Guide de l’utilisateur

AWS PCS

AWS PCS: Guide de l’utilisateur

Table of Contents

Qu'est-ce que le service de calcul AWS parallèle ?

AWS Parallel Computing Service (AWS PCS) est un service géré qui facilite l'exécution et le dimensionnement des charges de travail de calcul haute performance (HPC), ainsi que la création de modèles scientifiques et d'ingénierie basés sur AWS l'utilisation de Slurm. AWS PCSÀ utiliser pour créer des clusters de calcul qui intègrent les meilleurs systèmes de AWS calcul, de stockage, de mise en réseau et de visualisation. Exécutez des simulations ou créez des modèles scientifiques et techniques. Rationalisez et simplifiez les opérations de votre cluster à l'aide de fonctionnalités de gestion et d'observabilité intégrées. Donnez à vos utilisateurs les moyens de se concentrer sur la recherche et l'innovation en leur permettant d'exécuter leurs applications et leurs tâches dans un environnement familier.

Concepts clés

Un cluster AWS PCS contient une ou plusieurs files d'attente associées à au moins un groupe de nœuds de calcul. Les tâches sont soumises à des files d'attente et exécutées sur des EC2 instances définies par des groupes de nœuds de calcul. Vous pouvez utiliser ces bases pour implémenter des HPC architectures sophistiquées.

Cluster

Un cluster est une ressource permettant de gérer des ressources et d'exécuter des charges de travail. Un cluster est une AWS PCS ressource qui définit un ensemble de configurations de calcul, de mise en réseau, de stockage, d'identité et de planificateur de tâches. Vous créez un cluster en spécifiant le planificateur de tâches que vous souhaitez utiliser (Slurm actuellement), la configuration du planificateur que vous souhaitez, le contrôleur de service que vous souhaitez gérer le cluster et dans lequel VPC vous souhaitez que les ressources du cluster soient lancées. Le planificateur accepte et planifie les tâches, et lance également les nœuds de calcul (EC2instances) qui traitent ces tâches.

Groupe de nœuds de calcul

Un groupe de nœuds de calcul est un ensemble de nœuds de calcul AWS PCS utilisé pour exécuter des tâches ou fournir un accès interactif à un cluster. Lorsque vous définissez un groupe de nœuds de calcul, vous spécifiez des caractéristiques communes telles que les types d'EC2instances Amazon, le nombre d'instances minimal et maximal, les VPC sous-réseaux cibles, Amazon Machine Image (AMI), l'option d'achat et la configuration de lancement personnalisée. AWS PCSutilise ces

paramètres pour lancer, gérer et arrêter efficacement les nœuds de calcul d'un groupe de nœuds de calcul.

File d’attente

Lorsque vous souhaitez exécuter une tâche sur un cluster spécifique, vous la soumettez à une file d'attente spécifique (parfois appelée partition). La tâche reste dans la file d'attente jusqu'à ce AWS PCS qu'elle soit planifiée pour s'exécuter sur un groupe de nœuds de calcul. Vous associez un ou plusieurs groupes de nœuds de calcul à chaque file d'attente. Une file d'attente est requise pour planifier et exécuter des tâches sur les ressources du groupe de nœuds de calcul sous-jacents à l'aide des différentes politiques de planification proposées par le planificateur de tâches. Les utilisateurs ne soumettent pas de tâches directement à un nœud de calcul ou à un groupe de nœuds de calcul.

Administrateur système

Un administrateur système déploie, gère et exploite un cluster. Ils peuvent y accéder AWS PCS via le AWS Management Console AWS PCSAPI, et AWS SDK. Ils ont accès à des clusters spécifiques via SSH ou AWS Systems Manager, où ils peuvent exécuter des tâches administratives, exécuter des tâches, gérer des données et effectuer d'autres activités basées sur le shell. Pour plus d'informations, consultez la documentation AWS Systems Manager.

Utilisateur final

L'utilisateur final n'a pas day-to-day la responsabilité de déployer ou d'exploiter un cluster. Ils utilisent une interface de terminal (telle queSSH) pour accéder aux ressources du cluster, exécuter des tâches, gérer les données et effectuer d'autres activités basées sur le shell.

Configuration du service de calcul AWS parallèle

Effectuez les tâches suivantes pour configurer le service de calcul AWS parallèle (AWS PCS).

Rubriques

Inscrivez-vous pour un Compte AWS

Si vous n'en avez pas Compte AWS, procédez comme suit pour en créer un.

Pour vous inscrire à un Compte AWS

1. Ouvrez l'https://portal.aws.amazon.com/billing/inscription.
2. Suivez les instructions en ligne.

   Dans le cadre de la procédure d'inscription, vous recevrez un appel téléphonique et vous saisirez un code de vérification en utilisant le clavier numérique du téléphone.

   Lorsque vous vous inscrivez à un Compte AWS, un Utilisateur racine d'un compte AWSest créé. Par défaut, seul l'utilisateur racine a accès à l'ensemble des Services AWS et des ressources de ce compte. La meilleure pratique de sécurité consiste à attribuer un accès administratif à un utilisateur, et à utiliser uniquement l'utilisateur racine pour effectuer les tâches nécessitant un accès utilisateur racine.

AWS vous envoie un e-mail de confirmation une fois le processus d'inscription terminé. À tout moment, vous pouvez consulter l'activité actuelle de votre compte et gérer votre compte en accédant à https://aws.amazon.com/et en choisissant Mon compte.

Création d'un utilisateur doté d'un accès administratif

Une fois que vous vous êtes inscrit à un utilisateur administratif Compte AWS, que vous Utilisateur racine d'un compte AWS l'avez sécurisé AWS IAM Identity Center, que vous l'avez activé et que vous en avez créé un, afin de ne pas utiliser l'utilisateur root pour les tâches quotidiennes.

Sécurisez votre Utilisateur racine d'un compte AWS

1. Connectez-vous en AWS Management Consoletant que propriétaire du compte en choisissant Utilisateur root et en saisissant votre adresse Compte AWS e-mail. Sur la page suivante, saisissez votre mot de passe.

   Pour obtenir de l‘aide pour vous connecter en utilisant l‘utilisateur racine, consultez Connexion en tant qu‘utilisateur racine dans le Guide de l‘utilisateur Connexion à AWS .
2. Activez l'authentification multifactorielle (MFA) pour votre utilisateur root.

   Pour obtenir des instructions, voir Activer un MFA périphérique virtuel pour votre utilisateur Compte AWS root (console) dans le guide de IAM l'utilisateur.

Création d'un utilisateur doté d'un accès administratif

1. Activez IAM Identity Center.

   Pour obtenir des instructions, consultez Activation d’ AWS IAM Identity Center dans le Guide de l’utilisateur AWS IAM Identity Center .
2. Dans IAM Identity Center, accordez un accès administratif à un utilisateur.

   Pour un didacticiel sur l'utilisation du Répertoire IAM Identity Center comme source d'identité, voir Configurer l'accès utilisateur par défaut Répertoire IAM Identity Center dans le Guide de AWS IAM Identity Center l'utilisateur.

Connexion en tant qu‘utilisateur doté d'un accès administratif

- Pour vous connecter avec votre utilisateur IAM Identity Center, utilisez l'URLidentifiant envoyé à votre adresse e-mail lorsque vous avez créé l'utilisateur IAM Identity Center.

  Pour obtenir de l'aide pour vous connecter en utilisant un utilisateur d'IAMIdentity Center, consultez la section Connexion au portail AWS d'accès dans le guide de Connexion à AWS l'utilisateur.

Attribution d'un accès à d'autres utilisateurs

1. Dans IAM Identity Center, créez un ensemble d'autorisations conforme à la meilleure pratique consistant à appliquer les autorisations du moindre privilège.

Pour obtenir des instructions, consultez Création d'un ensemble d'autorisations dans le Guide de l'utilisateur AWS IAM Identity Center .

2. Attribuez des utilisateurs à un groupe, puis attribuez un accès par authentification unique au groupe.

   Pour obtenir des instructions, consultez Ajout de groupes dans le Guide de l'utilisateur AWS IAM Identity Center .

Installez le AWS CLI

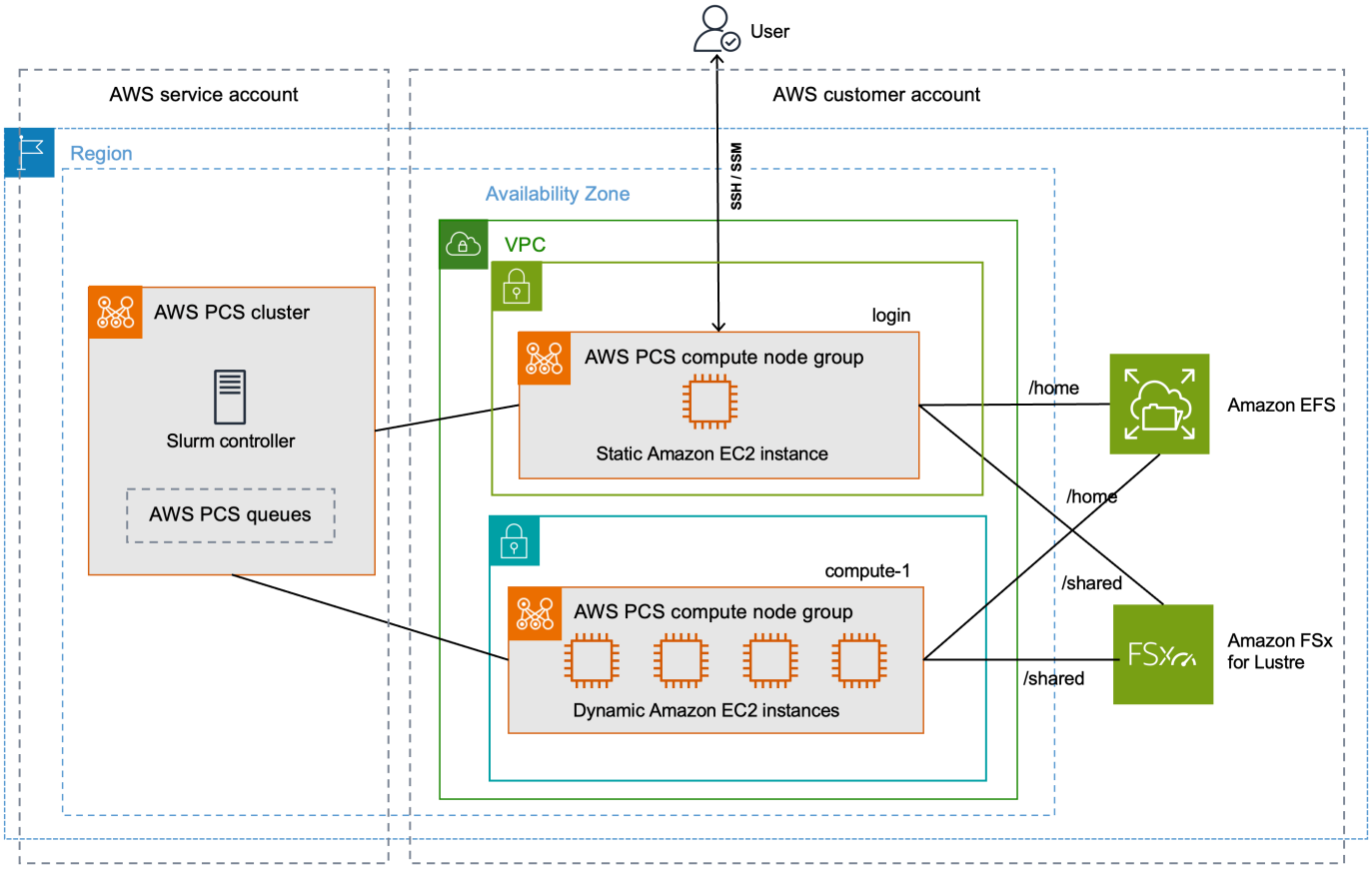
Vous devez utiliser la dernière version du AWS CLI. Pour plus d'informations, voir Installation ou mise à jour vers la dernière version du AWS CLI dans le guide de AWS Command Line Interface l'utilisateur de la version 2.

Entrez la commande suivante à une invite de commande pour vérifier votre AWS CLI ; elle devrait afficher des informations d'aide.

```
aws pcs help
```

Commencer avec AWS PCS

Il s'agit d'un didacticiel pour créer un cluster simple que vous pouvez utiliser pour essayer AWS PCS. La figure suivante montre la conception du cluster.



Le didacticiel sur la conception du cluster comporte les éléments clés suivants :

- A VPC et sous-réseaux qui répondent aux exigences du AWS PCS réseau.
- Un système de EFS fichiers Amazon, qui sera utilisé comme répertoire personnel partagé.
- Un système de fichiers Amazon FSx for Lustre, qui fournit un répertoire partagé à hautes performances.
- Un AWS PCS cluster, qui fournit un contrôleur Slurm.
- 2 groupes de nœuds de calcul.
  - Le groupe de `login` nœuds, qui fournit un accès interactif au système basé sur le shell.
  - Le groupe de `compute-1` nœuds fournit des instances évolutives de manière élastique pour exécuter des tâches.

- 1 file d'attente qui envoie des tâches aux EC2 instances du groupe de `compute-1` nœuds.

Le cluster nécessite des AWS ressources supplémentaires, telles que des groupes de sécurité, IAM des rôles et des modèles de EC2 lancement, qui ne sont pas illustrés dans le schéma.

Rubriques

Conditions préalables pour démarrer avec AWS PCS

Avant de commencer ce didacticiel, installez et configurez les outils et ressources suivants dont vous avez besoin pour créer et gérer un AWS PCS cluster.

- AWS CLI— Un outil de ligne de commande pour travailler avec les AWS services, y compris AWS PCS. Pour plus d'informations, voir Installation ou mise à jour vers la dernière version du AWS CLI dans le guide de AWS Command Line Interface l'utilisateur de la version 2. Après l'avoir installé AWS CLI, nous vous recommandons de le configurer également. Pour plus d'informations, voir Configurer le AWS CLI dans le guide de AWS Command Line Interface l'utilisateur de la version 2.
- IAMAutorisations requises : le principal de IAM sécurité que vous utilisez doit disposer des autorisations nécessaires pour utiliser les AWS PCS IAM rôles, les rôles liés aux services AWS CloudFormation VPC, a et les ressources associées. Pour plus d'informationsIdentity and Access Management pour le service de calcul AWS parallèle, reportez-vous à la section Création d'un rôle

lié à un service dans le Guide de l'AWS Identity and Access Management utilisateur. Vous devez effectuer toutes les étapes de ce guide avec le même utilisateur. Exécutez la commande suivante pour vérifier l'utilisateur actuel :

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

- Nous vous recommandons de suivre les étapes de ligne de commande décrites dans cette rubrique dans un shell Bash. Si vous n'utilisez pas de shell Bash, certaines commandes de script telles que les caractères de continuation de ligne et la façon dont les variables sont définies et utilisées nécessitent un ajustement pour votre shell. En outre, les règles de votre shell en matière de guillemets peuvent être différentes. Pour plus d'informations, voir Guillemets et littéraux avec chaînes AWS CLI dans le Guide de l'AWS Command Line Interface utilisateur de la version 2.

Créez un VPC et des sous-réseaux pour AWS PCS

Vous pouvez créer des sous-réseaux VPC et à l'aide d'un CloudFormation modèle. Utilisez ce qui suit URL pour télécharger le CloudFormation modèle, puis téléchargez-le dans la AWS CloudFormation console pour créer une nouvelle CloudFormation pile. Pour plus d'informations, consultez la section Utilisation de la AWS CloudFormation console dans le guide de AWS CloudFormation l'utilisateur.

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

Le modèle étant ouvert dans la AWS CloudFormation console, entrez les options suivantes. Vous pouvez utiliser les valeurs par défaut fournies dans le modèle.

- Sous Fournir un nom de pile :
  - Sous Nom de la pile, entrez :

    ```
    hpc-networking
    ```
- Sous Paramètres :
  - Sous VPC:
    - Sous CidrBlock, entrez :

      ```
      10.3.0.0/16
      ```
  - Sous les sous-réseaux A :

- Sous CidrPublicSubnetA, entrez :

```
10.3.0.0/20
```

- Sous CidrPrivateSubnetA, entrez :

```
10.3.128.0/20
```

- Sous les sous-réseaux B :
  - Sous CidrPublicSubnetB, entrez :

```
10.3.16.0/20
```

  - Sous CidrPrivateSubnetB, entrez :

```
10.3.144.0/20
```

- Sous les sous-réseaux C :
  - Pour ProvisionSubnetsC, sélectionnez Vrai
  - Sous CidrPublicSubnetC, entrez :

```
10.3.32.0/20
```

  - Sous CidrPrivateSubnetC, entrez :

```
10.3.160.0/20
```

- Sous Capacités :
  - Cochez la case « Je reconnais que cela AWS CloudFormation peut créer des IAM ressources ».

Surveillez l'état de la CloudFormation pile. Lorsqu'il atteint`CREATE_COMPLETE`, trouvez l'ID du groupe de sécurité par défaut dans le nouveauVPC. Vous utiliserez l'identifiant ultérieurement dans le didacticiel.

Trouvez le groupe de sécurité par défaut pour le cluster VPC

Pour trouver l'ID du groupe de sécurité par défaut dans le nouveauVPC, procédez comme suit :

- Accédez à la VPCconsole Amazon.

- Dans le VPCtableau de bord, sélectionnez Filtrer par VPC.
  - Choisissez l'VPCendroit où le nom commence par`hpc-networking`.
  - Sous Sécurité, sélectionnez Groupes de sécurité.
- Trouvez l'ID du groupe de sécurité pour le groupe nommé`default`. Il contient la description`default VPC security group`. Vous utiliserez l'ID ultérieurement pour configurer les modèles de EC2 lancement.

Créez des groupes de sécurité pour AWS PCS

AWS PCSs'appuie sur des groupes de sécurité pour gérer le trafic réseau entrant et sortant d'un cluster et de ses groupes de nœuds de calcul. Pour des informations détaillées sur ce sujet, voirExigences et considérations relatives aux groupes de sécurité.

Au cours de cette étape, vous allez utiliser un CloudFormation modèle pour deux groupes de sécurité.

- Groupe de sécurité du cluster, qui permet les communications entre le AWS PCS contrôleur, les nœuds de calcul et les nœuds de connexion.
- Un groupe SSH de sécurité entrant, que vous pouvez éventuellement ajouter à vos nœuds de connexion pour faciliter SSH l'accès

Créez les groupes de sécurité pour AWS PCS

Vous pouvez créer des sous-réseaux VPC et avec ce CloudFormation modèle. Utilisez ce qui suit URL pour télécharger le CloudFormation modèle, puis téléchargez-le dans la AWS CloudFormation console pour créer une nouvelle CloudFormation pile. Pour plus d'informations, consultez la section Utilisation de la AWS CloudFormation console dans le guide de AWS CloudFormation l'utilisateur.

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

Le modèle étant ouvert dans la AWS CloudFormation console, entrez les options suivantes. Notez que certaines options seront préremplies dans le modèle. Vous pouvez simplement les laisser comme valeurs par défaut.

