `CREATE_COMPLETE` la plantilla de lanzamiento estará lista para ser utilizada.

> **Note**
>
> Para ver todos los recursos que creó la CloudFormation plantilla, abre la AWS CloudFormation consola. Elija la pila `getstarted-lt` y, a continuación, elija la pestaña Resources (Recursos).

Cree un grupo de nodos de cómputo para los nodos de inicio de sesión en AWS PCS

Un grupo de nodos de cómputo es una colección virtual de nodos de cómputo (EC2instancias) que AWS PCS se lanza y administra. Al definir un grupo de nodos de cómputo, se especifican características comunes, como los tipos de EC2 instancias, el número mínimo y máximo de instancias, VPC las subredes de destino, la opción de compra preferida y la configuración de lanzamiento personalizada. AWS PCSLanza, administra y termina de manera eficiente los nodos de cómputo de un grupo de nodos de cómputo, de acuerdo con estos ajustes.

En este paso, lanzará un grupo de nodos de computación estáticos que proporciona acceso interactivo al clúster. Puede utilizar SSH Amazon EC2 Systems Manager (SSM) para iniciar sesión en él y, a continuación, ejecutar comandos de shell y gestionar los trabajos de Slurm.

Para crear el grupo de nodos de cómputo

- Abra la AWS PCSconsola y vaya a Clústeres.
- Seleccione el clúster denominado `get-started`
- Vaya a Compute nodes groups y elija Crear.
- En la sección de configuración del grupo de nodos de Compute, proporciona lo siguiente:
  - Nombre del grupo de nodos de cómputo: `login` introdúzcalo.
- En Configuración informática, introduzca o seleccione estos valores:
  - EC2plantilla de lanzamiento: elija la plantilla de lanzamiento donde está el nombre `login-getstarted-lt`
  - IAMperfil de instancia: elija el nombre del perfil de instancia `AWSPCS-getstarted-role`

  - Subredes: selecciona la subred por la que comienza el nombre. `hpc-networking:PublicSubnetA`
  - Instancias: seleccione. `c6i.xlarge`
  - Configuración de escalado: introduzca 1 el número mínimo de instancias. Para el recuento máximo de instancias, introduzca. 1
- En Configuración adicional, especifique lo siguiente:
  - AMIID: seleccione el AMI lugar por el que empieza el nombre `aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11`
- Seleccione Crear grupo de nodos de cómputo.

El campo Estado muestra la opción Crear mientras se aprovisiona el grupo de nodos de cómputo. Puede continuar con el siguiente paso del tutorial mientras está en curso.

Cree un grupo de nodos de cómputo para ejecutar trabajos de cómputo en AWS PCS

En este paso, lanzará un grupo de nodos de cómputo que se escale de forma elástica para ejecutar los trabajos enviados al clúster.

Para crear el grupo de nodos de cómputo

- Abra la AWS PCSconsola y vaya a Clústeres.
- Seleccione el clúster denominado `get-started`
- Vaya a Compute nodes groups y elija Crear.
- En la sección de configuración del grupo de nodos de Compute, proporciona lo siguiente:
  - Nombre del grupo de nodos de cómputo: `compute-1` introdúzcalo.
- En Configuración informática, introduzca o seleccione estos valores:
  - EC2plantilla de lanzamiento: elija la plantilla de lanzamiento donde está el nombre `compute-getstarted-lt`
  - IAMperfil de instancia: elija el nombre del perfil de instancia `AWSPCS-getstarted-role`
  - Subredes: selecciona la subred por la que comienza el nombre. `hpc-networking:PrivateSubnetA`
  - Instancias: seleccione. `c6i.xlarge`

  - Configuración de escalado: para el recuento mínimo de instancias, introduzca**0**. Para el recuento máximo de instancias, introduzca. 4
- En Configuración adicional, especifique lo siguiente:
  - AMIID: seleccione el AMI lugar por el que empieza el nombre`aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11`.
- Seleccione Crear grupo de nodos de cómputo.

El campo Estado muestra la opción Crear mientras se aprovisiona el grupo de nodos de cómputo.

> **Important**
>
> Espere a que el campo Estado muestre Activo antes de continuar con el siguiente paso de este tutorial.

Crear una cola para gestionar los trabajos en AWS PCS

Debe enviar un trabajo a una cola para ejecutarlo. El trabajo permanece en la cola hasta que se AWS PCS programe su ejecución en un grupo de nodos de cómputo. Cada cola está asociada a uno o más grupos de nodos de cómputo, que proporcionan las EC2 instancias necesarias para realizar el procesamiento.

En este paso, creará una cola que utilice el grupo de nodos de procesamiento para procesar los trabajos.

Para crear una cola

- Abre la AWS PCSconsola.
- Seleccione el clúster denominado `get-started`.
- Ve a Compute node groups y asegúrate de que el estado del `compute-1` grupo sea Activo.

> **Important**
>
> El estado del `compute-1` grupo debe ser Activo antes de continuar con el siguiente paso.

- Navegue hasta Colas y elija Crear cola.
  - En la sección de configuración de colas, proporciona los siguientes valores:

  - Nombre de la cola: introduzca lo siguiente: `demo`
  - Grupos de nodos de cómputo: seleccione el nombre `compute-1` del grupo de nodos de cómputo.
- Elija Crear cola.

El campo Estado muestra la opción Crear mientras se crea la cola.

> **Important**
>
> Espere a que el campo Estado muestre Activo antes de continuar con el siguiente paso de este tutorial.

Conéctese a su AWS PCS clúster

Cuando el estado del grupo de nodos de `login` cómputo pase a ser Activo, podrás conectarte a la EC2 instancia que creó.

Para conectarse al nodo de inicio de sesión

- Abra la AWS PCSconsola y vaya a Clusters.
- Seleccione el clúster denominado `get-started`.
- Seleccione Grupos de nodos de cómputo.
- Navegue hasta el grupo de nodos de cómputo denominado`login`.
- Busque el ID del grupo de nodos de cómputo.
- En otra ventana o pestaña del navegador, abre la EC2consola de Amazon.
  - Elija Instances.
  - Busca EC2 instancias con la siguiente etiqueta. Reemplazar *`node-group-id`* con el valor del ID del grupo de nodos de Compute del paso anterior. Debe haber 1 instancia.

    ```
    aws:pcs:compute-node-group-id=node-group-id
    ```
  - Conéctese a la EC2 instancia. Puede usar el Administrador de sesiones oSSH.

    Session Manager

    - Seleccione la instancia.

- Elija Conectar.
- En Conectar a la instancia, selecciona Administrador de sesiones.
- Elija Conectar.
- Elija Conectar. Se abrirá un terminal interactivo en el navegador.

SSH

- Seleccione la instancia.
- Elija Conectar.
- En Conectar a la instancia, selecciona SSHcliente.
- Sigue las instrucciones que te proporciona la consola.

> **Note**
>
> El nombre de usuario de la instancia **`ec2-user`**no lo es`root`.

Explore el entorno de clústeres en AWS PCS

Una vez que haya iniciado sesión en el clúster, podrá ejecutar comandos de shell. Por ejemplo, puede cambiar de usuario, trabajar con datos en sistemas de archivos compartidos e interactuar con Slurm.

Cambiar de usuario

Si ha iniciado sesión en el clúster mediante el Administrador de sesiones, es posible que esté conectado como`ssm-user`. Se trata de un usuario especial creado para el Administrador de sesiones. Cambie al usuario predeterminado en Amazon Linux 2 mediante el siguiente comando. No necesitará hacer esto si se conectó usandoSSH.

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

Trabaje con sistemas de archivos compartidos

Puede confirmar que el EFS sistema de archivos y los sistemas FSx de archivos Lustre están disponibles con el comando. `df -h` El resultado del clúster debe ser similar al siguiente:

```
[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:/zlshxbev  1.2T  7.5M  1.2T   1% /shared
tmpfs                      780M     0  780M   0% /run/user/0
tmpfs                      780M     0  780M   0% /run/user/1000
```

El `/home` sistema de archivos monta 127.0.0.1 y tiene una capacidad muy grande. Este es el sistema de EFS archivos que creó anteriormente en el tutorial. Todos los archivos que se escriban aquí estarán disponibles `/home` en todos los nodos del clúster.

El `/shared` sistema de archivos monta una IP privada y tiene una capacidad de 1,2 TB. Este es el sistema de archivos FSx para Lustre que creó anteriormente en el tutorial. Todos los archivos que se escriban aquí estarán disponibles `/shared` en todos los nodos del clúster.

Interactúa con Slurm

Temas

- Enumere las colas y los nodos
- Mostrar trabajos

Enumere las colas y los nodos

Puede enumerar las colas y los nodos a los que están asociadas. `sinfo` El resultado del clúster debe ser similar al siguiente:

```
[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 ~]$
```

Anote el nombre de la particiónde`demo`. Su estado es up y tiene un máximo de 4 nodos. Está asociado a los nodos del grupo de `compute-1` nodos. Si edita el grupo de nodos de cómputo y aumenta el número máximo de instancias a 8, el número de nodos se leerá 8 y la lista de nodos se

leerá`compute-1-[1-8]`. Si creara un segundo grupo de nodos de cómputo `test` con un nombre de 4 nodos y lo añadiera a la demo cola, esos nodos también aparecerían en la lista de nodos.

Mostrar trabajos

Puede enumerar todos los trabajos, en cualquier estado, del sistema con`squeue`. El resultado del clúster debe ser similar al siguiente:

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

Intente `squeue` volver a ejecutarlo más tarde, cuando tenga un trabajo de Slurm pendiente o en ejecución.

Ejecute un trabajo de un solo nodo en AWS PCS

Para ejecutar un trabajo con Slurm, debe preparar un script de envío que especifique los requisitos del trabajo y enviarlo a una cola con el comando. `sbatch` Por lo general, esto se hace desde un directorio compartido, de modo que los nodos de inicio de sesión y de procesamiento tengan un espacio común para acceder a los archivos.

Conéctese al nodo de inicio de sesión de su clúster y ejecute los siguientes comandos en su intérprete de comandos.

- Conviértete en el usuario predeterminado. Cambie al directorio compartido.

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

- Utilice los siguientes comandos para crear un ejemplo de script de trabajo:

  ```
  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
```

- Envíe el script de trabajo al programador de Slurm:

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

- Cuando se envíe el trabajo, devolverá una ID de trabajo en forma de número. Usa ese identificador para comprobar el estado del trabajo. Reemplazar *job-id* en el siguiente comando con el número devuelto desde`sbatch`.

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

  Example

```
squeue --job 1
```

  El `squeue` comando devuelve un resultado similar al siguiente:

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

- Continúe comprobando el estado del trabajo hasta que alcance el estado `R` (en ejecución). El trabajo está hecho cuando `squeue` no devuelve nada.
- Inspeccione el contenido del `/shared` directorio.

```
ls -alth /shared
```

  El resultado del comando es similar al siguiente:

```
-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
```

  Uno de los nodos de cómputo del clúster `single.1.err` asignó un nombre `single.1.out` a los archivos y los escribió. Como el trabajo se ejecutó en un directorio compartido (`/shared`), también están disponibles en su nodo de inicio de sesión. Por eso configuró un sistema de archivos FSx para Lustre para este clúster.

- Inspeccione el contenido del `single.1.out` archivo.

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

El resultado es similar al siguiente:

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

Ejecute un MPI trabajo de varios nodos con Slurm en AWS PCS

Estas instrucciones muestran cómo usar Slurm para ejecutar un trabajo de interfaz de paso de mensajes (). MPI AWS PCS

Ejecute los siguientes comandos en el intérprete de comandos de su nodo de inicio de sesión.

- Conviértete en el usuario predeterminado. Cambie a su directorio principal.

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

- Cree el código fuente en el lenguaje de programación 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
```

- Cargue el MPI módulo Open.

```
module load openmpi
```

- Compila el programa C.

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

- Escribe un guion de envío de trabajos de 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
```

- Cambie al directorio compartido.

```
cd /shared
```

- Envíe el script de trabajo.

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

- `squeue`Úselo para supervisar el trabajo hasta que esté terminado.
- Compruebe el contenido de`multi.out`:

```
cat multi.out
```

El resultado es similar al siguiente. Tenga en cuenta que cada rango tiene su propia dirección IP porque se ejecutó en un nodo diferente.

```
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
```

Elimine sus AWS recursos para AWS PCS

Una vez que haya terminado con los grupos de clústeres y nodos que creó para este tutorial, debe eliminar los recursos que creó.

> **Important**
>
> Recibirás cargos de facturación por todos los recursos que estén en funcionamiento en tu Cuenta de AWS

Para eliminar AWS PCS los recursos que creó para este tutorial

- Abre la AWS PCSconsola.
- Navegue hasta el clúster denominado get-started.
- Navegue hasta la sección Colas.
- Seleccione la cola llamada demo.
- Elija Eliminar.

> **Important**
>
> Espere a que se elimine la cola antes de continuar.

- Navegue hasta la sección Grupos de nodos de cómputo.
- Seleccione el grupo de nodos de cómputo denominado compute-1.
- Elija Eliminar.
- Seleccione el grupo de nodos de cómputo denominado login.
- Elija Eliminar.

> **Important**
>
> Espere a que se eliminen ambos grupos de nodos de procesamiento antes de continuar.

- En la página de detalles del clúster para empezar, selecciona Eliminar.

> **Important**
>
> Espere a que se elimine el clúster antes de continuar con los pasos siguientes.

Para eliminar otros AWS recursos que haya creado para este tutorial

- Abre la IAMconsola.
  - Elija Roles.
  - Seleccione el rol denominado AWSPCS-getstarted-role y, a continuación, elija Eliminar.
  - Una vez que se haya eliminado el rol, elija Políticas.
  - Seleccione la política denominada AWSPCS-getstarted-policy y, a continuación, elija Eliminar.
- Abra la consola de AWS CloudFormation.
  - Seleccione la pila denominada getstarted-lt.
  - Elija Eliminar.

> **Important**
>
> Espere a que se elimine la pila antes de continuar.

- Abre la EFSconsola de Amazon.
  - Elija Sistemas de archivos.
  - Seleccione el sistema de archivos denominado getstarted-efs.
  - Elija Eliminar.

> **Important**
>
> Espere a que el sistema de archivos elimine antes de continuar.

- Abre la FSxconsola de Amazon.
  - Elija Sistemas de archivos.
  - Selecciona el sistema de archivos llamado getstarted-fsx.
  - Elija Eliminar.

> **Important**
>
> Espere a que el sistema de archivos elimine antes de continuar.

- Abra la consola de AWS CloudFormation.
  - Seleccione la pila denominada getstarted-sg.
  - Elija Eliminar.
- Abra la consola de AWS CloudFormation.
  - Seleccione la pila denominada hpc-networking.
  - Elija Eliminar.

Trabajando con AWS PCS

En este capítulo se proporciona información y orientación para ayudarle a utilizar AWS PCS.

Temas

AWS PCSclústeres

Un AWS PCS clúster consta de los siguientes componentes:

- Instancias administradas del software programador HPC del sistema, como el daemon de control Slurm (). `slurmctld`
- Componentes que se integran con el programador HPC del sistema para aprovisionar y gestionar las EC2 instancias de Amazon.
- Componentes que se integran con el programador HPC del sistema para transmitir registros y métricas a Amazon CloudWatch.

Estos componentes se ejecutan en una cuenta gestionada por AWS. Trabajan juntos para gestionar las EC2 instancias de Amazon en tu cuenta de cliente. AWS PCSaprovisiona interfaces de red elásticas en su VPC subred de Amazon para proporcionar conectividad desde el software del programador a las EC2 instancias de Amazon (por ejemplo, para permitir la programación de trabajos por lotes en ellas y permitir a los usuarios ejecutar comandos del programador para enumerar y administrar esos trabajos).

Temas

Creación de un clúster en AWS Parallel Computing Service

En este tema se proporciona una descripción general de las opciones disponibles y se describe lo que se debe tener en cuenta al crear un clúster en AWS Parallel Computing Service (AWS PCS). Si es la primera vez que crea un AWS PCS clúster, le recomendamos que haga lo siguienteEmpezar con AWS PCS. El tutorial puede ayudarle a crear un HPC sistema funcional sin necesidad de ampliar todas las opciones y arquitecturas de sistema disponibles.

Requisitos previos

- Una red existente VPC y una subred que cumplen con los requisitosAWS PCSRedes. Antes de implementar un clúster para su uso en producción, le recomendamos que conozca a fondo los requisitos de la subred VPC y los de la subred. Para crear una subred VPC y, consulte. Crear un VPC para tu AWS PCS clúster
- Un IAMdirector con permisos para crear y administrar AWS PCS recursos. Para obtener más información, consulte Servicio de Gestión de Identidad y Acceso para Computación AWS Paralela.

Crea un AWS PCS clúster

Puede usar AWS Management Console o AWS CLI para crear un clúster.

AWS Management Console

Pasos para crear un clúster

1. Abre la AWS PCS consola en https://console.aws.amazon.com/pcs/home#/clusters y selecciona Crear clúster.
2. En la sección Configuración del clúster, introduce los siguientes campos:
   - Nombre del clúster: un nombre para el clúster. El nombre solo puede contener caracteres alfanuméricos (con distinción de mayúsculas y minúsculas) y guiones. Debe empezar por

un carácter alfabético y no puede tener más de 40 caracteres. El nombre debe ser único dentro del grupo Región de AWS y en el Cuenta de AWS que se va a crear el clúster.

- Planificador: elige un programador y una versión. AWS PCSactualmente es compatible con Slurm 23.11. Para obtener más información, consulte Versiones de Slurm en AWS PCS.
- Tamaño del mando: elige un tamaño para el mando. Esto determina cuántos trabajos y nodos de cómputo simultáneos puede administrar el AWS PCS clúster. Solo puede establecer el tamaño de la controladora cuando se crea el clúster. Para obtener más información sobre el tamaño, consulteElegir un tamaño AWS PCS de clúster.

3. En la sección Redes, seleccione valores para los siguientes campos:

   - VPC— Elija uno existente VPC que cumpla con AWS PCS los requisitos. Para obtener más información, consulte AWS PCSVPCy requisitos y consideraciones de subred. Tras crear el clúster, no podrá cambiarloVPC. VPCsSi no aparece ninguno, primero debe crear uno.
   - Subred: se muestran todas las subredes disponibles en la seleccionadaVPC. Elija dos en zonas de disponibilidad diferentes. Cada subred debe cumplir los requisitos de AWS PCS subred. Para obtener más información, consulte AWS PCSVPCy requisitos y consideraciones de subred. Le recomendamos que seleccione una subred privada para evitar exponer los puntos finales de su programador a la Internet pública.
   - Grupos de seguridad: especifique los grupos de seguridad que desea asociar AWS PCS a las interfaces de red que crea para su clúster. Debe seleccionar al menos un grupo de seguridad que permita la comunicación entre el clúster y sus nodos de procesamiento. Para obtener más información, consulte Requisitos y consideraciones sobre los grupos de seguridad.

4. (Opcional) En Cifrado, puede definir una clave personalizada para cifrar los datos de su controlador configurando estos campos:

   - KMSID de clave: deje que `aws/pcs` se utilice la KMS clave que PCS crea. Seleccione un alias de KMS clave existente para usar una KMS clave personalizada. Tenga en cuenta que la cuenta utilizada para crear el clúster debe tener `kms:Decrypt` privilegios en la KMS clave personalizada.

5. (Opcional) En la sección de configuración de Slurm, puede especificar las opciones de configuración de Slurm que anulan los valores predeterminados establecidos por: AWS PCS

   - Reduzca el tiempo de inactividad: esto controla cuánto tiempo permanecen activos los nodos de cómputo aprovisionados dinámicamente después de que se completen o

finalicen los trabajos que se les asignan. Si se establece en un valor más largo, es más probable que se ejecute un trabajo posterior en el nodo, pero puede aumentar los costes. Un valor más bajo reducirá los costes, pero puede aumentar la proporción de tiempo que el HPC sistema dedica a aprovisionar nodos en lugar de a ejecutar tareas en ellos.

- Prolog: se trata de una ruta totalmente cualificada para acceder a un directorio de scripts de prolog en las instancias del grupo de nodos de cómputo. Esto corresponde a la configuración de Prolog en Slurm. Tenga en cuenta que debe ser un directorio, no una ruta a un ejecutable específico.
- Epilog: esta es una ruta totalmente cualificada para acceder a un directorio de scripts de epilog en las instancias de su grupo de nodos de cómputo. Esto corresponde a la configuración de Epilog en Slurm. Tenga en cuenta que debe ser un directorio, no una ruta a un ejecutable específico.
- Seleccione los parámetros de tipo: esto ayuda a controlar el algoritmo de selección de recursos utilizado por Slurm. Si se establece este valor en, `CR_CPU_Memory` se activará la programación compatible con la memoria, mientras que si se establece en, se `CR_CPU` activará la programación únicamente. CPU Este parámetro corresponde a la SelectTypeParametersconfiguración de Slurm, donde se `SelectType` establece en. `select/cons_tres` AWS PCS

6. (Opcional) En Etiquetas, agrega cualquier etiqueta a tu AWS PCS clúster.
7. Elija Create cluster. El campo Estado `Creating` se muestra mientras AWS PCS crea el clúster. Este proceso puede tardar varios minutos.

> **Important**
>
> Solo puede haber 1 clúster en un `Creating` estado Región de AWS por cada uno Cuenta de AWS. AWS PCSdevuelve un error si ya hay un clúster en un `Creating` estado al intentar crear un clúster.

AWS CLI

Pasos para crear un clúster

1. Cree el clúster con el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:

- Reemplazar *region* con el ID del clúster en el Región de AWS que quieres crear el clúster, por ejemplous-east-1.
- Reemplazar *my-cluster* con un nombre para el clúster. El nombre solo puede contener caracteres alfanuméricos (con distinción de mayúsculas y minúsculas) y guiones. Debe empezar por un carácter alfabético y no puede tener más de 40 caracteres. El nombre debe ser único dentro del clúster Región de AWS y en el Cuenta de AWS lugar en el que vaya a crearlo.
- Reemplazar *23.11* con cualquier versión compatible de Slurm.

  > **Note**
  >
  > AWS PCSactualmente es compatible con Slurm 23.11.

- Reemplazar *SMALL* con cualquier tamaño de clúster compatible. Esto determina cuántos trabajos y nodos de cómputo simultáneos puede administrar el AWS PCS clúster. Solo se puede configurar cuando se crea el clúster. Para obtener más información sobre el tamaño, consulteElegir un tamaño AWS PCS de clúster.
- Sustituya el valor de por `subnetIds` el suyo propio. Le recomendamos que seleccione una subred privada para evitar exponer los puntos finales de su programador a la Internet pública.
- Especifique lo `securityGroupIds` que desea asociar AWS PCS a las interfaces de red que crea para su clúster. Los grupos de seguridad deben estar en el VPC mismo lugar que el clúster. Debe seleccionar al menos un grupo de seguridad que permita la comunicación entre el clúster y sus nodos de procesamiento. Para obtener más información, consulte Requisitos y consideraciones sobre los grupos de seguridad.
- Si lo desea, puede ajustar el comportamiento de Slurm añadiendo una opción. `--slurm-configration` Por ejemplo, puede establecer el tiempo de inactividad reducido en 60 minutos (3600 segundos) con. `--slurm configuration scaleDownIdeTime=3600`
- Si lo desea, puede proporcionar una KMS clave personalizada para cifrar los datos de su controlador. `--kms-key-id` *kms-key* *kms-key*Sustitúyala por un seudónimo o identificador clave existente KMSARN. Tenga en cuenta que la cuenta utilizada para crear el clúster debe tener `kms:Decrypt` privilegios en la KMS clave personalizada.

```
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. El aprovisionamiento del clúster puede tardar varios minutos. Puede consultar el estado del clúster con el siguiente comando. No proceda a crear colas ni a calcular grupos de nodos hasta que aparezca el campo de estado del clúster. `ACTIVE`

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

> **Important**
>
> Solo puede haber 1 clúster en un `Creating` estado Región de AWS por cada uno. Cuenta de AWS AWS PCSdevuelve un error si ya hay un clúster en un `Creating` estado al intentar crear un clúster.

Próximos pasos recomendados para su clúster

- Agregue grupos de nodos de cómputo.
- Añada colas.
- Habilitar el registro.

Eliminar un clúster en AWS PCS

En este tema se proporciona información general sobre cómo eliminar un AWS PCS clúster.

Consideraciones a la hora de eliminar un AWS PCS clúster

- Se deben eliminar todas las colas asociadas al clúster antes de poder eliminar el clúster. Para obtener más información, consulte Eliminar una cola en AWS PCS.
- Todos los grupos de nodos de procesamiento asociados al clúster deben eliminarse antes de poder eliminar el clúster. Para obtener más información, consulte Eliminar un grupo de nodos de cómputo en AWS PCS.

Elimine el clúster

Puede usar AWS Management Console o AWS CLI para eliminar un clúster.

AWS Management Console

Para eliminar un clúster

1. Abre la AWS PCSconsola.
2. Seleccione el clúster que desee eliminar.
3. Elija Eliminar.
4. Aparece el campo Estado del clúster`Deleting`. Puede tardar varios minutos en completarse.

AWS CLI

Para eliminar un clúster

1. Use el siguiente comando para eliminar un clúster, con estas sustituciones:
   - Reemplazar *`region-code`* con el Región de AWS clúster en el que se encuentra.
   - Reemplazar *`my-cluster`* con el nombre o el ID de su clúster.

   ```
   aws pcs delete-cluster --region region-code --cluster-identifier my-cluster
   ```
2. Eliminar el clúster puede tardar varios minutos. Puede comprobar el estado del clúster con el siguiente comando.

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

Elegir un tamaño AWS PCS de clúster

AWS PCSproporciona clústeres seguros y de alta disponibilidad, a la vez que automatiza tareas clave como la aplicación de parches, el aprovisionamiento de nodos y las actualizaciones.

Al crear un clúster, se selecciona su tamaño en función de dos factores:

- La cantidad de nodos de cómputo que administrará

- La cantidad de trabajos activos y en cola que espera ejecutar en el clúster

Tamaño del clúster de Slurm	Número de instancias administradas	Número de trabajos activos y en cola
Pequeña	Hasta 32	Hasta 256
Medio	Hasta 512	Hasta 8192
Grande	Hasta 2048	Hasta 16384

Ejemplos

- Si su clúster tendrá hasta 24 instancias administradas y ejecutará hasta 100 trabajos, elija Small.
- Si el clúster tendrá hasta 24 instancias administradas y ejecutará hasta 1000 trabajos, elija Medium.
- Si su clúster tendrá hasta 1000 instancias administradas y ejecutará hasta 100 trabajos, elija Grande.
- Si el clúster tendrá hasta 1000 instancias administradas y ejecutará hasta 10 000 trabajos, elija Grande.

Trabajar con secretos de clústeres en AWS PCS

Como parte de la creación de un clúster, AWS PCS crea un secreto de clúster que es necesario para conectarse al programador de tareas del clúster. También se crean grupos de nodos de AWS PCS cómputo, que definen conjuntos de instancias que se van a lanzar en respuesta a eventos de escalado. AWS PCSconfigura las instancias lanzadas por esos grupos de nodos de cómputo con el secreto del clúster para que puedan conectarse al programador de tareas. Hay casos en los que es posible que desee configurar los clientes de Slurm manualmente. Los ejemplos incluyen la creación de un nodo de inicio de sesión persistente o la configuración de un administrador de flujo de trabajo con capacidades de administración de trabajos.

AWS PCSalmacena el secreto del clúster como un secreto administrado con el prefijo `pcs!` in AWS Secrets Manager. El coste del secreto está incluido en el coste de su uso AWS PCS.

> **Warning**
>
> No modifique el secreto de su clúster. AWS PCSno podrás comunicarte con tu clúster si modificas el secreto de tu clúster. AWS PCSno admite la rotación del secreto del clúster. Debe crear un clúster nuevo si necesita modificar su secreto de clúster.

Contenido

Encuentra el secreto del clúster de Slurm

Puedes encontrar los secretos AWS PCS gestionados desde la AWS Secrets Manager consolaAPI, directamente desde o mediante AWS PCS etiquetas.

Se usa AWS Secrets Manager para encontrar el secreto del clúster

AWS Management Console

1. Vaya a la consola del administrador de secretos.
2. Selecciona Secretos y busca el `pcs!` prefijo.

> **Note**
>
> El secreto de un AWS PCS clúster tiene un nombre en el formato `pcs!slurm-secret-`*`cluster-id`* donde *`cluster-id`* está el ID del AWS PCS clúster.

AWS CLI

Cada secreto de AWS PCS clúster también está etiquetado con`aws:pcs:`*`cluster-id`*. Puede obtener el identificador secreto de un clúster con el siguiente comando. Realice estas sustituciones antes de ejecutar el comando:

- *region* Región de AWS Sustitúyalo por el para crear el clúster, por ejemplo`us-east-1`.
- *cluster-id*Sustitúyalo por el ID del AWS PCS clúster para buscar el secreto del clúster.

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

Se usa AWS PCS para encontrar el secreto del clúster

Puedes usar el AWS CLI para encontrar el ARN secreto de un AWS PCS clúster. Ingresa el siguiente comando y realiza las siguientes sustituciones:

- *region* Región de AWS Sustitúyalo por el para crear el clúster, por ejemplo`us-east-1`.
- *my-cluster*Sustitúyalo por el nombre o identificador del clúster.

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

El siguiente ejemplo de salida proviene del `get-cluster` comando. Puedes usar `secretArn` y `secretVersion` juntos para obtener el secreto.

```
{
    "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"
    }
}
```

Obtén el secreto del cúmulo de Slurm

Puede usar Secrets Manager para obtener la versión actual codificada en base64 de un secreto de clúster de Slurm. El siguiente ejemplo usa el. AWS CLI Realice las siguientes sustituciones antes de ejecutar el comando.

- *`region`* Región de AWS Sustitúyala por la para crear el clúster, por ejemplo`us-east-1`.
- *`secret-arn`*Sustitúyalo `secretArn` por el de un AWS PCS clúster.

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

Para obtener información sobre cómo utilizar el secreto del clúster de Slurm, consulte. Uso de instancias independientes como nodos de inicio de AWS PCS sesión

Permisos

Utiliza un IAM principal para obtener el secreto del clúster de Slurm. El IAM director debe tener permiso para leer el secreto. Para obtener más información, consulte los términos y conceptos de las funciones en la Guía del AWS Identity and Access Management usuario.

El siguiente ejemplo de IAM política permite el acceso a un ejemplo de secreto de clúster.

```
{
    "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 PCScalcular grupos de nodos

Un grupo de nodos de AWS PCS cómputo es un conjunto lógico de nodos (EC2instancias de Amazon). Estos nodos se pueden usar para ejecutar tareas informáticas, así como para proporcionar acceso interactivo a un HPC sistema basado en shell. Un grupo de nodos de cómputo consta de reglas para crear nodos, que incluyen qué tipos de instancias de Amazon EC2 usar, cuántas instancias ejecutar, si usar instancias puntuales o instancias bajo demanda, qué subredes y grupos de seguridad usar y cómo configurar cada instancia cuando se lanza. Cuando esas reglas se actualizan, AWS PCS actualiza los recursos asociados al grupo de nodos de cómputo para que coincidan.

Temas

- Crear un grupo de nodos de cómputo en AWS PCS
- Actualización de un grupo de nodos de AWS PCS cómputo
- Eliminar un grupo de nodos de cómputo en AWS PCS
- Búsqueda de instancias de grupos de nodos de cómputo en AWS PCS

Crear un grupo de nodos de cómputo en AWS PCS

En este tema se proporciona una descripción general de las opciones disponibles y se describe lo que se debe tener en cuenta al crear un grupo de nodos de procesamiento en AWS Parallel Computing Service (AWS PCS). Si es la primera vez que crea un grupo de nodos de cálculo en AWS PCS, le recomendamos que siga el tutorial que aparece enEmpezar con AWS PCS. El tutorial puede ayudarle a crear un HPC sistema funcional sin necesidad de ampliar todas las opciones y arquitecturas de sistema disponibles.