- Sous Fournir un nom de pile

- Sous Nom de la pile, entrez :

```
getstarted-sg
```

- Sous Paramètres
  - Sous VpcId, choisissez l'VPCendroit où le nom commence par`hpc-networking`.
  - (Facultatif) Sous ClientIpCidr, entrez une plage d'adresses IP plus restrictive pour le groupe SSH de sécurité entrant. Nous vous recommandons de limiter cela à votre propre adresse IP/sous-réseau (x.x.x.x/32 pour votre propre adresse IP ou x.x.x.x.x/24 pour la plage). Remplacez x.x.x.x par votre propre adresse IP. PUBLIC Vous pouvez obtenir votre adresse IP publique à l'aide d'outils tels que https://ifconfig.co/)

Surveillez l'état de la CloudFormation pile. Lorsqu'il atteint `CREATE_COMPLETE` le groupe de sécurité, les ressources sont prêtes.

Deux groupes de sécurité ont été créés, avec les noms suivants :

- `cluster-getstarted-sg`— il s'agit du groupe de sécurité du cluster
- `inbound-ssh-getstarted-sg`— il s'agit d'un groupe de sécurité permettant l'accès entrant SSH

Créez un cluster dans AWS PCS

Dans AWS PCS, un cluster est une ressource persistante permettant de gérer les ressources et d'exécuter les charges de travail. Vous créez un cluster pour un planificateur spécifique (AWS PCSactuellement compatible avec Slurm) dans un sous-réseau d'un nouveau ou d'un existant. VPC Le cluster accepte et planifie les tâches, et lance également les nœuds de calcul (EC2instances) qui traitent ces tâches.

Pour créer votre cluster

1. Ouvrez la AWS PCSconsole et choisissez Create cluster.
2. Dans la section Configuration du cluster, entrez les champs suivants :
   - Nom du cluster — Entrez `get-started`
   - Taille du contrôleur — Sélectionnez Petit
3. Dans la section Mise en réseau, sélectionnez des valeurs pour les champs suivants :

- VPC— Choisissez le VPC nom `hpc-networking:Large-Scale-HPC`
- Sous-réseau — Sélectionnez le sous-réseau dont le nom commence par `hpc-networking:PrivateSubnetA`
- Groupes de sécurité : sélectionnez le groupe de sécurité du cluster nommé `cluster-getstarted-sg`

4. Choisissez Créer un cluster.

> **Note**
>
> Le champ État indique Création pendant le provisionnement du cluster. La création d'un cluster peut prendre plusieurs minutes.

Création d'un espace de stockage partagé pour AWS PCS Amazon Elastic File System

Amazon Elastic File System (AmazonEFS) est un AWS service qui fournit un stockage de fichiers entièrement élastique sans serveur afin que vous puissiez partager des données de fichiers sans provisionner ni gérer la capacité et les performances de stockage. Pour plus d'informations, consultez Qu'est-ce qu'Amazon Elastic File System ? dans le guide de l'utilisateur d'Amazon Elastic File System.

Le cluster de AWS PCS démonstration utilise un système de EFS fichiers pour fournir un répertoire de base partagé entre les nœuds du cluster. Créez un système de EFS fichiers identique VPC à celui de votre cluster.

Pour créer votre système de EFS fichiers Amazon

1. Accédez à la EFSconsole Amazon.
2. Assurez-vous qu'il est réglé sur le même point que celui Région AWS où vous allez essayer AWS PCS.
3. Choisissez Create file system (Créer un système de fichiers).
4. Sur la page Créer un système de fichiers, définissez les paramètres suivants :

   - Pour Nom, saisissez `getstarted-efs`

- Sous Virtual Private Cloud (VPC), choisissez le VPC nom `hpc-networking:Large-Scale-HPC`
- Sélectionnez Create (Créer). Cela vous renvoie à la page Systèmes de fichiers.

5. Notez l'ID du système de fichiers du système de `getstarted-efs` fichiers. Vous aurez besoin de ces informations ultérieurement.

Création d'un espace de stockage partagé pour AWS PCS Amazon FSx for Lustre

Amazon FSx for Lustre permet de lancer et d'exécuter facilement et à moindre coût le système de fichiers Lustre, populaire et performant. Vous utilisez Lustre pour les charges de travail où la rapidité est importante, telles que l'apprentissage automatique, le calcul haute performance (HPC), le traitement vidéo et la modélisation financière. Pour plus d'informations, consultez Qu'est-ce qu'Amazon FSx pour Lustre ? dans le guide de l'utilisateur d'Amazon FSx for Lustre.

Le cluster de AWS PCS démonstration peut utiliser un système de fichiers FSx for Lustre pour fournir un répertoire partagé performant entre les nœuds du cluster. Créez un système de fichiers FSx pour Lustre identique VPC à celui de votre cluster.

Pour créer votre système de fichiers FSx for Lustre

1. Accédez à la FSxconsole Amazon.
2. Assurez-vous que la console est configurée pour utiliser la même chose Région AWS que votre cluster.
3. Choisissez Create file system (Créer un système de fichiers).
   - Pour Sélectionner le type de système de fichiers, choisissez Amazon FSx pour Lustre, puis Next.
4. Sur la page Spécifier les détails du système de fichiers, définissez les paramètres suivants :
   - Sous Détails du système de fichiers
     - Pour Nom, saisissez `getstarted-fsx`
     - Pour le type de déploiement et de stockage, choisissez Persistent, SSD
     - Pour le débit par unité de stockage, choisissez 125 Mo/s/TiB
     - Pour Capacité de stockage, entrez 1,2 TiB

    - Pour la configuration des métadonnées, choisissez Automatique
    - Pour le type de compression des données, sélectionnez LZ4
  - Sous Réseau et sécurité
    - Pour Virtual Private Cloud (VPC), choisissez le VPC nom `hpc-networking:Large-Scale-HPC`
    - Pour les groupes VPC de sécurité, laissez le groupe de sécurité nommé `default`
    - Pour Sous-réseau, choisissez le sous-réseau dont le nom commence par `hpc-networking:PrivateSubnetA`
  - Conservez les valeurs par défaut des autres options.
  - Choisissez Suivant.
5. Sur la page Réviser et créer, choisissez Créer un système de fichiers. Cela vous renvoie à la page Systèmes de fichiers.
6. Accédez à la page de détails du système de fichiers FSx for Lustre que vous avez créé.
7. Notez l'ID du système de fichiers et le nom du montage. Vous aurez besoin de ces informations ultérieurement.

> **Note**
>
> Le champ État indique Création pendant le provisionnement du système de fichiers. La création du système de fichiers peut prendre plusieurs minutes. Attendez qu'il soit terminé avant de poursuivre le reste du didacticiel.

Créez des groupes de nœuds de calcul dans AWS PCS

Un groupe de nœuds de calcul est un ensemble virtuel de nœuds de calcul (EC2instances) qui AWS PCS se lance et gère. Lorsque vous définissez un groupe de nœuds de calcul, vous spécifiez des caractéristiques communes telles que les types d'EC2instances, le nombre d'instances minimal et maximal, les VPC sous-réseaux cibles, l'option d'achat préférée et la configuration de lancement personnalisée. AWS PCSlance, gère et arrête efficacement les nœuds de calcul d'un groupe de nœuds de calcul, conformément à ces paramètres. Le cluster de démonstration utilise un groupe de nœuds de calcul pour fournir des nœuds de connexion pour l'accès des utilisateurs, et un groupe de nœuds de calcul distinct pour traiter les tâches. Les rubriques suivantes décrivent les procédures permettant de configurer ces groupes de nœuds de calcul dans votre cluster.

Rubriques

Créez un profil d'instance pour AWS PCS

Les groupes de nœuds de calcul nécessitent un profil d'instance lors de leur création. Si vous utilisez le AWS Management Console pour créer un rôle pour AmazonEC2, la console crée automatiquement un profil d'instance et lui donne le même nom que le rôle. Pour plus d'informations, consultez la section Utilisation des profils d'instance dans le Guide de AWS Identity and Access Management l'utilisateur.

Dans la procédure suivante, vous utiliserez le AWS Management Console pour créer un rôle pour AmazonEC2, qui crée également le profil d'instance pour vos groupes de nœuds de calcul.

Pour créer le profil de rôle et d'instance

- Accédez à la console IAM.
- Sous Access Management (Gestion des accès), choisissez Policies (politiques).
  - Choisissez Create Policy (Créer une politique).
  - Sous Spécifier les autorisations, dans Éditeur de politiques, sélectionnez JSON.
  - Remplacez le contenu de l'éditeur de texte par le suivant :

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

    - Choisissez Suivant.
    - Sous Vérifier et créer, dans le champ Nom de la politique, entrez`AWSPCS-getstarted-policy`.
    - Choisissez Create Policy (Créer une politique).
- Sous Access Management (Gestion des accès), choisissez Roles (Rôles).
- Sélectionnez Créer un rôle.
- Sous Sélectionner une entité de confiance :
    - Pour le type d'entité de confiance, sélectionnez AWS service
    - Sous Cas d'utilisation, sélectionnez EC2.
        - Ensuite, sous Choisir un cas d'utilisation pour le service spécifié, sélectionnez EC2.
    - Choisissez Suivant.
- Sous Ajouter des autorisations :
    - Dans Politiques d'autorisations, recherchez AWSPCS-getstarted-policy.
    - Cochez la case à côté de AWSPCS-getstarted-policy pour l'ajouter au rôle.
    - Dans Politiques d'autorisations, recherchez mazonSSMManagedInstanceCoreA.
    - Cochez la case à côté mazonSSMManaged InstanceCore de A pour l'ajouter au rôle.
    - Choisissez Suivant.
- Sous Nom, passez en revue et créez :
    - Sous Détails du rôle :
        - Pour le Nom du rôle, saisissez `AWSPCS-getstarted-role`.
    - Sélectionnez Créer un rôle.

Créez des modèles de lancement pour AWS PCS

Lorsque vous créez un groupe de nœuds de calcul, vous fournissez un modèle de EC2 lancement qui permet AWS PCS de configurer les EC2 instances qu'il lance. Cela inclut les paramètres tels que les groupes de sécurité et les scripts qui s'exécutent au lancement de l'instance.

Au cours de cette étape, un CloudFormation modèle sera utilisé pour créer deux modèles de EC2 lancement. Un modèle sera utilisé pour créer des nœuds de connexion et l'autre pour créer des nœuds de calcul. La principale différence entre eux est que les nœuds de connexion peuvent être configurés pour autoriser l'SSHaccès entrant.

Accédez au CloudFormation modèle

Utilisez ce qui suit URL pour télécharger le CloudFormation modèle, puis téléchargez-le dans la AWS CloudFormation console pour créer une nouvelle CloudFormation pile. Pour plus d'informations, consultez la section Utilisation de la AWS CloudFormation console dans le guide de AWS CloudFormation l'utilisateur.

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

Utiliser le CloudFormation modèle pour créer des modèles de EC2 lancement

Utilisez la procédure suivante pour compléter le CloudFormation modèle dans la AWS CloudFormation console

- Sous Fournir un nom de pile :
  - Sous Nom de la pile, entrez`getstarted-lt`.
- Sous Paramètres :
  - Sous Sécurité
    - Pour VpcSecurityGroupId, sélectionnez le groupe de sécurité nommé `default` dans votre clusterVPC.
    - Pour ClusterSecurityGroupId, sélectionnez le groupe nommé `cluster-getstarted-sg`
    - Pour SshSecurityGroupId, sélectionnez le groupe nommé `inbound-ssh-getstarted-sg`
    - Pour SshKeyName, sélectionnez votre paire de SSH clés préférée.
  - Sous Systèmes de fichiers
    - Pour EfsFilesystemId, entrez l'ID du système de fichiers à partir du système de EFS fichiers que vous avez créé plus tôt dans le didacticiel.
    - Pour FSxLustreFilesystemId, entrez l'ID du système de fichiers à partir du système de fichiers FSx for Lustre que vous avez créé plus tôt dans le didacticiel.
    - Pour FSxLustreFilesystemMountName, entrez le même FSx nom de montage pour le système de fichiers Lustre.
- Choisissez Next, puis de nouveau Next.
- Sélectionnez Envoyer.

Surveillez l'état de la CloudFormation pile. Lorsqu'il atteint, `CREATE_COMPLETE` le modèle de lancement est prêt à être utilisé.

> **Note**
>
> Pour voir toutes les ressources créées par le CloudFormation modèle, ouvrez la AWS CloudFormation console. Choisissez la pile `getstarted-lt`, puis choisissez l'onglet Ressources.

Créez un groupe de nœuds de calcul pour les nœuds de connexion dans AWS PCS

Un groupe de nœuds de calcul est un ensemble virtuel de nœuds de calcul (EC2instances) qui AWS PCS se lance et gère. Lorsque vous définissez un groupe de nœuds de calcul, vous spécifiez des caractéristiques communes telles que les types d'EC2instances, le nombre d'instances minimal et maximal, les VPC sous-réseaux cibles, l'option d'achat préférée et la configuration de lancement personnalisée. AWS PCSlance, gère et arrête efficacement les nœuds de calcul d'un groupe de nœuds de calcul, conformément à ces paramètres.

Au cours de cette étape, vous allez lancer un groupe de nœuds de calcul statique qui fournit un accès interactif au cluster. Vous pouvez utiliser SSH Amazon EC2 Systems Manager (SSM) pour vous y connecter, puis exécuter des commandes shell et gérer les tâches Slurm.

Pour créer le groupe de nœuds de calcul

- Ouvrez la AWS PCSconsole et accédez à Clusters.
- Sélectionnez le cluster nommé `get-started`
- Accédez à Compute node groups et choisissez Create.
- Dans la section Configuration du groupe de nœuds de calcul, fournissez les informations suivantes :
  - Nom du groupe de nœuds de calcul — Entrez`login`.
- Sous Configuration informatique, entrez ou sélectionnez les valeurs suivantes :
  - EC2modèle de lancement — Choisissez le modèle de lancement dont le nom est `login-getstarted-lt`
  - IAMprofil d'instance — Choisissez le profil d'instance nommé `AWSPCS-getstarted-role`

  - Sous-réseaux : sélectionnez le sous-réseau dont le nom commence par. `hpc-networking:PublicSubnetA`
  - Instances — Sélectionnez`c6i.xlarge`.
  - Configuration de mise à l'échelle : pour le nombre minimal d'instances, entrez**1**. Pour Nombre maximal d'instances, entrez**1**.
- Sous Paramètres supplémentaires, spécifiez les éléments suivants :
  - AMIID — Sélectionnez l'AMIendroit où le nom commence par `aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11`
- Choisissez Créer un groupe de nœuds de calcul.

Le champ Status indique Création pendant le provisionnement du groupe de nœuds de calcul. Vous pouvez passer à l'étape suivante du didacticiel pendant qu'il est en cours.

Créez un groupe de nœuds de calcul pour exécuter des tâches de calcul dans AWS PCS

Au cours de cette étape, vous allez lancer un groupe de nœuds de calcul qui évolue de manière élastique pour exécuter les tâches soumises au cluster.

Pour créer le groupe de nœuds de calcul

- Ouvrez la AWS PCSconsole et accédez à Clusters.
- Sélectionnez le cluster nommé `get-started`
- Accédez à Compute node groups et choisissez Create.
- Dans la section Configuration du groupe de nœuds de calcul, fournissez les informations suivantes :
  - Nom du groupe de nœuds de calcul — Entrez`compute-1`.
- Sous Configuration informatique, entrez ou sélectionnez les valeurs suivantes :
  - EC2modèle de lancement — Choisissez le modèle de lancement dont le nom est `compute-getstarted-lt`
  - IAMprofil d'instance — Choisissez le profil d'instance nommé `AWSPCS-getstarted-role`
  - Sous-réseaux : sélectionnez le sous-réseau dont le nom commence par. `hpc-networking:PrivateSubnetA`
  - Instances — Sélectionnez`c6i.xlarge`.

- Configuration de mise à l'échelle : pour le nombre minimal d'instances, entrez**0**. Pour Nombre maximal d'instances, entrez**4**.
- Sous Paramètres supplémentaires, spécifiez les éléments suivants :
  - AMIID — Sélectionnez le AMI point de départ du nom`aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11`.
- Choisissez Créer un groupe de nœuds de calcul.

Le champ Status indique Création pendant le provisionnement du groupe de nœuds de calcul.

> **Important**
>
> Attendez que le champ État indique Actif avant de passer à l'étape suivante de ce didacticiel.

Créez une file d'attente pour gérer les tâches dans AWS PCS

Vous soumettez une tâche à une file d'attente pour l'exécuter. La tâche reste dans la file d'attente jusqu'à ce AWS PCS qu'elle soit planifiée pour s'exécuter sur un groupe de nœuds de calcul. Chaque file d'attente est associée à un ou plusieurs groupes de nœuds de calcul, qui fournissent les EC2 instances nécessaires pour effectuer le traitement.

Au cours de cette étape, vous allez créer une file d'attente qui utilise le groupe de nœuds de calcul pour traiter les tâches.

Pour créer une file d'attente

- Ouvrez la AWS PCSconsole.
- Sélectionnez le cluster nommé`get-started`.
- Accédez à Calculer les groupes de nœuds et assurez-vous que le statut du `compute-1` groupe est Actif.

> **Important**
>
> Le statut du `compute-1` groupe doit être Actif avant de passer à l'étape suivante.

- Accédez à Files d'attente, puis choisissez Créer une file d'attente.

- Dans la section Configuration de la file d'attente, indiquez les valeurs suivantes :
  - Nom de la file d'attente — Entrez ce qui suit : `demo`
  - Groupes de nœuds de calcul : sélectionnez le groupe de nœuds de calcul nommé`compute-1`.
- Choisissez Créez une file d'attente.

Le champ État indique Création pendant la création de la file d'attente.

> **Important**
>
> Attendez que le champ État indique Actif avant de passer à l'étape suivante de ce didacticiel.

Connectez-vous à votre AWS PCS cluster

Lorsque le statut du groupe de nœuds de `login` calcul devient actif, vous pouvez vous connecter à l'EC2instance qu'il a créée.

Pour vous connecter au nœud de connexion

- Ouvrez la AWS PCSconsole et accédez à Clusters.
- Sélectionnez le cluster nommé`get-started`.
- Choisissez Compute node groups.
- Accédez au groupe de nœuds de calcul nommé`login`.
- Trouvez l'ID du groupe de nœuds Compute.
- Dans une autre fenêtre ou un autre onglet du navigateur, ouvrez la EC2console Amazon.
  - Choisissez Instances.
  - Recherchez des EC2 instances avec la balise suivante. Remplacez *`node-group-id`* avec la valeur de l'ID du groupe de nœuds de calcul de l'étape précédente. Il devrait y avoir une instance.

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

  - Connectez-vous à l'EC2instance. Vous pouvez utiliser le gestionnaire de session ouSSH.

    Session Manager

    - Sélectionnez l'instance.

- Choisissez Se connecter.
- Sous Connect to instance, sélectionnez Session Manager.
- Choisissez Se connecter.
- Choisissez Se connecter. Une borne interactive s'ouvre dans votre navigateur.

SSH

- Sélectionnez l'instance.
- Choisissez Se connecter.
- Sous Connect to instance, sélectionnez SSHclient.
- Suivez les instructions fournies par la console.

> **Note**
>
> Le nom d'utilisateur de l'instance ne l'est **`ec2-user`**pas`root`.

Explorez l'environnement du cluster dans AWS PCS

Une fois connecté au cluster, vous pouvez exécuter des commandes shell. Par exemple, vous pouvez changer d'utilisateur, travailler avec des données sur des systèmes de fichiers partagés et interagir avec Slurm.

Changer d'utilisateur

Si vous vous êtes connecté au cluster à l'aide du gestionnaire de session, vous êtes peut-être connecté en tant que`ssm-user`. Il s'agit d'un utilisateur spécial créé pour le gestionnaire de session. Passez à l'utilisateur par défaut sur Amazon Linux 2 à l'aide de la commande suivante. Vous n'aurez pas besoin de le faire si vous vous êtes connecté en utilisantSSH.

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

Travailler avec des systèmes de fichiers partagés

Vous pouvez vérifier que le système de EFS fichiers et FSx pour les systèmes de fichiers Lustre sont disponibles à l'aide de la commande`df -h`. La sortie de votre cluster doit ressembler à ce qui suit :

```
[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
```

Le système de `/home` fichiers monte 127.0.0.1 et possède une très grande capacité. Il s'agit du système de EFS fichiers que vous avez créé plus tôt dans le didacticiel. Tous les fichiers écrits ici seront disponibles `/home` sur tous les nœuds du cluster.

Le système de `/shared` fichiers monte une adresse IP privée et a une capacité de 1,2 To. Il s'agit du système de fichiers FSx for Lustre que vous avez créé plus tôt dans le didacticiel. Tous les fichiers écrits ici seront disponibles `/shared` sur tous les nœuds du cluster.

Interagir avec Slurm

Rubriques

- Répertorier les files d'attente et les nœuds
- Afficher les offres d'emploi

Répertorier les files d'attente et les nœuds

Vous pouvez répertorier les files d'attente et les nœuds auxquels elles sont associées à l'aide`sinfo`. La sortie de votre cluster doit ressembler à ce qui suit :

```
[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 ~]$
```

Notez le nom de la partition`demo`. Son statut est `up` et il dispose d'un maximum de 4 nœuds. Il est associé aux nœuds du groupe de `compute-1` nœuds. Si vous modifiez le groupe de nœuds de calcul et augmentez le nombre maximum d'instances à 8, le nombre de nœuds sera lu 8 et la liste des nœuds sera lu`compute-1-[1-8]`. Si vous avez créé un deuxième groupe de nœuds de

calcul nommé `test` avec 4 nœuds et que vous l'avez ajouté à la `demo` file d'attente, ces nœuds apparaîtront également dans la liste des nœuds.

Afficher les offres d'emploi

Vous pouvez répertorier toutes les tâches du système, quel que soit leur état, avec`squeue`. La sortie de votre cluster doit ressembler à ce qui suit :

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

Réessayez de l'exécuter `squeue` ultérieurement, lorsqu'une tâche Slurm est en attente ou en cours d'exécution.

Exécuter une tâche à nœud unique dans AWS PCS

Pour exécuter une tâche à l'aide de Slurm, vous devez préparer un script de soumission spécifiant les exigences de la tâche et le soumettre à une file d'attente avec la `sbatch` commande. Généralement, cela se fait à partir d'un répertoire partagé, de sorte que les nœuds de connexion et de calcul disposent d'un espace commun pour accéder aux fichiers.

Connectez-vous au nœud de connexion de votre cluster et exécutez les commandes suivantes à l'invite du shell.

- Devenez l'utilisateur par défaut. Accédez au répertoire partagé.

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

- Utilisez les commandes suivantes pour créer un exemple de script de tâche :

```
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
```

- Soumettez le script de tâche au planificateur Slurm :

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

- Lorsque la tâche est soumise, elle renvoie un identifiant de tâche sous forme de numéro. Utilisez cet identifiant pour vérifier le statut du travail. Remplacez *job-id* dans la commande suivante avec le numéro renvoyé par`sbatch`.

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

  Example

```
squeue --job 1
```

  La `squeue` commande renvoie un résultat similaire à ce qui suit :

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

- Continuez à vérifier l'état de la tâche jusqu'à ce qu'elle atteigne le statut R (en cours). Le travail est terminé quand il `squeue` ne renvoie rien.
- Inspectez le contenu du `/shared` répertoire.

```
ls -alth /shared
```

  Le résultat de la commande est similaire à ce qui suit :

```
-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
```

  Les fichiers `single.1.err` ont été nommés `single.1.out` et écrits par l'un des nœuds de calcul de votre cluster. La tâche ayant été exécutée dans un répertoire partagé (`/shared`), elles sont également disponibles sur votre nœud de connexion. C'est pourquoi vous avez configuré un système de fichiers FSx for Lustre pour ce cluster.

- Inspectez le contenu du `single.1.out` fichier.

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

La sortie est similaire à ce qui suit :

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

Exécuter une MPI tâche multi-nœuds avec Slurm dans AWS PCS

Ces instructions montrent comment utiliser Slurm pour exécuter une tâche interface (MPI) en transmettant un message. AWS PCS

Exécutez les commandes suivantes à l'invite du shell de votre nœud de connexion.

- Devenez l'utilisateur par défaut. Accédez à son répertoire personnel.

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

- Créez du code source dans le langage de programmation 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
```

- Chargez le MPI module Open.

```
module load openmpi
```

- Compilez le programme C.

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

- Rédigez un script de soumission de tâches 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
```

- Accédez au répertoire partagé.

```
cd /shared
```

- Soumettez le script de tâche.

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

- `squeue`À utiliser pour surveiller le travail jusqu'à ce qu'il soit terminé.
- Vérifiez le contenu de `multi.out` :

```
cat multi.out
```

La sortie est similaire à ce qui suit. Notez que chaque rang possède sa propre adresse IP car il s'est exécuté sur un nœud différent.

```
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
```

Supprimez vos AWS ressources pour AWS PCS

Une fois que vous avez terminé avec les groupes de clusters et de nœuds que vous avez créés pour ce didacticiel, vous devez supprimer les ressources que vous avez créées.

> **Important**
>
> Vous recevez des frais de facturation pour toutes les ressources utilisées dans votre Compte AWS

Pour supprimer les AWS PCS ressources que vous avez créées pour ce didacticiel

- Ouvrez la AWS PCSconsole.
- Accédez au cluster nommé get-started.
- Accédez à la section Files d'attente.
- Sélectionnez la file d'attente nommée demo.
- Sélectionnez Delete (Supprimer).

> **Important**
>
> Attendez que la file d'attente soit supprimée avant de continuer.

- Accédez à la section Groupes de nœuds de calcul.
- Sélectionnez le groupe de nœuds de calcul nommé compute-1.
- Sélectionnez Delete (Supprimer).
- Sélectionnez le groupe de nœuds de calcul nommé login.
- Sélectionnez Delete (Supprimer).

> **Important**
>
> Attendez que les deux groupes de nœuds de calcul aient été supprimés avant de continuer.

- Sur la page détaillée du cluster pour démarrer, choisissez Supprimer.

> **Important**
>
> Attendez que le cluster ait été supprimé avant de passer aux étapes suivantes.

Pour supprimer les autres AWS ressources que vous avez créées pour ce didacticiel

- Ouvrez la IAMconsole.
  - Sélectionnez Roles (Rôles).
  - Sélectionnez le rôle nommé AWSPCS-getstarted-role, puis choisissez Supprimer.
  - Une fois le rôle supprimé, choisissez Politiques.
  - Sélectionnez la politique nommée AWSPCS-getstarted-policy, puis choisissez Supprimer.
- Ouvrez la AWS CloudFormation console.
  - Sélectionnez la pile nommée getstarted-lt.
  - Sélectionnez Delete (Supprimer).

> **Important**
>
> Attendez que la pile soit supprimée avant de continuer.

- Ouvrez la EFSconsole Amazon.
  - Choisissez File Systems (Systèmes de fichiers).
  - Sélectionnez le système de fichiers nommé getstarted-efs.
  - Sélectionnez Delete (Supprimer).

> **Important**
>
> Attendez que le système de fichiers soit supprimé avant de continuer.

- Ouvrez la FSxconsole Amazon.
  - Choisissez File Systems (Systèmes de fichiers).
  - Sélectionnez le système de fichiers nommé getstarted-fsx.
  - Sélectionnez Delete (Supprimer).

> **Important**
>
> Attendez que le système de fichiers soit supprimé avant de continuer.

- Ouvrez la AWS CloudFormation console.
  - Sélectionnez la pile nommée getstarted-sg.
  - Sélectionnez Delete (Supprimer).
- Ouvrez la AWS CloudFormation console.
  - Sélectionnez la pile nommée hpc-networking.
  - Sélectionnez Supprimer.

Travailler avec AWS PCS

Ce chapitre fournit des informations et des conseils pour vous aider à utiliser AWS PCS.

Rubriques

AWS PCSclusters

Un AWS PCS cluster comprend les composants suivants :

- Instances gérées du logiciel de planification HPC du système, telles que le daemon de contrôle Slurm (). `slurmctld`
- Composants qui s'intègrent au planificateur du HPC système pour approvisionner et gérer les instances AmazonEC2.
- Composants qui s'intègrent au planificateur du HPC système pour transmettre les journaux et les métriques à Amazon. CloudWatch

Ces composants s'exécutent dans un compte géré par AWS. Ils travaillent ensemble pour gérer les EC2 instances Amazon de votre compte client. AWS PCSfournit des interfaces réseau élastiques dans votre VPC sous-réseau Amazon pour fournir une connectivité entre le logiciel de planification et les EC2 instances Amazon (par exemple, pour prendre en charge la planification de tâches par lots sur celles-ci et permettre aux utilisateurs d'exécuter des commandes du planificateur pour répertorier et gérer ces tâches).

Rubriques

Création d'un cluster dans AWS Parallel Computing Service

Cette rubrique fournit une vue d'ensemble des options disponibles et décrit les éléments à prendre en compte lors de la création d'un cluster dans AWS Parallel Computing Service (AWS PCS). Si c'est la première fois que vous créez un AWS PCS cluster, nous vous recommandons de suivreCommencer avec AWS PCS. Le didacticiel peut vous aider à créer un HPC système fonctionnel sans étendre toutes les options disponibles et les architectures système possibles.

Prérequis

- Un sous-réseau existant VPC qui répond aux AWS PCSRéseautage exigences. Avant de déployer un cluster à des fins de production, nous vous recommandons de bien connaître les exigences en matière de sous-réseau VPC et de sous-réseau. Pour créer un sous-réseau VPC et, consultezCréation d'un VPC pour votre AWS PCS cluster.
- Un IAMdirecteur autorisé à créer et à gérer AWS PCS des ressources. Pour de plus amples informations, veuillez consulter Identity and Access Management pour le service de calcul AWS parallèle.

Création d'un AWS PCS cluster

Vous pouvez utiliser le AWS Management Console ou AWS CLI pour créer un cluster.

AWS Management Console

Pour créer un cluster

1. Ouvrez la AWS PCS console à l'adresse https://console.aws.amazon.com/pcs/home#/clusters et choisissez Create cluster.
2. Dans la section Configuration du cluster, entrez les champs suivants :
   - Nom du cluster : nom de votre cluster. Un nom ne peut contenir que des caractères alphanumériques (sensibles à la casse) et des traits d'union. Il doit commencer par un

caractère alphabétique et ne doit pas comporter plus de 40 caractères. Le nom doit être unique dans le Région AWS et dans Compte AWS lequel vous créez le cluster.

- Planificateur : choisissez un planificateur et une version. AWS PCSsupporte actuellement Slurm 23.11. Pour de plus amples informations, veuillez consulter Versions Slurm en AWS PCS.
- Taille de la manette — Choisissez une taille pour votre manette. Cela détermine le nombre de tâches et de nœuds de calcul simultanés pouvant être gérés par le AWS PCS cluster. Vous ne pouvez définir la taille du contrôleur que lorsque le cluster est créé. Pour plus d'informations sur le dimensionnement, voirChoix d'une taille de AWS PCS cluster.

3. Dans la section Mise en réseau, sélectionnez des valeurs pour les champs suivants :

   - VPC— Choisissez un existant VPC qui répond aux AWS PCS exigences. Pour de plus amples informations, veuillez consulter AWS PCSVPCexigences et considérations relatives aux sous-réseaux. Après avoir créé le cluster, vous ne pouvez pas le modifierVPC. Si aucun VPCs n'est répertorié, vous devez d'abord en créer un.
   - Sous-réseau : tous les sous-réseaux disponibles dans le sous-réseau sélectionné VPC sont répertoriés. Choisissez-en deux dans différentes zones de disponibilité. Chaque sous-réseau doit répondre aux exigences du AWS PCS sous-réseau. Pour de plus amples informations, veuillez consulter AWS PCSVPCexigences et considérations relatives aux sous-réseaux. Nous vous recommandons de sélectionner un sous-réseau privé pour éviter d'exposer les points de terminaison de votre planificateur à l'Internet public.
   - Groupes de sécurité : spécifiez le ou les groupes de sécurité que vous AWS PCS souhaitez associer aux interfaces réseau qu'il crée pour votre cluster. Vous devez sélectionner au moins un groupe de sécurité qui autorise la communication entre votre cluster et ses nœuds de calcul. Pour de plus amples informations, veuillez consulter Exigences et considérations relatives aux groupes de sécurité.

4. (Facultatif) Sous Chiffrement, vous pouvez définir une clé personnalisée pour chiffrer les données de votre contrôleur en définissant les champs suivants :

   - KMSkey ID — Laissez comme si `aws/pcs` vous utilisiez la KMS clé qui PCS crée. Sélectionnez un alias de KMS clé existant pour utiliser une KMS clé personnalisée. Notez que le compte utilisé pour créer le cluster doit disposer de `kms:Decrypt` privilèges sur la KMS clé personnalisée.

5. (Facultatif) Dans la section Configuration de Slurm, vous pouvez spécifier les options de configuration de Slurm qui remplacent les valeurs par défaut définies par : AWS PCS

- Diminution du temps d'inactivité : cela permet de contrôler la durée pendant laquelle les nœuds de calcul provisionnés dynamiquement restent actifs après la fin ou la fin des tâches qui leur ont été confiées. Si vous définissez cette valeur sur une valeur plus longue, il est plus probable qu'une tâche ultérieure puisse être exécutée sur le nœud, mais cela peut entraîner une augmentation des coûts. Une valeur plus courte réduira les coûts, mais peut augmenter la proportion de temps que votre HPC système passe à provisionner des nœuds par rapport à l'exécution de tâches sur ceux-ci.
- Prolog — Il s'agit d'un chemin complet vers un répertoire de scripts prolog sur les instances de votre groupe de nœuds de calcul. Cela correspond au paramètre Prolog dans Slurm. Notez qu'il doit s'agir d'un répertoire et non d'un chemin d'accès à un exécutable spécifique.
- Epilog : il s'agit d'un chemin complet vers un répertoire de scripts epilog sur les instances de votre groupe de nœuds de calcul. Cela correspond au paramètre Epilog dans Slurm. Notez qu'il doit s'agir d'un répertoire et non d'un chemin d'accès à un exécutable spécifique.
- Paramètres du type de sélection : cela permet de contrôler l'algorithme de sélection des ressources utilisé par Slurm. Le fait de définir cette valeur sur `CR_CPU_Memory` activera la planification basée sur la mémoire, tandis que la définition sur `CR_CPU` CPU activera uniquement la planification. Ce paramètre correspond au SelectTypeParametersréglage dans Slurm où `SelectType` est défini sur by. `select/cons_tres` AWS PCS

6. (Facultatif) Sous Balises, ajoutez des balises à votre AWS PCS cluster.
7. Choisissez Créer un cluster. Le champ Status s'affiche `Creating` lors de AWS PCS la création du cluster. Ce processus peut prendre plusieurs minutes.

> **Important**
>
> Il ne peut y avoir qu'un seul cluster Région AWS par `Creating` état Compte AWS. AWS PCSrenvoie une erreur s'il existe déjà un cluster dans un `Creating` état lorsque vous essayez de créer un cluster.

AWS CLI

Pour créer un cluster

1. Créez votre cluster à l'aide de la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :

   - Remplacez *region* avec l'ID du cluster dans Région AWS lequel vous souhaitez créer votre cluster, tel que `us-east-1`.
   - Remplacez *my-cluster* avec un nom pour votre cluster. Un nom ne peut contenir que des caractères alphanumériques (sensibles à la casse) et des traits d'union. Il doit commencer par un caractère alphabétique et ne doit pas comporter plus de 40 caractères. Le nom doit être unique dans le cluster Région AWS et dans Compte AWS lequel vous créez le cluster.
   - Remplacez *23.11* avec n'importe quelle version prise en charge de Slurm.

     > **Note**
     >
     > AWS PCSsupporte actuellement Slurm 23.11.

   - Remplacez *SMALL* avec n'importe quelle taille de cluster prise en charge. Cela détermine le nombre de tâches et de nœuds de calcul simultanés pouvant être gérés par le AWS PCS cluster. Il ne peut être défini que lors de la création du cluster. Pour plus d'informations sur le dimensionnement, voirChoix d'une taille de AWS PCS cluster.
   - Remplacez la valeur de `subnetIds` par la vôtre. Nous vous recommandons de sélectionner un sous-réseau privé pour éviter d'exposer les points de terminaison de votre planificateur à l'Internet public.
   - Spécifiez `securityGroupIds` celles que vous AWS PCS souhaitez associer aux interfaces réseau qu'il crée pour votre cluster. Les groupes de sécurité doivent être identiques à VPC ceux du cluster. Vous devez sélectionner au moins un groupe de sécurité qui autorise la communication entre votre cluster et ses nœuds de calcul. Pour de plus amples informations, veuillez consulter Exigences et considérations relatives aux groupes de sécurité.
   - Vous pouvez éventuellement affiner le comportement de Slurm en ajoutant une option. `--slurm-configration` Par exemple, vous pouvez réduire le temps d'inactivité à 60 minutes (3 600 secondes) avec. `--slurm configuration scaleDownIdeTime=3600`