Requisitos previos

- Cuotas de servicio suficientes para lanzar el número deseado de EC2 instancias en su Región de AWS. Puede utilizarlas AWS Management Consolepara comprobar y solicitar aumentos en sus cuotas de servicio.
- Una red existente VPC y una subred que cumplen con los requisitos AWS PCS de red. Le recomendamos que comprenda detenidamente estos requisitos antes de implementar un clúster para su uso en producción. Para obtener más información, consulte AWS PCSVPCy requisitos y consideraciones de subred. También puede usar una CloudFormation plantilla para crear subredes VPC y. AWS proporciona una HPC receta para la CloudFormation plantilla. Para obtener más información, consulte aws-hpc-recipesen GitHub.
- Un perfil de IAM instancia con permisos para convocar la AWS PCS `RegisterComputeNodeGroupInstance` API acción y acceder a cualquier otro AWS recurso necesario para las instancias de tu grupo de nodos. Para obtener más información, consulte IAMperfiles de instancia para AWS Parallel Computing Service.
- Una plantilla de lanzamiento para las instancias de tu grupo de nodos. Para obtener más información, consulte Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS.
- Para crear un grupo de nodos de cómputo que utilice instancias de Amazon EC2 Spot, debe tener el rol AWSServiceRoleForEC2Spotvinculado al servicio en su. Cuenta de AWS Para obtener más información, consulte Función de Amazon EC2 Spot para AWS PCS.

Cree un grupo de nodos de cómputo en AWS PCS

Puede crear un grupo de nodos de cómputo mediante el AWS Management Console o el AWS CLI.

AWS Management Console

Para crear su grupo de nodos de cómputo mediante la consola

1. Abre la AWS PCSconsola.
2. Seleccione el clúster en el que desee crear un grupo de nodos de procesamiento. Diríjase a los grupos de nodos de cómputo y elija Crear.
3. En la sección de configuración del grupo de nodos de Compute, proporciona un nombre para el grupo de nodos. El nombre solo puede contener caracteres alfanuméricos y guiones que distingan mayúsculas de minúsculas. Debe empezar por un carácter alfabético y no puede tener más de 25 caracteres. El nombre debe ser único en el clúster.
4. En Configuración informática, introduzca o seleccione estos valores:

   a. EC2Plantilla de lanzamiento: seleccione una plantilla de lanzamiento personalizada para utilizarla en este grupo de nodos. Las plantillas de lanzamiento se pueden utilizar para personalizar la configuración de la red, como la subred y los grupos de seguridad, la configuración de supervisión y el almacenamiento a nivel de instancia. Si no tienes una plantilla de lanzamiento preparada, consulta Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS para aprender a crear una.

   > **Important**
   >
   > AWS PCScrea una plantilla de lanzamiento gestionada para cada grupo de nodos de procesamiento. Estos se denominan`pcs-`*`identifier`*`-do-not-delete`. No los seleccione cuando cree o actualice un grupo de nodos de procesamiento, o el grupo de nodos no funcionará correctamente.

   b. EC2versión de la plantilla de lanzamiento: selecciona una versión de tu plantilla de lanzamiento personalizada. Puede elegir una versión específica, lo que puede mejorar la reproducibilidad. Si cambia la versión más adelante, debe actualizar el grupo de nodos de procesamiento para detectar cambios en la plantilla de lanzamiento. Para obtener más información, consulte Actualización de un grupo de nodos de AWS PCS cómputo.

   c. AMIID: si tu plantilla de lanzamiento no incluye un AMI ID o si quieres anular el valor de la plantilla de lanzamiento, introduce un AMI ID aquí. Tenga en cuenta que el AMI utilizado para el grupo de nodos debe ser compatible con AWS PCS. También puede seleccionar una muestra AMI proporcionada por AWS. Para obtener más información sobre este tema, consulteAmazon Machine Images (AMIs) para AWS PCS.

d. IAMperfil de instancia: elija un perfil de instancia para el grupo de nodos. Un perfil de instancia otorga a la instancia permisos para acceder a AWS los recursos y servicios de forma segura. Si no tiene uno preparado, consulte IAMperfiles de instancia para AWS Parallel Computing Service para obtener información sobre cómo crear uno.

e. Subredes: elija una o más subredes en el VPC lugar donde se implementa el AWS PCS clúster. Si seleccionas varias subredes, EFA las comunicaciones no estarán disponibles entre los nodos y la comunicación entre los nodos de distintas subredes podría aumentar la latencia. Asegúrese de que las subredes que especifique aquí coincidan con las que haya definido en la EC2 plantilla de lanzamiento.

f. Instancias: elija uno o más tipos de instancias para cumplir con las solicitudes de escalado del grupo de nodos. Todos los tipos de instancias deben tener la misma arquitectura de procesador (x864_64 o arm64) y el mismo número de. vCPUs Si las instancias lo tienenGPUs, todos los tipos de instancias deben tener el mismo número de. GPUs

g. Configuración de escalado: especifique la cantidad mínima y máxima de instancias para el grupo de nodos. Puede definir una configuración estática, en la que hay un número fijo de nodos en ejecución, o una configuración dinámica, en la que se pueden ejecutar hasta el número máximo de nodos. Para una configuración estática, defina el mínimo y el máximo en el mismo número, superior a cero. Para una configuración dinámica, establece el número mínimo de instancias en cero y el máximo en un número superior a cero. AWS PCSno admite grupos de nodos de cómputo con una combinación de instancias estáticas y dinámicas.

5. (Opcional) En Configuración adicional, especifique lo siguiente:

   a. Opción de compra: seleccione entre instancias puntuales y bajo demanda.

   b. Estrategia de asignación: si ha seleccionado la opción de compra puntual, puede especificar cómo se eligen los grupos de capacidad puntuales al lanzar instancias en el grupo de nodos. Para obtener más información, consulte Estrategias de asignación para instancias puntuales en la Guía del usuario de Amazon Elastic Compute Cloud. Esta opción no tiene efecto si ha seleccionado la opción de compra bajo demanda.

6. (Opcional) En la sección de configuración Slurm personalizada, introduce estos valores:

   a. Peso: este valor establece la prioridad de los nodos del grupo para fines de programación. Los nodos con pesos más bajos tienen mayor prioridad y las unidades son arbitrarias. Para obtener más información, consulte Peso en la Slurm documentación.

b. Memoria real: este valor establece el tamaño (en GB) de la memoria real en los nodos del grupo de nodos. Está pensado para usarse junto con la `CR_CPU_Memory` opción de la Slurm configuración del clúster en AWS PCS. Para obtener más información, consulte RealMemoryla Slurm documentación.

7. (Opcional) En Etiquetas, agrega cualquier etiqueta a tu grupo de nodos de cómputo.
8. Selecciona Crear grupo de nodos de cómputo. El campo Estado muestra el grupo de nodos `Creating` mientras se AWS PCS aprovisiona. Esto puede tardar varios minutos.

Siguiente paso recomendado

- Añada su grupo de nodos a una cola AWS PCS para que pueda procesar los trabajos.

AWS CLI

Para crear su grupo de nodos de cómputo mediante AWS CLI

Cree su cola con el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:

1. Reemplazar *`region`*con el ID en el Región de AWS que se va a crear el clúster, por ejemplo`us-east-1`.
2. Reemplazar *`my-cluster`* con el nombre o el `clusterId` de tu clúster.
3. Reemplazar *`my-node-group`*con el nombre de su grupo de nodos de cómputo. El nombre solo puede contener caracteres alfanuméricos (con distinción de mayúsculas y minúsculas) y guiones. Debe empezar por un carácter alfabético y no puede tener más de 25 caracteres. El nombre debe ser único en el clúster.
4. Reemplazar *`subnet-ExampleID1`* con una o más subredes IDs de su clústerVPC.
5. Reemplazar *`lt-ExampleID1`* con el ID de tu plantilla de lanzamiento personalizada. Si no tienes una preparada, consulta Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS para aprender a crearla.

> **Important**
>
> AWS PCScrea una plantilla de lanzamiento gestionado para cada grupo de nodos de cómputo. Estos se denominan`pcs-`*`identifier`*`-do-not-delete`. No los

seleccione cuando cree o actualice un grupo de nodos de procesamiento, o el grupo de nodos no funcionará correctamente.

6. Reemplazar *launch-template-version* con una versión de plantilla de lanzamiento específica si quieres asociar tu grupo de nodos a una versión específica.
7. Reemplazar *arn:InstanceProfile*con el perfil ARN de tu IAM instancia. Si no tiene uno preparado, consulte Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS para obtener orientación.
8. Reemplazar *min-instances* y *max-instances* con valores enteros. Puede definir una configuración estática, en la que hay un número fijo de nodos en ejecución, o una configuración dinámica, en la que se puede ejecutar hasta el número máximo de nodos. Para una configuración estática, defina el mínimo y el máximo en el mismo número, superior a cero. Para una configuración dinámica, establece el número mínimo de instancias en cero y el máximo en un número superior a cero. AWS PCSno admite grupos de nodos de cómputo con una combinación de instancias estáticas y dinámicas.
9. Reemplazar *t3.large* con otro tipo de instancia. Puede añadir más tipos de instancias especificando una lista de `instanceType` ajustes. Por ejemplo: *--instance-configs instanceType=c6i.16xlarge,instanceType=c6a.16xlarge*. Todos los tipos de instancias deben tener la misma arquitectura de procesador (x864_64 o arm64) y el mismo número de. vCPUs Si las instancias lo tienenGPUs, todos los tipos de instancias deben tener el mismo número de. 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
```

Hay varios ajustes de configuración opcionales que puedes añadir al `create-compute-node-group` comando.

- Puede especificar `--amiId` si su plantilla de lanzamiento personalizada no incluye una referencia a un AMI valor o si desea anular ese valor. Ten en cuenta que la AMI utilizada para el grupo de nodos debe ser compatible con AWS PCS. También puede seleccionar

una muestra AMI proporcionada por AWS. Para obtener más información sobre este tema, consulteAmazon Machine Images (AMIs) para AWS PCS.

- Puede seleccionar entre instancias bajo demanda (`ONDEMAND`) y Spot (`SPOT`) utilizando`--purchase-option`. Bajo demanda es la opción predeterminada. Si elige las instancias puntuales, también puede utilizarlas `--allocation-strategy` para definir cómo se AWS PCS eligen los grupos de capacidad puntuales al lanzar instancias en el grupo de nodos. Para obtener más información, consulte Estrategias de asignación para instancias puntuales en la Guía del usuario de Amazon Elastic Compute Cloud.
- Es posible proporcionar opciones de Slurm configuración para los nodos del grupo de nodos mediante`--slurm-configuration`. Puede establecer el peso (prioridad de programación) y la memoria real. Los nodos con pesos más bajos tienen mayor prioridad y las unidades son arbitrarias. Para obtener más información, consulte Peso en la Slurm documentación. La memoria real establece el tamaño (en GB) de la memoria real en los nodos del grupo de nodos. Se ha diseñado para usarse junto con la `CR_CPU_Memory` opción para el clúster AWS PCS de su Slurm configuración. Para obtener más información, consulte RealMemoryla Slurm documentación.

> **Important**
>
> La creación del grupo de nodos de procesamiento puede tardar varios minutos.

Puede consultar el estado de su grupo de nodos con el siguiente comando. No podrás asociar el grupo de nodos a una cola hasta que se alcance `ACTIVE` su estado.

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

Actualización de un grupo de nodos de AWS PCS cómputo

En este tema se proporciona una descripción general de las opciones disponibles y se describe lo que se debe tener en cuenta al actualizar un grupo de AWS PCS nodos de procesamiento.

Opciones para actualizar un grupo de nodos de AWS PCS cómputo

La actualización de un grupo de AWS PCS nodos de procesamiento le permite cambiar las propiedades de las instancias lanzadas por AWSPCS, así como las reglas de lanzamiento de esas instancias. Por ejemplo, puedes reemplazar las instancias del grupo de nodos AMI por otras que tengan instalado un software diferente. O bien, puede actualizar los grupos de seguridad para cambiar la conectividad de red entrante o saliente. También puede cambiar la configuración de escalado o incluso cambiar la opción de compra preferida hacia o desde las instancias Spot.

La siguiente configuración del grupo de nodos no se puede modificar después de su creación:

- Nombre
- instancias

Consideraciones a la hora de actualizar un grupo de nodos de AWS PCS cómputo

Los grupos de nodos de cómputo definen las EC2 instancias que se utilizan para procesar trabajos, proporcionar acceso interactivo al shell y otras tareas. Suelen estar asociados a una o más AWS PCS colas. Al actualizar el grupo de nodos de procesamiento para cambiar su comportamiento (o el de sus nodos), tenga en cuenta lo siguiente:

- Los cambios en las propiedades del grupo de nodos de cómputo se hacen efectivos cuando el estado del grupo de nodos de cómputo cambia de Actualizado a Activo. Se lanzan nuevas instancias con las propiedades actualizadas.
- Las actualizaciones que no afectan a la configuración de nodos específicos no afectan a los nodos en ejecución. Por ejemplo, añadir una subred y cambiar la estrategia de asignación.
- Si actualiza la plantilla de lanzamiento de un grupo de nodos de cómputo, debe actualizar el grupo de nodos de cómputo para usar la nueva versión.
- Para añadir o eliminar un grupo de seguridad de los nodos de un grupo de nodos de procesamiento, edite su plantilla de lanzamiento y actualice el grupo de nodos de procesamiento. Se lanzan nuevas instancias con el conjunto actualizado de grupos de seguridad.
- Si editas directamente un grupo de seguridad utilizado por un grupo de nodos de procesamiento, esto tendrá efecto inmediato en las instancias en ejecución y en las futuras.
- Si agregas o eliminas permisos del perfil de IAM instancia utilizado por un grupo de nodos de cómputo, esto tendrá efecto inmediato en las instancias en ejecución y en las futuras.

- Para cambiar los AMI utilizados por las instancias de un grupo de nodos de cómputo, actualiza el grupo de nodos de cómputo (o su plantilla de lanzamiento) para usar el nuevo AMI y espera AWS PCS a que se reemplacen las instancias.
- AWS PCSreemplaza las instancias existentes en el grupo de nodos tras una operación de actualización del grupo de nodos. Si hay trabajos en ejecución en un nodo, se permite que dichos trabajos se completen antes de AWS PCS reemplazar el nodo. Los procesos de usuario interactivos (por ejemplo, en las instancias de nodos de inicio de sesión) finalizan. El estado del grupo de nodos vuelve al `Active` momento en que se AWS PCS marcan las instancias que se van a reemplazar, pero la sustitución real se produce cuando las instancias están inactivas.
- Si reduces el número máximo de instancias permitido en un grupo de nodos de cómputo, AWS PCS elimina los nodos de Slurm para cumplir con el nuevo máximo. AWS PCStermina las instancias en ejecución asociadas a los nodos de Slurm eliminados. Los trabajos en ejecución en los nodos eliminados fallan y vuelven a sus colas.
- AWS PCScrea una plantilla de lanzamiento gestionada para cada grupo de nodos de procesamiento. Se nombran`pcs-`*`identifier`*`-do-not-delete`. No los seleccione al crear o actualizar un grupo de nodos de procesamiento, o el grupo de nodos no funcionará correctamente.
- Si actualiza un grupo de nodos de cómputo para usar Spot como opción de compra, debe tener el rol AWSServiceRoleForEC2Spotvinculado al servicio en su cuenta. Para obtener más información, consulte Función de Amazon EC2 Spot para AWS PCS.

Para actualizar un grupo de nodos de AWS PCS cómputo

Puede actualizar un grupo de nodos mediante la consola AWS de administración o el AWSCLI.

AWS Management Console

Para actualizar un grupo de nodos de cómputo

1. Abra la AWS PCS consola en `https://console.aws.amazon.com/pcs/home#/clusters`
2. Seleccione el clúster en el que desee actualizar un grupo de nodos de cómputo.
3. Vaya a Grupos de nodos de cómputo, vaya al grupo de nodos que desee actualizar y, a continuación, seleccione Editar.
4. En las secciones Configuración informática, Ajustes adicionales y Ajustes de Slurmpersonalización, actualiza todos los valores excepto:

- Instancias: no puede cambiar las instancias de un grupo de nodos de cómputo.

5. Elija Actualizar. El campo Estado mostrará la actualización mientras se aplican los cambios.

> **Important**
>
> Las actualizaciones de los grupos de nodos de cómputo pueden tardar varios minutos.

AWS CLI

Para actualizar un grupo de nodos de cómputo

1. Actualice su grupo de nodos de cómputo con el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:

   a. Reemplazar *region-code* con la AWS región en la que quieres crear tu clúster.

   b. Reemplazar *my-node-group* con el nombre o `computeNodeGroupId` para su grupo de nodos de procesamiento.

   c. Reemplazar *my-cluster* con el nombre o el `clusterId` de su clúster.

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

2. Actualice todos los parámetros del grupo de nodos excepto`--instance-configs`. Por ejemplo, para establecer un nuevo AMI ID, introduce `--amiId my-custom-ami-id` dónde *my-custom-ami-id* se sustituye por el que elijas. AMI

> **Important**
>
> La actualización del grupo de nodos de procesamiento puede tardar varios minutos.

Puedes consultar el estado de tu grupo de nodos con el siguiente comando.

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

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

Eliminar un grupo de nodos de cómputo en AWS PCS

En este tema se proporciona una descripción general de las opciones disponibles y se describe lo que se debe tener en cuenta al eliminar un grupo de nodos de procesamiento en él AWS PCS.

Consideraciones a la hora de eliminar un grupo de nodos de cómputo

Los grupos de nodos de cómputo definen las EC2 instancias que se utilizan para procesar trabajos, proporcionar acceso interactivo al shell y otras tareas. Suelen estar asociados a una o más AWS PCS colas. Antes de eliminar un grupo de nodos de procesamiento, tenga en cuenta lo siguiente:

- Se cancelará cualquier EC2 instancia lanzada por el grupo de nodos de cómputo. Esto cancelará los trabajos que se estén ejecutando en estas instancias y pondrá fin a la ejecución de los procesos interactivos.
- Debe desasociar el grupo de nodos de cómputo de todas las colas antes de poder eliminarlo. Para obtener más información, consulte Actualización de una AWS PCS cola.

Elimine el grupo de nodos de cómputo

Puede usar AWS Management Console o AWS CLI para eliminar un grupo de nodos de procesamiento.

AWS Management Console

Para eliminar un grupo de nodos de cómputo

1. Abre la AWS PCSconsola.
2. Seleccione el clúster del grupo de nodos de procesamiento.
3. Vaya a Grupos de nodos de cómputo y seleccione el grupo de nodos de cómputo que desee eliminar.
4. Elija Eliminar.
5. Aparece el campo Estado`Deleting`. Puede tardar varios minutos en completarse.

> **Note**
>
> Puede usar los comandos nativos de su programador para confirmar que se ha eliminado el grupo de nodos de procesamiento. Por ejemplo, usa `sinfo` o `squeue` para Slurm.

AWS CLI

Para eliminar un grupo de nodos de cómputo

- Usa el siguiente comando para eliminar un grupo de nodos de cómputo, con estos reemplazos:
  - Reemplazar *region-code* con el Región de AWS clúster en el que se encuentra.
  - Reemplazar *my-node-group* con el nombre o el ID de su grupo de nodos de cómputo.
  - Reemplazar *my-cluster* con el nombre o el ID de su clúster.

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

Eliminar el grupo de nodos de procesamiento puede tardar varios minutos.

> **Note**
>
> Puede usar los comandos nativos de su programador para confirmar que se ha eliminado el grupo de nodos de procesamiento. Por ejemplo, usa `sinfo` o `squeue` para Slurm.

Búsqueda de instancias de grupos de nodos de cómputo en AWS PCS

Cada grupo de nodos de AWS PCS cómputo puede lanzar EC2 instancias con configuraciones compartidas. Puede usar EC2 etiquetas para buscar instancias en un grupo de nodos de cómputo en AWS Management Console o con AWS CLI.

AWS Management Console

Para encontrar las instancias de tu grupo de nodos de cómputo

1. Abre la AWS PCSconsola.
2. Seleccione el clúster.
3. Elija grupos de nodos de cómputo.
4. Busca el ID del grupo de nodos de inicio de sesión que creaste.
5. Navegue hasta la EC2consola y elija Instancias.
6. Busque las instancias con la siguiente etiqueta. Reemplazar *node-group-id* con el ID (no el nombre) de su grupo de nodos de cómputo.

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

7. (Opcional) Puede cambiar el valor del estado de la instancia en el campo de búsqueda para buscar las instancias que se estén configurando o que se hayan cancelado recientemente.
8. Busca el ID de instancia y la dirección IP de cada instancia en la lista de instancias etiquetadas.

AWS CLI

Para encontrar las instancias de tu grupo de nodos, usa los siguientes comandos. Antes de ejecutar los comandos, realiza las siguientes sustituciones:

- *region-code*Sustitúyalo por el Región de AWS de su clúster. Ejemplo: `us-east-1`
- *node-group-id*Sustitúyalo por el ID (no el nombre) de tu grupo de nodos de procesamiento.
- `running`Reemplácelo por otros estados de instancia, como `pending` o `terminated` para buscar EC2 instancias en otros estados.

```
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}
```

El comando devuelve un resultado similar al siguiente. El valor de `PublicIP` es `null` si la instancia está en una subred privada.

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

> **Note**
>
> Si espera `describe-instances` devolver una gran cantidad de instancias, debe usar las opciones para varias páginas. Para obtener más información, consulte DescribeInstancesla APIreferencia de Amazon Elastic Compute Cloud.

Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS

En AmazonEC2, una plantilla de lanzamiento puede almacenar un conjunto de preferencias para que no tengas que especificarlas de forma individual al lanzar instancias. AWS PCSincorpora plantillas de lanzamiento como una forma flexible de configurar grupos de nodos de cómputo. Al crear un grupo de nodos, se proporciona una plantilla de lanzamiento. AWS PCScrea una plantilla de lanzamiento derivada a partir de ella que incluye transformaciones para garantizar que funciona con el servicio.

Entender cuáles son las opciones y las consideraciones a tener en cuenta a la hora de escribir una plantilla de lanzamiento personalizada puede ayudarte a crear una que puedas utilizar con ella AWS PCS. Para obtener más información sobre las plantillas de lanzamiento, consulte Lanzar una instancia desde una plantilla de lanzamiento en la Guía del EC2 usuario de Amazon.

Temas

- Información general
- Creación de una plantilla de lanzamiento básica
- Trabajar con datos de EC2 usuario de Amazon

- Reservas de aforo en AWS PCS
- Parámetros útiles de la plantilla de lanzamiento

Información general

Hay más de 30 parámetros disponibles que puede incluir en una plantilla de EC2 lanzamiento, que controlan muchos aspectos de la configuración de las instancias. La mayoría son totalmente compatibles con AWS PCS, pero hay algunas excepciones.

Se ignorarán los siguientes parámetros de la plantilla de EC2 lanzamiento, AWS PCS ya que el servicio debe administrar directamente estas propiedades:

- Tipo de instancia/especifique los atributos del tipo de instancia (`InstanceRequirements`): no AWS PCS admite la selección de instancias basada en atributos.
- Tipo de instancia (`InstanceType`): especifique los tipos de instancias al crear un grupo de nodos.
- Detalles avanzados/perfil de IAM instancia (`IamInstanceProfile`): los proporcionas al crear o actualizar el grupo de nodos.
- Detalles avanzados/Desactivar la API terminación (`DisableApiTermination`): AWS PCS debe controlar el ciclo de vida de las instancias del grupo de nodos que lanza.
- Detalles avanzados/Disable API stop (`DisableApiStop`): AWS PCS debe controlar el ciclo de vida de las instancias del grupo de nodos que lanza.
- Detalles avanzados/Comportamiento de stop — Hibernate (`HibernationOptions`): AWS PCS no admite la hibernación de instancias.
- Advanced Details/Elastic GPU (`ElasticGpuSpecifications`): Amazon Elastic Graphics finalizó su vida útil el 8 de enero de 2024.
- Detalles avanzados/Inferencia elástica (`ElasticInferenceAccelerators`): Amazon Elastic Inference ya no está disponible para clientes nuevos.
- AAdvancedDetallar/especificar CPU opciones/hilos por núcleo (`ThreadsPerCore`): AWS PCS establece el número de hilos por núcleo en 1.

Estos parámetros tienen requisitos especiales que permiten la compatibilidad con: AWS PCS

- Datos de usuario (`UserData`): deben estar codificados en varias partes. Consulte Trabajar con datos de EC2 usuario de Amazon.

- Imágenes de la aplicación y del sistema operativo (`ImageId`): puede incluirlas. Sin embargo, si especificas un AMI ID al crear o actualizar el grupo de nodos, este anulará el valor de la plantilla de lanzamiento. El AMI que proporciones debe ser compatible con AWS PCS. Para obtener más información, consulte "Amazon Machine Images (AMIs) para AWS PCS.
- Configuración de red/Firewall (grupos de seguridad) (**`SecurityGroups`**): no se puede configurar una lista de nombres de grupos de seguridad en una AWS PCS plantilla de lanzamiento. Puede configurar una lista de grupos de seguridad IDs (`SecurityGroupIds`), a menos que defina las interfaces de red en la plantilla de lanzamiento. A continuación, debe especificar el grupo de seguridad IDs para cada interfaz. Para obtener más información, consulte Grupos de seguridad en AWS PCS.
- Configuración de red/Configuración de red avanzada (`NetworkInterfaces`): si utiliza EC2 instancias con una sola tarjeta de red y no necesita ninguna configuración de red especializada, AWS PCS puede configurar la red de instancias por usted. Para configurar varias tarjetas de red o habilitar el adaptador Elastic Fabric en sus instancias, utilice. `NetworkInterfaces` Cada interfaz de red debe tener una lista de grupos de seguridad IDs debajo`Groups`. Para obtener más información, consulte Varias interfaces de red en AWS PCS.
- Detalles avanzados/reserva de capacidad (`CapacityReservationSpecification`): se puede configurar, pero no puede hacer referencia a un dato específico `CapacityReservationId` cuando se trabaja con él. AWS PCS Sin embargo, puede hacer referencia a un grupo de reservas de capacidad, si ese grupo contiene una o más reservas de capacidad. Para obtener más información, consulte Reservas de aforo en AWS PCS.

Creación de una plantilla de lanzamiento básica

Puede crear una plantilla de lanzamiento mediante el AWS Management Console o el AWS CLI.

AWS Management Console

Para crear una plantilla de lanzamiento

1. Abre la EC2consola de Amazon y selecciona Launch templates.
2. Elija Crear plantilla de inicialización.
3. En Nombre y descripción de la plantilla de lanzamiento, introduce un nombre único y distintivo para el nombre de la plantilla de lanzamiento

4. En Par de claves (inicio de sesión) en Nombre del par de SSH claves, seleccione el par de claves que se usará para iniciar sesión en EC2 las instancias administradas por AWS PCS. Esto es opcional, pero recomendable.
5. En Configuración de red y, a continuación, en Firewall (grupos de seguridad), elija los grupos de seguridad que desee conectar a la interfaz de red. Todos los grupos de seguridad de la plantilla de lanzamiento deben pertenecer a su AWS PCS clústerVPC. Como mínimo, elija:

   - Un grupo de seguridad que permita la comunicación con el AWS PCS clúster
   - Un grupo de seguridad que permite la comunicación entre EC2 instancias lanzadas por AWS PCS
   - (Opcional) Un grupo de seguridad que permite el SSH acceso entrante a instancias interactivas
   - (Opcional) Un grupo de seguridad que permite a los nodos de cómputo establecer conexiones salientes a Internet
   - (Opcional) Grupos de seguridad que permiten el acceso a los recursos de la red, como los sistemas de archivos compartidos o un servidor de bases de datos.
6. Podrás acceder a tu nueva ID de plantilla de lanzamiento en la EC2 consola de Amazon, en Plantillas de lanzamiento. El identificador de la plantilla de lanzamiento incluirá el formulario`lt-0123456789abcdef01`.

Siguiente paso recomendado

- Utilice la nueva plantilla de lanzamiento para crear o actualizar un grupo de AWS PCS nodos de procesamiento.

AWS CLI

Para crear una plantilla de lanzamiento

Cree su plantilla de lanzamiento con el siguiente comando.

- Antes de ejecutar el comando, realice los siguientes reemplazos:

  a. Reemplazar *`region-code`* con el Región de AWS lugar con el que estás trabajando AWS PCS
  b. Reemplazar *`my-launch-template-name`* con un nombre para tu plantilla. Debe ser exclusivo del Cuenta de AWS y Región de AWS que está utilizando.

c. Reemplazar *my-ssh-key-name* con el nombre de la SSH clave que prefiera.

d. Reemplazar *sg-ExampleID1* y *sg-ExampleID2* con un grupo de seguridad IDs que permite la comunicación entre sus EC2 instancias y el programador y la comunicación entre EC2 instancias. Si solo tiene un grupo de seguridad que permite todo este tráfico, puede eliminarlo `sg-ExampleID2` y el carácter de coma que lo precede. También puede añadir más grupos IDs de seguridad. Todos los grupos de seguridad que incluya en la plantilla de lanzamiento deben pertenecer a su AWS PCS clústerVPC.

```
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"]}'
```

El resultado AWS CLI será un texto parecido al siguiente. El identificador de la plantilla de lanzamiento se encuentra en`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"
    }
}
```

Siguiente paso recomendado

- Utilice la nueva plantilla de lanzamiento para crear o actualizar un grupo de AWS PCS nodos de procesamiento.

Trabajar con datos de EC2 usuario de Amazon

Puede proporcionar datos de EC2 usuario en la plantilla de lanzamiento que `cloud-init` se ejecuta cuando se lanzan las instancias. Los bloques de datos de usuario con el tipo de contenido se `cloud-config` ejecutan antes de que la instancia se registre en el AWS PCSAPI, mientras que los

bloques de datos de usuario con ese tipo de contenido se `text/x-shellscript` ejecutan una vez finalizado el registro, pero antes de que se inicie el daemon de Slurm. Para obtener más información sobre los tipos de contenido, consulte la documentación de cloud-init.

nuestros datos de usuario pueden realizar escenarios de configuración comunes, incluidos, entre otros, los siguientes:

- Incluidos usuarios o grupos
- Instalación de paquetes
- Creación de particiones y sistemas de archivos
- Montaje de sistemas de archivos de red

Los datos de usuario de las plantillas de lanzamiento deben estar en formato de archivo de MIME varias partes. Esto se debe a que sus datos de usuario se combinan con otros datos de AWS PCS usuario necesarios para configurar los nodos de su grupo de nodos. Puede combinar varios bloques de datos de usuario en un único archivo de MIME varias partes.

Un archivo MIME de varias partes consta de los siguientes componentes:

- El tipo de contenido y declaración de límite de partes: `Content-Type: multipart/mixed; boundary="==BOUNDARY=="`
- La declaración de MIME versión: `MIME-Version: 1.0`
- Uno o más bloques de datos de usuario, que contienen los siguientes componentes:
  - El límite de apertura, que señala el inicio de un bloque de datos de usuario: `--==BOUNDARY==`. Debe dejar en blanco la línea anterior a este límite.
  - La declaración del tipo de contenido del bloque: `Content-Type: text/cloud-config; charset="us-ascii"` o`Content-Type: text/x-shellscript; charset="us-ascii"`. Debe dejar en blanco la línea que sigue a la declaración de tipo de contenido.
  - El contenido de los datos de usuario, por ejemplo, una lista de intérprete de comandos o políticas de `cloud-config`.
- El límite de cierre que señala el final del archivo de MIME varias partes:`--==BOUNDARY==--`. Debe dejar en blanco la línea anterior al límite de cierre.

> **Note**
>
> Si añades datos de usuario a una plantilla de lanzamiento en la EC2 consola de Amazon, puedes pegarlos como texto sin formato. O bien, puede cargarlos desde un archivo.
> Si utilizas AWS CLI o an AWS SDK, primero debes codificar en base64 los datos del usuario y enviar esa cadena como el valor del `UserData` parámetro cuando llames CreateLaunchTemplate, tal y como se muestra en este JSON archivo.

```
{
    "LaunchTemplateName": "base64-user-data",
    "LaunchTemplateData": {
        "UserData":
 "ewogICAgIkxhdW5jaFRlbXBsYXRlTmFtZSI6ICJpbmNyZWFzZS1jb250YWluZXItdm9sdW..."
    }
}
```

Ejemplos

- Ejemplo: instalar software desde un repositorio de paquetes
- Ejemplo: ejecutar scripts desde un bucket de S3
- Ejemplo: establecer variables de entorno globales
- Uso de sistemas de archivos de red con AWS PCS
- Ejemplo: utilice un sistema de EFS archivos como directorio principal compartido

Ejemplo: instalar software AWS PCS desde un repositorio de paquetes

Proporcione este script como el valor de `"userData"` en su plantilla de lanzamiento. Para obtener más información, consulte Trabajar con datos de EC2 usuario de Amazon.