- Vous pouvez éventuellement fournir une KMS clé personnalisée pour chiffrer les données de votre contrôleur à l'aide `--kms-key-id` *kms-key* de. *kms-key*Remplacez-le par un identifiant de clé ou un alias existant KMSARN. Notez que le compte utilisé pour créer le cluster doit disposer de `kms:Decrypt` privilèges sur la KMS clé personnalisée.

```
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. Le provisionnement du cluster peut prendre plusieurs minutes. Vous pouvez vérifier le statut de votre cluster avec la commande suivante. Ne créez pas de files d'attente ou de groupes de nœuds de calcul tant que le champ d'état du cluster n'est `ACTIVE` pas indiqué.

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

> **Important**
>
> Il ne peut y avoir qu'un seul cluster Région AWS par `Creating` état Compte AWS. AWS PCSrenvoie une erreur s'il existe déjà un cluster dans un `Creating` état lorsque vous essayez de créer un cluster.

Prochaines étapes recommandées pour votre cluster

- Ajoutez des groupes de nœuds de calcul.
- Ajoutez des files d'attente.
- Activez la journalisation

Supprimer un cluster dans AWS PCS

Cette rubrique fournit une vue d'ensemble de la procédure de suppression d'un AWS PCS cluster.

Considérations relatives à la suppression d'un AWS PCS cluster

- Toutes les files d'attente associées au cluster doivent être supprimées pour que le cluster puisse être supprimé. Pour de plus amples informations, veuillez consulter Supprimer une file d'attente dans AWS PCS.
- Tous les groupes de nœuds de calcul associés au cluster doivent être supprimés pour que le cluster puisse être supprimé. Pour de plus amples informations, veuillez consulter Suppression d'un groupe de nœuds de calcul dans AWS PCS.

Supprimer le cluster

Vous pouvez utiliser le AWS Management Console ou AWS CLI pour supprimer un cluster.

AWS Management Console

Pour supprimer un cluster

1. Ouvrez la AWS PCSconsole.
2. Sélectionnez le cluster à supprimer.
3. Sélectionnez Delete (Supprimer).
4. Le champ État du cluster s'afficheDeleting. Cela peut prendre plusieurs minutes.

AWS CLI

Pour supprimer un cluster

1. Utilisez la commande suivante pour supprimer un cluster, avec les remplacements suivants :
   - Remplacez *region-code* avec le nom dans lequel se trouve Région AWS votre cluster.
   - Remplacez *my-cluster* avec le nom ou l'ID de votre cluster.

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

2. La suppression du cluster peut prendre plusieurs minutes. Vous pouvez vérifier l'état de votre cluster à l'aide de la commande suivante.

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

Choix d'une taille de AWS PCS cluster

AWS PCSfournit des clusters sécurisés et hautement disponibles, tout en automatisant les tâches clés telles que l'application de correctifs, le provisionnement des nœuds et les mises à jour.

Lorsque vous créez un cluster, vous sélectionnez sa taille en fonction de deux facteurs :

- Le nombre de nœuds de calcul qu'il gérera
- Le nombre de tâches actives et de tâches en file d'attente que vous comptez exécuter sur le cluster

Taille du cluster Slurm	Nombre d'instances gérées	Nombre de tâches actives et en attente
Petite	Jusqu'à 32	Jusqu'à 256
Medium	Jusqu'à 512	Jusqu'à 8192
Large	Jusqu'à 2048	Jusqu'à 16384

Exemples

- Si votre cluster doit comporter jusqu'à 24 instances gérées et exécuter jusqu'à 100 tâches, choisissez Small.
- Si votre cluster doit comporter jusqu'à 24 instances gérées et exécuter jusqu'à 1 000 tâches, choisissez Medium.
- Si votre cluster doit comporter jusqu'à 1 000 instances gérées et exécuter jusqu'à 100 tâches, choisissez Large.
- Si votre cluster doit comporter jusqu'à 1 000 instances gérées et exécuter jusqu'à 10 000 tâches, choisissez Large.

Utilisation des secrets de cluster dans AWS PCS

Dans le cadre de la création d'un cluster, AWS PCS crée un secret de cluster requis pour se connecter au planificateur de tâches du cluster. Vous créez également des groupes de nœuds de AWS PCS calcul, qui définissent des ensembles d'instances à lancer en réponse à des événements

de dimensionnement. AWS PCSconfigure les instances lancées par ces groupes de nœuds de calcul avec le secret du cluster afin qu'ils puissent se connecter au planificateur de tâches. Dans certains cas, vous souhaiterez peut-être configurer les clients Slurm manuellement. Les exemples incluent la création d'un nœud de connexion permanent ou la configuration d'un gestionnaire de flux de travail doté de fonctionnalités de gestion des tâches.

AWS PCSstocke le secret du cluster en tant que secret géré avec le préfixe `pcs!` in AWS Secrets Manager. Le coût du secret est inclus dans les frais d'utilisation AWS PCS.

> **Warning**
>
> Ne modifiez pas le secret de votre cluster. AWS PCSne pourra pas communiquer avec votre cluster si vous modifiez le secret de votre cluster. AWS PCSne prend pas en charge la rotation du secret du cluster. Vous devez créer un nouveau cluster si vous devez modifier le secret de votre cluster.

Table des matières

Trouvez le secret du cluster Slurm

Vous pouvez trouver des secrets AWS PCS gérés à l'aide de la AWS Secrets Manager consoleAPI, directement à partir de AWS PCS ou à l'aide de balises.

AWS Secrets Manager À utiliser pour trouver le secret du cluster

AWS Management Console

1. Accédez à la console Secrets Manager.
2. Choisissez Secrets, puis recherchez le `pcs!` préfixe.

> **Note**
>
> Un secret de AWS PCS cluster porte un nom sous la forme `pcs!slurm-secret-`*`cluster-id`* où se *`cluster-id`* trouve l'ID du AWS PCS cluster.

AWS CLI

Chaque secret de AWS PCS cluster est également étiqueté avec`aws:pcs:`*`cluster-id`*. Vous pouvez obtenir l'ID secret d'un cluster à l'aide de la commande suivante. Effectuez les substitutions suivantes avant d'exécuter la commande :

- *`region`*Remplacez-le par le Région AWS pour créer votre cluster dans, par exemple`us-east-1`.
- Remplacez *`cluster-id`* par l'ID du AWS PCS cluster pour lequel vous souhaitez trouver le secret du cluster.

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

AWS PCSÀ utiliser pour trouver le secret du cluster

Vous pouvez utiliser le AWS CLI pour rechercher le ARN secret d'un AWS PCS cluster. Entrez la commande suivante en effectuant les substitutions suivantes :

- *`region`*Remplacez-le par le Région AWS pour créer votre cluster dans, par exemple`us-east-1`.
- Remplacez *`my-cluster`* par le nom ou l'identifiant de votre cluster.

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

L'exemple de sortie suivant provient de la `get-cluster` commande. Vous pouvez utiliser `secretArn` et `secretVersion` ensemble pour obtenir le secret.

```
{
```

```
    "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"
    }
}
```

Obtenez le secret du cluster Slurm

Vous pouvez utiliser Secrets Manager pour obtenir la version actuelle codée en base64 d'un secret de cluster Slurm. L'exemple suivant utilise le. AWS CLI Effectuez les substitutions suivantes avant d'exécuter la commande.

- *`region`*Remplacez-le par le Région AWS pour créer votre cluster dans, par exemple`us-east-1`.
- *`secret-arn`*Remplacez-le par le `secretArn` provenant d'un AWS PCS cluster.

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

Pour plus d'informations sur l'utilisation du secret du cluster Slurm, consultez. Utilisation d'instances autonomes comme nœuds de AWS PCS connexion

Autorisations

Vous utilisez un IAM principal pour obtenir le secret du cluster Slurm. Le IAM directeur doit avoir l'autorisation de lire le secret. Pour plus d'informations, consultez la section Termes et concepts relatifs aux rôles dans le guide de AWS Identity and Access Management l'utilisateur.

L'exemple de IAM politique suivant autorise l'accès à un exemple de secret de cluster.

```
{
    "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 PCSgroupes de nœuds de calcul

Un groupe de nœuds de AWS PCS calcul est un ensemble logique de nœuds (EC2instances Amazon). Ces nœuds peuvent être utilisés pour exécuter des tâches informatiques, ainsi que pour fournir un accès interactif basé sur un shell à un HPC système. Un groupe de nœuds de calcul comprend des règles pour créer des nœuds, notamment les types d'EC2instances Amazon à utiliser,

le nombre d'instances à exécuter, l'utilisation d'instances ponctuelles ou d'instances à la demande, les sous-réseaux et les groupes de sécurité à utiliser, et la manière de configurer chaque instance lors de son lancement. Lorsque ces règles sont AWS PCS mises à jour, les ressources associées au groupe de nœuds de calcul sont mises à jour pour qu'elles correspondent.

Rubriques

Création d'un groupe de nœuds de calcul dans AWS PCS

Cette rubrique fournit une vue d'ensemble des options disponibles et décrit les éléments à prendre en compte lorsque vous créez un groupe de nœuds de calcul dans AWS Parallel Computing Service (AWS PCS). Si c'est la première fois que vous créez un groupe de nœuds de calcul dans AWS PCS, nous vous recommandons de suivre le didacticiel dansCommencer avec AWS PCS. Le didacticiel peut vous aider à créer un HPC système fonctionnel sans étendre toutes les options disponibles et les architectures système possibles.

Prérequis

- Quotas de service suffisants pour lancer le nombre d'EC2instances souhaité dans votre Région AWS. Vous pouvez utiliser le AWS Management Consolepour vérifier et demander des augmentations de vos quotas de service.
- Un réseau existant VPC et un ou plusieurs sous-réseaux qui répondent aux exigences du AWS PCS réseau. Nous vous recommandons de bien comprendre ces exigences avant de déployer un cluster à des fins de production. Pour de plus amples informations, veuillez consulter AWS PCSVPCexigences et considérations relatives aux sous-réseaux. Vous pouvez également utiliser un CloudFormation modèle pour créer des sous-réseaux VPC et. AWS fournit une HPC recette pour le CloudFormation modèle. Pour plus d'informations, voir aws-hpc-recipesci-dessous GitHub.
- Un profil d'IAMinstance avec les autorisations nécessaires pour lancer l' AWS PCS`RegisterComputeNodeGroupInstance`APIaction et accéder à toutes les autres AWS ressources requises pour les instances de votre groupe de nœuds. Pour de plus amples informations, veuillez consulter IAMprofils d'instance pour AWS Parallel Computing Service.

- Un modèle de lancement pour les instances de votre groupe de nœuds. Pour de plus amples informations, veuillez consulter Utilisation des modèles EC2 de lancement Amazon avec AWS PCS.
- Pour créer un groupe de nœuds de calcul utilisant des instances Amazon EC2 Spot, vous devez avoir le rôle AWSServiceRoleForEC2Spotlié au service dans votre. Compte AWS Pour de plus amples informations, veuillez consulter Rôle Amazon EC2 Spot pour AWS PCS.

Créez un groupe de nœuds de calcul dans AWS PCS

Vous pouvez créer un groupe de nœuds de calcul à l'aide du AWS Management Console ou du AWS CLI.

AWS Management Console

Pour créer votre groupe de nœuds de calcul à l'aide de la console

1. Ouvrez la AWS PCSconsole.
2. Sélectionnez le cluster dans lequel vous souhaitez créer un groupe de nœuds de calcul. Accédez à Compute node groups et choisissez Create.
3. Dans la section Configuration du groupe de nœuds de calcul, donnez un nom à votre groupe de nœuds. Le nom ne peut contenir que des caractères alphanumériques et des tirets distinguant majuscules et minuscules. Il doit commencer par un caractère alphabétique et ne doit pas comporter plus de 25 caractères. Le nom doit être unique au sein du cluster.
4. Sous Configuration informatique, entrez ou sélectionnez les valeurs suivantes :

    a. EC2modèle de lancement : sélectionnez un modèle de lancement personnalisé à utiliser pour ce groupe de nœuds. Les modèles de lancement peuvent être utilisés pour personnaliser les paramètres réseau tels que les sous-réseaux et les groupes de sécurité, la configuration de surveillance et le stockage au niveau de l'instance. Si vous n'avez pas préparé de modèle de lancement, consultez la section Utilisation des modèles EC2 de lancement Amazon avec AWS PCS pour savoir comment en créer un.

    > **Important**
    >
    > AWS PCScrée un modèle de lancement géré pour chaque groupe de nœuds de calcul. Ils sont nomméspcs-*identifier*-do-not-delete. Ne les

sélectionnez pas lorsque vous créez ou mettez à jour un groupe de nœuds de calcul, sinon le groupe de nœuds ne fonctionnera pas correctement.

b. EC2version du modèle de lancement : sélectionnez une version de votre modèle de lancement personnalisé. Vous pouvez choisir une version spécifique, qui peut améliorer la reproductibilité. Si vous modifiez la version ultérieurement, vous devez mettre à jour le groupe de nœuds de calcul pour détecter les modifications apportées au modèle de lancement. Pour de plus amples informations, veuillez consulter Mise à jour d'un groupe AWS PCS de nœuds de calcul.

c. AMIID : si votre modèle de lancement n'inclut pas d'AMIidentifiant, ou si vous souhaitez remplacer la valeur du modèle de lancement, indiquez-en un AMI ici. Notez que le nom AMI utilisé pour le groupe de nœuds doit être compatible avec AWS PCS. Vous pouvez également sélectionner un échantillon AMI fourni par AWS. Pour plus d'informations sur ce sujet, consultezAmazon Machine Images (AMIs) pour AWS PCS.

d. IAMprofil d'instance — Choisissez un profil d'instance pour le groupe de nœuds. Un profil d'instance accorde à l'instance les autorisations nécessaires pour accéder aux AWS ressources et aux services en toute sécurité. Si vous n'en avez pas préparé un, consultez IAMprofils d'instance pour AWS Parallel Computing Service la section pour savoir comment en créer un.

e. Sous-réseaux — Choisissez un ou plusieurs sous-réseaux dans l'VPCendroit où votre AWS PCS cluster est déployé. Si vous sélectionnez plusieurs sous-réseaux, les EFA communications ne seront pas disponibles entre les nœuds et la communication entre les nœuds de différents sous-réseaux peut augmenter le temps de latence. Assurez-vous que les sous-réseaux que vous spécifiez ici correspondent à ceux que vous définissez dans le modèle de EC2 lancement.

f. Instances — Choisissez un ou plusieurs types d'instances pour répondre aux demandes de dimensionnement dans le groupe de nœuds. Tous les types d'instances doivent avoir la même architecture de processeur (x864_64 ou arm64) et le même nombre de. vCPUs Si les instances en ontGPUs, tous les types d'instances doivent avoir le même nombre deGPUs.

g. Configuration de dimensionnement : spécifiez le nombre minimum et maximum d'instances pour le groupe de nœuds. Vous pouvez définir soit une configuration statique, dans laquelle un nombre fixe de nœuds sont en cours d'exécution, soit une configuration dynamique, dans laquelle le nombre maximum de nœuds peut être exécuté. Pour une configuration statique, définissez le minimum et le maximum sur le même nombre, supérieur à zéro. Pour une configuration dynamique, définissez le

nombre minimum d'instances à zéro et le nombre maximal d'instances à un nombre supérieur à zéro. AWS PCSne prend pas en charge les groupes de nœuds de calcul composés d'une combinaison d'instances statiques et dynamiques.

5. (Facultatif) Sous Paramètres supplémentaires, spécifiez les éléments suivants :

   a. Option d'achat : choisissez entre les instances Spot et On-Demand.

   b. Stratégie d'allocation : si vous avez sélectionné l'option d'achat ponctuel, vous pouvez spécifier comment les pools de capacité ponctuels sont choisis lors du lancement des instances dans le groupe de nœuds. Pour plus d'informations, consultez la section Stratégies d'allocation pour les instances Spot dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud. Cette option n'a aucun effet si vous avez sélectionné l'option d'achat à la demande.

6. (Facultatif) Dans la section Slurmdes paramètres personnalisés, indiquez les valeurs suivantes :

   a. Poids : cette valeur définit la priorité des nœuds du groupe à des fins de planification. Les nœuds dont le poids est faible ont une priorité plus élevée et les unités sont arbitraires. Pour plus d'informations, consultez la section Poids dans la Slurm documentation.

   b. Mémoire réelle : cette valeur définit la taille (en Go) de la mémoire réelle sur les nœuds du groupe de nœuds. Il est destiné à être utilisé conjointement avec l'`CR_CPU_Memory`option de la Slurm configuration du cluster dans AWS PCS. Pour plus d'informations, consultez RealMemoryla Slurm documentation.

7. (Facultatif) Sous Balises, ajoutez des balises à votre groupe de nœuds de calcul.

8. Choisissez Créer un groupe de nœuds de calcul. Le champ Status s'affiche `Creating` lors du AWS PCS provisionnement du groupe de nœuds. Cela peut prendre plusieurs minutes.

Étape suivante recommandée

- Ajoutez votre groupe de nœuds à une file d'attente AWS PCS pour lui permettre de traiter les tâches.

AWS CLI

Pour créer votre groupe de nœuds de calcul à l'aide de AWS CLI

Créez votre file d'attente avec la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :

1. Remplacez *region*avec l'ID du dans Région AWS lequel créer votre cluster, tel que`us-east-1`.
2. Remplacez *my-cluster* avec le nom ou celui `clusterId` de votre cluster.
3. Remplacez *my-node-group*avec le nom de votre groupe de nœuds de calcul. Un nom ne peut contenir que des caractères alphanumériques (sensibles à la casse) et des traits d'union. Il doit commencer par un caractère alphabétique et ne doit pas comporter plus de 25 caractères. Le nom doit être unique au sein du cluster.
4. Remplacez *subnet-ExampleID1* avec un ou plusieurs sous-réseaux IDs de votre clusterVPC.
5. Remplacez *lt-ExampleID1* avec l'ID de votre modèle de lancement personnalisé. Si vous n'en avez pas préparé un, consultez Utilisation des modèles EC2 de lancement Amazon avec AWS PCS la section pour savoir comment en créer un.

   > **Important**
   >
   > AWS PCScrée un modèle de lancement géré pour chaque groupe de nœuds de calcul. Ils sont nomméspcs-*identifier*-do-not-delete. Ne les sélectionnez pas lorsque vous créez ou mettez à jour un groupe de nœuds de calcul, sinon le groupe de nœuds ne fonctionnera pas correctement.

6. Remplacez *launch-template-version* avec une version de modèle de lancement spécifique si vous souhaitez associer votre groupe de nœuds à une version spécifique.
7. Remplacez *arn:InstanceProfile*avec le profil ARN de votre IAM instance. Si vous n'en avez pas préparé un, consultez Utilisation des modèles EC2 de lancement Amazon avec AWS PCS pour obtenir des conseils.
8. Remplacez *min-instances* and *max-instances* avec des valeurs entières. Vous pouvez définir soit une configuration statique, dans laquelle un nombre fixe de nœuds sont en cours d'exécution, soit une configuration dynamique, dans laquelle le nombre maximum de nœuds peut être exécuté. Pour une configuration statique, définissez le minimum et le maximum sur le même nombre, supérieur à zéro. Pour une configuration dynamique, définissez le

nombre minimum d'instances à zéro et le nombre maximal d'instances à un nombre supérieur à zéro. AWS PCSne prend pas en charge les groupes de nœuds de calcul composés d'une combinaison d'instances statiques et dynamiques.

9. Remplacez *t3.large* avec un autre type d'instance. Vous pouvez ajouter d'autres types d'instances en spécifiant une liste de `instanceType` paramètres. Par exemple, *--instance-configs instanceType=c6i.16xlarge,instanceType=c6a.16xlarge*. Tous les types d'instances doivent avoir la même architecture de processeur (x864_64 ou arm64) et le même nombre de. vCPUs Si les instances en ontGPUs, tous les types d'instances doivent avoir le même nombre deGPUs.

```
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
```

Vous pouvez ajouter plusieurs paramètres de configuration facultatifs à la `create-compute-node-group` commande.

- Vous pouvez spécifier `--amiId` si votre modèle de lancement personnalisé n'inclut pas de référence à unAMI, ou si vous souhaitez remplacer cette valeur. Notez que le nom AMI utilisé pour le groupe de nœuds doit être compatible avec AWS PCS. Vous pouvez également sélectionner un échantillon AMI fourni par AWS. Pour plus d'informations sur ce sujet, consultezAmazon Machine Images (AMIs) pour AWS PCS.
- Vous pouvez choisir entre des instances à la demande (`ONDEMAND`) et des instances Spot (`SPOT`) à l'aide de`--purchase-option`. On-Demand est la valeur par défaut. Si vous choisissez des instances Spot, vous pouvez également les utiliser `--allocation-strategy` pour définir comment choisir AWS PCS les pools de capacité Spot lorsqu'il lance des instances dans le groupe de nœuds. Pour plus d'informations, consultez la section Stratégies d'allocation pour les instances Spot dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.
- Il est possible de fournir des options Slurm de configuration pour les nœuds du groupe de nœuds à l'aide de`--slurm-configuration`. Vous pouvez définir le poids (priorité de planification) et la mémoire réelle. Les nœuds dont le poids est faible ont une priorité plus élevée et les unités sont arbitraires. Pour plus d'informations, consultez la section Poids

dans la Slurm documentation. La mémoire réelle définit la taille (en Go) de la mémoire réelle sur les nœuds du groupe de nœuds. Il est destiné à être utilisé conjointement avec l'`CR_CPU_Memory`option pour le cluster AWS PCS dans votre Slurm configuration. Pour plus d'informations, consultez RealMemoryla Slurm documentation.

> **Important**
>
> La création du groupe de nœuds de calcul peut prendre plusieurs minutes.

Vous pouvez demander l'état de votre groupe de nœuds à l'aide de la commande suivante. Vous ne pourrez pas associer le groupe de nœuds à une file d'attente tant que son statut ne sera pas atteint`ACTIVE`.

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

Mise à jour d'un groupe AWS PCS de nœuds de calcul

Cette rubrique fournit une vue d'ensemble des options disponibles et décrit les éléments à prendre en compte lors de la mise à jour d'un groupe de nœuds de AWS PCS calcul.

Options de mise à jour d'un groupe AWS PCS de nœuds de calcul

La mise à jour d'un groupe de nœuds de AWS PCS calcul vous permet de modifier les propriétés des instances lancées par AWSPCS, ainsi que les règles régissant le lancement de ces instances. Par exemple, vous pouvez remplacer les instances AMI de groupes de nœuds par une autre instance sur laquelle un logiciel différent est installé. Vous pouvez également mettre à jour les groupes de sécurité pour modifier la connectivité réseau entrante ou sortante. Vous pouvez également modifier la configuration de dimensionnement ou même modifier l'option d'achat préférée vers ou depuis des instances Spot.

Les paramètres des groupes de nœuds suivants ne peuvent pas être modifiés après leur création :

- Nom
- instances

Considérations relatives à la mise à jour d'un groupe AWS PCS de nœuds de calcul

Les groupes de nœuds de calcul définissent les EC2 instances utilisées pour traiter les tâches, fournir un accès au shell interactif et effectuer d'autres tâches. Ils sont souvent associés à une ou plusieurs AWS PCS files d'attente. Lorsque vous mettez à jour votre groupe de nœuds de calcul pour modifier son comportement (ou celui de ses nœuds), tenez compte des points suivants :

- Les modifications apportées aux propriétés du groupe de nœuds de calcul entrent en vigueur lorsque le statut du groupe de nœuds de calcul passe de Mise à jour à Actif. Les nouvelles instances sont lancées avec les propriétés mises à jour.
- Les mises à jour qui n'ont aucun impact sur la configuration de nœuds spécifiques n'affectent pas les nœuds en cours d'exécution. Par exemple, ajouter un sous-réseau et modifier la stratégie d'allocation.
- Si vous mettez à jour le modèle de lancement d'un groupe de nœuds de calcul, vous devez mettre à jour le groupe de nœuds de calcul pour utiliser la nouvelle version.
- Pour ajouter ou supprimer un groupe de sécurité dans les nœuds d'un groupe de nœuds de calcul, modifiez son modèle de lancement et mettez à jour le groupe de nœuds de calcul. Les nouvelles instances sont lancées avec l'ensemble de groupes de sécurité mis à jour.
- Si vous modifiez directement un groupe de sécurité utilisé par un groupe de nœuds de calcul, cela a un effet immédiat sur les instances en cours d'exécution et les instances futures.
- Si vous ajoutez ou supprimez des autorisations dans le profil d'IAMinstance utilisé par un groupe de nœuds de calcul, cela a un effet immédiat sur les instances en cours d'exécution et les instances futures.
- Pour modifier le nombre AMI utilisé par les instances d'un groupe de nœuds de calcul, mettez à jour le groupe de nœuds de calcul (ou son modèle de lancement) AWS PCS pour utiliser le nouveau AMI et attendez le remplacement des instances.
- AWS PCSremplace les instances existantes du groupe de nœuds après une opération de mise à jour du groupe de nœuds. Si des tâches sont exécutées sur un nœud, elles sont autorisées à se terminer avant de AWS PCS remplacer le nœud. Les processus utilisateur interactifs (tels que sur les instances du nœud de connexion) sont interrompus. L'état du groupe de nœuds revient au `Active` AWS PCS moment où les instances sont marquées à remplacer, mais le remplacement réel a lieu lorsque les instances sont inactives.
- Si vous diminuez le nombre maximum d'instances autorisées dans un groupe de nœuds de calcul, AWS PCS supprimez des nœuds de Slurm pour atteindre le nouveau maximum. AWS PCSmet fin

à l'exécution des instances associées aux nœuds Slurm supprimés. Les tâches en cours sur les nœuds supprimés échouent et retournent dans leurs files d'attente.

- AWS PCScrée un modèle de lancement géré pour chaque groupe de nœuds de calcul. Ils sont nomméspcs-*identifier*-do-not-delete. Ne les sélectionnez pas lorsque vous créez ou mettez à jour un groupe de nœuds de calcul, sinon le groupe de nœuds ne fonctionnera pas correctement.
- Si vous mettez à jour un groupe de nœuds de calcul pour utiliser Spot pour son option d'achat, vous devez avoir le rôle AWSServiceRoleForEC2Spotlié au service dans votre compte. Pour de plus amples informations, veuillez consulter Rôle Amazon EC2 Spot pour AWS PCS.

Pour mettre à jour un groupe AWS PCS de nœuds de calcul

Vous pouvez mettre à jour un groupe de nœuds à l'aide de la console AWS de gestion ou du AWSCLI.

AWS Management Console

Pour mettre à jour un groupe de nœuds de calcul

1. Ouvrez la AWS PCS console à l'adresse `https://console.aws.amazon.com/pcs/home#/clusters`
2. Sélectionnez le cluster dans lequel vous souhaitez mettre à jour un groupe de nœuds de calcul.
3. Accédez à Calculer les groupes de nœuds, accédez au groupe de nœuds que vous souhaitez mettre à jour, puis sélectionnez Modifier.
4. Dans les sections Configuration informatique, Paramètres supplémentaires et Paramètres Slurmde personnalisation, mettez à jour toutes les valeurs sauf :
   - Instances : vous ne pouvez pas modifier les instances d'un groupe de nœuds de calcul.
5. Choisissez Mettre à jour. Le champ État affichera la mise à jour pendant que les modifications sont appliquées.

> **Important**
>
> Les mises à jour des groupes de nœuds de calcul peuvent prendre plusieurs minutes.

AWS CLI

Pour mettre à jour un groupe de nœuds de calcul

1. Mettez à jour votre groupe de nœuds de calcul à l'aide de la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :

   a. Remplacez *region-code* avec la AWS région dans laquelle vous souhaitez créer votre cluster.

   b. Remplacez *my-node-group* avec le nom ou `computeNodeGroupId` pour votre groupe de nœuds de calcul.

   c. Remplacez *my-cluster* avec le nom ou celui `clusterId` de votre cluster.

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

2. Mettez à jour les paramètres de tous les groupes de nœuds, à l'exception de`--instance-configs`. Par exemple, pour définir un nouvel AMI identifiant, passez `--amiId my-custom-ami-id` où *my-custom-ami-id* est remplacé par celui AMI de votre choix.

> **Important**
>
> La mise à jour du groupe de nœuds de calcul peut prendre plusieurs minutes.

Vous pouvez demander l'état de votre groupe de nœuds à l'aide de la commande suivante.

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

Suppression d'un groupe de nœuds de calcul dans AWS PCS

Cette rubrique fournit une vue d'ensemble des options disponibles et décrit les éléments à prendre en compte lorsque vous supprimez un groupe de nœuds de calcul dans AWS PCS.

Considérations relatives à la suppression d'un groupe de nœuds de calcul

Les groupes de nœuds de calcul définissent les EC2 instances utilisées pour traiter les tâches, fournir un accès au shell interactif et effectuer d'autres tâches. Ils sont souvent associés à une ou plusieurs AWS PCS files d'attente. Avant de supprimer un groupe de nœuds de calcul, tenez compte des points suivants :

- Toutes EC2 les instances lancées par le groupe de nœuds de calcul seront résiliées. Cela annulera les tâches en cours d'exécution sur ces instances et mettra fin à l'exécution des processus interactifs.
- Vous devez dissocier le groupe de nœuds de calcul de toutes les files d'attente avant de pouvoir le supprimer. Pour de plus amples informations, veuillez consulter Mettre à jour une AWS PCS file d'attente.

Supprimer le groupe de nœuds de calcul

Vous pouvez utiliser le AWS Management Console ou AWS CLI pour supprimer un groupe de nœuds de calcul.

AWS Management Console

Pour supprimer un groupe de nœuds de calcul

1. Ouvrez la AWS PCSconsole.
2. Sélectionnez le cluster du groupe de nœuds de calcul.
3. Accédez à Groupes de nœuds de calcul et sélectionnez le groupe de nœuds de calcul à supprimer.
4. Sélectionnez Delete (Supprimer).
5. Le champ État s'affiche`Deleting`. Cela peut prendre plusieurs minutes.

> **Note**
>
> Vous pouvez utiliser les commandes natives de votre planificateur pour confirmer que le groupe de nœuds de calcul est supprimé. Par exemple, utilisez `sinfo` ou `squeue` pour Slurm.

AWS CLI

Pour supprimer un groupe de nœuds de calcul

- Utilisez la commande suivante pour supprimer un groupe de nœuds de calcul, avec les remplacements suivants :
  - Remplacez *region-code* avec le nom dans lequel se trouve Région AWS votre cluster.
  - Remplacez *my-node-group* avec le nom ou l'ID de votre groupe de nœuds de calcul.
  - Remplacez *my-cluster* avec le nom ou l'ID de votre cluster.

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

La suppression du groupe de nœuds de calcul peut prendre plusieurs minutes.

> **Note**
>
> Vous pouvez utiliser les commandes natives de votre planificateur pour confirmer que le groupe de nœuds de calcul est supprimé. Par exemple, utilisez `sinfo` ou `squeue` pour Slurm.

Recherche d'instances de groupes de nœuds de calcul dans AWS PCS

Chaque groupe AWS PCS de nœuds de calcul peut lancer EC2 des instances avec des configurations partagées. Vous pouvez utiliser des EC2 balises pour rechercher des instances dans un groupe de nœuds de calcul dans AWS Management Console ou avec le AWS CLI.

AWS Management Console

Pour trouver les instances de votre groupe de nœuds de calcul

1. Ouvrez la AWS PCSconsole.
2. Sélectionnez le cluster .
3. Choisissez Compute node groups.

4. Trouvez l'ID du groupe de nœuds de connexion que vous avez créé.
5. Accédez à la EC2console et choisissez Instances.
6. Recherchez les instances avec la balise suivante. Remplacez *node-group-id* avec l'ID (et non le nom) de votre groupe de nœuds de calcul.

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

7. (Facultatif) Vous pouvez modifier la valeur de l'état de l'instance dans le champ de recherche pour trouver les instances en cours de configuration ou récemment résiliées.
8. Trouvez l'ID d'instance et l'adresse IP de chaque instance dans la liste des instances balisées.

AWS CLI

Pour rechercher les instances de votre groupe de nœuds, utilisez les commandes ci-dessous. Avant d'exécuter les commandes, effectuez les remplacements suivants :

- *region-code*Remplacez-le par celui Région AWS de votre cluster. Exemple : `us-east-1`
- *node-group-id*Remplacez-le par l'ID (et non le nom) de votre groupe de nœuds de calcul.
- `running`Remplacez-le par d'autres états d'instance tels que `pending` ou `terminated` pour rechercher des EC2 instances dans d'autres états.

```
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}
```

La commande renvoie un résultat semblable à ce qui suit. La valeur de `PublicIP` est `null` si l'instance se trouve dans un sous-réseau privé.

```
[
    [
        {
            "InstanceID": "i-0123456789abcdefa",
            "State": "running",
```

```
            "PublicIP": "18.189.32.188",
            "PrivateIP": "10.0.0.1"
        }
    ]
]
```

> **Note**
>
> Si vous prévoyez `describe-instances` de renvoyer un grand nombre d'instances, vous devez utiliser les options pour plusieurs pages. Pour plus d'informations, consultez DescribeInstancesle manuel Amazon Elastic Compute Cloud API Reference.

Utilisation des modèles EC2 de lancement Amazon avec AWS PCS

Sur AmazonEC2, un modèle de lancement peut stocker un ensemble de préférences afin que vous n'ayez pas à les spécifier individuellement lorsque vous lancez des instances. AWS PCSintègre des modèles de lancement comme moyen flexible de configurer des groupes de nœuds de calcul. Lorsque vous créez un groupe de nœuds, vous fournissez un modèle de lancement. AWS PCScrée à partir de celui-ci un modèle de lancement dérivé qui inclut des transformations pour garantir son fonctionnement avec le service.

Comprendre les options et les considérations à prendre en compte lors de la rédaction d'un modèle de lancement personnalisé peut vous aider à en créer un à utiliser AWS PCS. Pour plus d'informations sur les modèles de lancement, consultez Lancer une instance à partir d'un modèle de lancement Lancer une instance à partir d'un modèle de lancement dans le guide de EC2 l'utilisateur Amazon.

Rubriques

- Présentation
- Créer un modèle de lancement de base
- Utilisation des données EC2 utilisateur d'Amazon
- Réservations de capacité en AWS PCS
- Paramètres utiles du modèle de lancement

Présentation

Il existe plus de 30 paramètres disponibles que vous pouvez inclure dans un modèle de EC2 lancement, contrôlant de nombreux aspects de la configuration des instances. La plupart sont entièrement compatibles avec AWS PCS, à quelques exceptions près.

Les paramètres suivants du modèle de EC2 lancement seront ignorés AWS PCS car ces propriétés doivent être gérées directement par le service :

- Type d'instance/Spécifiez les attributs du type d'instance (`InstanceRequirements`) : AWS PCS ne prend pas en charge la sélection d'instance basée sur les attributs.
- Type d'instance (`InstanceType`) : spécifiez les types d'instances lorsque vous créez un groupe de nœuds.
- Détails avancés/profil d'IAMinstance (`IamInstanceProfile`) — Vous les fournissez lorsque vous créez ou mettez à jour le groupe de nœuds.
- Détails avancés/Disable API termination (`DisableApiTermination`) : AWS PCS doit contrôler le cycle de vie des instances du groupe de nœuds qu'il lance.
- Détails avancés/Disable API stop (`DisableApiStop`) : AWS PCS doit contrôler le cycle de vie des instances de groupes de nœuds qu'il lance.
- Détails avancés/Stop — Comportement d'hibernation (`HibernationOptions`) — AWS PCS ne prend pas en charge l'hibernation des instances.
- Détails avancés/Elastic GPU (`ElasticGpuSpecifications`) — Amazon Elastic Graphics a atteint la fin de son cycle de vie le 8 janvier 2024.
- Détails avancés/Inférence élastique (`ElasticInferenceAccelerators`) — Amazon Elastic Inference n'est plus disponible pour les nouveaux clients.
- AAdvancedDétails/Spécifier les CPU options/Threads par cœur (`ThreadsPerCore`) — AWS PCS définit le nombre de threads par cœur à 1.

Ces paramètres ont des exigences particulières qui garantissent la compatibilité avec AWS PCS :

- Données utilisateur (`UserData`) : elles doivent être codées en plusieurs parties. Consultez Utilisation des données EC2 utilisateur d'Amazon.
- Images de l'application et du système d'exploitation (`ImageId`) — Vous pouvez les inclure. Toutefois, si vous spécifiez un AMI ID lorsque vous créez ou mettez à jour le groupe de nœuds, il remplacera la valeur du modèle de lancement. Le AMI que vous fournissez doit être compatible

avec AWS PCS. Pour plus d'informations, reportez-vous à la section "Amazon Machine Images (AMIs) pour AWS PCS.

- Paramètres réseau/Pare-feu (groupes de sécurité) (`SecurityGroups`) — Il est impossible de définir une liste de noms de groupes de sécurité dans un modèle de AWS PCS lancement. Vous pouvez définir une liste de groupes de sécurité IDs (`SecurityGroupIds`), sauf si vous définissez des interfaces réseau dans le modèle de lancement. Vous devez ensuite spécifier le groupe de sécurité IDs pour chaque interface. Pour de plus amples informations, veuillez consulter Groupes de sécurité dans AWS PCS.
- Paramètres réseau/Configuration réseau avancée (`NetworkInterfaces`) — Si vous utilisez des EC2 instances avec une seule carte réseau et que vous n'avez pas besoin de configuration réseau spécialisée, vous AWS PCS pouvez configurer la mise en réseau des instances pour vous. Pour configurer plusieurs cartes réseau ou pour activer Elastic Fabric Adapter sur vos instances, utilisez`NetworkInterfaces`. Chaque interface réseau doit contenir une liste de groupes de IDs sécurité`Groups`. Pour de plus amples informations, veuillez consulter Plusieurs interfaces réseau dans AWS PCS.
- Détails avancés/réservation de capacité (`CapacityReservationSpecification`) — Cela peut être défini, mais vous ne pouvez pas faire référence à un élément spécifique `CapacityReservationId` lorsque vous travaillez avec. AWS PCS Vous pouvez toutefois faire référence à un groupe de réservation de capacité, lorsque ce groupe contient une ou plusieurs réservations de capacité. Pour de plus amples informations, veuillez consulter Réservations de capacité en AWS PCS.

Créer un modèle de lancement de base

Vous pouvez créer un modèle de lancement à l'aide du AWS Management Console ou du AWS CLI.

AWS Management Console

Pour créer un modèle de lancement

1. Ouvrez la EC2console Amazon et sélectionnez Launch templates.
2. Choisissez Créer un modèle de lancement.
3. Sous Nom et description du modèle de lancement, entrez un nom unique et distinctif pour le nom du modèle de lancement

4. Sous Paire de clés (connexion) sous Nom de la paire de SSH clés, sélectionnez la paire de clés qui sera utilisée pour se connecter aux EC2 instances gérées par AWS PCS. Cette action est facultative, mais recommandée.
5. Sous Paramètres réseau, puis Pare-feu (groupes de sécurité), choisissez les groupes de sécurité à associer à l'interface réseau. Tous les groupes de sécurité du modèle de lancement doivent provenir de votre AWS PCS clusterVPC. Choisissez au minimum :

   - Un groupe de sécurité qui permet la communication avec le AWS PCS cluster
   - Un groupe de sécurité qui permet la communication entre les EC2 instances lancées par AWS PCS
   - (Facultatif) Un groupe de sécurité qui autorise l'SSHaccès entrant aux instances interactives
   - (Facultatif) Un groupe de sécurité qui permet aux nœuds de calcul d'établir des connexions sortantes vers Internet
   - (Facultatif) Groupe (s) de sécurité qui autorisent l'accès à des ressources en réseau telles que des systèmes de fichiers partagés ou un serveur de base de données.
6. Votre nouvel identifiant de modèle de lancement sera accessible dans la EC2 console Amazon sous Modèles de lancement. L'ID du modèle de lancement contiendra le formulaire`lt-0123456789abcdef01`.

Étape suivante recommandée

- Utilisez le nouveau modèle de lancement pour créer ou mettre à jour un groupe de nœuds de AWS PCS calcul.

AWS CLI

Pour créer un modèle de lancement

Créez votre modèle de lancement à l'aide de la commande ci-dessous.