Este script usa cloud-config para instalar paquetes de software en instancias de grupos de nodos en el momento del lanzamiento. Para obtener más información, consulta los formatos de datos de usuario en la documentación de cloud-init. En este ejemplo, se instala `curl` y. `llvm`

> **Note**
>
> Sus instancias deben poder conectarse a sus repositorios de paquetes configurados.

```
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==--
```

Ejemplo: ejecutar scripts adicionales AWS PCS desde un bucket de S3

Proporcione este script como el valor de `"userData"` en su plantilla de lanzamiento. Para obtener más información, consulte Trabajar con datos de EC2 usuario de Amazon.

Este script usa cloud-config para importar un script de un bucket de S3 y ejecutarlo en instancias de grupos de nodos en el momento del lanzamiento. Para obtener más información, consulta los formatos de datos de usuario en la documentación de cloud-init.

Sustituya los siguientes valores de este script por sus propios detalles:

- *`my-bucket-name`* — El nombre de un bucket de S3 desde el que puede leer tu cuenta.
- *`path`* — La ruta relativa a la raíz del bucket de S3.
- *`shell`* — El shell de Linux que se utilizará para ejecutar el script, por ejemplo`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==--
```

El perfil de IAM instancia del grupo de nodos debe tener acceso al bucket. La siguiente IAM política es un ejemplo del depósito del script de datos de usuario anterior.

```
{
    "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/*"
            ]
        }
    ]
}
```

Ejemplo: establecer variables de entorno globales para AWS PCS

Proporcione este script como el valor de "`userData`" en su plantilla de lanzamiento. Para obtener más información, consulte Trabajar con datos de EC2 usuario de Amazon.

El siguiente ejemplo se utiliza `/etc/profile.d` para establecer variables globales en instancias de grupos de nodos.

```
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==--
```

Ejemplo: utilice un sistema de EFS archivos como directorio principal compartido para AWS PCS

Proporcione este script como el valor de `"userData"` en su plantilla de lanzamiento. Para obtener más información, consulte Trabajar con datos de EC2 usuario de Amazon.

En este ejemplo, se amplía el ejemplo de EFS montaje Uso de sistemas de archivos de red con AWS PCS para implementar un directorio principal compartido. Se hace una copia de seguridad del contenido de /home antes de montar el sistema de EFS archivos. Luego, el contenido se copia rápidamente en su lugar en el almacenamiento compartido una vez finalizado el montaje.

Sustituya los siguientes valores de este script por sus propios detalles:

- *`/mount-point-directory`* — La ruta de una instancia en la que desea montar el sistema de EFS archivos.
- *`filesystem-id`* — El ID del sistema de EFS archivos del sistema de archivos.

```
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==--
```

Habilitar la tecnología sin contraseña SSH

Puede basarse en el ejemplo del directorio principal compartido para implementar SSH conexiones entre instancias del clúster mediante SSH claves. Para cada usuario que utilice el sistema de archivos de inicio compartido, ejecute un script similar al siguiente:

```
#!/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**
>
> Las instancias deben usar un grupo de seguridad que permita SSH las conexiones entre los nodos del clúster.

Reservas de aforo en AWS PCS

Puedes reservar la EC2 capacidad de Amazon en una zona de disponibilidad específica y durante un período específico mediante reservas de capacidad bajo demanda o bloques de EC2 capacidad para asegurarte de que tienes la capacidad informática necesaria disponible cuando la necesites.

> **Note**
>
> AWS PCSadmite reservas de capacidad bajo demanda (ODCR), pero actualmente no admite bloques de capacidad para aprendizaje automático.

Utilizándolo ODCRs con AWS PCS

Puede elegir cómo AWS PCS consume sus instancias reservadas. Si crea una aperturaODCR, las instancias coincidentes lanzadas por AWS PCS u otros procesos de su cuenta se descontarán de la

reserva. Con una segmentaciónODCR, solo las instancias lanzadas con el identificador de reserva específico se tienen en cuenta para la reserva. En el caso de las cargas de trabajo urgentes, las segmentadas ODCRs son más habituales.

Puede configurar un grupo de AWS PCS nodos de procesamiento para que utilice un destino ODCR agregándolo a una plantilla de lanzamiento. Estos son los pasos para hacerlo:

1. Cree una reserva de capacidad específica y bajo demanda (ODCR).
2. ODCRAñádala a un grupo de reservas de capacidad.
3. Asocie el grupo de reserva de capacidad a una plantilla de lanzamiento.
4. Cree o actualice un grupo de AWS PCS nodos de procesamiento para usar la plantilla de lanzamiento.

Ejemplo: reserve y use instancias hpc6a.48xlarge con un objetivo ODCR

Este comando de ejemplo crea un destino para 32 instancias hpc6a.48xlarge. ODCR Para lanzar las instancias reservadas en un grupo de ubicación, agréguelas al comando. `--placement-group-arn` Puede definir una fecha de finalización con `--end-date` y`--end-date-type`, de lo contrario, la reserva continuará hasta que se finalice 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
```

El resultado de este comando será un ARN para el nuevoODCR. Para usarlo ODCR con AWS PCS, debe agregarse a un grupo de reserva de capacidad. Esto se debe a AWS PCS que no admite personasODCRs. Para obtener más información, consulte Grupos de reserva de capacidad en la Guía del usuario de Amazon Elastic Compute Cloud.

A continuación, se explica cómo agregarlos ODCR a un grupo de reserva de capacidad denominado`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
```

Una vez ODCR creado y agregado a un grupo de reserva de capacidad, ahora se puede conectar a un grupo de nodos de AWS PCS cómputo agregándolo a una plantilla de lanzamiento. A continuación, se muestra un ejemplo de plantilla de lanzamiento que hace referencia al grupo de reserva de capacidad.

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

Por último, cree o actualice un grupo de AWS PCS nodos de procesamiento para usar instancias hpc6a.48xlarge y use la plantilla de lanzamiento que hace referencia a ellas en su grupo de reserva de ODCR capacidad. En el caso de un grupo de nodos estático, establece el número mínimo y máximo de instancias según el tamaño de la reserva (32). Para un grupo de nodos dinámico, establece el número mínimo de instancias en 0 y el máximo en el tamaño de la reserva.

Este ejemplo es una implementación simple de un nodo único ODCR que se aprovisiona para un grupo de nodos de cómputo. Sin embargo, AWS PCS es compatible con muchos otros diseños. Por ejemplo, puede subdividir un grupo grande ODCR o de reserva de capacidad entre varios grupos de nodos de procesamiento. O bien, puedes usar la AWS cuenta ODCRs que otra ha creado y compartido con la tuya. La principal limitación es que ODCRs siempre debe estar incluida en un grupo de reserva de capacidad.

Para obtener más información, consulte Reservas de capacidad bajo demanda y bloques de capacidad para aprendizaje automático en la Guía del usuario de Amazon Elastic Compute Cloud.

Parámetros útiles de la plantilla de lanzamiento

En esta sección se describen algunos parámetros de la plantilla de lanzamiento que pueden resultar muy útiles AWS PCS.

Active la CloudWatch supervisión detallada

Puede habilitar la recopilación de CloudWatch métricas en un intervalo más corto mediante un parámetro de plantilla de lanzamiento.

AWS Management Console

En las páginas de la consola para crear o editar plantillas de lanzamiento, esta opción se encuentra en la sección de detalles avanzados. Configure la CloudWatch supervisión detallada en Activar.

YAML

```
Monitoring:
    Enabled: True
```

JSON

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

Para obtener más información, consulte Habilitar o desactivar la supervisión detallada de sus instancias en la Guía del usuario de Amazon Elastic Compute Cloud para instancias de Linux.

Instance Metadata Service, versión 2 (IMDSv2)

El uso IMDS de la versión 2 con EC2 las instancias ofrece importantes mejoras de seguridad y ayuda a mitigar los posibles riesgos asociados al acceso a los metadatos de las instancias en los AWS entornos.

AWS Management Console

En las páginas de la consola para crear o editar plantillas de lanzamiento, esta opción se encuentra en la sección de detalles avanzados. Configura los metadatos accesibles en Habilitados, la versión de metadatos en V2 únicamente (se requiere un token) y el límite de saltos de respuesta de los metadatos en 4.

YAML

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

JSON

```
{
```

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

.

AWS PCScolas

Una AWS PCS cola es una abstracción ligera de la implementación nativa del programador de una cola de trabajos. En el caso de Slurm, una AWS PCS cola equivale a una partición de Slurm.

Los usuarios envían los trabajos a una cola en la que residen hasta que puedan programarse para que se ejecuten en los nodos proporcionados por uno o más grupos de nodos de procesamiento. Un AWS PCS clúster puede tener varias colas de trabajos. Por ejemplo, puede crear una cola que utilice Amazon EC2 On-Demand Instances para los trabajos de alta prioridad y otra cola que utilice Amazon EC2 Spot Instances para los trabajos de baja prioridad.

Temas

- Crear una cola en AWS PCS
- Actualización de una AWS PCS cola
- Eliminar una cola en AWS PCS

Crear una cola en AWS PCS

En este tema se proporciona una visión general de las opciones disponibles y se describe lo que hay que tener en cuenta al crear una cola en. AWS PCS

Requisitos previos

- Un AWS PCS clúster: las colas solo se pueden crear en asociación con un clúster específicoPCS.
- Uno o más grupos de nodos de AWS PCS cómputo: una cola debe estar asociada a al menos un grupo de nodos de PCS cómputo.

Para crear una cola en AWS PCS

Puede crear una cola utilizando el AWS Management Console o el. AWS CLI

AWS Management Console

Para crear una cola mediante la consola

1. Abra la AWS PCS consola en `https://console.aws.amazon.com/pcs/home#/clusters`
2. Seleccione el clúster en el que desee crear una cola. Navegue hasta Colas y elija Crear cola.
3. En la sección de configuración de colas, proporciona los siguientes valores:
    a. Nombre de la cola: un nombre para la cola. El nombre solo puede contener caracteres alfanuméricos (con distinción de mayúsculas y minúsculas) y guiones. Debe empezar por un carácter alfabético y no puede tener más de 25 caracteres. El nombre debe ser único en el clúster.
    b. Grupos de nodos de cómputo: seleccione uno o más grupos de nodos de cómputo para dar servicio a esta cola. Un grupo de nodos de cómputo se puede asociar a más de una cola.
4. (Opcional) En Etiquetas, agrega cualquier etiqueta a tu cola AWS PCS
5. Elija Crear cola. El campo Estado mostrará la opción Crear mientras se está configurando la cola. La creación de la cola puede tardar varios minutos.

Siguiente paso recomendado

- Envía un trabajo a tu nueva lista

AWS CLI

Para crear una cola mediante AWS CLI

Cree su cola con el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:

1. Reemplazar *`region-code`* con la AWS región en la que quieres crear tu clúster.
2. Reemplazar *`my-queue`* con el nombre de la cola. El nombre solo puede contener caracteres alfanuméricos (con distinción de mayúsculas y minúsculas) y guiones. Debe empezar por un

carácter alfabético y no puede tener más de 25 caracteres. El nombre debe ser único en el clúster.

3. Reemplazar *my-cluster* con el nombre o el clusterId de su clúster.
4. Sustituya el valor de por `computeNodeGroupId` su propio identificador de grupo de nodos de procesamiento. Tenga en cuenta que no puede especificar los nombres de los grupos de nodos de cálculo al crear una cola.

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

La creación de la cola puede tardar varios minutos. Puede consultar el estado de la cola con el siguiente comando. No podrá enviar trabajos a la cola hasta que alcance su estado. `ACTIVE`

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

Siguiente paso recomendado

- Envía un trabajo a tu nueva lista

Actualización de una AWS PCS cola

En este tema se proporciona una descripción general de las opciones disponibles y se describe lo que se debe tener en cuenta al actualizar una AWS PCS cola.

Consideraciones a la hora de actualizar una cola AWS PCS

Las actualizaciones de las colas no afectarán a los trabajos en ejecución, pero es posible que el clúster no pueda aceptar nuevos trabajos mientras se actualiza la cola.

Para actualizar un grupo de nodos de AWS PCS cómputo

Puede actualizar un grupo de nodos mediante la consola AWS de administración o el AWSCLI.

AWS Management Console

Para actualizar una cola

1. Abra la AWS PCS consola en `https://console.aws.amazon.com/pcs/home#/clusters`
2. Seleccione el clúster en el que desee actualizar una cola.
3. Ve a Colas, ve a la cola que deseas actualizar y, a continuación, selecciona Editar.
4. En la sección de configuración de colas, actualiza cualquiera de los siguientes valores:
   - Grupos de nodos: agregue o elimine grupos de nodos de cómputo de la asociación con la cola.
   - Etiquetas: añada o elimine etiquetas para la cola.
5. Elija Actualizar. El campo Estado mostrará la actualización mientras se aplican los cambios.

> **Important**
>
> Las actualizaciones de las colas pueden tardar varios minutos.

AWS CLI

Para actualizar una cola

1. Actualice la cola con el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:
   a. Reemplazar *`region-code`* con la Región de AWS que quieres crear tu clúster.
   b. Reemplazar *`my-queue`* con el nombre o `computeNodeGroupId` para tu cola.
   c. Reemplazar *`my-cluster`* con el nombre o el `clusterId` de su clúster.
   d. Para cambiar las asociaciones de grupos de nodos de procesamiento, proporcione una lista actualizada de`--compute-node-group-configurations`.
      - Por ejemplo, para añadir un segundo grupo de nodos de cómputo`computeNodeGroupExampleID2`:

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

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

2. La actualización de la cola puede tardar varios minutos. Puede consultar el estado de la cola con el siguiente comando. No podrá enviar trabajos a la cola hasta que alcance su estado. `ACTIVE`

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

Próximos pasos recomendados

- Envía un trabajo a tu lista de espera actualizada.

Eliminar una cola en AWS PCS

En este tema se proporciona información general sobre cómo eliminar una cola en. AWS PCS

Consideraciones a la hora de eliminar una cola

- Si hay trabajos en ejecución en la cola, el programador los finalizará cuando se elimine la cola. Los trabajos pendientes de la cola se cancelarán. Considere la posibilidad de esperar a que terminen los trabajos de la cola o detenerlos o cancelarlos manualmente mediante los comandos nativos del programador (como los de Slurm). `scancel`

Eliminación de la cola

Puede usar o para eliminar una cola. AWS Management Console AWS CLI

AWS Management Console

Para eliminar una cola

1. Abre la AWS PCSconsola.
2. Seleccione el grupo de la cola.
3. Navegue hasta Colas y seleccione la cola que desee eliminar.
4. Elija Eliminar.
5. Aparece el campo Estado. `Deleting` Puede tardar varios minutos en completarse.

> **Note**
>
> Puede usar los comandos nativos de su programador para confirmar que la cola se ha eliminado. Por ejemplo, usa `sinfo` o `squeue` para Slurm.

AWS CLI

Para eliminar una cola

- Utilice el siguiente comando para eliminar una cola, con estas sustituciones:
  - Reemplazar *region-code* con el clúster en Región de AWS el que se encuentra.
  - Reemplazar *my-queue* con el nombre o el ID de tu cola.
  - Reemplazar *my-cluster* con el nombre o el ID de su clúster.

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

Eliminar la cola puede tardar varios minutos.

> **Note**
>
> Puede usar los comandos nativos de su programador para confirmar que la cola se ha eliminado. Por ejemplo, usa `sinfo` o `squeue` para Slurm.

AWS PCSnodos de inicio de sesión

Por lo general, un AWS PCS clúster necesita al menos un nodo de inicio de sesión para admitir el acceso interactivo y la administración de tareas. Una forma de lograrlo es con un grupo de nodos de AWS PCS computación estáticos configurado para la capacidad de nodos de inicio de sesión. También puede configurar una EC2 instancia independiente para que actúe como nodo de inicio de sesión.

Temas

Uso de un grupo de nodos de AWS PCS cómputo para proporcionar nodos de inicio de sesión

En este tema se proporciona una descripción general de las opciones de configuración sugeridas y se describe lo que se debe tener en cuenta al utilizar un grupo de AWS PCS nodos de procesamiento para proporcionar un acceso persistente e interactivo al clúster.

Crear un grupo de AWS PCS nodos de procesamiento para los nodos de inicio de sesión

Desde el punto de vista operativo, esto no es muy diferente de crear un grupo de nodos de cómputo normal. Sin embargo, hay que tomar algunas decisiones de configuración clave:

- Establezca una configuración de escalado estático de al menos una EC2 instancia del grupo de nodos de procesamiento.
- Elija la opción de compra bajo demanda para evitar que se recuperen sus instancias.
- Elija un nombre informativo para el grupo de nodos de cómputo, como el inicio de sesión.
- Si desea que se pueda acceder a las instancias del nodo de inicio de sesión desde fuera del suyoVPC, considere la posibilidad de utilizar una subred pública.
- Si pretende permitir el SSH acceso, la plantilla de lanzamiento necesitará tener un grupo de seguridad que exponga el SSH puerto a las direcciones IP que elija.
- El perfil de la IAM instancia debe tener solo los AWS permisos que desee que tengan sus usuarios finales. Para obtener más información, consulte IAMperfiles de instancia para AWS Parallel Computing Service.

- Considere permitir que AWS Systems Manager Session Manager administre sus instancias de inicio de sesión.
- Considere restringir el acceso a las AWS credenciales de la instancia únicamente a los usuarios administrativos
- Seleccione tipos de instancias menos costosos que los de los grupos de nodos de procesamiento normales, ya que los nodos de inicio de sesión se ejecutarán de forma continua.
- Usa el mismo (o un derivado) AMI que para tus otros grupos de nodos de cómputo para asegurarte de que todas las instancias tengan instalado el mismo software. Para obtener más información sobre la personalizaciónAMIs, consulte Amazon Machine Images (AMIs) para AWS PCS
- Configure el mismo sistema de archivos de red (AmazonEFS, Amazon FSx for Lustre, etc.) para montarlo en sus nodos de inicio de sesión que en sus instancias informáticas. Para obtener más información, consulte Uso de sistemas de archivos de red con AWS PCS.

Acceda a sus nodos de inicio de sesión

Cuando su nuevo grupo de nodos de cómputo alcance el ACTIVE estado, podrá encontrar las EC2 instancias que ha creado e iniciar sesión en ellas. Para obtener más información, consulte Búsqueda de instancias de grupos de nodos de cómputo en AWS PCS.

Actualización de un grupo de nodos de AWS PCS cómputo para los nodos de inicio de sesión

Puede actualizar un grupo de nodos de inicio de sesión mediante UpdateComputeNodeGroup. Como parte del proceso de actualización del grupo de nodos, se sustituirán las instancias en ejecución. Ten en cuenta que esto interrumpirá cualquier sesión de usuario o proceso activo en la instancia. Los trabajos de Slurm en ejecución o en cola no se verán afectados. Para obtener más información, consulte Actualización de un grupo de nodos de AWS PCS cómputo.

También puede editar la plantilla de lanzamiento utilizada por su grupo de nodos de cómputo. Debe utilizarla UpdateComputeNodeGroup para aplicar la plantilla de lanzamiento actualizada al grupo de nodos de procesamiento. EC2Las nuevas instancias lanzadas en el grupo de nodos de procesamiento utilizan la plantilla de lanzamiento actualizada. Para obtener más información, consulte Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS.

Eliminar un grupo de AWS PCS nodos de procesamiento para los nodos de inicio de sesión

Puede actualizar un grupo de nodos de inicio de sesión mediante el mecanismo de eliminación del grupo de nodos de cálculo de AWS PCS. Las instancias en ejecución se cancelarán como parte de la eliminación del grupo de nodos. Ten en cuenta que esto interrumpirá cualquier sesión de usuario o proceso activo en la instancia. Los trabajos de Slurm en ejecución o en cola no se verán afectados. Para obtener más información, consulte Eliminar un grupo de nodos de cómputo en AWS PCS.

Uso de instancias independientes como nodos de inicio de AWS PCS sesión

Puede configurar EC2 instancias independientes para que interactúen con el programador AWS PCS Slurm de un clúster. Esto resulta útil para crear nodos de inicio de sesión, estaciones de trabajo o hosts de administración de flujos de trabajo dedicados que funcionen con AWS PCS clústeres pero que operen fuera de la administración. AWS PCS Para ello, cada instancia independiente debe:

1. Tener instalada una versión de software Slurm compatible.
2. Poder conectarse al punto final AWS PCS Slurmctld del clúster.
3. Configure correctamente el Slurm Auth and Cred Kiosk Daemon (`sackd`) con el punto final y el secreto del clúster. AWS PCS Para obtener más información, consulte sackd en la documentación de Slurm.

Este tutorial le ayuda a configurar una instancia independiente que se conecta a un clúster. AWS PCS

Contenido

Paso 1: Recuperar la dirección y el secreto del AWS PCS clúster de destino

Recupere los detalles sobre el AWS PCS clúster de destino AWS CLI mediante el comando siguiente. Antes de ejecutar el comando, realice los siguientes reemplazos:

- Reemplazar *region-code* con el Región de AWS lugar en el que se ejecuta el clúster de destino.
- Reemplazar *cluster-ident* con el nombre o identificador del clúster de destino

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

El comando devolverá un resultado similar al de este ejemplo.

```
{
    "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-0123456789abdef"
            ],
            "securityGroupIds": [
                "sg-0123456789abdef"
            ]
        },
        "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"
        }
    }
}
```

En este ejemplo, el punto final del controlador Slurm del clúster tiene una dirección IP de `10.3.149.220` y se ejecuta en el puerto. `6817` Se `secretArn` usará en pasos posteriores para recuperar el secreto del clúster. La dirección IP y el puerto se utilizarán en pasos posteriores para configurar el `sackd` servicio.

Paso 2: lanzar una EC2 instancia

Para lanzar una instancia EC2

1. Abre la EC2consola de Amazon.
2. En el panel de navegación, elija Instancias y, a continuación, Iniciar instancias para abrir el nuevo asistente de inicialización de instancias.
3. (Opcional) En la sección Nombre y etiquetas, proporciona un nombre para la instancia, como`PCS-LoginNode`. El nombre se asigna a la instancia como etiqueta de recurso (`Name=PCS-LoginNode`).
4. En la sección Imágenes de aplicaciones y sistemas operativos, seleccione uno AMI para uno de los sistemas operativos compatibles con AWS PCS. Para obtener más información, consulte Sistemas operativos compatibles.
5. En la sección Tipo de instancia, seleccione un tipo de instancia compatible. Para obtener más información, consulte Tipos de instancias admitidas.
6. En la sección Par de claves, selecciona el par de SSH claves que quieres usar para la instancia.
7. En la sección Configuración de red:

    - Elija Editar.

        i. Seleccione el VPC de su AWS PCS clúster.

        ii. En Firewall (grupos de seguridad), elija Seleccionar un grupo de seguridad existente.

            A. Seleccione un grupo de seguridad que permita el tráfico entre la instancia y el controlador Slurm del AWS PCS clúster de destino. Para obtener más información, consulte Requisitos y consideraciones sobre los grupos de seguridad.

B. (Opcional) Seleccione un grupo de seguridad que permita el SSH acceso entrante a la instancia.

8. En la sección Almacenamiento, configure los volúmenes de almacenamiento según sea necesario. Asegúrese de configurar suficiente espacio para instalar aplicaciones y bibliotecas a fin de habilitar su caso de uso.
9. En Avanzado, elija un IAM rol que permita acceder al secreto del clúster. Para obtener más información, consulte Obtén el secreto del cúmulo de Slurm.
10. En el panel de resumen, selecciona Lanzar instancia.

Paso 3: Instala Slurm en la instancia

Cuando la instancia se haya lanzado y se active, conéctese a ella mediante el mecanismo que prefiera. Use el instalador de Slurm proporcionado por AWS para instalar Slurm en la instancia. Para obtener más información, consulte Instalador Slurm.

Descarga el instalador de Slurm, descomprímelo y usa el script para instalar Slurm. `installer.sh` Para obtener más información, consulte Paso 3: Instalar Slurm.

Paso 4: Recupere y almacene el secreto del clúster

Estas instrucciones requieren la AWS CLI. Para obtener más información, consulte Instalar o actualizar a la última versión de AWS CLI en la Guía del AWS Command Line Interface usuario de la versión 2.

Guarde el secreto del clúster con los siguientes comandos.

- Cree el directorio de configuración de Slurm.

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

- Recupere, decodifique y almacene el secreto del clúster. Antes de ejecutar este comando, sustituya *region-code* por la región en la que se ejecuta el clúster de destino y reemplace *secret-arn* por el valor `secretArn` obtenido en el paso 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**
>
> En un entorno multiusuario, cualquier usuario con acceso a la instancia podría obtener el secreto del clúster si puede acceder al servicio de metadatos de la instancia ()IMDS. Esto, a su vez, podría permitirles hacerse pasar por otros usuarios. Considere la posibilidad de restringir el acceso IMDS únicamente a los usuarios raíz o administrativos. Como alternativa, puedes usar un mecanismo diferente que no dependa del perfil de la instancia para obtener y configurar el secreto.

- Configura la propiedad y los permisos en el archivo de claves de Slurm.

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

> **Note**
>
> La clave Slurm debe ser propiedad del usuario y del grupo en los que se ejecuta el `sackd` servicio.

Paso 5: Configurar la conexión al clúster AWS PCS

Para establecer una conexión con el AWS PCS clúster, ejecútelo `sackd` como un servicio del sistema siguiendo estos pasos.

1. Configure el archivo de entorno del `sackd` servicio con el siguiente comando. Antes de ejecutar el comando, sustituya *`ip-address`* y *`port`* por los valores recuperados de los puntos finales del paso 1.

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

2. Cree un archivo `systemd` de servicio para gestionar el `sackd` proceso.

   ```
   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. Establezca la propiedad del archivo `sackd` de servicio.

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

4. Habilite el `sackd` servicio.

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

5. Inicie el servicio `sackd`.

```
sudo systemctl start sackd
```

Paso 6: (opcional) Pruebe la conexión

Confirme que el `sackd` servicio se esté ejecutando. A continuación, se muestra un resultado de ejemplo. Si hay errores, suelen aparecer aquí.

```
[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
```

Confirme que las conexiones al clúster funcionan mediante comandos del cliente de Slurm como `sinfo` y. `squeue` Este es un ejemplo de salida de. `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]
```

También deberías poder enviar trabajos. Por ejemplo, un comando similar a este ejemplo lanzaría un trabajo interactivo en un nodo del clúster.

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

AWS PCSRedes

Tu AWS PCS clúster se crea en AmazonVPC. En este capítulo se incluyen los siguientes temas sobre las redes para el programador y los nodos del clúster.

A excepción de elegir una subred para lanzar instancias, debe usar plantillas de EC2 lanzamiento para configurar las redes para los grupos de nodos de AWS PCS procesamiento. Para obtener más información acerca de las plantillas de inicialización, consulte Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS.

Temas

- AWS PCSVPCy requisitos y consideraciones de subred
- Crear un VPC para tu AWS PCS clúster
- Grupos de seguridad en AWS PCS
- Varias interfaces de red en AWS PCS

- Grupos de ubicación para EC2 instancias en AWS PCS
- Uso de un adaptador de tela elástica (EFA) con AWS PCS

AWS PCSVPCy requisitos y consideraciones de subred

Al crear un AWS PCS clúster, se especifica VPC una subred en él. VPC En este tema se proporciona una descripción general de los requisitos y consideraciones AWS PCS específicos para las subredes VPC y subredes que se utilizan con el clúster. Si no tienes una con VPC la que usarla AWS PCS, puedes crear una con una plantilla AWS proporcionada AWS CloudFormation . Para obtener más informaciónVPCs, consulte Nubes privadas virtuales (VPC) en la Guía del VPC usuario de Amazon.

VPCrequisitos y consideraciones

Al crear un clúster, el VPC que especifique debe cumplir los siguientes requisitos y consideraciones:

- VPCDebe tener un número suficiente de direcciones IP disponibles para el clúster, los nodos y otros recursos del clúster que desee crear. Para obtener más información, consulta el direccionamiento IP de tu red VPCs y de tus subredes en la Guía del VPC usuario de Amazon.
- VPCDebe tener un DNS nombre de host y soporte de DNS resolución. De lo contrario, los nodos no podrán registrar el clúster de clientes. Para obtener más información, consulta tus DNS atributos VPC en la Guía del VPC usuario de Amazon.
- VPCEs posible que requiera el uso de VPC puntos finales AWS PrivateLink para poder contactar con. AWS PCS API Para obtener más información, consulta Connect your VPC to services using AWS PrivateLink en la Guía del VPC usuario de Amazon.

Requisitos y consideraciones de la subred

Al crear un clúster de Slurm, AWS PCS crea una interfaz de red elástica (ENI) en la subred que especificó. Esta interfaz de red permite la comunicación entre el controlador del programador y el cliente. VPC La interfaz de red también permite a Slurm comunicarse con los componentes desplegados en la cuenta del cliente. Solo puede especificar la subred de un clúster en el momento de la creación.

Requisitos de la subred para los clústeres

La subred que especifique al crear un clúster debe cumplir los siguientes requisitos:

- La subred debe tener al menos una dirección IP para su uso. AWS PCS

- La subred no puede residir en AWS Outposts AWS Wavelength, o en una AWS zona local.
- La subred puede ser pública o privada. Le recomendamos que especifique una subred privada, si es posible. Una subred pública es una subred con una tabla de enrutamiento que incluye una ruta a una puerta de enlace a Internet; una subred privada es una subred con una tabla de enrutamiento que no incluye una ruta a una puerta de enlace a Internet.

Requisitos de la subred para los nodos

Puede implementar nodos y otros recursos del clúster en la subred que especifique al crear el AWS PCS clúster y en otras subredes de la misma. VPC

Cualquier subred en la que implementes nodos y recursos de clúster debe cumplir los siguientes requisitos:

- Debe asegurarse de que la subred tenga suficientes direcciones IP disponibles para implementar todos los nodos y los recursos del clúster.
- Si planea implementar nodos en una subred pública, esa subred debe asignar direcciones públicas automáticamenteIPv4.
- Si la subred en la que despliega los nodos es una subred privada y su tabla de enrutamiento no incluye una ruta a un dispositivo de traducción de direcciones de red (NAT) (IPv4), añada puntos VPC finales utilizando el cliente. AWS PrivateLink VPC VPCSe necesitan puntos de conexión para todos los AWS servicios con los que contactan los nodos. El único punto final obligatorio es AWS PCS permitir que el nodo inicie la `registerNodeGroupInstances` API acción.
- El estado de la subred pública o privada no afecta AWS PCS; los puntos finales necesarios deben estar accesibles.

Crear un VPC para tu AWS PCS clúster

Puede crear una Amazon Virtual Private Cloud (AmazonVPC) para sus clústeres dentro de AWS Parallel Computing Service (AWS PCS).

Usa Amazon VPC para lanzar VPC recursos a una red virtual que hayas definido. Esta red virtual es prácticamente idéntica a una red tradicional que podría operar en su propio centro de datos. Sin embargo, incluye los beneficios que supone utilizar la infraestructura escalable de Amazon Web Services. Te recomendamos que conozcas a fondo el VPC servicio de Amazon antes de implementar VPC clústeres de producción. Para obtener más información, consulta ¿Qué es AmazonVPC? en el modo visual de autor. Guía VPC del usuario de Amazon.

Un PCS clúster, nodos y recursos de apoyo (como sistemas de archivos y servicios de directorio) están desplegados en AmazonVPC. Si quieres utilizar un Amazon existente VPC conPCS, este debe cumplir los requisitos descritos enAWS PCSVPCy requisitos y consideraciones de subred. En este tema se describe cómo crear una VPC que cumpla con PCS los requisitos mediante una AWS AWS CloudFormation plantilla proporcionada. Una vez que haya implementado una plantilla, podrá ver los recursos creados por la plantilla para saber exactamente qué recursos creó y la configuración de esos recursos.

Requisitos previos

Para crear un Amazon VPC forPCS, debes tener los IAM permisos necesarios para crear VPC recursos de Amazon. Estos recursos son subredesVPCs, grupos de seguridad, tablas de enrutamiento y rutas, e Internet y NAT puertas de enlace. Para obtener más información, consulta Crear una VPC con una subred pública en la Guía del VPC usuario de Amazon. Para consultar la lista completa de AmazonEC2, consulta Acciones, recursos y claves de condición de Amazon EC2 en la Referencia de autorización de servicio.

Crea un Amazon VPC

Crea una VPC copiando y pegando la apropiada URL para el Región de AWS lugar donde la PCS usarás. También puedes descargar la AWS CloudFormation plantilla y subirla tú mismo a la AWS CloudFormation consola.

- EE.UU. Este (Virginia) (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
```

- EE.UU. Este (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
```

- EE.UU. Oeste (Oregón) (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 plantilla

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

Para crear un Amazon VPC para PCS

1. Abre la plantilla en la AWS CloudFormation consola.

   > **Note**
   >
   > Se rellenan previamente en la plantilla, por lo que puede simplemente dejarlos como valores predeterminados.

2. En Proporcione un nombre de pila y, a continuación, en Nombre de pila, introduzcahpc-networking.
3. En Parámetros, introduce los siguientes detalles:

   a. En VPC, a continuación CidrBlock, introduzca `10.3.0.0/16`

   b. En las subredes A:

      i. Luego CidrPublicSubnetA, introduzca `10.3.0.0/20`

      ii. Luego CidrPrivateSubnetA, ingresa `10.3.128.0/20`

   c. En las subredes B:

      i. A continuación, CidrPublicSubnetB, introduzca `10.3.16.0/20`

      ii. Luego CidrPrivateSubnetA, ingresa `10.3.144.0/20`

   d. En las subredes C:

      i. Para ProvisionSubnetsC, seleccione`True`.

      > **Note**
      >
      > Si va a crear una VPC en una región que tiene menos de tres zonas de disponibilidad, esta opción se ignorará si se establece en`True`.

      ii. A continuación, CidrPublicSubnetB, introduzca `10.3.32.0/20`

iii. Luego CidrPrivateSubnetA, ingresa `10.3.160.0/20`

4. En Capacidades, active la casilla Acepto que AWS CloudFormation podría crear IAM recursos.

Supervisa el estado de la AWS CloudFormation pila. Cuando llegue`CREATE_COMPLETE`, el VPC recurso estará listo para su uso.

> **Note**
>
> Para ver todos los recursos que creó la AWS CloudFormation plantilla, abre la AWS CloudFormation consola. Elija la pila `hpc-networking` y, a continuación, elija la pestaña Resources (Recursos).

Grupos de seguridad en AWS PCS

Los grupos de seguridad de Amazon EC2 actúan como firewalls virtuales para controlar el tráfico entrante y saliente a las instancias. Usa una plantilla de lanzamiento para que un grupo de nodos de AWS PCS cómputo añada o elimine grupos de seguridad en sus instancias. Si la plantilla de lanzamiento no contiene ninguna interfaz de red, úsala `SecurityGroupIds` para proporcionar una lista de grupos de seguridad. Si la plantilla de lanzamiento define las interfaces de red, debe usar el `Groups` parámetro para asignar grupos de seguridad a cada interfaz de red. Para obtener más información acerca de las plantillas de inicialización, consulte Uso de plantillas de EC2 lanzamiento de Amazon con AWS PCS.

> **Note**
>
> Los cambios en la configuración del grupo de seguridad en la plantilla de lanzamiento solo afectan a las nuevas instancias lanzadas una vez actualizado el grupo de nodos de procesamiento.

Requisitos y consideraciones sobre los grupos de seguridad

AWS PCScrea una interfaz de red elástica (ENI) multicuenta en la subred que especifique al crear un clúster. Esto proporciona al HPC programador, que se ejecuta en una cuenta administrada por AWS, una ruta para comunicarse con las EC2 instancias lanzadas por. AWS PCS Debe proporcionar

un grupo de seguridad ENI que permita la comunicación bidireccional entre el programador ENI y las instancias del clúster. EC2

Una forma sencilla de lograrlo consiste en crear un grupo de seguridad autorreferenciado y permisivo que permita el tráfico TCP /IP en todos los puertos entre todos los miembros del grupo. Puede adjuntarlo tanto al clúster como a las instancias del grupo de nodos. EC2

Ejemplo de configuración permisiva de un grupo de seguridad

Tipo de regla	Protocolos	Puertos	Origen	Destino
Entrada	Todos	Todos	Auto	
Salida	Todos	Todos		0.0.0.0/0
Salida	Todos	Todos		Auto

Estas reglas permiten que todo el tráfico fluya libremente entre el controlador Slurm y los nodos, permiten que todo el tráfico saliente se dirija a cualquier destino y habilitan el tráfico. EFA

Ejemplo de configuración restrictiva de un grupo de seguridad

También puede limitar los puertos abiertos entre el clúster y sus nodos de procesamiento. En el caso del programador de Slurm, el grupo de seguridad adjunto al clúster debe permitir los siguientes puertos:

- 6817: habilita las conexiones entrantes desde y hacia las instancias `slurmctld` EC2
- 6818: habilita las conexiones salientes desde `slurmctld` y hasta que se ejecuten en las instancias `slurmd` EC2

El grupo de seguridad adjunto a sus nodos de procesamiento debe permitir los siguientes puertos:

- 6817: habilita las conexiones salientes `slurmctld` desde EC2 las instancias.
- 6818: habilita las conexiones entrantes y salientes desde y hacia las instancias `slurmd` del grupo de `slurmctld` nodos `slurmd`
- 60001—63000: conexiones entrantes y salientes entre instancias de grupos de nodos para admitir `srun`

- EFAtráfico entre instancias de grupos de nodos. Para obtener más información, consulte Preparar un grupo de seguridad EFA habilitado en la Guía del usuario de instancias de Linux
- Cualquier otro tráfico entre nodos que requiera su carga de trabajo

Varias interfaces de red en AWS PCS

Algunas EC2 instancias tienen varias tarjetas de red. Esto les permite ofrecer un mayor rendimiento de la red, incluidas capacidades de ancho de banda superiores a 100 Gbps y una mejor gestión de paquetes. Para obtener más información sobre las instancias con varias tarjetas de red, consulte Interfaces de red elásticas en la Guía del usuario de Amazon Elastic Compute Cloud.

Configure tarjetas de red adicionales para las instancias de un grupo de nodos de AWS PCS cómputo añadiendo interfaces de red a su plantilla de EC2 lanzamiento. A continuación, se muestra un ejemplo de plantilla de lanzamiento que habilita dos tarjetas de red, como las que se encuentran en una `hpc7a.96xlarge` instancia. Tenga en cuenta los siguientes detalles:

- La subred de cada interfaz de red debe ser la misma que la elegida al configurar el grupo de AWS PCS nodos de procesamiento que utilizará la plantilla de lanzamiento.
- El dispositivo de red principal, donde se producirá la comunicación de red rutinaria, como SSH el HTTPS tráfico, se establece configurando una `DeviceIndex` de`0`. Otras interfaces de red tienen un `DeviceIndex` de`1`. Solo puede haber una interfaz de red principal; todas las demás interfaces son secundarias.
- Todas las interfaces de red deben tener una única. `NetworkCardIndex` Una práctica recomendada es numerarlas secuencialmente tal como se definen en la plantilla de lanzamiento.
- Los grupos de seguridad para cada interfaz de red se configuran mediante`Groups`. En este ejemplo, se agrega un grupo de SSH seguridad entrante (`sg-`*`SshSecurityGroupId`*) a la interfaz de red principal, así como el grupo de seguridad que permite las comunicaciones dentro del clúster (). `sg-`*`ClusterSecurityGroupId`* Por último, se agrega un grupo de seguridad que permite las conexiones salientes a Internet (`sg-`*`InternetOutboundSecurityGroupId`*) a las interfaces principal y secundaria.

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

Grupos de ubicación para EC2 instancias en AWS PCS

Puede utilizar un grupo de ubicación para influir en la ubicación de las EC2 instancias y adaptarlas a las necesidades de la carga de trabajo que se ejecuta en ellas.

Tipos de grupos de colocación

- Clúster: agrupa las instancias juntas en una zona de disponibilidad para optimizar la comunicación de baja latencia.
- Partición: distribuye las instancias entre las particiones lógicas para ayudar a maximizar la resiliencia.
- Distribución: exige estrictamente que un número reducido de instancias se lance en un hardware distinto, lo que también contribuye a aumentar la resiliencia.

Para obtener más información, consulta los grupos de ubicación para tus EC2 instancias de Amazon en la Guía del usuario de Amazon Elastic Compute Cloud.

Le recomendamos que incluya un grupo de ubicación de clústeres al configurar un grupo de nodos de AWS PCS cómputo para usar Elastic Fabric Adapter (EFA).

Para crear un grupo de ubicación de clústeres que funcione con EFA

1. Cree un grupo de ubicación con el tipo cluster para el grupo de nodos de procesamiento.
   - Utilice el siguiente AWS CLI comando:

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

- También puede utilizar una CloudFormation plantilla para crear un grupo de ubicaciones. Para obtener más información, consulte Trabajar con CloudFormation plantillas en la Guía del AWS CloudFormation usuario. Descargue la plantilla de las siguientes opciones URL y cárguela en la CloudFormation consola.

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

2. Incluya el grupo de ubicación en la plantilla de EC2 lanzamiento del grupo de AWS PCS nodos de procesamiento.

Uso de un adaptador de tela elástica (EFA) con AWS PCS

El adaptador Elastic Fabric (EFA) es una interconexión de red avanzada de alto rendimiento AWS que se puede conectar a la EC2 instancia para acelerar las aplicaciones de computación de alto rendimiento (HPC) y aprendizaje automático. Para habilitar las aplicaciones que se ejecutan en un AWS PCS clúster, es necesario configurar las instancias del grupo de nodos de AWS PCS cómputo para que se utilicen de la EFA siguiente manera. EFA

Contenido

Instálelo EFA en un dispositivo AWS PCS compatible AMI

Los que AMI se utilizan en el grupo de nodos de AWS PCS cómputo deben tener el EFA controlador instalado y cargado. Para obtener información sobre cómo crear una personalizada AMI con el EFA software instalado, consulteImágenes personalizadas de Amazon Machine (AMIs) para AWS PCS.

Identifique las EFA instancias habilitadas EC2

Para poder EFA utilizarlos, todos los tipos de instancias permitidos para un grupo de AWS PCS cómputo deben ser compatibles EFA y tener el mismo número de instancias vCPUs (GPUssi corresponde). Para obtener una lista de instancias EFA habilitadas, consulta el adaptador de Elastic Fabric para HPC cargas de trabajo de aprendizaje automático EC2 en Amazon en la Guía del usuario de Amazon Elastic Compute Cloud. También puedes usarlo AWS CLI para ver una lista de los tipos de instancias compatibles. EFA Reemplazar *region-code* con el Región de AWS lugar donde usas AWS PCS, por ejemplo`us-east-1`.

```
aws ec2 describe-instance-types \
   --region region-code \
   --filters Name=network-info.efa-supported,Values=true \
   --query "InstanceTypes[*].[InstanceType]" \
   --output text | sort
```

Determine cuántas interfaces de red están disponibles

Algunas EC2 instancias tienen varias tarjetas de red. Esto les permite tener variasEFAs. Para obtener más información, consulte Varias interfaces de red en AWS PCS.

Cree un grupo de seguridad para respaldar EFA las comunicaciones

AWS CLI

Puede usar el siguiente AWS CLI comando para crear un grupo de seguridad que lo admitaEFA. El comando genera un ID de grupo de seguridad. Realice las siguientes sustituciones:

- *region-code*— Especifique el Región de AWS lugar donde va a utilizar AWS PCS, por ejemplo. `us-east-1`
- *vpc-id*— Especifique el ID del VPC que va a utilizar AWS PCS.
- *efa-group-name*— Proporcione el nombre que haya elegido para el grupo de seguridad.

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

Utilice los siguientes comandos para adjuntar las reglas de los grupos de seguridad entrantes y salientes. Realice la siguiente sustitución:

- *efa-secgroup-id*— Proporcione el ID del grupo de EFA seguridad que acaba de crear.

```
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

Puede usar una CloudFormation plantilla para crear un grupo de seguridad que sea compatibleEFA. Descargue la plantilla de las siguientes opciones yURL, a continuación, cárguela en la AWS CloudFormation consola.

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

Con la plantilla abierta en la AWS CloudFormation consola, introduce las siguientes opciones.

- En Proporcione un nombre de pila
  - En Nombre de pila, introduce un nombre como`efa-sg-stack`.
- En Parámetros
  - En SecurityGroupName, introduzca un nombre como`efa-sg`.
  - En VPC, selecciona el VPC lugar que vas a usar AWS PCS.

Termine de crear la CloudFormation pila y supervise su estado. Cuando llegue, `CREATE_COMPLETE` el grupo EFA de seguridad estará listo para su uso.

(Opcional) Cree un grupo de ubicación

Se recomienda lanzar todas las instancias que se utilicen EFA en un grupo de ubicación en clúster para minimizar la distancia física entre ellas. Le recomendamos que cree un grupo de ubicación para cada grupo de nodos de procesamiento que vaya a utilizarEFA. Consulte esta Grupos de ubicación para EC2 instancias en AWS PCS sección para crear un grupo de ubicación para su grupo de nodos de cómputo.

Cree o actualice una plantilla de EC2 lanzamiento

EFALas interfaces de red se configuran en la plantilla de EC2 lanzamiento de un grupo de nodos de AWS PCS cómputo. Si hay varias tarjetas de red, se EFAs pueden configurar varias. El grupo EFA de seguridad y el grupo de ubicación opcional también se incluyen en la plantilla de lanzamiento.

A continuación, se muestra un ejemplo de plantilla de lanzamiento para instancias con dos tarjetas de red, como hpc7a.96xlarge. Las instancias se lanzarán en un grupo de ubicación en clústeres. `subnet-`*`SubnetID1`* `pg-`*`PlacementGroupId1`*

Los grupos de seguridad se deben agregar específicamente a cada EFA interfaz. Cada uno EFA necesita el grupo de seguridad que habilita EFA el tráfico (`sg-`*`EfaSecGroupId`*). Otros grupos de seguridad, especialmente los que gestionan el tráfico normalHTTPS, como SSH o, solo necesitan estar conectados a la interfaz de red principal (designada con un `DeviceIndex` de`0`). Las plantillas de lanzamiento en las que se definen las interfaces de red no admiten la configuración de grupos de seguridad mediante el `SecurityGroupIds` parámetro; debe establecer un valor para cada interfaz `Groups` de red que configure.

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

Cree o actualice un grupo de nodos de cómputo

Cree o actualice un grupo de nodos de AWS PCS cómputo con instancias que tengan el mismo número y la misma arquitectura de vCPUs procesador y que todas sean compatiblesEFA. Configure el grupo de nodos de cómputo para usarlo AMI con el EFA software instalado y para usar la plantilla de lanzamiento que configura las interfaces de red EFA habilitadas.

Prueba (opcional) EFA

Para demostrar la comunicación EFA habilitada entre dos nodos de un grupo de nodos de procesamiento, ejecute el `fi_pingpong` programa, que se incluye en la instalación del EFA software. Si la prueba se realiza correctamente, es probable que EFA esté configurada correctamente.

Para empezar, necesitas dos instancias en ejecución en el grupo de nodos de cómputo. Si tu grupo de nodos de cómputo usa capacidad estática, ya debería haber instancias disponibles. En el caso de un grupo de nodos de cómputo que utilice capacidad dinámica, puede lanzar dos nodos mediante el `salloc` comando. A continuación, se muestra un ejemplo de un clúster con un nombre de grupo de nodos dinámico `hpc7g` asociado a una cola denominada`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
```

Averigüe la dirección IP de los dos nodos asignados mediante`scontrol`. En el siguiente ejemplo, las direcciones son `10.3.140.69` para `hpc7g-1` y `10.3.132.211` para`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
```

Conéctese a uno de los nodos (en este caso de ejemplohpc7g-1) mediante SSH (oSSM). Tenga en cuenta que se trata de una dirección IP interna, por lo que puede que necesite conectarse desde uno de sus nodos de inicio de sesión si la utilizaSSH. Ten en cuenta también que la instancia debe configurarse con una SSH clave mediante la plantilla de lanzamiento del grupo de nodos de cómputo.

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

Ahora, fi_pingpong ejecútala en modo servidor.

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

Conéctese a la segunda instancia (hpc7g-2).

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

Se ejecuta fi_pingpong en modo cliente, conectándose al servidor activadohpc7g-1. Debería ver un resultado parecido al del ejemplo siguiente.

```
% /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
```

(Opcional) Usa una CloudFormation plantilla para crear una plantilla de lanzamiento EFA habilitada

Como hay que configurar varias dependenciasEFA, se ha proporcionado una CloudFormation plantilla que puede utilizar para configurar un grupo de nodos de cómputo. Admite instancias con hasta cuatro tarjetas de red. Para obtener más información sobre las instancias con varias tarjetas de red, consulte las interfaces de red elásticas en la Guía del usuario de Amazon Elastic Compute Cloud.

Descarga la CloudFormation plantilla de las siguientes opciones URL y, a continuación, cárgala en la CloudFormation consola Región de AWS donde la utilices AWS PCS.

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

Con la plantilla abierta en la AWS CloudFormation consola, introduce los siguientes valores. Tenga en cuenta que la plantilla proporcionará algunos valores de parámetros predeterminados; puede dejarlos como valores predeterminados.

- En Indique un nombre de pila
  - En Nombre de pila, introduce un nombre descriptivo. Le recomendamos que incorpore el nombre que elija para su grupo de nodos de AWS PCS cómputo, por ejemplo*NODEGROUPNAME*`-efa-lt`.
- En Parámetros
  - En NumberOfNetworkCards, selecciona el número de tarjetas de red en las instancias que estarán en tu grupo de nodos.
  - En VpcId, elige el VPC lugar en el que se implementará el AWS PCS clúster.
  - En NodeGroupSubnetId, elige la subred del clúster en la que VPC se lanzarán las instancias EFA habilitadas.
  - En esta sección PlacementGroupName, deja el campo en blanco para crear un nuevo grupo de ubicación de clústeres para el grupo de nodos. Si ya tiene un grupo de ubicación que quiere usar, introduzca su nombre aquí.
  - En ClusterSecurityGroupId, elija el grupo de seguridad que va a utilizar para permitir el acceso a otras instancias del clúster y al AWS PCSAPI. Muchos clientes eligen el grupo de seguridad predeterminado de su clústerVPC.
  - En la sección SshSecurityGroupId, proporcione el ID del grupo de seguridad que esté utilizando para permitir el SSH acceso entrante a los nodos del clúster.
  - Para SshKeyName, seleccione el SSH par de claves para acceder a los nodos del clúster.
  - Para LaunchTemplateName, introduzca un nombre descriptivo para la plantilla de lanzamiento, por ejemplo*NODEGROUPNAME*`-efa-lt`. El nombre debe ser único para usted Cuenta de AWS en el Región de AWS lugar donde lo vaya a utilizar AWS PCS.
- En Capacidades
  - Marque la casilla de «Reconozco que AWS CloudFormation podría crear IAM recursos».

Supervisa el estado de la CloudFormation pila. Cuando llegue a `CREATE_COMPLETE` la plantilla de lanzamiento estará lista para ser utilizada. Úselo con un grupo de nodos de AWS PCS cómputo, tal y como se describe anteriormente enCree o actualice un grupo de nodos de cómputo.

Uso de sistemas de archivos de red con AWS PCS

Puede adjuntar volúmenes de almacenamiento de red a los nodos lanzados en un grupo de nodos de cómputo de AWS Parallel Computing Service (AWS PCS) para proporcionar una ubicación persistente en la que se puedan escribir los datos y los archivos y acceder a ellos. Puede utilizar los volúmenes proporcionados por AWS los servicios. Los volúmenes incluyen Amazon Elastic File System (AmazonEFS), Amazon FSx for NetApp ONTAP, Amazon FSx for Open ZFS, Amazon FSx for Lustre y Amazon File Cache. También puede utilizar volúmenes autogestionados, como NFS servidores.

En este tema se describen algunas consideraciones y ejemplos del uso de sistemas de archivos en red con. AWS PCS

Consideraciones sobre el uso de sistemas de archivos de red

Los detalles de implementación de los distintos sistemas de archivos son diferentes, pero hay algunas consideraciones comunes.

- El software del sistema de archivos correspondiente debe estar instalado en la instancia. Por ejemplo, para usar Amazon FSx for Lustre, debe estar presente el Lustre paquete adecuado. Esto se puede lograr incluyéndolo en el grupo de nodos de cómputo AMI o utilizando un script que se ejecute al arrancar la instancia.
- Debe haber una ruta de red entre el volumen de almacenamiento compartido y las instancias del grupo de nodos de cómputo.
- Las reglas del grupo de seguridad, tanto para el volumen de almacenamiento compartido como para las instancias del grupo de nodos de procesamiento, deben permitir las conexiones a los puertos correspondientes.
- Debe mantener un espacio de nombres de POSIX usuario y grupo coherente en todos los recursos que acceden a los sistemas de archivos. De lo contrario, los trabajos y los procesos interactivos que se ejecutan en el PCS clúster pueden producir errores de permisos.
- Los montajes del sistema de archivos se realizan mediante plantillas de EC2 lanzamiento. Los errores o los tiempos de espera al montar un sistema de archivos de red pueden impedir que las instancias estén disponibles para ejecutar tareas. Esto, a su vez, puede generar costes inesperados. Para obtener más información sobre la depuración de plantillas de lanzamiento, consulteUso de plantillas de EC2 lanzamiento de Amazon con AWS PCS.

Ejemplos de montajes de red

Puede crear sistemas de archivos con AmazonEFS, Amazon FSx for Lustre, Amazon FSx for Open ZFS y Amazon File Cache. Amplíe la sección correspondiente a continuación para ver un ejemplo de cada montaje de red.

Amazon EFS

Configuración del sistema de archivos

Crea un sistema de EFS archivos de Amazon. Asegúrese de que tenga un objetivo de montaje en cada zona de disponibilidad en la que vaya a lanzar instancias de grupos de nodos de PCS cómputo. Asegúrese también de que cada destino de montaje esté asociado a un grupo de seguridad que permita el acceso entrante y saliente desde las instancias del grupo de PCS nodos de procesamiento. Para obtener más información, consulte Montar objetivos y grupos de seguridad en la Guía del usuario de Amazon Elastic File System.

Plantilla de lanzamiento

Añada los grupos de seguridad de la configuración del sistema de archivos a la plantilla de lanzamiento que utilizará para el grupo de nodos de cómputo.

Incluya los datos del usuario que utilizan el `cloud-config` mecanismo para montar el sistema de EFS archivos de Amazon. Sustituya los siguientes valores de este script por sus propios detalles:

- *`mount-point-directory`*— La ruta en cada instancia en la que montarás Amazon EFS
- *`filesystem-id`*— El ID del sistema de EFS archivos del sistema de archivos

```
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 para Lustre

Configuración del sistema de archivos

Cree un sistema de archivos FSx para Lustre en el VPC lugar donde lo vaya a utilizar AWS PCS. Para minimizar las transferencias entre zonas, impleméntelo en una subred de la misma zona de disponibilidad en la que lanzará la mayoría de las instancias del grupo de nodos de PCS cómputo. Asegúrese de que el sistema de archivos esté asociado a un grupo de seguridad que permita el acceso entrante y saliente desde las instancias del grupo de nodos de PCS procesamiento. Para obtener más información sobre los grupos de seguridad, consulte Control de acceso al sistema de archivos con Amazon VPC en la Guía del usuario de Amazon FSx for Lustre.

Plantilla de lanzamiento

Incluya los datos de usuario que se utilizan `cloud-config` para montar el sistema FSx de archivos de Lustre. Sustituya los siguientes valores de este script por sus propios detalles:

- *`mount-point-directory`*— La ruta de la instancia en la que quieres montarla FSx para Lustre
- *`filesystem-id`*— El ID del sistema de archivos del sistema de archivos FSx de Lustre
- *`mount-name`*— El nombre de montaje del sistema de FSx archivos para Lustre
- *`region-code`*— El Región de AWS lugar donde se implementa el sistema de archivos FSx for Lustre (debe ser el mismo que el de su AWS PCS sistema)
- (Opcional)*`latest`*: cualquier versión de Lustre Lustre FSx compatible

```
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 para Open ZFS

Configuración del sistema de archivos

Cree un sistema de ZFS archivos FSx para abrir en el VPC lugar donde lo va a utilizar AWS PCS. Para minimizar las transferencias entre zonas, impleméntelas en una subred de la misma zona de disponibilidad en la que lanzará la mayoría de las instancias del grupo de nodos de AWS PCS cómputo. Asegúrese de que el sistema de archivos esté asociado a un grupo de seguridad que permita el acceso entrante y saliente desde las instancias del grupo de nodos de AWS PCS procesamiento. Para obtener más información sobre los grupos de seguridad, consulte Administrar el acceso al sistema de archivos con Amazon VPC en la Guía del ZFS usuario de FSx for Open.

Plantilla de lanzamiento

Incluya los datos de usuario que `cloud-config` se utilizan para montar el volumen raíz de un FSx sistema de ZFS archivos de Open. Sustituya los siguientes valores de este script por sus propios detalles:

- *`mount-point-directory`*— La ruta de una instancia en la que quieres montar tu ZFS recurso compartido FSx para Open
- *`filesystem-id`*— El ID del sistema de archivos del sistema FSx de ZFS archivos Open
- *`region-code`*— El Región de AWS lugar donde se implementa el sistema de ZFS archivos FSx for Open (debe ser el mismo que el de su AWS PCS sistema)

```
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,wsize=1048576 filesystem-
id.fsx.region-code.amazonaws.com:/fsx/ /mount-point-directory
```

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

Amazon File Cache

Configuración del sistema de archivos

Crea una caché de archivos de Amazon en el VPC lugar donde la usarás AWS PCS. Para minimizar las transferencias entre zonas, elija una subred en la misma zona de disponibilidad en la que lanzará la mayoría de las instancias del grupo de nodos de PCS cómputo. Asegúrese de que la caché de archivos esté asociada a un grupo de seguridad que permita el tráfico entrante y saliente en el puerto 988 entre las PCS instancias y la caché de archivos. Para obtener más información sobre los grupos de seguridad, consulte Control de acceso a caché con Amazon VPC en la Guía del usuario de Amazon File Cache.

Plantilla de lanzamiento

Añada los grupos de seguridad de la configuración de su sistema de archivos a la plantilla de lanzamiento que utilizará para el grupo de nodos de cómputo.

Incluye los datos de usuario que se utilizan `cloud-config` para montar la caché de archivos de Amazon. Sustituya los siguientes valores de este script por sus propios detalles:

- *`mount-point-directory`*— La ruta de la instancia en la que quieres montarla FSx para Lustre
- *`cache-dns-name`*— El nombre del sistema de nombres de dominio (DNS) de la caché de archivos
- *`mount-name`*— El nombre de montaje de la caché de archivos

```
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) para AWS PCS

AWS PCSfunciona con lo AMIs que usted proporciona, lo que ofrece una gran flexibilidad en el software y la configuración que se encuentran en los nodos de su clúster. Si está realizando una prueba AWS PCS, puede utilizar una muestra AMI proporcionada y mantenida por AWS. Si lo está utilizando AWS PCS en producción, le recomendamos que construya el suyo propioAMIs. En este tema se explica cómo descubrir y utilizar la muestraAMIs, así como cómo crear y utilizar la suya propia personalizadaAMIs.

Temas

Uso de Amazon Machine Images (AMIs) de muestra con AWS PCS

AWSproporciona un ejemplo AMIs que puede utilizar como punto de partida para trabajar con él AWS PCS.

> **Important**
>
> AMIsLos ejemplos son para fines de demostración y no se recomiendan para cargas de trabajo de producción.

Encuentre la muestra actual AWS PCS AMIs

AWS Management Console

AWSPCSlas muestras AMIs tienen la siguiente convención de nomenclatura:

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

Valores aceptados

- *OS* – amzn2
- *architecture* — x86_64 o arm64

- *scheduler* – slurm
- *scheduler-major-version* – 23.11

Para encontrar AWS PCS una muestra AMIs

1. Abre la EC2consola de Amazon.
2. Navega hasta AMIs.
3. Seleccione Imágenes públicas.
4. En Buscar AMI por atributo o etiqueta, busque y AMI utilice el nombre de la plantilla.

   Ejemplos

   - Slurm 23.11 es compatible con Graviton AMI

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

   - Ejemplo de instancias x86 AMI

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

   > **Note**
   >
   > Si hay variasAMIs, utilícelas AMI con la marca de tiempo más reciente.

5. Usa el AMI ID cuando crees o actualices un grupo de nodos de cómputo.

AWS CLI

Puedes encontrar el AWS PCS ejemplo más reciente AMI con los siguientes comandos. Reemplazar *region-code* con el Región de AWS lugar que utilice AWS PCS, por ejemplous-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
```

- Arm64

```
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
```

Use el AMI ID cuando cree o actualice un grupo de nodos de cómputo.

Más información sobre el AWS PCS ejemplo AMIs

Para ver el contenido y los detalles de configuración de las versiones actuales y anteriores del AWS PCS ejemploAMIs, consulteNotas de publicación para AWS PCS muestra AMIs.

Cree la suya propia AMIs compatible con AWS PCS

Para aprender a crear los tuyos propios con AMIs los que puedas trabajar AWS PCS, consultaImágenes personalizadas de Amazon Machine (AMIs) para AWS PCS.

Imágenes personalizadas de Amazon Machine (AMIs) para AWS PCS

AWS PCSestá diseñado para funcionar con Amazon Machine Images (AMI) que usted incorpore al servicio. AMIsPueden tener instalados software y configuraciones arbitrarios, siempre que tengan el AWS PCS agente y una versión compatible de Slurm instalados y configurados correctamente. Debe utilizar los AWS instaladores proporcionados para instalar el AWS PCS software en su versión personalizada. AMI Le recomendamos que utilice los AWS instaladores proporcionados para instalar Slurm de forma personalizada, AMI pero puede instalar Slurm por su cuenta si lo prefiere (no se recomienda).

> **Note**
>
> Si quiere intentarlo AWS PCS sin crear una personalizadaAMI, puede utilizar un ejemplo proporcionado por. AMI AWS Para obtener más información, consulte Uso de Amazon Machine Images (AMIs) de muestra con AWS PCS.

Este tutorial le ayuda a crear una AMI que pueda usarse con grupos de nodos de PCS cómputo para impulsar sus cargas de trabajo y las de inteligencia artificial HPC y aprendizaje automático.

Temas

Paso 1: lanza una instancia temporal

Lance una instancia temporal que pueda usar para instalar y configurar el AWS PCS software y el programador de Slurm. Utiliza esta instancia para crear una AMI compatible con. AWS PCS

Para iniciar una instancia temporal

1. Abre la EC2consola de Amazon.
2. En el panel de navegación, selecciona Instances y, a continuación, selecciona Launch instances para abrir el asistente de nuevas instancias de lanzamiento.
3. (Opcional) En la sección Nombre y etiquetas, proporciona un nombre para la instancia, por ejemploPCS-AMI-instance. El nombre se asigna a la instancia como etiqueta de recurso (Name=PCS-AMI-instance).
4. En la sección Imágenes de aplicaciones y sistemas operativos, seleccione uno AMI para uno de los sistemas operativos compatibles.
5. En la sección Tipo de instancia, seleccione el tipo de instancia admitida.
6. En la sección Par de claves, seleccione el par de claves que desea utilizar en la instancia.
7. En la sección Configuración de red:
    - En Firewall (grupos de seguridad), selecciona Seleccionar un grupo de seguridad existente y, a continuación, selecciona un grupo de seguridad que permita el SSH acceso entrante a la instancia.

8. En la sección Almacenamiento, configure los volúmenes según sea necesario. Asegúrese de configurar suficiente espacio para instalar sus propias aplicaciones y bibliotecas.
9. En el panel Resumen, elija Iniciar instancia.

Paso 2: Instalar el AWS PCS agente

Instale el agente que configura las instancias lanzadas por AWS PCS para su uso con Slurm.

Instalación del agente de AWS PCS

1. Conéctese a la instancia que lanzó. Para obtener más información, consulte Conexión con la instancia de Linux.
2. (Opcional) Para asegurarse de que todos los paquetes de software estén actualizados, realice una actualización rápida del software de la instancia. Este proceso puede demorar unos minutos.

   - 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. Reinicie la instancia y vuelva a conectarse a ella.
4. Descargue los archivos de instalación del AWS PCS agente. Los archivos de instalación se empaquetan en un archivo tarball (`.tar.gz`) comprimido. Para descargar la última versión estable, utilice el comando siguiente. Sustituya *`region`* con el Región de AWS lugar en el que lanzaste tu instancia temporal, por ejemplo`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
   ```

   También puede obtener la última versión sustituyendo el número de versión por `latest` el del comando anterior (por ejemplo:`aws-pcs-agent-v1-latest.tar.gz`).

> **Note**
>
> Esto podría cambiar en futuras versiones del software del AWS PCS agente.