- Avant d'exécuter la commande, effectuez les remplacements suivants :

  a. Remplacez *`region-code`* avec l' Région AWS endroit avec lequel vous travaillez AWS PCS
  b. Remplacez *`my-launch-template-name`* avec un nom pour votre modèle. Il doit être unique au Compte AWS et Région AWS que vous utilisez.

c. Remplacez *my-ssh-key-name* avec le nom de votre SSH clé préférée.

d. Remplacez *sg-ExampleID1* and *sg-ExampleID2* avec un groupe de sécurité IDs qui permet la communication entre vos EC2 instances et le planificateur ainsi que la communication entre les EC2 instances. Si vous ne disposez que d'un seul groupe de sécurité qui autorise tout ce trafic, vous pouvez supprimer `sg-ExampleID2` la virgule qui la précède. Vous pouvez également ajouter d'autres groupes de sécuritéIDs. Tous les groupes de sécurité que vous incluez dans le modèle de lancement doivent provenir de votre AWS PCS clusterVPC.

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

Le texte affiché AWS CLI ressemblera à ce qui suit. L'ID du modèle de lancement se trouve dans`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"
    }
}
```

Étape suivante recommandée

- Utilisez le nouveau modèle de lancement pour créer ou mettre à jour un groupe de nœuds de AWS PCS calcul.

Utilisation des données EC2 utilisateur d'Amazon

Vous pouvez fournir des données EC2 utilisateur dans votre modèle de lancement qui `cloud-init` s'exécute lors du lancement de vos instances. Les blocs de données utilisateur avec le

type de contenu `cloud-config` s'exécutent avant que l'instance ne s'enregistre auprès du AWS PCSAPI, tandis que les blocs de données utilisateur avec le type de contenu `text/x-shellscript` s'exécutent une fois l'enregistrement terminé, mais avant le démarrage du démon Slurm. Pour plus d'informations sur les types de contenus, consultez la documentation sur Cloud-Init.

nos données utilisateur peuvent exécuter des scénarios de configuration courants, y compris, mais sans s'y limiter, les suivants :

- Inclure des utilisateurs ou des groupes
- Installation de packages
- Création de partitions et de systèmes de fichiers
- Montage de systèmes de fichiers réseau

Les données utilisateur figurant dans les modèles de lancement doivent être au format d'archive en MIME plusieurs parties. Cela est dû au fait que vos données utilisateur sont fusionnées avec d'autres données AWS PCS utilisateur requises pour configurer les nœuds de votre groupe de nœuds. Vous pouvez combiner plusieurs blocs de données utilisateur dans un seul fichier en MIME plusieurs parties.

Un fichier en MIME plusieurs parties comprend les éléments suivants :

- Le type de contenu et la déclaration de limite : `Content-Type: multipart/mixed; boundary="==BOUNDARY=="`
- La déclaration de MIME version : `MIME-Version: 1.0`
- Un ou plusieurs blocs de données utilisateur contenant les composants suivants :
  - La limite d'ouverture qui indique le début d'un bloc de données utilisateur :`--==BOUNDARY==`. Vous devez laisser la ligne avant cette limite vide.
  - La déclaration du type de contenu pour le bloc : `Content-Type: text/cloud-config; charset="us-ascii"` ou`Content-Type: text/x-shellscript; charset="us-ascii"`. Vous devez laisser la ligne après la déclaration de type de contenu vide.
  - Le contenu des données utilisateur, tel qu'une liste de commandes ou de `cloud-config` directives du shell.
- La limite de fermeture qui indique la fin du fichier en MIME plusieurs parties :`--==BOUNDARY==--`. Vous devez laisser la ligne avant la limite de fermeture vide.

> **Note**
>
> Si vous ajoutez des données utilisateur à un modèle de lancement dans la EC2 console Amazon, vous pouvez les coller sous forme de texte brut. Vous pouvez également le télécharger à partir d'un fichier. Si vous utilisez le AWS CLI ou an AWS SDK, vous devez d'abord encoder les données utilisateur en base64 et envoyer cette chaîne comme valeur du `UserData` paramètre lorsque vous appelez CreateLaunchTemplate, comme indiqué dans ce JSON fichier.

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

Exemples

- Exemple : installation d'un logiciel à partir d'un référentiel de packages
- Exemple : exécution de scripts à partir d'un compartiment S3
- Exemple : définir des variables d'environnement globales
- Utilisation de systèmes de fichiers réseau avec AWS PCS
- Exemple : utilisation d'un système de EFS fichiers comme répertoire de base partagé

Exemple : installation d'un logiciel AWS PCS depuis un dépôt de packages

Indiquez ce script comme valeur de `"userData"` dans votre modèle de lancement. Pour de plus amples informations, veuillez consulter Utilisation des données EC2 utilisateur d'Amazon.

Ce script utilise cloud-config pour installer des packages logiciels sur les instances de groupes de nœuds lors du lancement. Pour plus d'informations, consultez les formats de données utilisateur dans la documentation de cloud-init. Cet exemple installe `curl` et. `llvm`

> **Note**
>
> Vos instances doivent être en mesure de se connecter à leurs référentiels de packages configurés.

```
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==--
```

Exemple : exécution de scripts supplémentaires à AWS PCS partir d'un compartiment S3

Indiquez ce script comme valeur de `"userData"` dans votre modèle de lancement. Pour de plus amples informations, veuillez consulter Utilisation des données EC2 utilisateur d'Amazon.

Ce script utilise cloud-config pour importer un script depuis un compartiment S3 et l'exécuter sur des instances de groupes de nœuds lors du lancement. Pour plus d'informations, consultez les formats de données utilisateur dans la documentation de cloud-init.

Remplacez les valeurs suivantes dans ce script par vos propres informations :

- *`my-bucket-name`* — Le nom d'un compartiment S3 que votre compte peut lire.
- *`path`* — Le chemin relatif à la racine du compartiment S3.
- *`shell`* — Le shell Linux à utiliser pour exécuter le script, tel que`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==--
```

Le profil d'IAMinstance du groupe de nœuds doit avoir accès au bucket. La IAM politique suivante est un exemple pour le bucket dans le script de données utilisateur ci-dessus.

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

Exemple : définir des variables d'environnement globales pour AWS PCS

Indiquez ce script comme valeur de `"userData"` dans votre modèle de lancement. Pour de plus amples informations, veuillez consulter Utilisation des données EC2 utilisateur d'Amazon.

L'exemple suivant permet `/etc/profile.d` de définir des variables globales sur des instances de groupes de nœuds.

```
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==--
```

Exemple : utiliser un système de EFS fichiers comme répertoire de base partagé pour AWS PCS

Indiquez ce script comme valeur de `"userData"` dans votre modèle de lancement. Pour de plus amples informations, veuillez consulter Utilisation des données EC2 utilisateur d'Amazon.

Cet exemple étend l'exemple de EFS montage Utilisation de systèmes de fichiers réseau avec AWS PCS pour implémenter un répertoire de base partagé. Le contenu de /home est sauvegardé avant le montage du système de EFS fichiers. Le contenu est ensuite rapidement copié sur place sur le stockage partagé une fois le montage terminé.

Remplacez les valeurs suivantes dans ce script par vos propres informations :

- *`/mount-point-directory`* — Le chemin d'une instance sur laquelle vous souhaitez monter le système de EFS fichiers.
- *`filesystem-id`* — L'ID du système de fichiers pour le système de EFS fichiers.

```
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==--
```

Activation du mode sans mot de passe SSH

Vous pouvez vous appuyer sur l'exemple du répertoire de base partagé pour implémenter des SSH connexions entre des instances de cluster à l'aide de SSH clés. Pour chaque utilisateur utilisant le système de fichiers d'accueil partagé, exécutez un script semblable au suivant :

```
#!/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**
>
> Les instances doivent utiliser un groupe de sécurité qui autorise SSH les connexions entre les nœuds du cluster.

Réservations de capacité en AWS PCS

Vous pouvez réserver des EC2 capacités Amazon dans une zone de disponibilité spécifique et pour une durée spécifique à l'aide de réservations de capacité à la demande ou de blocs de EC2 capacité afin de vous assurer de disposer de la capacité de calcul nécessaire lorsque vous en avez besoin.

> **Note**
>
> AWS PCSprend en charge les réservations de capacité à la demande (ODCR) mais ne prend pas actuellement en charge les blocs de capacité pour le ML.

Utilisation ODCRs avec AWS PCS

Vous pouvez choisir le mode de AWS PCS consommation de vos instances réservées. Si vous créez une ouvertureODCR, toutes les instances correspondantes lancées par votre compte AWS PCS ou tout autre processus associé à celui-ci sont prises en compte dans la réservation. Avec un identifiant de réservation cibléODCR, seules les instances lancées avec le numéro de réservation spécifique sont prises en compte dans la réservation. Pour les charges de travail sensibles au facteur temps, les tâches ciblées ODCRs sont plus courantes.

Vous pouvez configurer un groupe de nœuds de AWS PCS calcul pour utiliser une cible ODCR en l'ajoutant à un modèle de lancement. Voici les étapes à suivre pour ce faire :

1. Créez une réservation de capacité ciblée à la demande (ODCR).
2. Ajoutez-le ODCR à un groupe de réservation de capacité.
3. Associez le groupe de réservation de capacité à un modèle de lancement.
4. Créez ou mettez à jour un groupe de nœuds de AWS PCS calcul pour utiliser le modèle de lancement.

Exemple : réservez et utilisez des instances hpc6a.48xlarge avec un ODCR

Cet exemple de commande crée une cible ODCR pour 32 instances hpc6a.48xlarge. Pour lancer les instances réservées dans un groupe de placement, ajoutez `--placement-group-arn` à la commande. Vous pouvez définir une date de fin avec `--end-date` et`--end-date-type`, sinon, la réservation se poursuivra jusqu'à ce qu'elle soit annulée manuellement.

```
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
```

Le résultat de cette commande sera un ARN pour le nouveauODCR. Pour utiliser le ODCR with AWS PCS, il doit être ajouté à un groupe de réservation de capacité. C'est parce qu'il AWS PCS ne prend pas en charge les individusODCRs. Pour plus d'informations, consultez la section Groupes de réservation de capacité dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.

Voici comment ajouter le ODCR à un groupe de réservation de capacité nomméEXAMPLE-CR-GROUP.

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

Une fois ODCR créé et ajouté à un groupe de réservation de capacité, il est désormais possible de le connecter à un groupe de nœuds de AWS PCS calcul en l'ajoutant à un modèle de lancement. Voici un exemple de modèle de lancement qui fait référence au groupe de réservation de capacité.

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

Enfin, créez ou mettez à jour un groupe de nœuds de AWS PCS calcul pour utiliser les instances hpc6a.48xlarge et utilisez le modèle de lancement qui y fait référence dans son groupe de réservation de ODCR capacité. Pour un groupe de nœuds statique, définissez les instances minimale et maximale en fonction de la taille de la réservation (32). Pour un groupe de nœuds dynamique, définissez le nombre minimum d'instances sur 0 et le maximum jusqu'à la taille de réservation.

Cet exemple est une implémentation simple d'un nœud configuré pour un groupe de nœuds de calcul. ODCR Mais, AWS PCS prend en charge de nombreux autres modèles. Par exemple, vous pouvez subdiviser un grand groupe ODCR ou un groupe de réservation de capacité entre plusieurs groupes de nœuds de calcul. Vous pouvez également utiliser ODCRs ce qu'un autre AWS compte a créé et partagé avec le vôtre. La principale contrainte est qu'il doit ODCRs toujours être contenu dans un groupe de réservation de capacité.

Pour plus d'informations, consultez la section Réservations de capacité à la demande et blocs de capacité pour le ML dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.

Paramètres utiles du modèle de lancement

Cette section décrit certains paramètres du modèle de lancement qui peuvent être très utiles avec AWS PCS.

Activez la CloudWatch surveillance détaillée

Vous pouvez activer la collecte de CloudWatch métriques à un intervalle plus court à l'aide d'un paramètre de modèle de lancement.

AWS Management Console

Sur les pages de console permettant de créer ou de modifier des modèles de lancement, cette option se trouve dans la section Détails avancés. Définissez la CloudWatch surveillance détaillée sur Activer.

YAML

```
Monitoring:
    Enabled: True
```

JSON

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

Pour plus d'informations, consultez Activer ou désactiver la surveillance détaillée de vos instances dans le guide de l'utilisateur Amazon Elastic Compute Cloud pour les instances Linux.

Service de métadonnées d'instance, version 2 (IMDSv2)

L'utilisation IMDS de la version 2 avec EC2 les instances améliore considérablement la sécurité et contribue à atténuer les risques potentiels associés à l'accès aux métadonnées des instances dans AWS les environnements.

AWS Management Console

Sur les pages de console permettant de créer ou de modifier des modèles de lancement, cette option se trouve dans la section Détails avancés. Définissez les métadonnées accessibles sur Activé, la version des métadonnées sur V2 uniquement (jeton requis) et la limite de sauts de réponse des métadonnées sur 4.

YAML

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

JSON