5. (Opcional) Compruebe la autenticidad e integridad del archivo tar del AWS PCS software. Le recomendamos que lo haga para verificar la identidad del editor de software y para verificar que el archivo no se haya modificado ni dañado desde que se publicó.

   a. Descarga la GPG clave pública AWS PCS e impórtala a tu llavero. Sustituya *region* con el Región de AWS lugar en el que lanzaste tu instancia temporal. El comando debe devolver un valor de clave. Registra el valor clave y úsalo en el siguiente paso.

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

   b. Ejecute el siguiente comando para verificar la huella digital de la GPG clave.

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

   El comando debería devolver una huella digital idéntica a la siguiente:

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

   > **Important**
   >
   > No ejecute el script de instalación del AWS PCS agente si la huella digital no coincide. Contacte con AWS Support.

   c. Descargue el archivo de firma y compruebe la firma del archivo tar del AWS PCS software. Reemplazar *region* con el Región de AWS lugar donde lanzaste tu instancia temporal, por ejemplo. `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
   ```

   El resultado debería ser similar al siguiente:

```
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
```

Si el resultado incluye `Good signature` y la huella digital coincide con la huella digital devuelta en el paso anterior, continúe con el paso siguiente.

> **Important**
>
> No ejecute el script de instalación del AWS PCS software si la huella digital no coincide. Contacte con AWS Support.

6. Extraiga los archivos del `.tar.gz` archivo comprimido y navegue hasta el directorio extraído.

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

7. Instale el software AWS PCS.

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

8. Compruebe el archivo de la versión del AWS PCS software para confirmar que la instalación se ha realizado correctamente.

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

El resultado debería ser similar al siguiente:

```
AGENT_INSTALL_DATE='Mon Aug 12 12:28:43 UTC 2024'
AGENT_VERSION='1.0.0'
AGENT_RELEASE='1'
```

Paso 3: Instalar Slurm

Instale una versión de Slurm que sea compatible con. AWS PCS

Para instalar Slurm

1. Conéctese a la misma instancia temporal en la que instaló el AWS PCS software.
2. Descargue el software de instalación de Slurm. El instalador de Slurm está empaquetado en un archivo tarball () comprimido. `.tar.gz` Para descargar la última versión estable, utilice el comando siguiente. Sustituya *`region`* con la Región de AWS de su instancia temporal, como`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
   ```

   También puede obtener la última versión sustituyendo el número de versión por `latest` el del comando anterior (por ejemplo:`aws-pcs-slurm-23.11-installer-latest.tar.gz`).

   > **Note**
   >
   > Esto podría cambiar en futuras versiones del software de instalación de Slurm.

3. (Opcional) Compruebe la autenticidad e integridad del archivo tar del instalador de Slurm. Le recomendamos que lo haga para verificar la identidad del editor de software y para verificar que el archivo no se haya modificado ni dañado desde que se publicó.

   a. Descarga la GPG clave pública AWS PCS e impórtala a tu conjunto de claves. Sustituya *`region`* con el Región de AWS lugar en el que lanzaste tu instancia temporal. El comando debe devolver un valor de clave. Registra el valor clave y úsalo en el siguiente paso.

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

   b. Ejecute el siguiente comando para verificar la huella digital de la GPG clave.

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

El comando debería devolver una huella digital idéntica a la siguiente:

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

> **Important**
>
> No ejecute el script de instalación de Slurm si la huella digital no coincide. Contacte con AWS Support.

c. Descargue el archivo de firma y verifique la firma del archivo tarball del instalador de Slurm. Reemplazar *region* con el Región de AWS lugar donde lanzaste tu instancia temporal, por ejemplo. `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
```

El resultado debería ser similar al siguiente:

```
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
```

Si el resultado incluye `Good signature` y la huella digital coincide con la huella digital devuelta en el paso anterior, continúe con el paso siguiente.

> **Important**
>
> No ejecute el script de instalación de Slurm si la huella digital no coincide. Contacte con AWS Support.

4. Extraiga los archivos desde el archivo `.tar.gz` comprimido y acceda al directorio extraído.

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

5. Instale Slurm. El instalador descarga, compila e instala Slurm y sus dependencias. Tarda varios minutos, según las especificaciones de la instancia temporal que haya seleccionado.

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

6. Compruebe el archivo de versión del programador para confirmar la instalación.

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

El resultado debería ser similar al siguiente:

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

Paso 4: (opcional) Instalar controladores, bibliotecas y software de aplicación adicionales

Instale controladores, bibliotecas y software de aplicación adicionales en la instancia temporal. Los procedimientos de instalación variarán en función de las aplicaciones y bibliotecas específicas.
Si no ha creado una personalizada AMI AWS PCS anteriormente, le recomendamos que primero cree y pruebe una AMI con solo el AWS PCS software y Slurm instalados y, a continuación, añada gradualmente su propio software y configuraciones una vez que haya confirmado su éxito inicial.

Ejemplos

- Software Elastic Fabric Adapter ()EFA. Para obtener más información, consulta Cómo empezar con EFA y MPI para HPC las cargas de trabajo en Amazon EC2 en la Guía del usuario de Amazon Elastic Compute Cloud.
- Cliente Amazon Elastic File System (AmazonEFS). Para obtener más información, consulte Instalación manual del EFS cliente de Amazon en la Guía del usuario de Amazon Elastic File System.
- Cliente Lustre, para usar Amazon FSx for Lustre y Amazon File Cache. Para obtener más información, consulte Instalación del cliente de Lustre en la Guía del usuario FSxde Lustre.

- CloudWatch Agente de Amazon, para usar CloudWatch registros y métricas. Para obtener más información, consulte Instalar el CloudWatch agente en la Guía del CloudWatch usuario de Amazon.
- AWS Neuron, para usar los tipos de instancia trn* e inf*. Para obtener más información, consulte la documentación de Neuron.AWS
- NVIDIAControlador y CUDADCGM, para usar los tipos de instancia p* o g*.

Paso 5: Crea una compatible con AMI AWS PCS

Una vez instalados los componentes de software necesarios, debe crear uno AMI que pueda reutilizar para lanzar instancias en grupos de nodos de AWS PCS cómputo.

Para crear una a AMI partir de tu instancia temporal

1. Abre la EC2consola de Amazon.
2. En el panel de navegación, seleccione Instances (Instancia[s]).
3. Selecciona la instancia temporal que has creado. Selecciona Acciones, Imagen y Crear imagen.
4. En Crear imagen, realice lo siguiente:
    a. En Nombre de imagen, introduzca un nombre descriptivo paraAMI.
    b. (Opcional) En la descripción de la imagen, introduzca una breve descripción del propósito de laAMI.
    c. Elija Crear imagen.
5. En el panel de navegación, elija AMIs.
6. Busque el AMI que creó en la lista. Espere a que su estado cambie de Pendiente a Disponible y utilícelo con un grupo de AWS PCS nodos de procesamiento.

Paso 6: Usa la configuración personalizada AMI con un grupo de nodos de AWS PCS cómputo

Puede usar su configuración personalizada AMI con un grupo de nodos de AWS PCS procesamiento nuevo o existente.

New compute node group

Para usar el personalizado AMI

1. Abre la AWS PCSconsola.
2. En el panel de navegación, seleccione Clusters (Clústeres).
3. Elija el clúster en el que utilizará la configuración personalizada y, a continuaciónAMI, seleccione Grupos de nodos de cómputo.
4. Cree un nuevo grupo de nodos de procesamiento. Para obtener más información, consulte Crear un grupo de nodos de cómputo en AWS PCS. En AMIID, busca el nombre o el ID de la AMI personalización que quieres usar. Termine de configurar el grupo de nodos de cómputo y, a continuación, seleccione Crear grupo de nodos de cómputo.
5. (Opcional) Confirme que AMI admite los lanzamientos de instancias. Lanza una instancia en el grupo de nodos de cómputo. Para ello, configura el grupo de nodos de cómputo para que tenga una única instancia estática, o puedes enviar un trabajo a una cola que utilice el grupo de nodos de cómputo.

   a. Comprueba la EC2 consola de Amazon hasta que aparezca una instancia etiquetada con el nuevo ID de grupo de nodos de cómputo. Para obtener más información al respecto, consultaBúsqueda de instancias de grupos de nodos de cómputo en AWS PCS..

   b. Cuando veas que una instancia se lanza y completa su proceso de arranque, confirma que está utilizando lo esperadoAMI. Para ello, selecciona la instancia y, a continuación, inspecciona el AMIID en Detalles. Debe coincidir con lo AMI que configuraste en la configuración del grupo de nodos de cómputo.

   c. (Opcional) Actualice la configuración de escalado del grupo de nodos de cómputo a sus valores preferidos.

Existing compute node group

Para usar la personalizada AMI

1. Abre la AWS PCSconsola.
2. En el panel de navegación, seleccione Clusters (Clústeres).
3. Elija el clúster en el que utilizará la configuración personalizada y, a continuaciónAMI, seleccione Grupos de nodos de cómputo.

4. Seleccione el grupo de nodos que desee configurar y elija Editar. En AMIID, busca el nombre o el ID de la AMI personalización que quieres usar. Termine de configurar el grupo de nodos de procesamiento y, a continuación, seleccione Actualizar. Las nuevas instancias lanzadas en el grupo de nodos de cómputo utilizarán el AMI ID actualizado. Las instancias existentes seguirán usando las antiguas AMI hasta que las AWS PCS reemplacen. Para obtener más información, consulte Actualización de un grupo de nodos de AWS PCS cómputo.
5. (Opcional) Confirma que las instancias AMI compatibles se lanzan. Lanza una instancia en el grupo de nodos de cómputo. Para ello, configura el grupo de nodos de cómputo para que tenga una única instancia estática, o puedes enviar un trabajo a una cola que utilice el grupo de nodos de cómputo.

    a. Comprueba la EC2 consola de Amazon hasta que aparezca una instancia etiquetada con el nuevo ID de grupo de nodos de cómputo. Para obtener más información al respecto, consultaBúsqueda de instancias de grupos de nodos de cómputo en AWS PCS..

    b. Cuando veas que una instancia se lanza y completa su proceso de arranque, confirma que está utilizando lo esperadoAMI. Para ello, selecciona la instancia y, a continuación, inspecciona el AMIID en Detalles. Debe coincidir con lo AMI que configuraste en la configuración del grupo de nodos de cómputo.

    c. (Opcional) Actualice la configuración de escalado del grupo de nodos de cómputo a sus valores preferidos.

Paso 7: Finalizar la instancia temporal

Una vez que haya confirmado que AMI funciona según lo previsto AWS PCS, puede cancelar la instancia temporal para dejar de incurrir en cargos por ella.

Para terminar la instancia temporal

1. Abre la EC2consola de Amazon.
2. En el panel de navegación, seleccione Instances (Instancia[s]).
3. Selecciona la instancia temporal que has creado y selecciona Acciones, Estado de la instancia y Finalizar instancia.
4. Cuando se le pida que confirme, elija Finalizar.

Instaladores de software para crear programas personalizados AMIs AWS PCS

AWS proporciona un archivo descargable que puede instalar el AWS PCS software en una instancia. AWS también proporciona software que puede descargar, compilar e instalar las versiones pertinentes de Slurm y sus dependencias. Puede usar estas instrucciones para crear métodos personalizados AMIs para usarlos con ellos AWS PCS o puede usar sus propios métodos.

Contenido

AWS PCSinstalador de software

El instalador de AWS PCS software configura una instancia para que funcione AWS PCS durante el proceso de arranque de la instancia. Debe utilizar los AWS instaladores proporcionados para instalar el AWS PCS software de forma personalizada. AMI

Instalador Slurm

El instalador de Slurm descarga, compila e instala las versiones relevantes de Slurm y sus dependencias. Puede usar el instalador de Slurm para crear versiones personalizadas. AMIs AWS PCS También puede utilizar sus propios mecanismos si son coherentes con la configuración de software que proporciona el instalador de Slurm.

El software AWS proporcionado instala lo siguiente:

- Utilice la versión principal y de mantenimiento solicitada (actualmente la versión 23.11.8): licencia 2 GPL
  - Slurm está construido con un conjunto de `--sysconfdir /etc/slurm`
  - Slurm está diseñado con la opción y `--enable-pam --without-munge`

  - Slurm se construye con la opción `--sharedstatedir=/run/slurm/`
  - Slurm está construido con un soporte PMIX JWT
  - Slurm está instalado en `/opt/aws/pcs/schedulers/slurm-23.11`
- Open PMIX (versión 4.2.6) — Licencia
  - Open PMIX se instala como un subdirectorio de `/opt/aws/pcs/scheduler/`
- libjwt (versión 1.15.3) — Licencia -2.0 MPL
  - libjwt se instala como un subdirectorio de `/opt/aws/pcs/scheduler/`

El software AWS suministrado cambia la configuración del sistema de la siguiente manera:

- El `systemd` archivo Slurm creado por la compilación se copia `/etc/systemd/system/` con el nombre del archivo. `slurmd-23.11.service`
- Si no existen, se crean un usuario y un grupo de Slurm (`slurm:slurm`) con/de. UID GID `401`
- En Amazon Linux 2 y Rocky Linux 9, la instalación añade el EPEL repositorio para instalar el software necesario para compilar Slurm o sus dependencias.
- Durante RHEL9 la instalación, se habilitará `codeready-builder-for-rhel-9-rhui-rpms` y `epel-release-latest-9` se instalará el software necesario `fedoraproject` para compilar Slurm o sus dependencias.

Sistemas operativos compatibles

El AWS PCS software y los instaladores de Slurm son compatibles con los siguientes sistemas operativos:

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

> **Note**
>
> AWS Deep Learning AMIs (DLAMI) las versiones basadas en Amazon Linux 2 y Ubuntu 22.04 deberían ser compatibles con el AWS PCS software y los instaladores de

Slurm. Para obtener más información, consulte Choosing Your DLAMI en la Guía para desarrolladores.AWS Deep Learning AMIs

Tipos de instancias admitidas

AWS PCSLos instaladores de software y Slurm admiten cualquier tipo de instancia x86_64 o arm64 que pueda ejecutar uno de los sistemas operativos compatibles.

Versiones de Slurm compatibles

Se admiten las siguientes versiones principales de Slurm:

- Slurm 23.11

Verifique los instaladores mediante una suma de verificación

Puede utilizar SHA256 sumas de comprobación para comprobar los archivos tar (.tar.gz) del instalador. Le recomendamos que lo haga para verificar la identidad del editor de software y para comprobar que la aplicación no se ha modificado ni dañado desde que se publicó.

Para verificar un tarball

Utilice la utilidad sha256sum para la suma de SHA256 comprobación y especifique el nombre del archivo tarball. Debe ejecutar el comando desde el directorio en el que guardó el archivo tarball.

- SHA256

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

El comando debe devolver un valor de suma de comprobación con el siguiente formato.

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

Compare el valor de la suma de verificación devuelto por el comando con el valor de la suma de verificación que se proporciona en la siguiente tabla. Si las sumas de comprobación coinciden, es seguro ejecutar el script de instalación.

> **Important**
>
> Si las sumas de comprobación no coinciden, no ejecute el script de instalación. Póngase en contacto con AWS Support.

Por ejemplo, el siguiente comando genera la SHA256 suma de comprobación del tarball de Slurm 23.11.9.

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

Ejemplo de salida:

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

En la siguiente tabla se enumeran las sumas de comprobación de las versiones recientes de los instaladores. Reemplazar *us-east-1* con el Región de AWS lugar donde se usa. AWS PCS

Installer (Instalador)	Descarga URL	SHA256suma de comprobación
Slurm 23.11.9	`https://aws-pcs-repo- us-east-1 .s3.amazonaws.com/aws-pcs-slurm/aws-pcs-slurm-23.11-installer-23.11.9-1.tar.gz`	`1de7d919c8632fe8e2806611bed4fde1005a4fadc795412456e935c7bba2a9b8`
AWS PCSagente 1.0.0	`https://aws-pcs-repo- us-east-1 .s3.amazonaws.com/aws-pcs-agent/aws-pcs-agent-v1.0.0-1.tar.gz`	`d2d3d68d00c685435c38af471d7e2492dde5ce9eb222d7b6ef0042144b134ce0`

Versiones de Slurm en AWS PCS

SchedMD mejora continuamente Slurm con nuevas capacidades, optimizaciones y parches de seguridad. SchedMD lanza una nueva versión principal a intervalos regulares y planea admitir hasta 3 versiones en un momento dado. AWS PCSinicialmente es compatible con Slurm 23.11. Puede actualizar su versión principal de Slurm después del lanzamiento de una nueva versión. AWS PCSestá diseñado para actualizar automáticamente el controlador Slurm con versiones de parches.

Cuando SchedMD finaliza el soporte para una versión principal en particular, AWS PCS también deja de dar soporte a esa versión principal. AWS PCSenvía un aviso anticipado si una versión principal de Slurm está cerca del final de su vida útil, para que los clientes sepan cuándo actualizar sus clústeres a una versión compatible más reciente.

Le recomendamos que utilice la última versión compatible de Slurm para implementar su clúster y acceder a los avances y mejoras más recientes.

Preguntas frecuentes sobre las versiones de Slurm

¿Durante cuánto tiempo es AWS PCS compatible una versión de Slurm?

AWS PCSsigue los ciclos de soporte de SchedMD para las principales versiones. AWS PCSadmite hasta 3 versiones principales en un momento dado. Cuando SchedMD publique una nueva versión principal, AWS PCS retira la versión compatible más antigua. AWS PCSpublique una nueva versión principal de Slurm lo antes posible, pero es posible que haya un retraso entre el lanzamiento de SchedMD y su disponibilidad en. AWS PCS

¿Cuándo me AWS PCS avisarán sobre el fin de la vida útil (EOSL) de las versiones de Slurm?

AWS PCSte lo notifica varias veces, con una cadencia predeterminada, antes de la fecha. EOSL

¿Qué debo hacer cuando se acerca una versión de Slurm? EOSL

Debe actualizar sus versiones de Slurm antes EOSL para ayudar a mantener un entorno seguro y compatible.

¿Cómo puedo actualizar mis clústeres para usar una nueva versión principal de Slurm?

Para actualizar la versión de Slurm, debe crear un clúster nuevo. También debe actualizarse al AWS PCS software equivalente del suyo AMI y usarlo para crear los grupos de nodos de cómputo del nuevo clúster.

¿Cómo obtendrán mis clústeres las nuevas versiones de parches de Slurm?

AWS PCSestá diseñado para aplicar parches automáticamente para abordar las vulnerabilidades y exposiciones más comunes de Slurm (). CVEs AWS PCSaplica los parches a los controladores de clúster que se ejecutan en las cuentas internas propiedad del servicio. Debe usar las AWS PCS API acciones AWS Management Console o para instalar parches en las EC2 instancias de su. Cuenta de AWS

¿Qué pasa si no actualizo Slurm antes de esa fecha? EOSL

AWS PCSestá diseñado para detener los clústeres que tienen una versión de Slurm no compatible. Debe actualizar la versión principal de Slurm del controlador de clúster y el AWS PCS software instalado en los grupos de nodos de procesamiento.

¿Cuántas versiones de Slurm admite? AWS PCS

AWS PCSadmite hasta 3 versiones principales de Slurm en un momento dado, incluidas la versión principal actual y las 2 versiones principales anteriores.

¿Qué actualizaciones de la versión de Slurm debo aplicar?

Le recomendamos encarecidamente que utilice la misma versión principal en todos los componentes del clúster e instale los parches más recientes tan pronto como se publiquen. Los grupos AMIs de nodos de cómputo deben usar una versión del software Slurm compatible con la versión Slurm del controlador de clúster. La versión principal de Slurm AMIs debe estar dentro de las dos versiones de la versión principal de Slurm en el controlador de clúster. La versión de Slurm instalada en las EC2 instancias en ejecución del clúster AMI y en las mismas no puede ser más reciente que la versión de Slurm del controlador de clúster. Para mantener la compatibilidad con su clúster, AMIs debe usar una versión de software compatible. AWS PCS

¿Qué sucede si actualizo la versión principal de Slurm pero utilizo un software de Slurm más antiguo en mis AMI grupos de nodos de cómputo?

Debe actualizar el AWS PCS software a la misma versión para utilizar la nueva funcionalidad de Slurm. Para obtener un AWS PCS soporte completo, todos los componentes de Slurm deben usar versiones compatibles. En resumen:

- Podemos ofrecer un soporte completo cuando el controlador de clúster y todos los componentes (AWS PCSpaquetes) del mismo utilizan las Cuenta de AWS versiones compatibles.
- AWS PCSestá diseñado para detener un clúster si llega la versión Slurm de su controlador. EOSL

- Si tienes la versión Slurm de los componentes Cuenta de AWS a tu alcanceEOSL, tu clúster no será compatible.

¿En qué orden debo actualizar los componentes de mi clúster?

Debe actualizar la versión de Slurm del controlador de clúster antes de utilizar una versión de AMI Slurm más reciente. Actualice un grupo de nodos de cómputo para usar el. AMI AWS PCSusa el AMI para lanzar nuevas EC2 instancias en el grupo de nodos de cómputo. AWS PCSno actualiza las EC2 instancias existentes que tienen tareas en ejecución; AWS PCS está diseñada para terminar esas instancias una vez finalizadas sus tareas.

¿ AWS PCSOfrece soporte ampliado para las versiones de Slurm?

No. Le proporcionaremos información detallada sobre las opciones de soporte extendido, incluidos los costes adicionales y la cobertura de soporte específica proporcionada.

Servicio de seguridad en computación AWS paralela

La seguridad en la nube AWS es la máxima prioridad. Como AWS cliente, usted se beneficia de los centros de datos y las arquitecturas de red diseñados para cumplir con los requisitos de las organizaciones más sensibles a la seguridad.

La seguridad es una responsabilidad compartida entre AWS usted y usted. El modelo de responsabilidad compartida la describe como seguridad de la nube y seguridad en la nube:

- Seguridad de la nube: AWS es responsable de proteger la infraestructura que ejecuta AWS los servicios en la Nube de AWS. AWS también le proporciona servicios que puede utilizar de forma segura. Los auditores externos prueban y verifican periódicamente la eficacia de nuestra seguridad como parte de los AWS programas de de . Para obtener más información sobre los programas de conformidad que se aplican al Servicio de Computación AWS Paralela, consulte AWS el apartado AWS Servicios incluidos .
- Seguridad en la nube: su responsabilidad viene determinada por el AWS servicio que utilice. Usted también es responsable de otros factores, incluida la confidencialidad de los datos, los requisitos de la empresa y la legislación y los reglamentos aplicables.

Esta documentación le ayuda a comprender cómo aplicar el modelo de responsabilidad compartida cuando se utiliza AWS PCS. Los siguientes temas muestran cómo configurarlo AWS PCS para cumplir sus objetivos de seguridad y conformidad. También aprenderá a utilizar otros AWS servicios que le ayudan a supervisar y proteger sus AWS PCS recursos.

Temas

- Prácticas recomendadas de seguridad para AWS Parallel Computing Service

Protección de datos en AWS Parallel Computing Service

El modelo de se aplica a protección de datos en AWS Parallel Computing Service. Como se describe en este modelo, AWS es responsable de proteger la infraestructura global que ejecuta todos los Nube de AWS. Usted es responsable de mantener el control sobre el contenido alojado en esta infraestructura. Usted también es responsable de las tareas de administración y configuración de seguridad para los Servicios de AWS que utiliza. Para obtener más información sobre la privacidad de los datos, consulte la sección Privacidad de datos FAQ. Para obtener información sobre la protección de datos en Europa, consulte el modelo de responsabilidad AWS compartida y la entrada del GDPR blog sobre AWS seguridad.

Con fines de protección de datos, le recomendamos que proteja Cuenta de AWS las credenciales y configure los usuarios individuales con AWS IAM Identity Center o AWS Identity and Access Management (IAM). De esta manera, solo se otorgan a cada usuario los permisos necesarios para cumplir sus obligaciones laborales. También recomendamos proteger sus datos de la siguiente manera:

- Utilice la autenticación multifactorial (MFA) con cada cuenta.
- UseSSL/TLSpara comunicarse con AWS los recursos. Necesitamos TLS 1.2 y recomendamos TLS 1.3.
- Configure API y registre la actividad del usuario con AWS CloudTrail.
- Utilice soluciones de AWS cifrado, junto con todos los controles de seguridad predeterminados Servicios de AWS.
- Utilice servicios de seguridad administrados avanzados, como Amazon Macie, que lo ayuden a detectar y proteger los datos confidenciales almacenados en Amazon S3.
- Si necesita entre FIPS 140 y 3 módulos criptográficos validados para acceder a AWS través de una interfaz de línea de comandos o unaAPI, utilice un FIPS terminal. Para obtener más información sobre los FIPS puntos finales disponibles, consulte la Norma federal de procesamiento de información () FIPS 140-3.

Se recomienda encarecidamente no introducir nunca información confidencial o sensible, como, por ejemplo, direcciones de correo electrónico de clientes, en etiquetas o campos de formato libre, tales como el campo Nombre. Esto incluye cuando trabaja con AWS PCS o Servicios de AWS utiliza la consola, API AWS CLI, o. AWS SDKs Cualquier dato que ingrese en etiquetas o campos de formato

libre utilizados para nombres se puede emplear para los registros de facturación o diagnóstico. Si proporciona una URL a un servidor externo, le recomendamos encarecidamente que no incluya la información sobre las credenciales URL para validar la solicitud a ese servidor.

Cifrado en reposo

El cifrado está activado de forma predeterminada para los datos en reposo al crear un clúster de AWS Parallel Computing Service (AWS PCS) con AWS Management Console AWS CLI, AWS PCSAPI, o AWS SDKs. AWS PCSutiliza una AWS KMSclave propia para cifrar los datos en reposo. Para obtener más información, consulte las claves y AWS claves del cliente en la Guía para AWS KMS desarrolladores. El secreto del clúster se almacena en la KMS clave gestionada por Secrets Manager AWS Secrets Manager y se cifra con ella. Para obtener más información, consulte Trabajar con secretos de clústeres en AWS PCS.

En un AWS PCS clúster, los siguientes datos están en reposo:

- Estado del programador: incluye datos sobre las tareas en ejecución y los nodos aprovisionados en el clúster. Estos son los datos en los que Slurm persiste según lo definido en su. `StateSaveLocation` `slurm.conf` Para obtener más información, consulte la descripción de la documentación StateSaveLocationde Slurm. AWS PCSelimina los datos del trabajo una vez finalizado un trabajo.
- Secreto de autenticación del programador: lo AWS PCS usa para autenticar todas las comunicaciones del programador en el clúster.

Para obtener información sobre el estado del programador, cifra AWS PCS automáticamente los datos y los metadatos antes de escribirlos en el sistema de archivos. El sistema de archivos cifrados utiliza el algoritmo de cifrado AES -256 estándar del sector para los datos en reposo.

Cifrado en tránsito

Sus conexiones al sistema AWS PCS API utilizan el TLS cifrado con el proceso de firma de la versión 4 de Signature, independientemente de si utiliza AWS Command Line Interface (AWS CLI) o. AWS SDKs Para obtener más información, consulte Firmar AWS API solicitudes en la Guía del AWS Identity and Access Management usuario. AWS gestiona el control de acceso mediante API las IAM políticas de las credenciales de seguridad que se utilizan para conectarse.

AWS PCSTLSse utiliza para conectarse a otros AWS servicios.

Dentro de un clúster de Slurm, el programador se configura con el complemento de autenticación que proporciona la `auth/slurm` autenticación para todas las comunicaciones del programador. Slurm no proporciona cifrado a nivel de aplicación para sus comunicaciones, ya que todos los datos que fluyen entre las instancias del clúster permanecen locales EC2 VPC y, por lo tanto, están sujetos a VPC cifrado si esas instancias admiten el cifrado en tránsito. Para obtener más información, consulte Cifrado en tránsito en la Guía del usuario de Amazon Elastic Compute Cloud. La comunicación se cifra entre el controlador (aprovisionado en una cuenta de servicio) y los nodos del clúster de su cuenta.

Administración de claves

AWS PCSutiliza una KMSclave AWS propia para cifrar los datos. Para obtener más información, consulte las claves y AWS claves del cliente en la Guía para AWS KMS desarrolladores. El secreto del clúster se almacena en la KMS clave gestionada por Secrets Manager AWS Secrets Manager y se cifra con ella. Para obtener más información, consulte Trabajar con secretos de clústeres en AWS PCS.

Privacidad del tráfico entre redes

AWS PCSLos recursos de cómputo de un clúster se encuentran dentro de 1 VPC en la cuenta del cliente. Por lo tanto, todo el tráfico de AWS PCS servicio interno de un clúster permanece dentro de la AWS red y no viaja a través de Internet. La comunicación entre el usuario y AWS PCS los nodos puede viajar a través de Internet y recomendamos usar SSH nuestro Systems Manager para conectarse a los nodos. Para obtener más información, consulte ¿Qué es AWS Systems Manager? en la Guía AWS Systems Manager del usuario.

También puede utilizar las siguientes ofertas para conectar su red local a AWS:

- AWS Site-to-Site VPN. Para obtener más información, consulte ¿Qué es AWS Site-to-Site VPN? en la Guía AWS Site-to-Site VPN del usuario.
- Un AWS Direct Connect. Para obtener más información, consulte ¿Qué es AWS Direct Connect? en la Guía AWS Direct Connect del usuario.

Puede acceder al AWS PCS API para realizar tareas administrativas del servicio. Usted y sus usuarios acceden a los puertos de punto final de Slurm para interactuar directamente con el programador.

Cifrar el tráfico API

Para acceder a AWS PCSAPI, los clientes deben ser compatibles con Transport Layer Security (TLS) 1.2 o una versión posterior. Necesitamos la versión TLS 1.2 y recomendamos la TLS versión 1.3. Los clientes también deben admitir conjuntos de cifrado con Perfect Forward Secrecy (PFS), como Ephemeral Diffie-Hellman () o Elliptic Curve Diffie-Hellman Ephemeral (DHE). ECDHE La mayoría de los sistemas modernos como Java 7 y posteriores son compatibles con estos modos. Además, las solicitudes deben firmarse con un identificador de clave de acceso y una clave de acceso secreta que estén asociadas a un director. IAM También puedes usar AWS Security Token Service (AWS STS) para generar credenciales de seguridad temporales para firmar las solicitudes.

Cifrado del tráfico de datos

El cifrado de los datos en tránsito se habilita desde EC2 las instancias compatibles que acceden al punto final del programador y entre ComputeNodeGroup instancias desde dentro del Nube de AWS. Para obtener más información, consulte Cifrado en tránsito.

Acceda al servicio de computación AWS paralela mediante un punto final de interfaz ()AWS PrivateLink

Puede utilizarla AWS PrivateLink para crear una conexión privada entre usted VPC y AWS Parallel Computing Service (AWS PCS). Puede acceder AWS PCS como si estuviera en su casaVPC, sin el uso de una puerta de enlace, NAT dispositivo, VPN conexión o AWS Direct Connect conexión a Internet. Las instancias VPC que tengas no necesitan direcciones IP públicas para acceder AWS PCS.

Esta conexión privada se establece mediante la creación de un punto de conexión de interfaz alimentado por AWS PrivateLink. Creamos una interfaz de red de punto de conexión en cada subred habilitada para el punto de conexión de interfaz. Se trata de interfaces de red administradas por el solicitante que sirven como punto de entrada para el tráfico destinado. AWS PCS

Para obtener más información, consulte Acceso directo AWS PrivateLink en Servicios de AWS la Guía.AWS PrivateLink

Consideraciones sobre AWS PCS

Antes de configurar un punto final de interfaz para AWS PCS, consulte Acceder a un AWS servicio mediante un VPC punto final de interfaz en la AWS PrivateLink Guía.

AWS PCSpermite realizar llamadas a todas sus API acciones a través del punto final de la interfaz.

Si VPC no tiene acceso directo a Internet, debe configurar un VPC punto final para permitir que las instancias del grupo de nodos de cómputo convoquen la AWS PCS `RegisterComputeNodeGroupInstance`APIacción.

Cree un punto final de interfaz para AWS PCS

Puede crear un punto final de interfaz para AWS PCS usar la VPC consola de Amazon o AWS Command Line Interface (AWS CLI). Para obtener más información, consulte Creación de un punto de conexión de interfaz en la Guía de AWS PrivateLink .

Cree un punto final de interfaz para AWS PCS usar el siguiente nombre de servicio:

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

Reemplazar *`region`* con el ID del en el Región de AWS que se va a crear el punto final, por ejemplo`us-east-1`.

Si habilita la opción privada DNS para el punto final de la interfaz, puede realizar API solicitudes AWS PCS utilizando su DNS nombre regional predeterminado. Por ejemplo, `pcs.us-east-1.amazonaws.com`.

Creación de una política de puntos de conexión para el punto de conexión de interfaz

Una política de punto final es un IAM recurso que se puede adjuntar a un punto final de interfaz. La política de puntos finales predeterminada permite el acceso total a AWS PCS través del punto final de la interfaz. Para controlar el acceso permitido AWS PCS desde su punto de conexiónVPC, adjunte una política de punto final personalizada al punto final de la interfaz.

Una política de punto de conexión especifica la siguiente información:

- Los principales que pueden realizar acciones (Cuentas de AWS IAMusuarios y IAM roles).
- Las acciones que se pueden realizar.
- El recurso en el que se pueden realizar las acciones.

Para obtener más información, consulte Control del acceso a los servicios con políticas de punto de conexión en la Guía del usuario de AWS PrivateLink .

Ejemplo: política de VPC puntos finales para las acciones AWS PCS

El siguiente es un ejemplo de una política de un punto de conexión personalizado. Al adjuntar esta política al punto final de la interfaz, se concede acceso a las AWS PCS acciones enumeradas a todos los principales del clúster con las especificadas *`cluster-id`*. Reemplazar *`region`* por el ID Región de AWS del clúster, por ejemplo`us-east-1`. Reemplazar *`account-id`* con el Cuenta de AWS número del clúster.

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

Servicio de Gestión de Identidad y Acceso para Computación AWS Paralela

AWS Identity and Access Management (IAM) es una Servicio de AWS que ayuda al administrador a controlar de forma segura el acceso a AWS los recursos. IAMlos administradores controlan quién puede autenticarse (iniciar sesión) y quién puede autorizarse (tener permisos) para usar AWS PCS los recursos. IAMes una Servicio de AWS que puede utilizar sin coste adicional.

Temas

- Público
- Autenticación con identidades

- Administración de acceso mediante políticas
- Cómo funciona AWS Parallel Computing Service con IAM
- Ejemplos de políticas basadas en la identidad para Parallel Computing Service AWS
- AWS políticas administradas para AWS Parallel Computing Service
- Roles vinculados a servicios de AWS PCS
- Función de Amazon EC2 Spot para AWS PCS
- Permisos mínimos para AWS PCS
- IAMperfiles de instancia para AWS Parallel Computing Service
- Solución de problemas de identidad y acceso a AWS Parallel Computing Service

Público

La forma de usar AWS Identity and Access Management (IAM) varía según el trabajo en el que se realice AWS PCS.

Usuario del servicio: si utiliza el AWS PCS servicio para realizar su trabajo, el administrador le proporcionará las credenciales y los permisos que necesita. A medida que vaya utilizando más AWS PCS funciones para realizar su trabajo, es posible que necesite permisos adicionales. Entender cómo se administra el acceso puede ayudarlo a solicitar los permisos correctos al administrador. Si no puede acceder a una función en AWS PCS, consulteSolución de problemas de identidad y acceso a AWS Parallel Computing Service.

Administrador de servicios: si está a cargo de AWS PCS los recursos de su empresa, probablemente tenga acceso total a ellos AWS PCS. Su trabajo consiste en determinar a qué AWS PCS funciones y recursos deben acceder los usuarios del servicio. A continuación, debe enviar solicitudes a su IAM administrador para cambiar los permisos de los usuarios del servicio. Revise la información de esta página para comprender los conceptos básicos deIAM. Para obtener más información sobre cómo su empresa puede utilizar IAM con AWS PCS, consulteCómo funciona AWS Parallel Computing Service con IAM.

IAMadministrador: si es IAM administrador, puede que desee obtener más información sobre cómo puede redactar políticas para administrar el acceso a ellas AWS PCS. Para ver ejemplos de políticas AWS PCS basadas en la identidad que puede utilizarIAM, consulte. Ejemplos de políticas basadas en la identidad para Parallel Computing Service AWS

Autenticación con identidades

La autenticación es la forma de iniciar sesión AWS con sus credenciales de identidad. Debe estar autenticado (con quien haya iniciado sesión AWS) como IAM usuario o asumiendo un IAM rol. Usuario raíz de la cuenta de AWS

Puede iniciar sesión AWS como una identidad federada mediante las credenciales proporcionadas a través de una fuente de identidad. AWS IAM Identity Center Los usuarios (IAMIdentity Center), la autenticación de inicio de sesión único de su empresa y sus credenciales de Google o Facebook son ejemplos de identidades federadas. Al iniciar sesión como una identidad federada, el administrador configuró previamente la federación de identidades mediante roles. IAM Cuando accede AWS mediante la federación, asume indirectamente un rol.

Según el tipo de usuario que sea, puede iniciar sesión en el portal AWS Management Console o en el de AWS acceso. Para obtener más información sobre cómo iniciar sesión AWS, consulte Cómo iniciar sesión Cuenta de AWS en su Guía del AWS Sign-In usuario.

Si accede AWS mediante programación, AWS incluye un kit de desarrollo de software (SDK) y una interfaz de línea de comandos (CLI) para firmar criptográficamente sus solicitudes con sus credenciales. Si no utilizas AWS herramientas, debes firmar las solicitudes tú mismo. Para obtener más información sobre cómo usar el método recomendado para firmar las solicitudes usted mismo, consulte Firmar AWS API las solicitudes en la Guía del IAM usuario.

Independientemente del método de autenticación que use, es posible que deba proporcionar información de seguridad adicional. Por ejemplo, le AWS recomienda que utilice la autenticación multifactorial (MFA) para aumentar la seguridad de su cuenta. Para obtener más información, consulte Autenticación multifactorial en la Guía del AWS IAM Identity Center usuario y Uso de la autenticación multifactorial (MFA) AWS en la Guía del IAM usuario.

Cuenta de AWS usuario root

Al crear una Cuenta de AWS, comienza con una identidad de inicio de sesión que tiene acceso completo a todos Servicios de AWS los recursos de la cuenta. Esta identidad se denomina usuario Cuenta de AWS raíz y se accede a ella iniciando sesión con la dirección de correo electrónico y la contraseña que utilizaste para crear la cuenta. Recomendamos encarecidamente que no utilice el usuario raíz para sus tareas diarias. Proteja las credenciales del usuario raíz y utilícelas solo para las tareas que solo el usuario raíz pueda realizar. Para ver la lista completa de tareas que requieren que inicie sesión como usuario root, consulte Tareas que requieren credenciales de usuario root en la Guía del IAM usuario.

Identidad federada

Como práctica recomendada, exija a los usuarios humanos, incluidos los que requieren acceso de administrador, que utilicen la federación con un proveedor de identidades para acceder Servicios de AWS mediante credenciales temporales.

Una identidad federada es un usuario del directorio de usuarios de su empresa, un proveedor de identidades web AWS Directory Service, el directorio del Centro de Identidad o cualquier usuario al que acceda Servicios de AWS mediante las credenciales proporcionadas a través de una fuente de identidad. Cuando las identidades federadas acceden Cuentas de AWS, asumen funciones y las funciones proporcionan credenciales temporales.

Para una administración de acceso centralizada, le recomendamos que utilice AWS IAM Identity Center. Puede crear usuarios y grupos en IAM Identity Center, o puede conectarse y sincronizarse con un conjunto de usuarios y grupos de su propia fuente de identidad para usarlos en todas sus aplicaciones Cuentas de AWS . Para obtener información sobre IAM Identity Center, consulte ¿Qué es IAM Identity Center? en la Guía AWS IAM Identity Center del usuario.

Usuarios y grupos de IAM

Un IAMusuario es una identidad propia Cuenta de AWS que tiene permisos específicos para una sola persona o aplicación. Siempre que sea posible, recomendamos utilizar credenciales temporales en lugar de crear IAM usuarios con credenciales de larga duración, como contraseñas y claves de acceso. Sin embargo, si tiene casos de uso específicos que requieren credenciales a largo plazo con IAM los usuarios, le recomendamos que rote las claves de acceso. Para obtener más información, consulte Rotar las claves de acceso con regularidad para los casos de uso que requieran credenciales de larga duración en la Guía del IAM usuario.

Un IAMgrupo es una identidad que especifica un conjunto de IAM usuarios. No puede iniciar sesión como grupo. Puede usar los grupos para especificar permisos para varios usuarios a la vez. Los grupos facilitan la administración de los permisos para grandes conjuntos de usuarios. Por ejemplo, puede asignar un nombre a un grupo IAMAdminsy concederle permisos para administrar IAM los recursos.

Los usuarios son diferentes de los roles. Un usuario se asocia exclusivamente a una persona o aplicación, pero la intención es que cualquier usuario pueda asumir un rol que necesite. Los usuarios tienen credenciales de larga duración permanentes; no obstante, los roles proporcionan credenciales temporales. Para obtener más información, consulte Cuándo crear un IAM usuario (en lugar de un rol) en la Guía del IAM usuario.

IAMroles

Un IAMrol es una identidad dentro de ti Cuenta de AWS que tiene permisos específicos. Es similar a un IAM usuario, pero no está asociado a una persona específica. Puede asumir temporalmente un IAM rol en el AWS Management Console cambiando de rol. Puede asumir un rol llamando a una AWS API operación AWS CLI o o utilizando una operación personalizadaURL. Para obtener más información sobre los métodos de uso de roles, consulte Uso de IAM roles en la Guía del IAM usuario.

IAMlos roles con credenciales temporales son útiles en las siguientes situaciones:

- Acceso de usuario federado: para asignar permisos a una identidad federada, puede crear un rol y definir sus permisos. Cuando se autentica una identidad federada, se asocia la identidad al rol y se le conceden los permisos define el rol. Para obtener información sobre los roles para la federación, consulte Creación de un rol para un proveedor de identidad externo en la Guía del IAM usuario. Si usa IAM Identity Center, configura un conjunto de permisos. Para controlar a qué pueden acceder sus identidades después de autenticarse, IAM Identity Center correlaciona el conjunto de permisos con un rol en. IAM Para obtener información acerca de los conjuntos de permisos, consulte Conjuntos de permisos en la Guía del usuario de AWS IAM Identity Center .
- Permisos IAM de usuario temporales: un IAM usuario o rol puede asumir un IAM rol para asumir temporalmente diferentes permisos para una tarea específica.
- Acceso multicuenta: puedes usar un IAM rol para permitir que alguien (un responsable de confianza) de una cuenta diferente acceda a los recursos de tu cuenta. Los roles son la forma principal de conceder acceso entre cuentas. Sin embargo, con algunos Servicios de AWS, puedes adjuntar una política directamente a un recurso (en lugar de usar un rol como proxy). Para conocer la diferencia entre las funciones y las políticas basadas en recursos para el acceso multicuenta, consulta el tema sobre el acceso a los recursos entre cuentas IAM en la Guía del IAM usuario.
- Acceso entre servicios: algunos Servicios de AWS utilizan funciones en otros. Servicios de AWS Por ejemplo, cuando realizas una llamada en un servicio, es habitual que ese servicio ejecute aplicaciones en Amazon EC2 o almacene objetos en Amazon S3. Es posible que un servicio haga esto usando los permisos de la entidad principal, usando un rol de servicio o usando un rol vinculado al servicio.
  - Sesiones de acceso directo (FAS): cuando utilizas un IAM usuario o un rol para realizar acciones en AWS ellas, se te considera director. Cuando utiliza algunos servicios, es posible que realice una acción que desencadene otra acción en un servicio diferente. FASutiliza los permisos del principal que llama a an Servicio de AWS, junto con los que solicitan, Servicio de AWS para realizar solicitudes a los servicios descendentes. FASlas solicitudes solo se realizan cuando un

servicio recibe una solicitud que requiere interacciones con otros recursos Servicios de AWS o para completarse. En este caso, debe tener permisos para realizar ambas acciones. Para obtener información detallada sobre la política a la hora de realizar FAS solicitudes, consulte Reenviar las sesiones de acceso.

- Función de servicio: una función de servicio es una IAMfunción que un servicio asume para realizar acciones en su nombre. Un IAM administrador puede crear, modificar y eliminar un rol de servicio desde dentroIAM. Para obtener más información, consulte Crear un rol para delegar permisos Servicio de AWS en un rol en el IAMManual del usuario.
- Función vinculada a un servicio: una función vinculada a un servicio es un tipo de función de servicio que está vinculada a un. Servicio de AWS El servicio puede asumir el rol para realizar una acción en su nombre. Los roles vinculados al servicio aparecen en usted Cuenta de AWS y son propiedad del servicio. Un IAM administrador puede ver los permisos de los roles vinculados al servicio, pero no editarlos.

- Aplicaciones que se ejecutan en Amazon EC2: puedes usar un IAM rol para administrar las credenciales temporales de las aplicaciones que se ejecutan en una EC2 instancia y que realizan AWS CLI o AWS API solicitan. Esto es preferible a almacenar las claves de acceso dentro de la EC2 instancia. Para asignar un AWS rol a una EC2 instancia y ponerlo a disposición de todas sus aplicaciones, debe crear un perfil de instancia adjunto a la instancia. Un perfil de instancia contiene el rol y permite que los programas que se ejecutan en la EC2 instancia obtengan credenciales temporales. Para obtener más información, consulte Uso de un IAM rol para conceder permisos a aplicaciones que se ejecutan en EC2 instancias de Amazon en la Guía del IAM usuario.

Para saber si se deben usar IAM roles o IAM usuarios, consulte Cuándo crear un IAM rol (en lugar de un usuario) en la Guía del IAM usuario.

Administración de acceso mediante políticas

El acceso se controla AWS creando políticas y adjuntándolas a AWS identidades o recursos. Una política es un objeto AWS que, cuando se asocia a una identidad o un recurso, define sus permisos. AWS evalúa estas políticas cuando un director (usuario, usuario raíz o sesión de rol) realiza una solicitud. Los permisos en las políticas determinan si la solicitud se permite o se deniega. La mayoría de las políticas se almacenan AWS como JSON documentos. Para obtener más información sobre la estructura y el contenido de los documentos de JSON políticas, consulte Descripción general de JSON las políticas en la Guía del IAM usuario.

Los administradores pueden usar AWS JSON las políticas para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y en qué condiciones.

De forma predeterminada, los usuarios y los roles no tienen permisos. Para conceder a los usuarios permiso para realizar acciones en los recursos que necesitan, un IAM administrador puede crear IAM políticas. A continuación, el administrador puede añadir las IAM políticas a las funciones y los usuarios pueden asumir las funciones.

IAMlas políticas definen los permisos para una acción independientemente del método que se utilice para realizar la operación. Por ejemplo, suponga que dispone de una política que permite la acción `iam:GetRole`. Un usuario con esa política puede obtener información sobre el rol de AWS Management Console AWS CLI, el o el AWS API.

Políticas basadas en identidad

Las políticas basadas en la identidad son documentos de política de JSON permisos que se pueden adjuntar a una identidad, como un IAM usuario, un grupo de usuarios o un rol. Estas políticas controlan qué acciones pueden realizar los usuarios y los roles, en qué recursos y en qué condiciones. Para obtener información sobre cómo crear una política basada en la identidad, consulte Creación de IAM políticas en la Guía del usuario. IAM

Las políticas basadas en identidades pueden clasificarse además como políticas insertadas o políticas administradas. Las políticas insertadas se integran directamente en un único usuario, grupo o rol. Las políticas administradas son políticas independientes que puede adjuntar a varios usuarios, grupos y funciones de su empresa. Cuenta de AWS Las políticas administradas incluyen políticas AWS administradas y políticas administradas por el cliente. Para saber cómo elegir entre una política gestionada o una política integrada, consulte Elegir entre políticas gestionadas y políticas integradas en la Guía del IAM usuario.

Políticas basadas en recursos

Las políticas basadas en recursos son documentos de JSON política que se adjuntan a un recurso. Algunos ejemplos de políticas basadas en recursos son las políticas de confianza de IAM roles y las políticas de bucket de Amazon S3. En los servicios que admiten políticas basadas en recursos, los administradores de servicios pueden utilizarlos para controlar el acceso a un recurso específico. Para el recurso al que se asocia la política, la política define qué acciones puede realizar una entidad principal especificada en ese recurso y en qué condiciones. Debe especificar una entidad principal en una política en función de recursos. Los principales pueden incluir cuentas, usuarios, roles, usuarios federados o. Servicios de AWS

Las políticas basadas en recursos son políticas insertadas que se encuentran en ese servicio. No puede usar políticas AWS administradas desde una política IAM basada en recursos.

Listas de control de acceso () ACLs

Las listas de control de acceso (ACLs) controlan qué responsables (miembros de la cuenta, usuarios o roles) tienen permisos para acceder a un recurso. ACLsson similares a las políticas basadas en recursos, aunque no utilizan el formato de documento de JSON políticas.

Amazon S3 AWS WAF y Amazon VPC son ejemplos de servicios compatiblesACLs. Para obtener más informaciónACLs, consulte la descripción general de la lista de control de acceso (ACL) en la Guía para desarrolladores de Amazon Simple Storage Service.

Otros tipos de políticas

AWS admite tipos de políticas adicionales y menos comunes. Estos tipos de políticas pueden establecer el máximo de permisos que los tipos de políticas más frecuentes le conceden.

- Límites de permisos: un límite de permisos es una función avanzada en la que se establecen los permisos máximos que una política basada en la identidad puede conceder a una IAM entidad (IAMusuario o rol). Puede establecer un límite de permisos para una entidad. Los permisos resultantes son la intersección de las políticas basadas en la identidad de la entidad y los límites de permisos. Las políticas basadas en recursos que especifiquen el usuario o rol en el campo `Principal` no estarán restringidas por el límite de permisos. Una denegación explícita en cualquiera de estas políticas anulará el permiso. Para obtener más información sobre los límites de los permisos, consulte los límites de los permisos para IAM las entidades en la Guía del IAMusuario.
- Políticas de control de servicios (SCPs): SCPs son JSON políticas que especifican los permisos máximos para una organización o unidad organizativa (OU) en AWS Organizations. AWS Organizations es un servicio para agrupar y administrar de forma centralizada varios de los Cuentas de AWS que son propiedad de su empresa. Si habilitas todas las funciones de una organización, puedes aplicar políticas de control de servicios (SCPs) a una o a todas tus cuentas. SCPLimita los permisos de las entidades en las cuentas de los miembros, incluidas las de cada una Usuario raíz de la cuenta de AWS. Para obtener más información sobre OrganizationsSCPs, consulte las políticas de control de servicios en la Guía del AWS Organizations usuario.
- Políticas de sesión: las políticas de sesión son políticas avanzadas que se pasan como parámetro cuando se crea una sesión temporal mediante programación para un rol o un usuario federado. Los permisos de la sesión resultantes son la intersección de las políticas basadas en identidades del rol y las políticas de la sesión. Los permisos también pueden proceder de una política en función de recursos. Una denegación explícita en cualquiera de estas políticas anulará el permiso. Para obtener más información, consulte las políticas de sesión en la Guía del IAM usuario.

Varios tipos de políticas

Cuando se aplican varios tipos de políticas a una solicitud, los permisos resultantes son más complicados de entender. Para saber cómo se AWS determina si se debe permitir una solicitud cuando se trata de varios tipos de políticas, consulte la lógica de evaluación de políticas en la Guía del IAM usuario.

Cómo funciona AWS Parallel Computing Service con IAM

Antes de administrar el IAM acceso a AWS PCS, infórmese sobre IAM las funciones disponibles para su uso AWS PCS.

IAMfunciones que puede utilizar con AWS Parallel Computing Service

IAMcaracterística	AWS PCSapoyo
Políticas basadas en identidades	Sí
Políticas basadas en recursos	No
Acciones de políticas	Sí
Recursos de políticas	Sí
Claves de condición de política (específicas del servicio)	Sí
ACLs	No
ABAC(etiquetas en las políticas)	Sí
Credenciales temporales	Sí
Permisos de entidades principales	Sí
Roles de servicio	No
Roles vinculados al servicio	Sí

Para obtener una visión general de cómo AWS PCS funcionan otros AWS servicios con la mayoría de las IAM funciones, consulte AWS los servicios con los que funcionan IAM en la Guía del IAM usuario.

Políticas basadas en la identidad para AWS PCS

Compatibilidad con las políticas basadas en identidad: sí

Las políticas basadas en la identidad son documentos de política de JSON permisos que se pueden adjuntar a una identidad, como un IAM usuario, un grupo de usuarios o un rol. Estas políticas controlan qué acciones pueden realizar los usuarios y los roles, en qué recursos y en qué condiciones. Para obtener información sobre cómo crear una política basada en la identidad, consulte Creación de IAM políticas en la Guía del usuario. IAM

Con las políticas IAM basadas en la identidad, puede especificar las acciones y los recursos permitidos o denegados, así como las condiciones en las que se permiten o deniegan las acciones. No es posible especificar la entidad principal en una política basada en identidad porque se aplica al usuario o rol al que está adjunto. Para obtener más información sobre todos los elementos que puede utilizar en una JSON política, consulte la referencia sobre los elementos de la IAM JSON política en la Guía del IAMusuario.

Ejemplos de políticas basadas en la identidad para AWS PCS

Para ver ejemplos de políticas AWS PCS basadas en la identidad, consulte. Ejemplos de políticas basadas en la identidad para Parallel Computing Service AWS

Políticas basadas en recursos dentro de AWS PCS

Admite políticas basadas en recursos: no

Las políticas basadas en recursos son documentos de JSON política que se adjuntan a un recurso. Algunos ejemplos de políticas basadas en recursos son las políticas de confianza de IAM roles y las políticas de bucket de Amazon S3. En los servicios que admiten políticas basadas en recursos, los administradores de servicios pueden utilizarlos para controlar el acceso a un recurso específico. Para el recurso al que se asocia la política, la política define qué acciones puede realizar una entidad principal especificada en ese recurso y en qué condiciones. Debe especificar una entidad principal en una política en función de recursos. Los principales pueden incluir cuentas, usuarios, roles, usuarios federados o. Servicios de AWS

Para habilitar el acceso entre cuentas, puede especificar una cuenta completa o IAM entidades de otra cuenta como principales en una política basada en recursos. Añadir a una política en función de recursos una entidad principal entre cuentas es solo una parte del establecimiento de una relación de confianza. Cuando el principal y el recurso son diferentes Cuentas de AWS, el IAM administrador de la cuenta de confianza también debe conceder permiso a la entidad principal (usuario o rol) para acceder al recurso. Para conceder el permiso, adjunte la entidad a una política basada en identidad. Sin embargo, si la política en función de recursos concede el acceso a una entidad principal de la misma cuenta, no es necesaria una política basada en identidad adicional. Para obtener más información, consulte el tema Acceso a recursos entre cuentas IAM en la Guía del IAM usuario.

Acciones políticas para AWS PCS

Compatibilidad con las acciones de política: sí

Los administradores pueden usar AWS JSON políticas para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y en qué condiciones.

El `Action` elemento de una JSON política describe las acciones que puede utilizar para permitir o denegar el acceso en una política. Las acciones de política suelen tener el mismo nombre que la AWS API operación asociada. Hay algunas excepciones, como las acciones que solo permiten permisos y que no tienen una operación coincidente. API También hay algunas operaciones que requieren varias acciones en una política. Estas acciones adicionales se denominan acciones dependientes.

Incluya acciones en una política para conceder permisos y así llevar a cabo la operación asociada.

Para ver una lista de AWS PCS acciones, consulte las acciones definidas por AWS Parallel Computing Service en la Referencia de autorización del servicio.

Las acciones políticas AWS PCS utilizan el siguiente prefijo antes de la acción:

```
pcs
```

Para especificar varias acciones en una única instrucción, sepárelas con comas.

```
"Action": [
      "pcs:action1",
      "pcs:action2"
         ]
```

Para ver ejemplos de políticas AWS PCS basadas en la identidad, consulte. Ejemplos de políticas basadas en la identidad para Parallel Computing Service AWS

Recursos de políticas para AWS PCS

Compatibilidad con los recursos de políticas: sí

Los administradores pueden usar AWS JSON políticas para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y en qué condiciones.

El elemento `Resource` JSON de política especifica el objeto o los objetos a los que se aplica la acción. Las instrucciones deben contener un elemento `Resource` o `NotResource`. Como práctica recomendada, especifique un recurso mediante su nombre de recurso de Amazon (ARN). Puede hacerlo para acciones que admitan un tipo de recurso específico, conocido como permisos de nivel de recurso.

Para las acciones que no admiten permisos de nivel de recurso, como las operaciones de descripción, utilice un carácter comodín (*) para indicar que la instrucción se aplica a todos los recursos.

```
"Resource": "*"
```

Para ver una lista de los tipos de AWS PCS recursos y sus respectivos tiposARNs, consulte Recursos definidos por AWS Parallel Computing Service en la Referencia de autorización de servicios. Para saber con qué acciones puede especificar cada recurso, consulte Acciones definidas por AWS Parallel Computing Service. ARN

Para ver ejemplos de políticas AWS PCS basadas en la identidad, consulte. Ejemplos de políticas basadas en la identidad para Parallel Computing Service AWS

Claves de condición de la política para AWS PCS

Compatibilidad con claves de condición de políticas específicas del servicio: sí

Los administradores pueden usar AWS JSON políticas para especificar quién tiene acceso a qué. Es decir, qué entidad principal puede realizar acciones en qué recursos y en qué condiciones.

El elemento `Condition` (o bloque de `Condition`) permite especificar condiciones en las que entra en vigor una instrucción. El elemento `Condition` es opcional. Puede crear expresiones condicionales que utilicen operadores de condición, tales como igual o menor que, para que la condición de la política coincida con los valores de la solicitud.

Si especifica varios elementos de `Condition` en una instrucción o varias claves en un único elemento de `Condition`, AWS las evalúa mediante una operación `AND` lógica. Si especifica varios valores para una única clave de condición, AWS evalúa la condición mediante una `OR` operación lógica. Se deben cumplir todas las condiciones antes de que se concedan los permisos de la instrucción.

También puede utilizar variables de marcador de posición al especificar condiciones. Por ejemplo, puede conceder a un IAM usuario permiso para acceder a un recurso solo si está etiquetado con su nombre de IAM usuario. Para obtener más información, consulte los elementos IAM de la política: variables y etiquetas en la Guía del IAM usuario.

AWS admite claves de condición globales y claves de condición específicas del servicio. Para ver todas las claves de condición AWS globales, consulte las claves de contexto de condición AWS globales en la Guía del IAMusuario.

Para ver una lista de claves de AWS PCS condición, consulte las claves de condición del servicio de computación AWS paralela en la Referencia de autorización de servicios. Para saber con qué acciones y recursos puede utilizar una clave de condición, consulte Acciones definidas por AWS Parallel Computing Service.

Para ver ejemplos de políticas AWS PCS basadas en la identidad, consulte. Ejemplos de políticas basadas en la identidad para Parallel Computing Service AWS

ACLsen AWS PCS

SoportesACLs: No

Las listas de control de acceso (ACLs) controlan qué directores (miembros de la cuenta, usuarios o roles) tienen permisos para acceder a un recurso. ACLsson similares a las políticas basadas en recursos, aunque no utilizan el formato de documento de JSON políticas.

ABACcon AWS PCS

Soportes ABAC (etiquetas en las políticas): Sí

El control de acceso basado en atributos (ABAC) es una estrategia de autorización que define los permisos en función de los atributos. En AWS, estos atributos se denominan etiquetas. Puede adjuntar etiquetas a IAM entidades (usuarios o roles) y a muchos AWS recursos. Etiquetar entidades y recursos es el primer paso deABAC. Luego, diseñe ABAC políticas para permitir las operaciones cuando la etiqueta del principal coincida con la etiqueta del recurso al que está intentando acceder.

ABACes útil en entornos de rápido crecimiento y ayuda en situaciones en las que la administración de políticas se vuelve engorrosa.

Para controlar el acceso en función de etiquetas, debe proporcionar información de las etiquetas en el elemento de condición de una política utilizando las claves de condición `aws:ResourceTag/key-name`, `aws:RequestTag/key-name` o `aws:TagKeys`.

Si un servicio admite las tres claves de condición para cada tipo de recurso, el valor es Sí para el servicio. Si un servicio admite las tres claves de condición solo para algunos tipos de recursos, el valor es Parcial.

Para obtener más información al respectoABAC, consulte ¿Qué es? ABAC en la Guía IAM del usuario. Para ver un tutorial con los pasos de configuraciónABAC, consulte Usar el control de acceso basado en atributos (ABAC) en la Guía del IAMusuario.

Usar credenciales temporales con AWS PCS

Compatibilidad con credenciales temporales: sí

Algunas Servicios de AWS no funcionan cuando inicias sesión con credenciales temporales. Para obtener información adicional, incluida la información sobre cuáles Servicios de AWS funcionan con credenciales temporales, consulta la sección Servicios de AWS Cómo trabajar con credenciales temporales IAM en la Guía del IAM usuario.

Está utilizando credenciales temporales si inicia sesión en ellas AWS Management Console mediante cualquier método excepto un nombre de usuario y una contraseña. Por ejemplo, cuando accedes AWS mediante el enlace de inicio de sesión único (SSO) de tu empresa, ese proceso crea automáticamente credenciales temporales. También crea credenciales temporales de forma automática cuando inicia sesión en la consola como usuario y luego cambia de rol. Para obtener más información sobre el cambio de rol, consulte Cambiar a un rol (consola) en la Guía del IAMusuario.

Puede crear credenciales temporales manualmente con la tecla AWS CLI o AWS API. A continuación, puede utilizar esas credenciales temporales para acceder AWS. AWS recomienda

generar credenciales temporales de forma dinámica en lugar de utilizar claves de acceso a largo plazo. Para obtener más información, consulte Credenciales de seguridad temporales en IAM.

Permisos principales entre servicios para AWS PCS

Admite sesiones de acceso directo (FAS): Sí

Cuando utilizas un IAM usuario o un rol para realizar acciones en AWSél, se te considera director. Cuando utiliza algunos servicios, es posible que realice una acción que desencadene otra acción en un servicio diferente. FASutiliza los permisos del principal que llama a una Servicio de AWS, junto con los que solicita, Servicio de AWS para realizar solicitudes a los servicios descendentes. FASlas solicitudes solo se realizan cuando un servicio recibe una solicitud que requiere interacciones con otros recursos Servicios de AWS o para completarse. En este caso, debe tener permisos para realizar ambas acciones. Para obtener detalles sobre la política a la hora de realizar FAS solicitudes, consulte Reenviar sesiones de acceso.

Roles de servicio para AWS PCS

Compatible con roles de servicio: No

Una función de servicio es una IAMfunción que un servicio asume para realizar acciones en su nombre. Un IAM administrador puede crear, modificar y eliminar un rol de servicio desde dentroIAM. Para obtener más información, consulte Crear un rol para delegar permisos Servicio de AWS en un rol en el IAMManual del usuario.

> **Warning**
>
> Cambiar los permisos de un rol de servicio podría interrumpir AWS PCS la funcionalidad. Edite las funciones de servicio solo cuando se AWS PCS proporcionen instrucciones para hacerlo.

Funciones vinculadas al servicio para AWS PCS

Admite roles vinculados al servicio: sí

Un rol vinculado a un servicio es un tipo de rol de servicio que está vinculado a un. Servicio de AWS El servicio puede asumir el rol para realizar una acción en su nombre. Los roles vinculados al servicio aparecen en usted Cuenta de AWS y son propiedad del servicio. Un IAM administrador puede ver los permisos de los roles vinculados al servicio, pero no editarlos.

Para obtener más información sobre la creación o la administración de funciones vinculadas a un servicio, consulte los AWS servicios que funcionan con. IAM Busque un servicio en la tabla que incluya `Yes` en la columna Rol vinculado a un servicio. Seleccione el vínculo Sí para ver la documentación acerca del rol vinculado a servicios para ese servicio.

Ejemplos de políticas basadas en la identidad para Parallel Computing Service AWS

De forma predeterminada, los usuarios y los roles no tienen permiso para crear o modificar AWS PCS recursos. Tampoco pueden realizar tareas con AWS Management Console, AWS Command Line Interface (AWS CLI) o AWS API. Para conceder a los usuarios permiso para realizar acciones en los recursos que necesitan, un IAM administrador puede crear IAM políticas. A continuación, el administrador puede añadir las IAM políticas a las funciones y los usuarios pueden asumir las funciones.

Para obtener información sobre cómo crear una política IAM basada en la identidad mediante estos documentos de JSON política de ejemplo, consulte Creación de IAM políticas en la Guía del IAMusuario.

Para obtener más información sobre las acciones y los tipos de recursos definidos por cada uno de ellos AWS PCS, incluido el ARNs formato de cada uno de ellos, consulte Acciones, recursos y claves de condición del servicio de computación AWS paralela en la referencia de autorización de servicios.

Temas

- Prácticas recomendadas sobre las políticas
- Uso de la AWS PCS consola
- Cómo permitir a los usuarios consultar sus propios permisos

Prácticas recomendadas sobre las políticas

Las políticas basadas en la identidad determinan si alguien puede crear AWS PCS recursos de su cuenta, acceder a ellos o eliminarlos. Estas acciones pueden generar costos adicionales para su Cuenta de AWS. Siga estas directrices y recomendaciones al crear o editar políticas basadas en identidades:

- Comience con las políticas AWS administradas y avance hacia los permisos con privilegios mínimos: para empezar a conceder permisos a sus usuarios y cargas de trabajo, utilice las

políticas AWS administradas que otorgan permisos para muchos casos de uso comunes. Están disponibles en su. Cuenta de AWS Le recomendamos que reduzca aún más los permisos definiendo políticas administradas por el AWS cliente que sean específicas para sus casos de uso. Para obtener más información, consulte las políticas AWS gestionadas o las políticas AWS gestionadas para las funciones laborales en la Guía del IAM usuario.

- Aplique permisos con privilegios mínimos: cuando establezca permisos con IAM políticas, conceda solo los permisos necesarios para realizar una tarea. Para ello, debe definir las acciones que se pueden llevar a cabo en determinados recursos en condiciones específicas, también conocidos como permisos de privilegios mínimos. Para obtener más información sobre cómo IAM aplicar permisos, consulte Políticas y permisos IAM en la IAM Guía del usuario.
- Utilice las condiciones en IAM las políticas para restringir aún más el acceso: puede añadir una condición a sus políticas para limitar el acceso a las acciones y los recursos. Por ejemplo, puede escribir una condición de política para especificar que todas las solicitudes deben enviarse medianteSSL. También puedes usar condiciones para conceder el acceso a las acciones del servicio si se utilizan a través de una acción específica Servicio de AWS, por ejemplo AWS CloudFormation. Para obtener más información, consulte los elementos IAM JSON de la política: Condición en la Guía del IAM usuario.
- Utilice IAM Access Analyzer para validar sus IAM políticas y garantizar permisos seguros y funcionales: IAM Access Analyzer valida las políticas nuevas y existentes para que se ajusten al lenguaje de las políticas (JSON) y IAM a las IAM mejores prácticas. IAMAccess Analyzer proporciona más de 100 comprobaciones de políticas y recomendaciones prácticas para ayudarle a crear políticas seguras y funcionales. Para obtener más información, consulte la validación de políticas de IAM Access Analyzer en la Guía del IAM usuario.
- Requerir autenticación multifactorial (MFA): si se encuentra en una situación en la que se requieren IAM usuarios o un usuario raíz Cuenta de AWS, actívela MFA para aumentar la seguridad. Para solicitarlo MFA cuando se convoque a API las operaciones, añada MFA condiciones a sus políticas. Para obtener más información, consulte Configuración del API acceso MFA protegido en la Guía del IAM usuario.

Para obtener más información sobre las prácticas recomendadasIAM, consulte las prácticas recomendadas de seguridad IAM en la Guía del IAM usuario.

Uso de la AWS PCS consola

Para acceder a la consola del Servicio de Computación AWS Paralela, debe tener un conjunto mínimo de permisos. Estos permisos deben permitirle enumerar y ver detalles sobre los AWS PCS

recursos de su cuenta Cuenta de AWS. Si crea una política basada en identidades que sea más restrictiva que el mínimo de permisos necesarios, la consola no funcionará del modo esperado para las entidades (usuarios o roles) que tengan esa política.

No es necesario que concedas permisos mínimos de consola a los usuarios que realicen llamadas únicamente al AWS CLI o al AWS API. En su lugar, permita el acceso únicamente a las acciones que coincidan con la API operación que están intentando realizar.

Para obtener más información sobre los permisos mínimos necesarios para usar la AWS PCS consola, consultePermisos mínimos para AWS PCS.

Cómo permitir a los usuarios consultar sus propios permisos

En este ejemplo se muestra cómo se puede crear una política que permita a IAM los usuarios ver las políticas integradas y administradas asociadas a su identidad de usuario. Esta política incluye permisos para completar esta acción en la consola o mediante programación mediante la tecla o. AWS CLI AWS API

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ViewOwnUserInfo",
            "Effect": "Allow",
            "Action": [
                "iam:GetUserPolicy",
                "iam:ListGroupsForUser",
                "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 políticas administradas para AWS Parallel Computing Service

Una política AWS administrada es una política independiente creada y administrada por AWS. AWS Las políticas administradas están diseñadas para proporcionar permisos para muchos casos de uso comunes, de modo que pueda empezar a asignar permisos a usuarios, grupos y funciones.

Ten en cuenta que es posible que las políticas AWS administradas no otorguen permisos con privilegios mínimos para tus casos de uso específicos, ya que están disponibles para que los usen todos los AWS clientes. Se recomienda definir políticas administradas por el cliente específicas para sus casos de uso a fin de reducir aún más los permisos.

No puedes cambiar los permisos definidos en AWS las políticas administradas. Si AWS actualiza los permisos definidos en una política AWS administrada, la actualización afecta a todas las identidades principales (usuarios, grupos y roles) a las que está asociada la política. AWS es más probable que actualice una política AWS administrada cuando Servicio de AWS se lance una nueva o cuando haya nuevas API operaciones disponibles para los servicios existentes.

Para obtener más información, consulte las políticas AWS administradas en la Guía del IAM usuario.

AWS política gestionada: AWSPCSServiceRolePolicy

No puede adjuntarse AWSPCSServiceRolePolicy a sus IAM entidades. Esta política está asociada a un rol vinculado al servicio que te permite AWS PCS realizar acciones en tu nombre. Para obtener más información, consulte Roles vinculados a servicios de AWS PCS.

Detalles de los permisos

Esta política incluye los siguientes permisos.

- `ec2`— Permite AWS PCS crear y gestionar EC2 los recursos de Amazon.
- `iam`— Permite AWS PCS crear un rol vinculado al servicio para la EC2 flota de Amazon y pasarlo a Amazon. EC2
- `cloudwatch`— Permite AWS PCS publicar métricas de servicio en Amazon CloudWatch.
- `secretsmanager`— Permite AWS PCS gestionar los secretos de los recursos AWS PCS del clúster.

```
{
    "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": "PermissionsToTerminatePCSManagedInstances",
            "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 PCSactualizaciones de las políticas AWS gestionadas

Consulte los detalles sobre las actualizaciones de las políticas AWS administradas AWS PCS desde que este servicio comenzó a realizar el seguimiento de estos cambios. Para recibir alertas

automáticas sobre los cambios en esta página, suscríbase al RSS feed de la página del historial del AWS PCS documento.

Cambio	Descripción	Fecha
AWS PCS comenzó a hacer el seguimiento de los cambios	AWS PCScomenzó a realizar un seguimiento de los cambios de sus políticas AWS gestionadas.	28 de agosto de 2024

Roles vinculados a servicios de AWS PCS

AWS Parallel Computing Service usa AWS Identity and Access Management (IAM) roles vinculados al servicio. Un rol vinculado a un servicio es un tipo único de IAM rol al que se vincula directamente. AWS PCS Los roles vinculados al servicio están predefinidos AWS PCS e incluyen todos los permisos que el servicio requiere para llamar a otros AWS servicios en tu nombre.

Un rol vinculado a un servicio facilita la configuración AWS PCS, ya que no es necesario añadir manualmente los permisos necesarios. AWS PCSdefine los permisos de sus funciones vinculadas al servicio y, a menos que se defina lo contrario, solo AWS PCS puede asumir sus funciones. Los permisos definidos incluyen la política de confianza y la política de permisos, y esa política de permisos no se puede adjuntar a ninguna otra IAM entidad.

Solo puede eliminar un rol vinculado a servicios después de eliminar sus recursos relacionados. Esto protege tus AWS PCS recursos porque no puedes eliminar accidentalmente el permiso de acceso a los recursos.

Para obtener información sobre otros servicios que admiten funciones vinculadas a servicios, consulta AWS los servicios que funcionan con ellas IAM y busca los servicios con la palabra Sí en la columna Funciones vinculadas a servicios. Elija una opción Sí con un enlace para ver la documentación acerca del rol vinculado a servicios en cuestión.

Permisos de roles vinculados al servicio para AWS PCS

AWS PCSutiliza el rol vinculado al servicio denominado AWSServiceRoleForPCS— Permitir gestionar AWS PCS los recursos de AmazonEC2.

El rol AWSServiceRoleForPCS vinculado al servicio confía en los siguientes servicios para asumir el rol:

- `pcs.amazonaws.com`

La política de permisos de roles denominada AWSPCSServiceRolePolicypermite AWS PCS completar acciones en recursos específicos.

Debe configurar los permisos para permitir a sus usuarios, grupos o funciones, crear, editar o eliminar la descripción de un rol vinculado al servicio. Para obtener más información, consulte los permisos de roles vinculados a un servicio en la Guía del IAMusuario.

Crear un rol vinculado a un servicio para AWS PCS

No es necesario crear manualmente un rol vinculado a un servicio. AWS PCScrea un rol vinculado a un servicio para usted al crear un clúster.

Edición de un rol vinculado a un servicio para AWS PCS

AWS PCSno permite editar el rol vinculado al AWSServiceRoleForPCS servicio. Después de crear un rol vinculado a un servicio, no podrá cambiar el nombre del rol, ya que varias entidades podrían hacer referencia al mismo. Sin embargo, puede editar la descripción del rol utilizando. IAM Para obtener más información, consulte Edición de un rol vinculado a un servicio en la Guía del IAMusuario.

Eliminar un rol vinculado a un servicio para AWS PCS

Si ya no necesita usar una característica o servicio que requieran un rol vinculado a un servicio, le recomendamos que elimine dicho rol. Así no tendrá una entidad no utilizada que no se monitorice ni mantenga de forma activa. Sin embargo, debe limpiar los recursos de su rol vinculado al servicio antes de eliminarlo manualmente.

> **Note**
>
> Si el AWS PCS servicio utiliza el rol al intentar eliminar los recursos, es posible que la eliminación no se realice correctamente. En tal caso, espere unos minutos e intente de nuevo la operación.

Para eliminar AWS PCS los recursos utilizados por el AWSServiceRoleForPCS

Debe eliminar todos los clústeres para eliminar la función AWSServiceRoleForPCS vinculada al servicio. Para obtener más información, consulte Eliminar un clúster.

Para eliminar manualmente el rol vinculado al servicio mediante IAM

Utilice la IAM consola AWS CLI, la o la AWS API para eliminar la función vinculada al AWSServiceRoleForPCS servicio. Para obtener más información, consulte Eliminar un rol vinculado a un servicio en la Guía del usuario. IAM

Regiones admitidas para los roles vinculados a un servicio de AWS PCS

AWS PCSadmite el uso de funciones vinculadas al servicio en todas las regiones en las que el servicio está disponible. Para obtener más información, consulte Puntos de conexión y Regiones de AWS.

Función de Amazon EC2 Spot para AWS PCS

Si desea crear un grupo de nodos de AWS PCS cómputo que utilice Spot como opción de compra, también debe tener en su cuenta la función AWSServiceRoleForEC2Spotvinculada al servicio. Cuenta de AWS Puede usar el siguiente AWS CLI comando para crear el rol. Para obtener más información, consulte Crear un rol vinculado a un servicio y Crear un rol para delegar permisos a un AWS servicio en la Guía del AWS Identity and Access Management usuario.

```
aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

> **Note**
>
> Aparece el siguiente error si Cuenta de AWS ya tiene un `AWSServiceRoleForEC2Spot` IAM rol.
>
> ```
> An error occurred (InvalidInput) when calling the CreateServiceLinkedRole
> operation: Service role name AWSServiceRoleForEC2Spot has been taken in this
> account, please try a different suffix.
> ```

Permisos mínimos para AWS PCS

En esta sección se describen los IAM permisos mínimos necesarios para que una IAM identidad (usuario, grupo o rol) utilice el servicio.

Contenido

Permisos mínimos para usar API acciones

APIacción	Permisos mínimos	Permisos adicionales para la consola
CreateCluster	`ec2:CreateNetworkInterface, ec2:DescribeVpcs, ec2:DescribeSubnets, ec2:DescribeSecurityGroups, ec2:GetSecurityGroupsForVpc, iam:CreateServiceLinkedRole, secretsmanager:CreateSecret, secretsmanager:TagResource, pcs:CreateCluster`	
ListClusters	`pcs:ListClusters`	
GetCluster	`pcs:GetCluster`	`ec2:DescribeSubnets`

APIacción	Permisos mínimos	Permisos adicionales para la consola
DeleteCluster	`pcs:DeleteCluster`	
CreateComputeNodeGroup	`ec2:DescribeVpcs, ec2:DescribeSubnets, ec2:DescribeSecurityGroups, ec2:DescribeLaunchTemplates, ec2:DescribeLaunchTemplateVersions, ec2:DescribeInstanceTypes, ec2:RunInstances, ec2:CreateFleet, ec2:CreateTags, iam:PassRole, iam:GetInstanceProfile, pcs:CreateComputeNodeGroup`	`iam:ListInstanceProfiles, ec2:DescribeImages, pcs:GetCluster`
ListComputerNodeGroups	`pcs:ListComputeNodeGroups`	`pcs:GetCluster`
GetComputeNodeGroup	`pcs:GetComputeNodeGroup`	`ec2:DescribeSubnets`

APIacción	Permisos mínimos	Permisos adicionales para la consola
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`
DeleteComputeNodeGroup	`pcs:DeleteComputeNodeGroup`	
CreateQueue	`pcs:CreateQueue`	`pcs:ListComputeNodeGroups, pcs:GetCluster`
ListQueues	`pcs:ListQueues`	`pcs:GetCluster`
GetQueue	`pcs:GetQueue`	
UpdateQueue	`pcs:UpdateQueue`	`pcs:ListComputeNodeGroups, pcs:GetQueue`

APIacción	Permisos mínimos	Permisos adicionales para la consola
DeleteQueue	`pcs:DeleteQueue`	

Se requieren permisos mínimos para usar etiquetas

Se requieren los siguientes permisos para usar etiquetas con los recursos incluidos AWS PCS.

```
pcs:ListTagsForResource
pcs:TagResource
pcs:UntagResource
```

Se requieren permisos mínimos para admitir registros

AWS PCSenvía los datos de registro a Amazon CloudWatch Logs (CloudWatch Logs). Debe asegurarse de que su identidad tiene los permisos mínimos para usar CloudWatch Logs. Para obtener más información, consulte Descripción general de la gestión de los permisos de acceso a sus recursos de CloudWatch Logs en la Guía del usuario de Amazon CloudWatch Logs.

Para obtener información sobre los permisos necesarios para que un servicio envíe CloudWatch registros a Logs, consulte Habilitar el registro desde AWS los servicios en la Guía del usuario de Amazon CloudWatch Logs.

Permisos mínimos para un administrador de servicios

La siguiente IAM política especifica los permisos mínimos necesarios para que una IAM identidad (usuario, grupo o rol) configure y administre el AWS PCS servicio.

> **Note**
>
> Los usuarios que no configuran ni administran el servicio no necesitan estos permisos. Los usuarios que solo ejecutan trabajos utilizan secure shell (SSH) para conectarse al clúster. AWS Identity and Access Management (IAM) no gestiona la autenticación o la autorización deSSH.

```
{
```

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

Puede excluir los siguientes permisos de la política y, en su lugar, utilizar la política gestionada correspondiente enIAM:

- `"firehose:*"`

  ```
  AmazonKinesisFirehoseFullAccess
  ```

- `"kms:*"`

  ```
  AWSKeyManagementServicePowerUser
  ```

- `"logs:*"`

  ```
  CloudWatchLogsFullAccess
  ```

- `"s3:*"`

```
AmazonS3FullAccess
```

IAMperfiles de instancia para AWS Parallel Computing Service

Las aplicaciones que se ejecutan en una EC2 instancia deben incluir AWS credenciales en todas AWS API las solicitudes que realicen. Te recomendamos que utilices un IAM rol para administrar las credenciales temporales de la EC2 instancia. Para ello, puedes definir un perfil de instancia y adjuntarlo a tus instancias. Para obtener más información, consulta las IAMfunciones de Amazon EC2 en la Guía del usuario de Amazon Elastic Compute Cloud.

> **Note**
>
> Cuando utilizas el AWS Management Console para crear un IAM rol para AmazonEC2, la consola crea un perfil de instancia automáticamente y le da el mismo nombre que el IAM rol. Si usa las AWS API acciones AWS CLI, o an AWS SDK para crear el IAM rol, crea el perfil de instancia como una acción independiente. Para obtener más información, consulte Perfiles de instancia en la Guía del usuario de Amazon Elastic Compute Cloud.

Debe especificar el perfil ARN de una instancia al crear un grupo de nodos de cómputo. Puede elegir diferentes perfiles de instancia para algunos o todos los grupos de nodos de cómputo.

Requisitos del perfil de instancia

Nombre del perfil de instancia

El perfil de IAM instancia ARN debe empezar por su ruta `AWSPCS` o `/aws-pcs/` incluirla en su ruta.

Example

- `arn:aws:iam::*:instance-profile/AWSPCS-example-role-1` y
- `arn:aws:iam::*:instance-profile/aws-pcs/example-role-2`.

Permisos

Como mínimo, el perfil de instancia AWS PCS debe incluir la siguiente política. Permite que los nodos de cómputo notifiquen al AWS PCS servicio cuando estén operativos.

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

Políticas adicionales

Puede considerar la posibilidad de añadir políticas administradas al perfil de la instancia. Por ejemplo:

- AmazonS3 ReadOnlyAccess proporciona acceso de solo lectura a todos los buckets de S3.
- A mazonSSMManaged InstanceCore habilita las funciones principales del servicio AWS Systems Manager, como el acceso remoto directamente desde Amazon Management Console.
- CloudWatchAgentServerPolicycontiene los permisos necesarios para su uso AmazonCloudWatchAgent en los servidores.

También puede incluir sus propias IAM políticas que respalden su caso de uso específico.

Creación de un perfil de instancia

Puedes crear un perfil de instancia directamente desde la EC2 consola de Amazon. Para obtener más información, consulta Cómo usar perfiles de instancia en la Guía del AWS Identity and Access Management usuario.

Solución de problemas de identidad y acceso a AWS Parallel Computing Service

Utilice la siguiente información como ayuda para diagnosticar y solucionar los problemas más comunes que pueden surgir al trabajar con AWS PCS yIAM.

Temas

No estoy autorizado a realizar ninguna acción en AWS PCS

Si recibe un error que indica que no tiene autorización para realizar una acción, las políticas se deben actualizar para permitirle realizar la acción.

El siguiente ejemplo de error se produce cuando el `mateojackson` IAM usuario intenta usar la consola para ver los detalles de un *`my-example-widget`* recurso ficticio, pero no tiene los `pcs:`*`GetWidget`* permisos ficticios.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 pcs:GetWidget on resource: my-example-widget
```

En este caso, la política del usuario `mateojackson` debe actualizarse para permitir el acceso al recurso *`my-example-widget`* mediante la acción `pcs:`*`GetWidget`*.

Si necesita ayuda, póngase en contacto con AWS el administrador. El administrador es la persona que le proporcionó las credenciales de inicio de sesión.

No estoy autorizado a realizar tareas como: PassRole

Si recibes un mensaje de error que indica que no estás autorizado a realizar la `iam:PassRole` acción, debes actualizar tus políticas para que puedas transferirle AWS PCS una función.

Algunas Servicios de AWS permiten transferir una función existente a ese servicio en lugar de crear una nueva función de servicio o una función vinculada al servicio. Para ello, debe tener permisos para transferir el rol al servicio.

El siguiente ejemplo de error se produce cuando un IAM usuario denominado `marymajor` intenta utilizar la consola para realizar una acción en ella. AWS PCS Sin embargo, la acción requiere que el servicio cuente con permisos que otorguen un rol de servicio. Mary no tiene permisos para transferir el rol al servicio.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

En este caso, las políticas de Mary se deben actualizar para permitirle realizar la acción `iam:PassRole`.

Si necesita ayuda, póngase en contacto con AWS el administrador. El administrador es la persona que le proporcionó las credenciales de inicio de sesión.

Quiero permitir que personas ajenas a mí accedan Cuenta de AWS a mis AWS PCS recursos

Puede crear un rol que los usuarios de otras cuentas o las personas externas a la organización puedan utilizar para acceder a sus recursos. Puede especificar una persona de confianza para que asuma el rol. En el caso de los servicios que respaldan las políticas basadas en recursos o las listas de control de acceso (ACLs), puedes usar esas políticas para permitir que las personas accedan a tus recursos.

Para más información, consulte lo siguiente:

- Para saber si AWS PCS es compatible con estas funciones, consulte. Cómo funciona AWS Parallel Computing Service con IAM
- Para obtener información sobre cómo proporcionar acceso a los recursos de su propiedad, consulte Proporcionar acceso a un IAM usuario en otro Cuenta de AWS de su propiedad en la Guía del IAM usuario. Cuentas de AWS
- Para obtener información sobre cómo proporcionar acceso a tus recursos a terceros Cuentas de AWS, consulta Cómo permitir el acceso a recursos que Cuentas de AWS son propiedad de terceros en la Guía del IAM usuario.
- Para obtener información sobre cómo proporcionar acceso mediante la federación de identidades, consulte Proporcionar acceso a usuarios autenticados externamente (federación de identidades) en la Guía del IAM usuario.
- Para saber la diferencia entre el uso de roles y políticas basadas en recursos para el acceso entre cuentas, consulte el acceso a recursos entre cuentas IAM en la Guía del usuario. IAM

Validación del cumplimiento del servicio de computación paralela AWS

Para saber si uno Servicio de AWS está dentro del ámbito de aplicación de programas de cumplimiento específicos, consulte Servicios de AWS Alcance por programa de cumplimiento

Servicios de AWS de cumplimiento y elija el programa de cumplimiento que le interese. Para obtener información general, consulte Programas de AWS cumplimiento > Programas AWS .

Puede descargar informes de auditoría de terceros utilizando AWS Artifact. Para obtener más información, consulte Descarga de informes en AWS Artifact .

Su responsabilidad de cumplimiento al Servicios de AWS utilizarlos viene determinada por la confidencialidad de sus datos, los objetivos de cumplimiento de su empresa y las leyes y reglamentos aplicables. AWS proporciona los siguientes recursos para ayudar con el cumplimiento:

- Guías de inicio rápido sobre seguridad y cumplimiento: estas guías de implementación analizan las consideraciones arquitectónicas y proporcionan los pasos para implementar entornos básicos centrados en AWS la seguridad y el cumplimiento.
- Diseñando una arquitectura basada en la HIPAA seguridad y el cumplimiento en Amazon Web Services: en este documento técnico se describe cómo pueden utilizar las empresas AWS para crear HIPAA aplicaciones aptas.

  > **Note**
  >
  > No todos son aptos. Servicios de AWS HIPAA Para obtener más información, consulta la Referencia de servicios HIPAA aptos.

- AWS Recursos de de cumplimiento: esta colección de libros de trabajo y guías puede aplicarse a su industria y ubicación.
- AWS Guías de cumplimiento para clientes: comprenda el modelo de responsabilidad compartida desde el punto de vista del cumplimiento. En las guías se resumen las mejores prácticas para garantizar la seguridad Servicios de AWS y se orientan a los controles de seguridad en varios marcos (incluidos el Instituto Nacional de Estándares y Tecnología (NIST), el Consejo de Normas de Seguridad del Sector de Tarjetas de Pago (PCI) y la Organización Internacional de Normalización (ISO)).
- Evaluación de los recursos con reglas en la guía para AWS Config desarrolladores: el AWS Config servicio evalúa en qué medida las configuraciones de los recursos cumplen con las prácticas internas, las directrices del sector y las normas.
- AWS Security Hub— Esto Servicio de AWS proporciona una visión completa del estado de su seguridad interior AWS. Security Hub utiliza controles de seguridad para evaluar sus recursos de AWS y comprobar su cumplimiento con los estándares y las prácticas recomendadas del

sector de la seguridad. Para obtener una lista de los servicios y controles compatibles, consulte la Referencia de controles de Security Hub.

- Amazon GuardDuty: Servicio de AWS detecta posibles amenazas para sus cargas de trabajo Cuentas de AWS, contenedores y datos mediante la supervisión de su entorno para detectar actividades sospechosas y maliciosas. GuardDuty puede ayudarlo a cumplir con varios requisitos de conformidad, por ejemplo PCIDSS, cumpliendo con los requisitos de detección de intrusiones exigidos por ciertos marcos de cumplimiento.
- AWS Audit Manager— Esto le Servicio de AWS ayuda a auditar continuamente su AWS consumo para simplificar la gestión del riesgo y el cumplimiento de las normativas y los estándares del sector.

Servicio de resiliencia en la computación AWS paralela

La infraestructura AWS global se basa en zonas Regiones de AWS de disponibilidad. Regiones de AWS proporcionan varias zonas de disponibilidad aisladas y separadas físicamente, que están conectadas mediante redes de baja latencia, alto rendimiento y alta redundancia. Con las zonas de disponibilidad, puede diseñar y utilizar aplicaciones y bases de datos que realizan una conmutación por error automática entre las zonas sin interrupciones. Las zonas de disponibilidad tienen una mayor disponibilidad, tolerancia a errores y escalabilidad que las infraestructuras tradicionales de uno o varios centros de datos.

Para obtener más información sobre las zonas de disponibilidad Regiones de AWS y las zonas de disponibilidad, consulte Infraestructura global.AWS

Servicio de seguridad de infraestructura en computación AWS paralela

Como servicio gestionado, AWS Parallel Computing Service está protegido por la seguridad de la red AWS global. Para obtener información sobre los servicios AWS de seguridad y cómo se AWS protege la infraestructura, consulte Seguridad AWS en la nube. Para diseñar su AWS entorno utilizando las mejores prácticas de seguridad de la infraestructura, consulte Protección de infraestructuras en un marco de buena AWS arquitectura basado en el pilar de la seguridad.

Utiliza las API llamadas AWS publicadas para acceder a AWS PCS través de la red. Los clientes deben admitir lo siguiente:

- Seguridad de la capa de transporte (TLS). Necesitamos TLS 1.2 y recomendamos TLS 1.3.

- Cifre suites con perfecto secreto (PFS), como (Ephemeral Diffie-Hellman) o DHE ECDHE (Elliptic Curve Ephemeral Diffie-Hellman). La mayoría de los sistemas modernos como Java 7 y posteriores son compatibles con estos modos.

Además, las solicitudes deben firmarse con un identificador de clave de acceso y una clave de acceso secreta que esté asociada a un director. IAM También puede utilizar AWS Security Token Service (AWS STS) para generar credenciales de seguridad temporales para firmar solicitudes.

Al AWS PCS crear un clúster, el servicio lanza el controlador Slurm en una cuenta propiedad del servicio, independiente de los nodos de procesamiento de su cuenta. Para unir la comunicación entre el controlador y los nodos de cómputo, AWS PCS crea una interfaz de red elástica () entre cuentas en su. ENI VPC El controlador Slurm lo utiliza ENI para administrar y comunicarse con los diferentes nodos de cómputo, lo que mantiene la seguridad y el aislamiento de los recursos y Cuentas de AWS, al mismo tiempo, facilita la eficiencia de las operaciones de inteligencia artificial HPC y aprendizaje automático.

Análisis y gestión de vulnerabilidades en Parallel Computing Service AWS

La configuración y los controles de TI son una responsabilidad compartida entre usted AWS y usted. Para obtener más información, consulte el modelo de responsabilidad AWS compartida. AWS gestiona las tareas de seguridad básicas de la infraestructura subyacente de la cuenta de servicio, como la aplicación de parches al sistema operativo en las instancias del controlador, la configuración del firewall y la recuperación ante desastres de la AWS infraestructura. Estos procedimientos han sido revisados y certificados por los terceros pertinentes. Para obtener más información, consulte las Prácticas recomendadas sobre seguridad, identidad y conformidad.

Usted es responsable de la seguridad de la infraestructura subyacente de sus Cuenta de AWS:

- Mantenga su código, incluidas las actualizaciones y los parches de seguridad.
- Aplica parches y actualiza el sistema operativo en las instancias de grupos de nodos.
- Actualice el programador para mantenerlo dentro de las versiones compatibles.
- Autentique y cifre la comunicación entre los clientes usuarios y los nodos a los que se conectan.

Prevención de la sustitución confusa entre servicios

El problema de la sustitución confusa es un problema de seguridad en el que una entidad que no tiene permiso para realizar una acción puede obligar a una entidad con más privilegios a realizar la acción. En AWS, la suplantación de identidad entre servicios puede provocar el confuso problema de un diputado. La suplantación entre servicios puede producirse cuando un servicio (el servicio que lleva a cabo las llamadas) llama a otro servicio (el servicio al que se llama). El servicio que lleva a cabo las llamadas se puede manipular para utilizar sus permisos a fin de actuar en función de los recursos de otro cliente de una manera en la que no debe tener permiso para acceder. Para evitarlo, AWS proporciona herramientas que lo ayudan a proteger sus datos para todos los servicios con entidades principales de servicio a las que se les ha dado acceso a los recursos de su cuenta.

Se recomienda utilizar las claves de contexto de condición `aws:SourceAccount`global `aws:SourceArn`y las claves de contexto en las políticas de recursos para limitar los permisos que AWS Parallel Computing Service (AWS PCS) concede a otro servicio al recurso. Utilice `aws:SourceArn` si desea que solo se asocie un recurso al acceso entre servicios. Utilice `aws:SourceAccount` si quiere permitir que cualquier recurso de esa cuenta se asocie al uso entre servicios.

La forma más eficaz de protegerse contra el confuso problema de los adjuntos es utilizar la clave de contexto ARN de la condición `aws:SourceArn` global con todo el recurso. Si no conoce la totalidad ARN del recurso o si está especificando varios recursos, utilice la clave de condición del contexto `aws:SourceArn` global con caracteres comodín (*) para las partes desconocidas delARN. Por ejemplo, `arn:aws:`*`servicename`*`:*:`*`123456789012`*`:*`.

Si el `aws:SourceArn` valor no contiene el ID de la cuenta, como un bucket de Amazon S3ARN, debe usar ambas claves de contexto de condición global para limitar los permisos.

El valor de `aws:SourceArn` debe ser un clústerARN.

El siguiente ejemplo muestra cómo puede utilizar las claves de contexto de condición `aws:SourceAccount` global `aws:SourceArn` y las claves contextuales AWS PCS para evitar el confuso problema de los diputados.

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

IAMrol para las EC2 instancias de Amazon aprovisionadas como parte de un grupo de nodos de cómputo

AWS PCSorganiza automáticamente la EC2 capacidad de Amazon para cada uno de los grupos de nodos de procesamiento configurados de un clúster. Al crear un grupo de nodos de cómputo, los usuarios deben proporcionar un perfil de IAM instancia a través del `iamInstanceProfileArn` campo. El perfil de instancia especifica los permisos asociados a las EC2 instancias aprovisionadas. AWS PCSacepta cualquier rol que tenga `AWSPCS` como prefijo de nombre de rol o `/aws-pcs/` como parte de la ruta del rol. El `iam:PassRole` permiso es obligatorio para la IAM identidad (usuario o rol) que crea o actualiza un grupo de nodos de procesamiento. Cuando un usuario llama a la `UpdateComputeNodeGroup` API acción `CreateComputeNodeGroup` o, AWS PCS comprueba si el usuario tiene permiso para realizar la `iam:PassRole` acción.