```
{
```

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

.

AWS PCSfiles d'attente

Une AWS PCS file d'attente est une abstraction légère de l'implémentation native d'une file de travail par le planificateur. Dans le cas de Slurm, une AWS PCS file d'attente est équivalente à une partition Slurm.

Les utilisateurs soumettent les tâches à une file d'attente où elles résident jusqu'à ce qu'elles puissent être planifiées pour s'exécuter sur des nœuds fournis par un ou plusieurs groupes de nœuds de calcul. Un AWS PCS cluster peut comporter plusieurs files d'attente de tâches. Par exemple, vous pouvez créer une file d'attente qui utilise les instances Amazon EC2 On-Demand pour les tâches hautement prioritaires et une autre file d'attente qui utilise les instances Amazon EC2 Spot pour les tâches peu prioritaires.

Rubriques

- Création d'une file d'attente dans AWS PCS
- Mettre à jour une AWS PCS file d'attente
- Supprimer une file d'attente dans AWS PCS

Création d'une file d'attente dans AWS PCS

Cette rubrique fournit une vue d'ensemble des options disponibles et décrit les éléments à prendre en compte lors de la création d'une file d'attente dans AWS PCS.

Prérequis

- Un AWS PCS cluster : les files d'attente ne peuvent être créées qu'en association avec un PCS cluster spécifique.

- Un ou plusieurs groupes de nœuds de AWS PCS calcul : une file d'attente doit être associée à au moins un groupe de nœuds de PCS calcul.

Pour créer une file d'attente dans AWS PCS

Vous pouvez créer une file d'attente à l'aide du AWS Management Console ou du AWS CLI.

AWS Management Console

Pour créer une file d'attente à l'aide de la console

1. Ouvrez la AWS PCS console à l'adresse `https://console.aws.amazon.com/pcs/home#/clusters`
2. Sélectionnez le cluster dans lequel vous souhaitez créer une file d'attente. Accédez à Files d'attente et choisissez Créer une file d'attente.
3. Dans la section Configuration de la file d'attente, indiquez les valeurs suivantes :
    a. Nom de la file d'attente : nom de votre file d'attente. Un nom ne peut contenir que des caractères alphanumériques (sensibles à la casse) et des traits d'union. Il doit commencer par un caractère alphabétique et ne doit pas dépasser 25 caractères. Le nom doit être unique au sein du cluster.
    b. Groupes de nœuds de calcul : sélectionnez un ou plusieurs groupes de nœuds de calcul pour desservir cette file d'attente. Un groupe de nœuds de calcul peut être associé à plusieurs files d'attente.
4. (Facultatif) Sous Balises, ajoutez des balises à votre AWS PCS file d'attente
5. Choisissez Créez une file d'attente. Le champ État indiquera Création pendant la configuration de la file d'attente. La création d'une file d'attente peut prendre plusieurs minutes.

Étape suivante recommandée

- Soumettre une tâche à votre nouvelle file d'attente

AWS CLI

Pour créer une file d'attente à l'aide de AWS CLI

Créez votre file d'attente avec la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :

1. Remplacez *region-code* avec la AWS région dans laquelle vous souhaitez créer votre cluster.
2. Remplacez *my-queue* avec le nom de votre file d'attente. Un nom ne peut contenir que des caractères alphanumériques (sensibles à la casse) et des traits d'union. Il doit commencer par un caractère alphabétique et ne doit pas dépasser 25 caractères. Le nom doit être unique au sein du cluster.
3. Remplacez *my-cluster* avec le nom ou celui clusterId de votre cluster.
4. Remplacez la valeur pour `computeNodeGroupId` par votre propre identifiant de groupe de nœuds de calcul. Notez que vous ne pouvez pas spécifier les noms des groupes de nœuds de calcul lors de la création d'une file d'attente.

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

La création de la file d'attente peut prendre plusieurs minutes. Vous pouvez demander l'état de votre file d'attente à l'aide de la commande suivante. Vous ne pourrez pas soumettre de tâches à la file d'attente tant que son statut n'aura pas été atteint`ACTIVE`.

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

Étape suivante recommandée

- Soumettre une tâche à votre nouvelle file d'attente

Mettre à jour une AWS PCS file d'attente

Cette rubrique fournit une vue d'ensemble des options disponibles et décrit les éléments à prendre en compte lors de la mise à jour AWS PCS d'une file d'attente.

Considérations relatives à la mise à jour d'une AWS PCS file

Les mises à jour de file d'attente n'auront aucun impact sur les tâches en cours, mais le cluster risque de ne pas être en mesure d'accepter de nouvelles tâches pendant la mise à jour de la file d'attente.

Pour mettre à jour un groupe AWS PCS de nœuds de calcul

Vous pouvez mettre à jour un groupe de nœuds à l'aide de la console AWS de gestion ou du AWSCLI.

AWS Management Console

Pour mettre à jour une file d'attente

1. Ouvrez la AWS PCS console à l'adresse `https://console.aws.amazon.com/pcs/home#/clusters`
2. Sélectionnez le cluster dans lequel vous souhaitez mettre à jour une file d'attente.
3. Accédez aux files d'attente, accédez à la file que vous souhaitez mettre à jour, puis sélectionnez Modifier.
4. Dans la section de configuration de la file d'attente, mettez à jour l'une des valeurs suivantes :
   - Groupes de nœuds : ajoutez ou supprimez des groupes de nœuds de calcul associés à la file d'attente.
   - Balises : ajoutez ou supprimez des balises pour la file d'attente.
5. Choisissez Mettre à jour. Le champ État affichera la mise à jour pendant que les modifications sont appliquées.

> ⚠ Important
>
> Les mises à jour des files d'attente peuvent prendre plusieurs minutes.

AWS CLI

Pour mettre à jour une file d'attente

1. Mettez à jour votre file d'attente avec la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :

   a. Remplacez *region-code* avec Région AWS celui dans lequel vous souhaitez créer votre cluster.

   b. Remplacez *my-queue* avec le nom ou `computeNodeGroupId` pour votre file d'attente.

   c. Remplacez *my-cluster* avec le nom ou celui `clusterId` de votre cluster.

   d. Pour modifier les associations de groupes de nœuds de calcul, fournissez une liste mise à jour pour`--compute-node-group-configurations`.

      - Par exemple, pour ajouter un deuxième groupe de nœuds de calcul `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 mise à jour de la file d'attente peut prendre plusieurs minutes. Vous pouvez demander l'état de votre file d'attente à l'aide de la commande suivante. Vous ne pourrez pas soumettre de tâches à la file d'attente tant que son statut n'aura pas été atteint`ACTIVE`.

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

Prochaines étapes recommandées

- Soumettez une tâche à votre file d'attente mise à jour.

Supprimer une file d'attente dans AWS PCS

Cette rubrique fournit une vue d'ensemble de la procédure de suppression d'une file d'attente dans AWS PCS.

Considérations relatives à la suppression d'une file d'attente

- Si des tâches sont en cours d'exécution dans la file d'attente, elles seront interrompues par le planificateur lorsque la file d'attente sera supprimée. Les tâches en attente dans la file d'attente seront annulées. Envisagez d'attendre que les tâches de la file d'attente soient terminées ou de les arrêter/annuler manuellement à l'aide des commandes natives du planificateur (comme `scancel` pour Slurm).

Supprimer la file d'attente

Vous pouvez utiliser le AWS Management Console ou AWS CLI pour supprimer une file d'attente.

AWS Management Console

Pour supprimer une file d'attente

1. Ouvrez la AWS PCSconsole.
2. Sélectionnez le cluster de la file d'attente.
3. Accédez à Files d'attente et sélectionnez la file d'attente à supprimer.
4. Sélectionnez Delete (Supprimer).
5. Le champ État s'affiche`Deleting`. Cela peut prendre plusieurs minutes.

> **Note**
>
> Vous pouvez utiliser les commandes natives de votre planificateur pour confirmer que la file d'attente est supprimée. Par exemple, utilisez `sinfo` ou `squeue` pour Slurm.

AWS CLI

Pour supprimer une file d'attente

- Utilisez la commande suivante pour supprimer une file d'attente, avec les remplacements suivants :
  - Remplacez *region-code* avec le nom dans lequel se trouve Région AWS votre cluster.
  - Remplacez *my-queue* avec le nom ou l'ID de votre file d'attente.
  - Remplacez *my-cluster* avec le nom ou l'ID de votre cluster.

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

La suppression de la file d'attente peut prendre plusieurs minutes.

> **Note**
>
> Vous pouvez utiliser les commandes natives de votre planificateur pour confirmer que la file d'attente est supprimée. Par exemple, utilisez `sinfo` ou `squeue` pour Slurm.

AWS PCSnœuds de connexion

Un AWS PCS cluster a généralement besoin d'au moins un nœud de connexion pour prendre en charge l'accès interactif et la gestion des tâches. Un moyen d'y parvenir consiste à utiliser un groupe de nœuds de AWS PCS calcul statique configuré pour la capacité du nœud de connexion. Vous pouvez également configurer une EC2 instance autonome pour qu'elle fasse office de nœud de connexion.

Rubriques

- Utilisation d'un groupe AWS PCS de nœuds de calcul pour fournir des nœuds de connexion
- Utilisation d'instances autonomes comme nœuds de AWS PCS connexion

Utilisation d'un groupe AWS PCS de nœuds de calcul pour fournir des nœuds de connexion

Cette rubrique fournit une vue d'ensemble des options de configuration suggérées et décrit les éléments à prendre en compte lorsque vous utilisez un groupe de nœuds de AWS PCS calcul pour fournir un accès permanent et interactif à votre cluster.

Création d'un groupe AWS PCS de nœuds de calcul pour les nœuds de connexion

Sur le plan opérationnel, cela n'est pas très différent de la création d'un groupe de nœuds de calcul normal. Cependant, certains choix de configuration clés sont à effectuer :

- Définissez une configuration de dimensionnement statique pour au moins une EC2 instance du groupe de nœuds de calcul.
- Choisissez l'option d'achat à la demande pour éviter que vos instances ne soient récupérées.
- Choisissez un nom informatif pour le groupe de nœuds de calcul, tel que login.
- Si vous souhaitez que les instances du nœud de connexion soient accessibles en dehors de vousVPC, pensez à utiliser un sous-réseau public.
- Si vous avez l'intention SSH d'autoriser l'accès, le modèle de lancement doit disposer d'un groupe de sécurité qui expose le SSH port aux adresses IP de votre choix.
- Le profil d'IAMinstance ne doit comporter que les AWS autorisations que vous souhaitez accorder à vos utilisateurs finaux. Consultez IAMprofils d'instance pour AWS Parallel Computing Service pour plus de détails.
- Envisagez AWS d'autoriser le gestionnaire de session Systems Manager à gérer vos instances de connexion.
- Envisagez de restreindre l'accès aux AWS informations d'identification de l'instance aux seuls utilisateurs administratifs
- Sélectionnez des types d'instance moins coûteux que pour les groupes de nœuds de calcul classiques, car le ou les nœuds de connexion fonctionneront en continu.
- Utilisez le même (ou un dérivé) AMI que pour vos autres groupes de nœuds de calcul afin de garantir que le même logiciel est installé sur toutes les instances. Pour plus d'informations sur la personnalisationAMIs, voir Amazon Machine Images (AMIs) pour AWS PCS
- Configurez les mêmes montages de système de fichiers réseau (AmazonEFS, Amazon FSx pour Lustre, etc.) sur vos nœuds de connexion que sur vos instances de calcul. Pour de plus amples informations, veuillez consulter Utilisation de systèmes de fichiers réseau avec AWS PCS.

Accédez à vos nœuds de connexion

Une fois que votre nouveau groupe de nœuds de calcul atteint ACTIVE son statut, vous pouvez trouver les EC2 instances qu'il a créées et vous y connecter. Pour de plus amples informations, veuillez consulter Recherche d'instances de groupes de nœuds de calcul dans AWS PCS.

Mise à jour d'un groupe de nœuds de AWS PCS calcul pour les nœuds de connexion

Vous pouvez mettre à jour un groupe de nœuds de connexion à l'aide de UpdateComputeNodeGroup. Dans le cadre du processus de mise à jour du groupe de nœuds, les instances en cours d'exécution seront remplacées. Notez que cela interrompra toutes les sessions utilisateur ou tous les processus actifs sur l'instance. Les tâches Slurm en cours d'exécution ou mises en file d'attente ne seront pas affectées. Pour de plus amples informations, veuillez consulter Mise à jour d'un groupe AWS PCS de nœuds de calcul.

Vous pouvez également modifier le modèle de lancement utilisé par votre groupe de nœuds de calcul. Vous devez l'utiliser UpdateComputeNodeGroup pour appliquer le modèle de lancement mis à jour au groupe de nœuds de calcul. EC2Les nouvelles instances lancées dans le groupe de nœuds de calcul utilisent le modèle de lancement mis à jour. Pour de plus amples informations, veuillez consulter Utilisation des modèles EC2 de lancement Amazon avec AWS PCS.

Suppression d'un groupe AWS PCS de nœuds de calcul pour les nœuds de connexion

Vous pouvez mettre à jour un groupe de nœuds de connexion à l'aide du mécanisme de suppression du groupe de nœuds de calcul dans AWS PCS. Les instances en cours d'exécution seront interrompues dans le cadre de la suppression du groupe de nœuds. Notez que cela interrompra toutes les sessions utilisateur ou tous les processus actifs sur l'instance. Les tâches Slurm en cours d'exécution ou mises en file d'attente ne seront pas affectées. Pour de plus amples informations, veuillez consulter Suppression d'un groupe de nœuds de calcul dans AWS PCS.

Utilisation d'instances autonomes comme nœuds de AWS PCS connexion

Vous pouvez configurer des EC2 instances indépendantes pour interagir avec le planificateur Slurm d'un AWS PCS cluster. Cela est utile pour créer des nœuds de connexion, des postes de travail ou des hôtes de gestion de flux de travail dédiés qui fonctionnent avec des AWS PCS clusters mais fonctionnent en dehors de la AWS PCS gestion. Pour ce faire, chaque instance autonome doit :

1. Installez une version compatible du logiciel Slurm.
2. Être capable de se connecter au point de terminaison Slurmctld du AWS PCS cluster.

3. Configurez correctement les démons Slurm Auth et Cred Kiosk (`sackd`) avec le point de terminaison et le secret du AWS PCS cluster. Pour plus d'informations, consultez sackd dans la documentation de Slurm.

Ce didacticiel vous aide à configurer une instance indépendante qui se connecte à un AWS PCS cluster.

Table des matières

Étape 1 — Récupérez l'adresse et le secret du AWS PCS cluster cible

Récupérez les détails du AWS PCS cluster cible à l' AWS CLI aide de la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :

- Remplacez *`region-code`* avec l' Région AWS endroit où s'exécute le cluster cible.
- Remplacez *`cluster-ident`* avec le nom ou l'identifiant du cluster cible

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

La commande renverra une sortie similaire à celle de cet exemple.

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

Dans cet exemple, le point de terminaison du contrôleur Slurm du cluster possède une adresse IP de `10.3.149.220` et s'exécute sur le port. `6817` Il `secretArn` sera utilisé dans les étapes ultérieures pour récupérer le secret du cluster. L'adresse IP et le port seront utilisés ultérieurement pour configurer le `sackd` service.

Étape 2 — Lancer une EC2 instance

Pour lancer une instance EC2

1. Ouvrez la EC2console Amazon.
2. Dans le volet de navigation, choisissez Instances, puis Launch Instances (Lancer des instances) pour ouvrir le nouvel assistant de lancement d'instance.

3. (Facultatif) Dans la section Nom et balises, saisissez un nom pour l'instance, par exemple`PCS-LoginNode`. Le nom est attribué à l'instance en tant qu'identification de ressource (`Name=PCS-LoginNode`).
4. Dans la section Images de l'application et du système d'exploitation, sélectionnez un AMI pour l'un des systèmes d'exploitation pris en charge par AWS PCS. Pour de plus amples informations, veuillez consulter Systèmes d'exploitation pris en charge.
5. Dans la section Type d'instance, sélectionnez un type d'instance pris en charge. Pour de plus amples informations, veuillez consulter Types d'instance pris en charge.
6. Dans la section Paire de clés, sélectionnez la paire de SSH clés à utiliser pour l'instance.
7. Dans la section Paramètres réseau :
   - Choisissez Modifier.
     i. Sélectionnez celui VPC de votre AWS PCS cluster.
     ii. Pour Pare-feu (groupes de sécurité), choisissez Sélectionner un groupe de sécurité existant.
       A. Sélectionnez un groupe de sécurité qui autorise le trafic entre l'instance et le contrôleur Slurm du AWS PCS cluster cible. Pour de plus amples informations, veuillez consulter Exigences et considérations relatives aux groupes de sécurité.
       B. (Facultatif) Sélectionnez un groupe de sécurité qui autorise l'SSHaccès entrant à votre instance.
8. Dans la section Stockage, configurez les volumes de stockage selon vos besoins. Assurez-vous de configurer suffisamment d'espace pour installer les applications et les bibliothèques afin d'activer votre cas d'utilisation.
9. Sous Avancé, choisissez un IAM rôle qui autorise l'accès au secret du cluster. Pour de plus amples informations, veuillez consulter Obtenez le secret du cluster Slurm.
10. Dans le volet Résumé, choisissez Launch instance.

Étape 3 — Installation de Slurm sur l'instance

Lorsque l'instance est lancée et devient active, connectez-vous à celle-ci en utilisant le mécanisme de votre choix. Utilisez le programme d'installation de Slurm fourni par AWS pour installer Slurm sur l'instance. Pour de plus amples informations, veuillez consulter Installateur Slurm.

Téléchargez le programme d'installation de Slurm, décompressez-le et utilisez le `installer.sh` script pour installer Slurm. Pour de plus amples informations, veuillez consulter Étape 3 — Installation de Slurm.

Étape 4 — Récupérez et stockez le secret du cluster

Ces instructions nécessitent le AWS CLI. Pour plus d'informations, voir Installation ou mise à jour vers la dernière version du AWS CLI dans le guide de AWS Command Line Interface l'utilisateur de la version 2.

Enregistrez le secret du cluster à l'aide des commandes suivantes.

- Créez le répertoire de configuration pour Slurm.

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

- Récupérez, décodez et stockez le secret du cluster. Avant d'exécuter cette commande, remplacez *region-code* avec la région dans laquelle le cluster cible est exécuté, et remplacez *secret-arn* avec la valeur `secretArn` récupérée à l'étape 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**
>
> Dans un environnement multi-utilisateurs, tout utilisateur ayant accès à l'instance peut être en mesure de récupérer le secret du cluster s'il peut accéder au service de métadonnées de l'instance ()IMDS. Cela pourrait à son tour leur permettre de se faire passer pour d'autres utilisateurs. Envisagez de restreindre l'accès IMDS aux utilisateurs root ou administratifs uniquement. Vous pouvez également envisager d'utiliser un mécanisme différent qui ne repose pas sur le profil de l'instance pour récupérer et configurer le secret.

- Définissez la propriété et les autorisations sur le fichier clé de Slurm.

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

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

> **Note**
>
> La clé Slurm doit appartenir à l'utilisateur et au groupe sous lesquels le `sackd` service s'exécute.

Étape 5 — Configuration de la connexion au AWS PCS cluster

Pour établir une connexion au AWS PCS cluster, lancez-le `sackd` en tant que service système en suivant ces étapes.

1. Configurez le fichier d'environnement du `sackd` service à l'aide de la commande suivante. Avant d'exécuter la commande, remplacez *`ip-address`* and *`port`* avec les valeurs extraites des points de terminaison à l'étape 1.

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

2. Créez un fichier `systemd` de service pour gérer le `sackd` processus.

   ```
   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. Définissez la propriété du fichier `sackd` de service.

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

4. Activez le `sackd` service.

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

5. Lancez le service `sackd`.

```
sudo systemctl start sackd
```

Étape 6 — (Facultatif) Testez la connexion

Vérifiez que le `sackd` service est en cours d'exécution. Vous trouverez ci-après un exemple de sortie. S'il y a des erreurs, elles apparaissent généralement ici.

```
[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
```

Vérifiez que les connexions au cluster fonctionnent à l'aide des commandes du client Slurm telles que `sinfo` et. `squeue` Voici un exemple de sortie 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]
```

Vous devriez également être en mesure de soumettre des offres d'emploi. Par exemple, une commande similaire à cet exemple lancerait une tâche interactive sur un nœud du cluster.

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

AWS PCSRéseautage

Votre AWS PCS cluster est créé dans un AmazonVPC. Ce chapitre inclut les rubriques suivantes concernant la mise en réseau pour le planificateur et les nœuds de votre cluster.

À l'exception du choix d'un sous-réseau dans lequel lancer les instances, vous devez utiliser des modèles de EC2 lancement pour configurer la mise en réseau des groupes de nœuds de AWS PCS calcul. Pour plus d'informations sur les modèles de lancement, consultezUtilisation des modèles EC2 de lancement Amazon avec AWS PCS.

Rubriques

- AWS PCSVPCexigences et considérations relatives aux sous-réseaux
- Création d'un VPC pour votre AWS PCS cluster
- Groupes de sécurité dans AWS PCS
- Plusieurs interfaces réseau dans AWS PCS
- Groupes de placement pour les EC2 instances dans AWS PCS
- Utilisation d'Elastic Fabric Adapter (EFA) avec AWS PCS

AWS PCSVPCexigences et considérations relatives aux sous-réseaux

Lorsque vous créez un AWS PCS cluster, vous y spécifiez VPC un sous-réseau. VPC Cette rubrique fournit une vue d'ensemble AWS PCS des exigences et considérations spécifiques relatives au VPC ou aux sous-réseaux que vous utilisez avec votre cluster. Si vous n'en avez pas VPC à utiliser AWS PCS, vous pouvez en créer un à l'aide d'un AWS CloudFormation modèle AWS fourni. Pour plus d'informationsVPCs, consultez la section Virtual Private Clouds (VPC) dans le guide de VPC l'utilisateur Amazon.

VPCexigences et considérations

Lorsque vous créez un cluster, le cluster VPC que vous spécifiez doit répondre aux exigences et considérations suivantes :

- Vous VPC devez disposer d'un nombre suffisant d'adresses IP disponibles pour le cluster, les nœuds et les autres ressources de cluster que vous souhaitez créer. Pour plus d'informations, consultez la section Adressage IP pour vos sous-réseaux VPCs et sous-réseaux dans le guide de VPC l'utilisateur Amazon.
- Ils VPC doivent avoir un DNS nom d'hôte et un support DNS de résolution. Sinon, les nœuds ne peuvent pas enregistrer le cluster client. Pour plus d'informations, consultez DNSles attributs correspondants VPC dans le guide de VPC l'utilisateur Amazon.
- Ils VPC peuvent nécessiter l'utilisation de VPC points de terminaison AWS PrivateLink pour pouvoir contacter le AWS PCSAPI. Pour plus d'informations, consultez Connect your VPC to services using AWS PrivateLink dans le guide de VPC l'utilisateur Amazon.

Exigences et considérations requises pour les sous-réseaux

Lorsque vous créez un cluster Slurm, il AWS PCS crée une interface réseau élastique (ENI) dans le sous-réseau que vous avez spécifié. Cette interface réseau permet la communication entre le contrôleur du planificateur et le client. VPC L'interface réseau permet également à Slurm de communiquer avec les composants déployés dans le compte client. Vous ne pouvez spécifier le sous-réseau d'un cluster qu'au moment de sa création.

Exigences requises pour les sous-réseaux des clusters

Le sous-réseau que vous spécifiez lorsque vous créez un cluster doit répondre aux exigences suivantes :

- Le sous-réseau doit avoir au moins une adresse IP à utiliser par AWS PCS.
- Le sous-réseau ne peut pas résider dans AWS Outposts une zone AWS locale ou dans une telle zone. AWS Wavelength
- Le sous-réseau peut être public ou privé. Nous vous recommandons de spécifier un sous-réseau privé, si possible. Un sous-réseau public est un sous-réseau avec une table de routage qui inclut une route vers une passerelle Internet ; un sous-réseau privé est un sous-réseau avec une table de routage qui n'inclut pas de route vers une passerelle Internet.

Exigences requises pour les sous-réseaux des nœuds

Vous pouvez déployer des nœuds et d'autres ressources de cluster sur le sous-réseau que vous spécifiez lors de la création de votre AWS PCS cluster, ainsi que sur d'autres sous-réseaux de ce dernier. VPC

Tout sous-réseau sur lequel vous déployez des nœuds et des ressources de cluster doit répondre aux exigences suivantes :

- Vous devez vous assurer que le sous-réseau dispose de suffisamment d'adresses IP disponibles pour déployer tous les nœuds et ressources du cluster.
- Si vous prévoyez de déployer des nœuds sur un sous-réseau public, ce sous-réseau doit attribuer automatiquement IPv4 des adresses publiques.
- Si le sous-réseau sur lequel vous déployez des nœuds est un sous-réseau privé et que sa table de routage n'inclut pas de route vers un périphérique de traduction d'adresses réseau (NAT) (IPv4), ajoutez des VPC points de terminaison AWS PrivateLink à l'aide du client. VPC VPCdes points de terminaison sont nécessaires pour tous les AWS services contactés par les nœuds. Le seul point de terminaison requis est AWS PCS de permettre au nœud d'appeler l'`registerNodeGroupInstances`APIaction.
- L'état du sous-réseau public ou privé n'a aucun impact AWS PCS ; les points de terminaison requis doivent être accessibles.

Création d'un VPC pour votre AWS PCS cluster

Vous pouvez créer un Amazon Virtual Private Cloud (AmazonVPC) pour vos clusters dans AWS Parallel Computing Service (AWS PCS).

Utilisez Amazon VPC pour lancer VPC des ressources sur un réseau virtuel que vous avez défini. Ce réseau virtuel ressemble beaucoup à un réseau traditionnel que vous pourriez exécuter dans votre propre centre de données. Toutefois, il présente l'avantage d'utiliser l'infrastructure évolutive d'Amazon Web Services. Nous vous recommandons de bien comprendre le VPC service Amazon avant de déployer des VPC clusters de production. Pour plus d'informations, consultez Qu'est-ce qu'Amazon VPC ? dans le mode visuel de l'auteur. Guide de VPC l'utilisateur Amazon.

Un PCS cluster, des nœuds et des ressources de support (telles que des systèmes de fichiers et des services d'annuaire) sont déployés au sein de votre AmazonVPC. Si vous souhaitez utiliser un Amazon existant VPC avecPCS, celui-ci doit répondre aux exigences décrites dansAWS

PCSVPCexigences et considérations relatives aux sous-réseaux. Cette rubrique explique comment créer un modèle VPC répondant aux PCS exigences à l'aide d'un AWS AWS CloudFormation modèle fourni. Une fois que vous avez déployé un modèle, vous pouvez consulter les ressources créées par le modèle pour savoir exactement quelles ressources il a créées, et la configuration de ces ressources.

Prérequis

Pour créer un Amazon VPC pourPCS, vous devez disposer des IAM autorisations nécessaires pour créer des VPC ressources Amazon. Ces ressources sont les sous-réseauxVPCs, les groupes de sécurité, les tables de routage et les itinéraires, ainsi qu'Internet et les NAT passerelles. Pour plus d'informations, consultez la section Créer un sous-réseau VPC avec un sous-réseau public dans le guide de l'VPCutilisateur Amazon. Pour consulter la liste complète d'AmazonEC2, consultez la section Actions, ressources et clés de condition pour Amazon EC2 dans le Service Authorization Reference.

Créez un Amazon VPC

Créez un VPC en copiant et collant celui qui URL convient à l' Région AWS endroit où vous allez l'utiliserPCS. Vous pouvez également télécharger le AWS CloudFormation modèle et le charger vous-même sur la AWS CloudFormation console.

- USA Est (Virginie du Nord) (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
```