El siguiente ejemplo de política concede permisos para transferir únicamente IAM las funciones cuyo nombre comience por`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"
                    ]
                }
            }
        }
    ]
}
```

Prácticas recomendadas de seguridad para AWS Parallel Computing Service

En esta sección se describen las mejores prácticas de seguridad específicas de AWS Parallel Computing Service (AWS PCS). Para obtener más información sobre las prácticas recomendadas de seguridad AWS, consulte Prácticas recomendadas en materia de seguridad, identidad y conformidad.

AMIrelacionada con la seguridad

- No utilice la AWS PCS muestra AMIs para las cargas de trabajo de producción. Las muestras no AMIs son compatibles y solo están destinadas a ser probadas.
- Actualice periódicamente el sistema operativo y el software de las AWS PCS instancias para mitigar las vulnerabilidades.
- Úselo AWS Systems Manager para automatizar la aplicación de parches y mantener el cumplimiento de sus políticas de seguridad.
- Utilice únicamente AWS PCS paquetes oficiales autenticados descargados de fuentes oficiales AWS .
- Actualice periódicamente AWS PCS los paquetes en los nodos de cómputo para recibir mejoras y parches de seguridad. Considere la posibilidad de automatizar este proceso para minimizar las vulnerabilidades.

Seguridad de Slurm Workload Manager

- Implemente controles de acceso y restricciones de red para proteger los nodos de control y cómputo de Slurm. Permita que solo los usuarios y sistemas de confianza envíen trabajos y accedan a los comandos de administración de Slurm.

- Utilice las funciones de seguridad integradas de Slurm, como la autenticación de Slurm, para garantizar que las solicitudes de trabajo y las comunicaciones estén autenticadas.
- Actualice las versiones de Slurm para mantener un funcionamiento fluido y la compatibilidad con clústeres.

> **Important**
>
> Cualquier clúster que utilice una versión de Slurm que haya llegado al final de su vida útil (EOSL) se detendrá inmediatamente. Utilice el enlace que aparece en la parte superior de las páginas de la guía del usuario para suscribirse al RSS feed de AWS PCS documentación y recibir notificaciones cuando se acerque una versión de Slurm. EOSL

Supervisión y registro

- Use Amazon CloudWatch Logs AWS CloudTrail para monitorear y registrar las acciones en sus clústeres y Cuenta de AWS. Utilice los datos para solucionar problemas y realizar auditorías.

Seguridad de la red

- Implemente sus AWS PCS clústeres de forma independiente VPC para aislar su HPC entorno del resto del tráfico de la red.
- Utilice grupos de seguridad y listas de control de acceso a la red (ACLs) para controlar el tráfico entrante y saliente a AWS PCS las instancias y subredes.
- Utilice nuestros AWS PrivateLink VPC puntos finales para mantener el tráfico de red entre los clústeres y otros AWS servicios de la red. AWS

Registro y supervisión de AWS PCS

El monitoreo es una parte importante del mantenimiento de la confiabilidad, la disponibilidad y el rendimiento de AWS PCS sus demás AWS recursos. AWSproporciona las siguientes herramientas de monitoreo para observar AWS PCS, informar cuando algo anda mal y tomar medidas automáticas cuando sea apropiado:

- Amazon CloudWatch monitorea tus AWS recursos y las aplicaciones en las que AWS ejecutas en tiempo real. Puede recopilar métricas y realizar un seguimiento de las métricas, crear paneles personalizados y definir alarmas que le advierten o que toman medidas cuando una métrica determinada alcanza el umbral que se especifique. Por ejemplo, puedes CloudWatch hacer un seguimiento CPU del uso u otras métricas de tus EC2 instancias de Amazon y lanzar automáticamente nuevas instancias cuando sea necesario. Para obtener más información, consulta la Guía del CloudWatch usuario de Amazon.
- Amazon CloudWatch Logs le permite supervisar, almacenar y acceder a sus archivos de registro desde EC2 instancias de Amazon y otras fuentes. CloudTrail CloudWatch Los registros pueden monitorear la información de los archivos de registro y notificarle cuando se alcanzan ciertos umbrales. También se pueden archivar los datos del registro en un almacenamiento de larga duración. Para obtener más información, consulta la Guía del usuario CloudWatch de Amazon Logs.
- AWS CloudTrailcaptura API las llamadas y los eventos relacionados realizados por su AWS cuenta o en su nombre y envía los archivos de registro a un bucket de Amazon S3 que especifique. Puede identificar qué usuarios y cuentas llamaron AWS, la dirección IP de origen desde la que se realizaron las llamadas y cuándo se produjeron. Para obtener más información, consulte la AWS CloudTrail Guía del usuario de .

AWS PCSregistros del planificador

Puede configurarlo AWS PCS para enviar datos de registro detallados desde el programador de clústeres a Amazon CloudWatch Logs, Amazon Simple Storage Service (Amazon S3) y Amazon Data Firehose. Esto puede ayudar con la supervisión y la solución de problemas. Puede configurar los registros del AWS PCS programador mediante la AWS PCS consola, así como mediante programación mediante la tecla o. AWS CLI SDK

Contenido

Requisitos previos

El IAM principal utilizado para administrar el AWS PCS clúster debe permitirlo. `pcs:AllowVendedLogDeliveryForResource` Este es un ejemplo de AWS IAM política que lo habilita.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PcsAllowVendedLogsDelivery",
            "Effect": "Allow",
            "Action": ["pcs:AllowVendedLogDeliveryForResource"],
            "Resource": [
                "arn:aws:pcs:::cluster/*"
            ]
        }
    ]
}
```

Configuración de los registros del programador mediante la consola AWS PCS

Para configurar los registros del AWS PCS programador en la consola, sigue estos pasos:

1. Abre la AWS PCSconsola.
2. Elija Clústeres y vaya a la página de detalles del AWS PCS clúster, donde habilitará el registro.
3. Seleccione Logs (Registros).
4. En Entregas de registros: Scheduler Logs (opcional)

a. Agregue hasta tres destinos de entrega de registros. Las opciones incluyen CloudWatch Logs, Amazon S3 o Firehose.
b. Selecciona Actualizar entregas de registros.

Para volver a configurar, añadir o eliminar las entregas de registros, vuelva a visitar esta página.

Configurar los registros del programador mediante el AWS CLI

Para ello, necesita al menos un destino de entrega, una fuente de entrega (el PCS clúster) y una entrega, que es una relación que conecta una fuente con un destino.

Crea un destino de entrega

Necesita al menos un destino de entrega para recibir los registros del programador de un AWS PCS clúster. Puede obtener más información sobre este tema en la PutDeliveryDestination sección de la Guía del CloudWatch API usuario.

Para crear un destino de entrega mediante el AWS CLI

- Cree un destino con el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:
  - Reemplazar *region-code* con el Región de AWS lugar donde crearás tu destino. Por lo general, será la misma región en la que se implementó el AWS PCS clúster.
  - Reemplazar *pcs-logs-destination* con el nombre que prefiera. Debe ser único para todos los destinos de entrega de tu cuenta.
  - Reemplazar *resource-arn* con el ARN para un grupo de CloudWatch registros existente en Logs, un bucket de S3 o un flujo de entrega en Firehose. Entre los ejemplos se incluyen:
    - CloudWatch Grupo de registros

      ```
      arn:aws:logs:region-code:account-id:log-group:/log-group-name:*
      ```

    - S3 bucket

      ```
      arn:aws:s3:::bucket-name
      ```

    - Flujo de entrega de 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
```

Toma nota del nuevo destino ARN de entrega, ya que lo necesitarás para configurar las entregas.

Habilite el AWS PCS clúster como fuente de entrega

Para recopilar los registros del programador AWSPCS, configure el clúster como fuente de entrega. Para obtener más información, consulta PutDeliverySourcela APIreferencia de Amazon CloudWatch Logs.

Para configurar un clúster como fuente de entrega mediante el AWS CLI

- Habilite la entrega de registros desde su clúster con el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:
  - Reemplazar *region-code* con el Región de AWS lugar en el que está desplegado el clúster.
  - Reemplazar *cluster-logs-source-name* con un nombre para esta fuente. Debe ser único para todas las fuentes de entrega de su Cuenta de AWS. Considere la posibilidad de incorporar el nombre o el ID del AWS PCS clúster.
  - Reemplazar *cluster-arn* con el ARN para su AWS PCS clúster

```
aws logs put-delivery-source \
    --region region-code \
    --name cluster-logs-source-name \
    --resource-arn cluster-arn \
    --log-type PCS_SCHEDULER_LOGS
```

Conecte la fuente de entrega del clúster al destino de entrega

Para que los datos de registro del programador fluyan del clúster al destino, debe configurar una entrega que los conecte. Para obtener más información, consulta CreateDeliverylа APIreferencia de Amazon CloudWatch Logs.

Para crear una entrega mediante el AWS CLI

- Cree una entrega mediante el siguiente comando. Antes de ejecutar el comando, realice los siguientes reemplazos:

- Reemplazar *region-code* con el Región de AWS lugar donde se encuentran su origen y su destino.
- Reemplazar *cluster-logs-source-name* con el nombre de la fuente de entrega indicada arriba.
- Reemplazar *destination-arn* con el ARN de un destino de entrega al que desee que se entreguen los troncos.

```
aws logs create-delivery \
    --region region-code \
    --delivery-source-name cluster-logs-source \
    --delivery-destination-arn destination-arn
```

Rutas y nombres de los flujos de registro del programador

La ruta y el nombre de los registros del AWS PCS programador dependen del tipo de destino.

- CloudWatch Registros
  - Una transmisión CloudWatch de registros sigue esta convención de nomenclatura.

    ```
    AWSLogs/PCS/${cluster_id}/${log_name}_${scheduler_major_version}.log
    ```

    Example

    ```
    AWSLogs/PCS/abcdef0123/slurmctld_24.05.log
    ```
- S3 bucket
  - La ruta de salida de un bucket de S3 sigue esta convención de nomenclatura:

    ```
    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.
    ```
  - El nombre de un objeto de S3 sigue esta convención:

```
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
```

Ejemplo de registro del AWS PCS programador

AWSPCSlos registros del planificador están estructurados. Incluyen campos como el identificador del clúster, el tipo de programador y las versiones principales y de parche, además del mensaje de registro emitido por el proceso del controlador Slurm. A continuación se muestra un ejemplo.

```
{
    "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"
}
```

Servicio de monitorización de computación AWS paralela con Amazon CloudWatch

Amazon CloudWatch supervisa el estado y el rendimiento del clúster de AWS Parallel Computing Service (AWS PCS) mediante la recopilación de métricas del clúster a intervalos. Estas métricas se conservan, lo que le permite acceder a los datos históricos y obtener información sobre el rendimiento de su clúster a lo largo del tiempo.

CloudWatch también le permite monitorear las EC2 instancias lanzadas por AWS PCS para cumplir con sus requisitos de escalado. Si bien puede inspeccionar los registros de las instancias en ejecución, CloudWatch las métricas y los datos de registro normalmente se eliminan una vez que

las instancias se cierran. Sin embargo, puede configurar el CloudWatch agente en las instancias mediante una plantilla de EC2 lanzamiento para conservar las métricas y los registros incluso después de la finalización de la instancia, lo que permite la supervisión y el análisis a largo plazo.

Explore los temas de esta sección para obtener más información sobre la supervisión del AWS PCS uso CloudWatch.

Temas

- Supervise AWS PCS las métricas mediante CloudWatch
- Monitorización de AWS PCS instancias con Amazon CloudWatch

Supervise AWS PCS las métricas mediante CloudWatch

Puedes monitorizar el estado AWS PCS del clúster con Amazon CloudWatch, que recopila datos de tu clúster y los convierte en métricas prácticamente en tiempo real. Estas estadísticas se conservan durante un período de 15 meses, para que pueda acceder a la información histórica y obtener una mejor perspectiva del rendimiento de su clúster. Las métricas del clúster se envían CloudWatch en períodos de 1 minuto. Para obtener más información CloudWatch, consulta ¿Qué es Amazon CloudWatch? en la Guía del CloudWatch usuario de Amazon.

AWS PCSpublica las siguientes métricas en el espacio de PCS nombres AWS/de. CloudWatch Tienen una sola dimensión,. `ClusterId`

Nombre	Descripción	Unidades
ActualCapacity	IdleCapacity + UtilizedCapacity	Recuento
CapacityUtilization	UtilizedCapacity / ActualCapacity	Recuento
DesiredCapacity	ActualCapacity + PendingCapacity	Recuento
IdleCapacity	Recuento de instancias que se están ejecutando pero que no están asignadas a trabajos	Recuento

Nombre	Descripción	Unidades
UtilizedCapacity	Recuento de instancias que se están ejecutando y asignadas a trabajos	Recuento

Monitorización de AWS PCS instancias con Amazon CloudWatch

AWSPCSlanza EC2 las instancias de Amazon según sea necesario para cumplir con los requisitos de escalado definidos en sus grupos de nodos de PCS cómputo. Puedes monitorizar estas instancias mientras se ejecutan con Amazon CloudWatch. Puede inspeccionar los registros de las instancias en ejecución iniciando sesión en ellas y utilizando herramientas de línea de comandos interactivas. Sin embargo, de forma predeterminada, los datos de CloudWatch las métricas solo se conservan durante un período limitado una vez que se cierra una instancia y, por lo general, los registros de la instancia se eliminan junto con los EBS volúmenes que respaldan la instancia. Para conservar las métricas o los datos de registro de las instancias lanzadas una PCS vez finalizadas, puedes configurar el CloudWatch agente de tus instancias con una plantilla de EC2 lanzamiento. En este tema se proporciona información general sobre la supervisión de las instancias en ejecución y se proporcionan ejemplos de cómo configurar las métricas y los registros de las instancias persistentes.

Supervisión de instancias en ejecución

Búsqueda de AWS PCS instancias

Para supervisar las instancias lanzadas porPCS, busca las instancias en ejecución asociadas a un clúster o grupo de nodos de cómputo. A continuación, en la EC2 consola de una instancia determinada, inspeccione las secciones de estado y alarmas y de supervisión. Si el acceso de inicio de sesión está configurado para esas instancias, puede conectarse a ellas e inspeccionar los distintos archivos de registro de las instancias. Para obtener más información sobre cómo identificar las instancias que administranPCS, consulte[Búsqueda de instancias de grupos de nodos de cómputo en AWS PCS](#).

Habilitar métricas detalladas

De forma predeterminada, las métricas de las instancias se recopilan en intervalos de 5 minutos. Para recopilar métricas en intervalos de un minuto, habilita la CloudWatch supervisión detallada en la plantilla de lanzamiento de tu grupo de nodos de cómputo. Para obtener más información, consulte [Active la CloudWatch supervisión detallada](#).

Configurar las métricas y los registros de las instancias persistentes

Puedes conservar las métricas y los registros de tus instancias instalando y configurando el CloudWatch agente de Amazon en ellas. Consta de tres pasos principales:

1. Cree una configuración de CloudWatch agente.
2. Guarde la configuración en un lugar donde las PCS instancias puedan recuperarla.
3. Escriba una plantilla de EC2 lanzamiento que instale el software del CloudWatch agente, busque la configuración e inicie el CloudWatch agente con la configuración.

Para obtener más información, consulte Recopilar métricas, registros y seguimientos con el CloudWatch agente en la Guía del CloudWatch usuario de Amazon yUso de plantillas de EC2 lanzamiento de Amazon con AWS PCS.

Cree una configuración CloudWatch de agente

Antes de implementar el CloudWatch agente en las instancias, debe generar un archivo de JSON configuración que especifique las métricas, los registros y los seguimientos que se van a recopilar. Los archivos de configuración se pueden crear mediante un asistente o manualmente, mediante un editor de texto. El archivo de configuración se creará manualmente para esta demostración.

En el ordenador en el que lo tenga AWS CLI instalado, cree un archivo de CloudWatch configuración denominado config.json con el siguiente contenido. También puede usar lo siguiente URL para descargar una copia del archivo.

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/cloudwatch/assets/config.json
```

Notas

- Las rutas de registro del archivo de ejemplo son para Amazon Linux 2. Si sus instancias utilizarán un sistema operativo base diferente, cambie las rutas según corresponda.
- Para capturar otros registros, añada entradas adicionales en`collect_list`.
- Los valores de `{brackets}` son variables modeladas. Para ver la lista completa de variables admitidas, consulte Crear o editar manualmente el archivo de configuración del CloudWatch agente en la Guía del CloudWatch usuario de Amazon.
- Puede optar por omitir `logs` `metrics` o no recopilar estos tipos de información.

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

Este archivo indica al CloudWatch agente que supervise varios archivos que pueden ser útiles para diagnosticar errores, como el arranque, la autenticación y el inicio de sesión, y otros ámbitos de solución de problemas. Entre ellos se incluyen:

- `/var/log/cloud-init.log`— Resultado de la fase inicial de configuración de la instancia
- `/var/log/cloud-init-output.log`— Resultado de los comandos que se ejecutan durante la configuración de la instancia
- `/var/log/amazon/pcs/bootstrap.log`— Resultado de operaciones PCS específicas que se ejecutan durante la configuración de la instancia
- `/var/log/slurmd.log`— Resultado del daemon slurmd del administrador de cargas de trabajo de Slurm

- `/var/log/messages`— Mensajes del sistema desde el núcleo, los servicios del sistema y las aplicaciones
- `/var/log/secure`— Registros relacionados con los intentos de autenticaciónSSH, como el sudo y otros eventos de seguridad

Los archivos de registro se envían a un grupo de CloudWatch registros denominado`/PCSLogs/instances`. Los flujos de registro son una combinación del ID de instancia y el nombre base del archivo de registro. El grupo de registros tiene un tiempo de retención de 30 días.

Además, el archivo indica al CloudWatch agente que recopile varias métricas comunes y las agregue por ID de instancia.

Guarde la configuración

El archivo de configuración del CloudWatch agente debe almacenarse en un lugar donde las instancias de los nodos de PCS procesamiento puedan acceder a él. Existen dos formas comunes de hacerlo. Puede cargarlo en un bucket de Amazon S3 al que las instancias de su grupo de nodos de cómputo tendrán acceso a través de su perfil de instancia. También puede almacenarlo como un SSM parámetro en el almacén de parámetros de Amazon Systems Manager.

Súbelo a un bucket de S3

Para almacenar el archivo en S3, utilice los siguientes AWS CLI comandos. Antes de ejecutar el comando, realice las siguientes sustituciones:

- Reemplazar *DOC-EXAMPLE-BUCKET* con tu propio nombre de bucket de S3

En primer lugar (esto es opcional si tiene un depósito existente), cree un depósito para almacenar sus archivos de configuración.

```
aws s3 mb s3://DOC-EXAMPLE-BUCKET
```

A continuación, sube el archivo al depósito.

```
aws s3 cp ./config.json s3://DOC-EXAMPLE-BUCKET/
```

Almacenar como SSM parámetro

Para almacenar el archivo como SSM parámetro, utilice el siguiente comando. Antes de ejecutar el comando, realice las siguientes sustituciones:

- Reemplazar *region-code* con la AWS región en la que está trabajando AWSPCS.
- (Opcional) Reemplace *AmazonCloudWatch-PCS* con su propio nombre para el parámetro. Ten en cuenta que si cambias el prefijo del nombre, `AmazonCloudWatch-` tendrás que añadir específicamente el acceso de lectura al SSM parámetro en el perfil de instancia de tu grupo de nodos.

```
aws ssm put-parameter \
   --region region-code \
   --name "AmazonCloudWatch-PCS" \
   --type String \
   --value file://config.json
```

Escribe una plantilla de EC2 lanzamiento

Los detalles específicos de la plantilla de lanzamiento dependen de si el archivo de configuración está almacenado en S3 oSSM.

Utilice una configuración almacenada en S3

Este script instala el CloudWatch agente, importa un archivo de configuración de un bucket de S3 e inicia el CloudWatch agente con él. Sustituya los siguientes valores de este script por sus propios detalles:

- *DOC-EXAMPLE-BUCKET* — El nombre de un bucket de S3 desde el que puede leer su cuenta
- */config.json* — Ruta relativa a la raíz del bucket de S3 donde se almacena la configuración

```
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==--
```

El perfil de IAM instancia del grupo de nodos debe tener acceso al bucket. Este es un ejemplo IAM de política para el depósito en el script de datos de usuario anterior.

```
{
    "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/*"
            ]
        }
    ]
}
```

Tenga en cuenta también que las instancias deben permitir el tráfico saliente al S3 y a los puntos CloudWatch finales. Esto se puede lograr mediante grupos de seguridad o VPC puntos finales, según la arquitectura del clúster.

Utilice una configuración almacenada en SSM

Este script instala el CloudWatch agente, importa un archivo de configuración desde un SSM parámetro e inicia el CloudWatch agente con él. Sustituya los siguientes valores de este script por sus propios detalles:

- (Opcional) Reemplace *AmazonCloudWatch-PCS* con su propio nombre para el parámetro.

```
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 política de IAM instancias del grupo de nodos debe tener la directiva CloudWatchAgentServerPolicyadjunta.

Si el nombre de tu parámetro no empieza por, `AmazonCloudWatch-` tendrás que añadir específicamente el acceso de lectura al SSM parámetro en el perfil de instancia de tu grupo de nodos. Este es un ejemplo IAM de política que ilustra esto para el prefijo *DOC-EXAMPLE-PREFIX*.

```
{
  "Version" : "2012-10-17",
  "Statement" : [
    {
      "Sid" : "CustomCwSsmMParamReadOnly",
      "Effect" : "Allow",
      "Action" : [
        "ssm:GetParameter"
      ],
      "Resource" : "arn:aws:ssm:*:*:parameter/DOC-EXAMPLE-PREFIX*"
    }
  ]
}
```

Tenga en cuenta también que las instancias deben permitir el tráfico saliente hacia los puntos finales SSM y CloudWatch. Esto se puede lograr mediante grupos de seguridad o VPC puntos finales, según la arquitectura del clúster.

Registro de llamadas al Servicio de Computación AWS API Paralela mediante AWS CloudTrail

AWS PCSestá integrado con AWS CloudTrail un servicio que proporciona un registro de las acciones realizadas por un usuario, rol o AWS servicio en AWS PCS. CloudTrail captura todas API las llamadas AWS PCS como eventos. Las llamadas capturadas incluyen las llamadas desde la AWS PCS consola y las llamadas en código a las AWS PCS API operaciones. Si crea una ruta, puede habilitar la entrega continua de CloudTrail eventos a un bucket de Amazon S3, incluidos los eventos para AWS PCS. Si no configura una ruta, podrá ver los eventos más recientes en la CloudTrail consola, en el historial de eventos. Con la información recopilada por usted CloudTrail, puede determinar a AWS PCS qué dirección IP se realizó la solicitud, quién la realizó, cuándo se realizó y detalles adicionales.

Para obtener más información CloudTrail, consulte la Guía AWS CloudTrail del usuario.

AWS PCSinformación en CloudTrail

CloudTrail está habilitada en tu cuenta Cuenta de AWS al crear la cuenta. Cuando se produce una actividad en AWS PCS, esa actividad se registra en un CloudTrail evento junto con otros eventos de AWS servicio en el historial de eventos. Puede ver, buscar y descargar eventos recientes en su Cuenta de AWS. Para obtener más información, consulte Visualización de eventos con el historial de CloudTrail eventos.

Para tener un registro continuo de tus eventos Cuenta de AWS, incluidos los eventos para AWS PCS ti, crea una ruta. Un rastro permite CloudTrail entregar archivos de registro a un bucket de Amazon S3. De forma predeterminada, cuando se crea un registro de seguimiento en la consola, el registro de seguimiento se aplica a todas las Regiones de AWS. La ruta registra los eventos de todas las regiones de la AWS partición y envía los archivos de registro al bucket de Amazon S3 que especifique. Además, puede configurar otros AWS servicios para analizar más a fondo los datos de eventos recopilados en los CloudTrail registros y actuar en función de ellos. Para más información, consulte los siguientes temas:

- Introducción a la creación de registros de seguimiento
- CloudTrail servicios e integraciones compatibles
- Configuración de SNS las notificaciones de Amazon para CloudTrail
- Recibir archivos de CloudTrail registro de varias regiones y recibir archivos de CloudTrail registro de varias cuentas

Todas AWS PCS las acciones se registran CloudTrail y se documentan en la APIReferencia del Servicio de Computación AWS Paralela. Por ejemplo, las llamadas a las `CreateComputeNodeGroup` `DeleteCluster` acciones y las llamadas generan entradas en los archivos de CloudTrail registro. `UpdateQueue`

Cada entrada de registro o evento contiene información sobre quién generó la solicitud. La información de identidad del usuario lo ayuda a determinar lo siguiente:

- Si la solicitud se realizó con credenciales de usuario root o AWS Identity and Access Management (IAM).
- Si la solicitud se realizó con credenciales de seguridad temporales de un rol o fue un usuario federado.
- Si la solicitud la realizó otro AWS servicio.

Para obtener más información, consulte el CloudTrail userIdentityelemento.

Descripción de las entradas de los archivos de CloudTrail registro procedentes de AWS PCS

Un registro es una configuración que permite la entrega de eventos como archivos de registro a un bucket de S3 que usted especifique. CloudTrail Los archivos de registro contienen una o más entradas de registro. Un evento representa una solicitud única de cualquier fuente e incluye información sobre la acción solicitada, la fecha y la hora de la acción, los parámetros de la solicitud, etc. CloudTrail Los archivos de registro no son un registro ordenado de las API llamadas públicas, por lo que no aparecen en ningún orden específico.

En el siguiente ejemplo, se muestra una entrada de CloudTrail registro para una `CreateQueue` acción.

```
{
    "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": "AROAY36PTPIEEXAMPLE",
                "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": [
            {
                "computeNodeGroupId": "abcdef0123"
            }
        ],
        "queueName": "all"
    },
    "responseElements": {
        "queue": {
            "arn": "arn:aws:pcs:us-east-1:609783872011:cluster/abcdef0123/queue/
abcdef0123",
            "clusterId": "abcdef0123",
            "computeNodeGroupConfigurations": [
                {
                    "computeNodeGroupId": "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"
}
```

Puntos finales y cuotas de servicio para AWS PCS

En las siguientes secciones se describen los puntos finales y las cuotas de servicio de AWS Parallel Computing Service ()AWS PCS. Las cuotas de servicio, anteriormente denominadas límites, son la cantidad máxima de recursos u operaciones de servicio para usted Cuenta de AWS.

Cuenta de AWS Tiene cuotas predeterminadas para cada AWS servicio. A menos que se indique lo contrario, cada cuota es específica de la región de . Puede solicitar el aumento de algunas cuotas, pero otras no se pueden aumentar.

Para obtener más información, consulte AWS service quotas en la Referencia general de AWS .

Contenido

Puntos de conexión de servicio

Nombres de las regiones	Región	Punto de conexión	Protocolo
Este de EE. UU. (Norte de Virginia)	us-east-1	pcs.us-east-1.amazonaws.com	HTTPS
Este de EE. UU. (Ohio)	us-east-2	pcs.us-east-2.amazonaws.com	HTTPS
Oeste de EE. UU. (Oregón)	us-west-2	pcs.us-west-2.amazonaws.com	HTTPS
Asia-Pacífico (Singapur)	ap-southeast-1	pcs.ap-southeast-1.amazonaws.com	HTTPS

Nombres de las regiones	Región	Punto de conexión	Protocolo
Asia Pacífico (Sídney)	ap-southeast-2	pcs.ap-southeast-2.amazonaws.com	HTTPS
Asia-Pacífico (Tokio)	ap-northeast-1	pcs.ap-northeast-1.amazonaws.com	HTTPS
Europe (Fráncfort)	eu-central-1	pcs.eu-central-1.amazonaws.com	HTTPS
Europa (Irlanda)	eu-west-1	pcs.eu-west-1.amazonaws.com	HTTPS
Europa (Estocolmo)	eu-north-1	pcs.eu-north-1.amazonaws.com	HTTPS

Service Quotas

Nombre	Predeterminado	Ajustable	Descripción
Clústeres	5	Sí	El número máximo de clústeres por Región de AWS.

> **Note**
>
> Los valores predeterminados son las cuotas iniciales establecidas por AWS. Estos valores predeterminados son independientes de los valores reales de la cuota aplicada y de las cuotas de servicio máximas posibles. Para obtener más información, consulte Terminología de Service Quotas en la Guía del usuario de Service Quotas.

Estas cuotas de servicio se muestran en AWS Parallel Computing Service (PCS) en la AWS Management Console. Para solicitar un aumento de cuota para los valores que se muestran como ajustables, consulte Solicitar un aumento de cuota en la Guía del usuario de Service Quotas.

> **Important**
>
> Recuerde comprobar la Región de AWS configuración actual en AWS Management Console.

Cuotas internas

Las siguientes cuotas son internas y no ajustables.

Nombre	Predeterminado	Ajustable	Descripción
Creación simultánea de clústeres	1	No	El número máximo de clústeres en el `Creating` estado por. Región de AWS

Cuotas relevantes para otros AWS servicios

AWS PCSutiliza otros AWS servicios. Sus cuotas de servicio para esos servicios afectan al uso que haga de ellos AWS PCS.

Cuotas EC2 de servicio de Amazon que afectan AWS PCS

- Solicitudes de instancias puntuales
- Ejecución de instancias bajo demanda
- Plantillas de inicialización
- Versiones de plantillas de lanzamiento
- EC2APISolicitudes de Amazon

Para obtener más información, consulta las cuotas de EC2 servicio de Amazon en la Guía del usuario de Amazon Elastic Compute Cloud.

Notas de publicación para AWS PCS muestra AMIs

AWS PCSPor ejemplo, AMIs tienen una cadencia de publicación nocturna de los parches de seguridad. Estos parches de seguridad incrementales no están incluidos en las notas de lanzamiento oficiales.

> **Important**
>
> AMIsLos ejemplos son para fines de demostración y no se recomiendan para cargas de trabajo de producción.

Contenido

AWS PCSejemplo de x86_64 AMI para Slurm 23.11 (Amazon Linux 2)

Este documento describe los cambios, las adiciones, los problemas conocidos y las correcciones más recientes de AWS PCS Sample x86_64 (AMIAmazon Linux 2).

- Fecha de creación: 15 de julio de 2024
- Fecha de lanzamiento: 22 de agosto de 2024
- Última actualización: 22 de agosto de 2024

AMInombre

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

EC2Instancias compatibles

- Todas las instancias con un procesador x86 de 64 bits. Para encontrar instancias compatibles, navega a la EC2consola de Amazon. Selecciona Tipos de instancia y, a continuación, buscaArchitectures=x86_64.

AMIcontenido

- AWS Servicio compatible: AWS PCS
- Sistema operativo: Amazon Linux 2
- Arquitectura informática: x86_64
- Núcleo de Linux: 5.10.220-209.867.amzn2.x86_64
- EBStipo de volumen: gp2
- AWS PCSInstalador Slurm 23.11:23.11.9-1
- AWS PCSinstalador de software: 1.0.0-1
- EFAInstalador: 1.33.0
- GDRCopy: 2.4
- NVIDIAControlador: 535.154.05
- NVIDIACUDA: 12.2.2_535.104.05

Avisos

- Ninguna

Fecha de lanzamiento: 22-08-2020

Actualizado

- Ninguna. Primera versión.

Se añadió

- Ninguna. Primera versión.

Eliminaciones

- Ninguna. Primera versión.

AWS PCSejemplo de Arm64 AMI para Slurm 23.11 (Amazon Linux 2)

Este documento describe los cambios, las adiciones, los problemas conocidos y las correcciones más recientes de AWS PCS Sample Arm64 AMI (Amazon Linux 2).

- Fecha de creación: 15 de julio de 2024
- Fecha de lanzamiento: 22 de agosto de 2024
- Última actualización: 22 de agosto de 2024

AMInombre

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

EC2Instancias compatibles

- Todas las instancias con un procesador Arm de 64 bits. Para encontrar instancias compatibles, navega a la EC2consola de Amazon. Selecciona Tipos de instancia y, a continuación, busca`Architectures=arm64`.

AMIcontenido

- AWS Servicio compatible: AWS PCS
- Sistema operativo: Amazon Linux 2
- Arquitectura informática: arm64
- Núcleo de Linux: 5.10.220-209.867.amzn2.aarch64
- EBStipo de volumen: gp2
- AWS PCSInstalador Slurm 23.11:23.11.9-1
- AWS PCSinstalador de software: 1.0.0-1
- EFAInstalador: 1.33.0
- GDRCopy: 2.4
- NVIDIAControlador: 535.154.05
- NVIDIACUDA: 12.2.2_535.104.05

Avisos

- Ninguna

Fecha de lanzamiento: 22-08-2020

Actualizado

- Ninguna. Primera versión.

Se añadió

- Ninguna. Primera versión.

Eliminaciones

- Ninguna. Primera versión.

Historial de documentos para la guía del usuario de AWS PCS

En la siguiente tabla se describen las versiones de la documentación de AWS PCS.

Date	Cambio	Actualizaciones de la documentación	APIversiones actualizadas
28 de agosto de 2024	Se agregó la página de políticas administradas	Para obtener más información, consulte AWS políticas administradas para AWS Parallel Computing Service.	N/A
28 de agosto de 2024	AWS PCSlanzamiento	Versión inicial de la guía AWS PCS del usuario.	AWS SDK: 28 de agosto de 2024

AWS Glosario

Para obtener la AWS terminología más reciente, consulte el AWS glosario de la Glosario de AWS Referencia.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la version original de inglés, prevalecerá la version en inglés.