- USA Est (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
```

- USA Ouest (Oregon) (us-west-2)

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

- Modèle uniquement

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

Pour créer un Amazon VPC pour PCS

1. Ouvrez le modèle dans la AWS CloudFormation console.

   > **Note**
   >
   > Elles sont préremplies dans le modèle afin que vous puissiez simplement les laisser comme valeurs par défaut.

2. Sous Fournir un nom de pile, puis Nom de pile, entrez`hpc-networking`.
3. Dans la section Paramètres, entrez les informations suivantes :
   a. Sous VPCCidrBlock, puis entrez `10.3.0.0/16`
   b. Sous les sous-réseaux A :
      i. Puis CidrPublicSubnetA, entrez `10.3.0.0/20`
      ii. Puis CidrPrivateSubnetA, entrez `10.3.128.0/20`
   c. Sous les sous-réseaux B :
      i. Puis CidrPublicSubnetB, entrez `10.3.16.0/20`
      ii. Puis CidrPrivateSubnetA, entrez `10.3.144.0/20`
   d. Sous les sous-réseaux C :
      i. Pour ProvisionSubnetsC, sélectionnez`True`.

      > **Note**
      >
      > Si vous créez un VPC dans une région qui compte moins de trois zones de disponibilité, cette option sera ignorée si elle est définie sur`True`.

      ii. Puis CidrPublicSubnetB, entrez `10.3.32.0/20`
      iii. Puis CidrPrivateSubnetA, entrez `10.3.160.0/20`

4. Sous Capacités, cochez la case « Je reconnais que cela AWS CloudFormation peut créer des IAM ressources ».

Surveillez l'état de la AWS CloudFormation pile. Lorsqu'elle est `CREATE_COMPLETE` atteinte, la VPC ressource est prête à être utilisée.

> **Note**
>
> Pour voir toutes les ressources créées par le AWS CloudFormation modèle, ouvrez la AWS CloudFormation console. Choisissez la pile `hpc-networking`, puis choisissez l'onglet Ressources.

Groupes de sécurité dans AWS PCS

Les groupes de sécurité d'Amazon EC2 agissent comme des pare-feux virtuels pour contrôler le trafic entrant et sortant vers les instances. Utilisez un modèle de lancement pour un groupe de nœuds de AWS PCS calcul afin d'ajouter ou de supprimer des groupes de sécurité à ses instances. Si votre modèle de lancement ne contient aucune interface réseau, utilisez-le `SecurityGroupIds` pour fournir une liste de groupes de sécurité. Si votre modèle de lancement définit des interfaces réseau, vous devez utiliser le `Groups` paramètre pour attribuer des groupes de sécurité à chaque interface réseau. Pour plus d'informations sur les modèles de lancement, consultezUtilisation des modèles EC2 de lancement Amazon avec AWS PCS.

> **Note**
>
> Les modifications apportées à la configuration du groupe de sécurité dans le modèle de lancement concernent uniquement les nouvelles instances lancées après la mise à jour du groupe de nœuds de calcul.

Exigences et considérations relatives aux groupes de sécurité

AWS PCScrée une interface réseau élastique (ENI) entre comptes dans le sous-réseau que vous spécifiez lors de la création d'un cluster. Cela fournit au HPC planificateur, qui s'exécute dans un compte géré par AWS, un chemin pour communiquer avec les EC2 instances lancées par. AWS PCS Pour cela, vous devez fournir un groupe de sécurité ENI qui autorise la communication bidirectionnelle entre le planificateur ENI et vos instances de cluster. EC2

Une méthode simple consiste à créer un groupe de sécurité autoréférencé permissif qui autorise le trafic TCP /IP sur tous les ports entre tous les membres du groupe. Vous pouvez l'associer à la fois au cluster et aux EC2 instances du groupe de nœuds.

Exemple de configuration de groupe de sécurité permissive

Type de règle	Protocoles	Ports	Source	Destination
Entrant	Tous	Tous	Auto-utilisateur	
Sortant	Tous	Tous		0.0.0.0/0
Sortant	Tous	Tous		Auto-utilisateur

Ces règles permettent à tout le trafic de circuler librement entre le contrôleur Slurm et les nœuds, autorisent tout le trafic sortant vers n'importe quelle destination et autorisent le trafic. EFA

Exemple de configuration restrictive d'un groupe de sécurité

Vous pouvez également limiter les ports ouverts entre le cluster et ses nœuds de calcul. Pour le planificateur Slurm, le groupe de sécurité rattaché à votre cluster doit autoriser les ports suivants :

- 6817 — activer les connexions entrantes vers les instances depuis `slurmctld` EC2
- 6818 — active les connexions sortantes depuis et en cours d'`slurmd`exécution `slurmctld` sur les instances EC2

Le groupe de sécurité attaché à vos nœuds de calcul doit autoriser les ports suivants :

- 6817 — active les connexions sortantes vers des instances à `slurmctld` partir EC2 d'instances.
- 6818 — activer les connexions entrantes et sortantes vers `slurmctld` et `slurmd` depuis des instances de groupes `slurmd` de nœuds
- 60001—63000 — connexions entrantes et sortantes entre les instances de groupes de nœuds à prendre en charge `srun`
- EFAtrafic entre les instances de groupes de nœuds. Pour plus d'informations, voir Préparer un groupe de sécurité EFA activé dans le Guide de l'utilisateur pour les instances Linux
- Tout autre trafic inter-nœuds requis par votre charge de travail

Plusieurs interfaces réseau dans AWS PCS

Certaines EC2 instances possèdent plusieurs cartes réseau. Cela leur permet de fournir des performances réseau supérieures, notamment des capacités de bande passante supérieures à 100 Gbit/s et une meilleure gestion des paquets. Pour plus d'informations sur les instances dotées de plusieurs cartes réseau, consultez les interfaces réseau Elastic dans le guide de l'utilisateur Amazon Elastic Compute Cloud.

Configurez des cartes réseau supplémentaires pour les instances d'un groupe de nœuds de AWS PCS calcul en ajoutant des interfaces réseau à son modèle de EC2 lancement. Vous trouverez ci-dessous un exemple de modèle de lancement qui active deux cartes réseau, comme celles que l'on trouve sur une `hpc7a.96xlarge` instance. Notez les informations suivantes :

- Le sous-réseau de chaque interface réseau doit être le même que celui que vous avez choisi lors de la configuration du groupe de nœuds de AWS PCS calcul qui utilisera le modèle de lancement.
- Le périphérique réseau principal, sur lequel les communications réseau de routine telles que SSH HTTPS le trafic auront lieu, est établi en définissant un `DeviceIndex` de`0`. Les autres interfaces réseau ont un `DeviceIndex` de`1`. Il ne peut y avoir qu'une seule interface réseau principale ; toutes les autres interfaces sont secondaires.
- Toutes les interfaces réseau doivent avoir un identifiant unique`NetworkCardIndex`. Il est recommandé de les numéroter de manière séquentielle selon leur définition dans le modèle de lancement.
- Les groupes de sécurité pour chaque interface réseau sont définis à l'aide de`Groups`. Dans cet exemple, un groupe de SSH sécurité entrant (`sg-`*`SshSecurityGroupId`*) est ajouté à l'interface réseau principale, ainsi que le groupe de sécurité permettant les communications au sein du cluster (). `sg-`*`ClusterSecurityGroupId`* Enfin, un groupe de sécurité autorisant les connexions sortantes vers Internet (`sg-`*`InternetOutboundSecurityGroupId`*) est ajouté aux interfaces principale et secondaire.

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

Groupes de placement pour les EC2 instances dans AWS PCS

Vous pouvez utiliser un groupe de placement pour influencer le placement des EC2 instances afin de répondre aux besoins de la charge de travail qui les exécute.

Types de groupes de placement

- Cluster — Regroupe les instances les unes aux autres dans une zone de disponibilité afin d'optimiser les communications à faible latence.
- Partition : répartit les instances sur des partitions logiques pour optimiser la résilience.
- Spread : impose strictement le lancement d'un petit nombre d'instances sur du matériel distinct, ce qui peut également contribuer à la résilience.

Pour plus d'informations, consultez la section Groupes de placement pour vos EC2 instances Amazon dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.

Nous vous recommandons d'inclure un groupe de placement de clusters lorsque vous configurez un groupe de nœuds de AWS PCS calcul pour utiliser Elastic Fabric Adapter (EFA).

Pour créer un groupe de placement de clusters qui fonctionne avec EFA

1. Créez un groupe de placement avec le type de cluster pour le groupe de nœuds de calcul.
   - Utilisez la AWS CLI commande suivante :

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

- Vous pouvez également utiliser un CloudFormation modèle pour créer un groupe de placement. Pour plus d'informations, consultez la section Utilisation des CloudFormation modèles dans le Guide de AWS CloudFormation l'utilisateur. Téléchargez le modèle ci-dessous URL et chargez-le dans la CloudFormation console.

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

2. Incluez le groupe de placement dans le modèle de EC2 lancement du groupe de nœuds de AWS PCS calcul.

Utilisation d'Elastic Fabric Adapter (EFA) avec AWS PCS

Elastic Fabric Adapter (EFA) est une interconnexion réseau avancée à hautes performances AWS que vous pouvez associer à votre EC2 instance pour accélérer le calcul haute performance (HPC) et les applications d'apprentissage automatique. L'activation de vos applications exécutées sur un AWS PCS cluster EFA implique de configurer les instances du groupe de nœuds de AWS PCS calcul à utiliser EFA comme suit.

Table des matières

Installation EFA sur un AWS PCS -compatible AMI

Le EFA pilote du groupe de nœuds de AWS PCS calcul AMI utilisé doit être installé et chargé. Pour plus d'informations sur la création d'une version personnalisée AMI avec EFA le logiciel installé, consultezImages Amazon Machine personnalisées (AMIs) pour AWS PCS.

Identifier les EFA instances activées EC2

Pour être utiliséEFA, tous les types d'instances autorisés pour un groupe de AWS PCS calcul doivent prendre en charge EFA le même nombre de vCPUs (et le cas GPUs échéant). Pour obtenir la liste des instances EFA activées, consultez Elastic Fabric Adapter for HPC and ML workloads on Amazon EC2 dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud. Vous pouvez également utiliser le AWS CLI pour afficher la liste des types d'instances compatiblesEFA. Remplacez *region-code* avec l' Région AWS endroit où vous l'utilisez AWS PCS, par exemple`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
```

Déterminer le nombre d'interfaces réseau disponibles

Certaines EC2 instances disposent de plusieurs cartes réseau. Cela leur permet d'en avoir plusieursEFAs. Pour de plus amples informations, veuillez consulter Plusieurs interfaces réseau dans AWS PCS.

Création d'un groupe de sécurité pour prendre en charge EFA les communications

AWS CLI

Vous pouvez utiliser la AWS CLI commande suivante pour créer un groupe de sécurité qui prend en chargeEFA. La commande génère un ID de groupe de sécurité. Procédez aux remplacements suivants :

- *region-code*— Spécifiez l' Région AWS endroit où vous l'utilisez AWS PCS, par exemple`us-east-1`.
- *vpc-id*— Spécifiez l'ID du pour VPC lequel vous l'utilisez AWS PCS.
- *efa-group-name*— Indiquez le nom que vous avez choisi pour le groupe de sécurité.

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

```
    --region region-code
```

Utilisez les commandes suivantes pour associer des règles de groupe de sécurité entrantes et sortantes. Effectuez le remplacement suivant :

- *efa-secgroup-id*— Indiquez l'ID du groupe de EFA sécurité que vous venez de créer.

```
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

Vous pouvez utiliser un CloudFormation modèle pour créer un groupe de sécurité qui prend en chargeEFA. Téléchargez le modèle ci-dessousURL, puis chargez-le dans la AWS CloudFormation console.

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

Le modèle étant ouvert dans la AWS CloudFormation console, entrez les options suivantes.

- Sous Fournir un nom de pile
  - Sous Nom de la pile, entrez un nom tel que`efa-sg-stack`.
- Sous Paramètres
  - Sous SecurityGroupName, entrez un nom tel que`efa-sg`.
  - Sous VPC, sélectionnez l'VPCendroit où vous allez utiliser AWS PCS.

Terminez la création de la CloudFormation pile et surveillez son état. Lorsqu'il atteint `CREATE_COMPLETE` le groupe EFA de sécurité, il est prêt à être utilisé.

(Facultatif) Créez un groupe de placement

Il est recommandé de lancer toutes les instances utilisées EFA dans un groupe de placement de clusters afin de minimiser la distance physique entre elles. Nous vous recommandons de créer un groupe de placement pour chaque groupe de nœuds de calcul que vous utiliserezEFA. Consultez Groupes de placement pour les EC2 instances dans AWS PCS la section pour créer un groupe de placement pour votre groupe de nœuds de calcul.

Création ou mise à jour d'un modèle de EC2 lancement

EFAles interfaces réseau sont configurées dans le modèle de EC2 lancement d'un groupe de nœuds de AWS PCS calcul. S'il existe plusieurs cartes réseau, plusieurs EFAs peuvent être configurées. Le groupe EFA de sécurité et le groupe de placement facultatif sont également inclus dans le modèle de lancement.

Voici un exemple de modèle de lancement pour les instances dotées de deux cartes réseau, telles que hpc7a.96xlarge. Les instances seront lancées `subnet-`*`SubnetID1`* dans un groupe de placement en clusterpg-*`PlacementGroupId1`*.

Les groupes de sécurité doivent être ajoutés spécifiquement à chaque EFA interface. Chacun EFA a besoin du groupe de sécurité qui active EFA le trafic (`sg-`*`EfaSecGroupId`*). Les autres groupes de sécurité, en particulier ceux qui gèrent le trafic normal tel que SSH ouHTTPS, doivent uniquement être attachés à l'interface réseau principale (désignée par un `DeviceIndex` de`0`). Les modèles de lancement dans lesquels des interfaces réseau sont définies ne permettent pas de définir des groupes de sécurité à l'aide de `SecurityGroupIds` ce paramètre. Vous devez définir une valeur pour `Groups` chaque interface réseau que vous configurez.

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

Création ou mise à jour d'un groupe de nœuds de calcul

Créez ou mettez à jour un groupe de nœuds de AWS PCS calcul avec des instances possédant le même nombre de nœudsvCPUs, la même architecture de processeur et toutes compatiblesEFA. Configurez le groupe de nœuds de calcul pour qu'il utilise le EFA logiciel installé dessus et pour utiliser le modèle de lancement qui configure les interfaces réseau EFA activées. AMI

Test (facultatif) EFA

Vous pouvez démontrer l'EFAactivation de la communication entre deux nœuds d'un groupe de nœuds de calcul en exécutant le `fi_pingpong` programme, qui est inclus dans l'installation du EFA logiciel. Si ce test est réussi, il est probable qu'il EFA soit correctement configuré.

Pour commencer, vous avez besoin de deux instances actives dans le groupe de nœuds de calcul. Si votre groupe de nœuds de calcul utilise une capacité statique, des instances devraient déjà être disponibles. Pour un groupe de nœuds de calcul qui utilise la capacité dynamique, vous pouvez lancer deux nœuds à l'aide de la `salloc` commande. Voici un exemple d'un cluster avec un groupe de nœuds dynamique nommé `hpc7g` associé à une file d'attente nommée`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
```

Découvrez l'adresse IP des deux nœuds alloués à l'aide de`scontrol`. Dans l'exemple qui suit, les adresses sont `10.3.140.69` pour `hpc7g-1` et `10.3.132.211` pour`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
```

Connectez-vous à l'un des nœuds (dans ce cas,`hpc7g-1`) à l'aide de SSH (ou`SSM`). Notez qu'il s'agit d'une adresse IP interne, vous devrez peut-être vous connecter depuis l'un de vos nœuds de connexion si vous en utilisez`SSH`. Sachez également que l'instance doit être configurée avec une SSH clé via le modèle de lancement du groupe de nœuds de calcul.

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

Maintenant, `fi_pingpong` lancez-vous en mode serveur.

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

Connectez-vous à la deuxième instance (`hpc7g-2`).

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

Exécuter `fi_pingpong` en mode client, en se connectant au serveur sur`hpc7g-1`. Vous devriez voir une sortie similaire à l'exemple ci-dessous.

```
% /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
```

(Facultatif) Utilisez un CloudFormation modèle pour créer un modèle de lancement EFA compatible

Comme la configuration comporte plusieurs dépendancesEFA, un CloudFormation modèle a été fourni que vous pouvez utiliser pour configurer un groupe de nœuds de calcul. Il prend en charge les instances dotées d'un maximum de quatre cartes réseau. Pour en savoir plus sur les instances dotées de plusieurs cartes réseau, consultez les interfaces réseau élastiques dans le guide de l'utilisateur Amazon Elastic Compute Cloud.

Téléchargez le CloudFormation modèle ci-dessousURL, puis chargez-le sur la CloudFormation console Région AWS où vous l'utilisez AWS PCS.

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

Le modèle étant ouvert dans la AWS CloudFormation console, entrez les valeurs suivantes. Notez que le modèle fournit certaines valeurs de paramètres par défaut. Vous pouvez les conserver comme valeurs par défaut.

- Sous Fournir un nom de pile
  - Sous Nom de la pile, entrez un nom descriptif. Nous vous recommandons d'incorporer le nom que vous choisirez pour votre groupe de nœuds de AWS PCS calcul, tel que*NODEGROUPNAME*`-efa-lt`.
- Sous Paramètres
  - Sous NumberOfNetworkCards, choisissez le nombre de cartes réseau dans les instances qui figureront dans votre groupe de nœuds.
  - Sous VpcId, choisissez l'VPCendroit où votre AWS PCS cluster est déployé.
  - Sous NodeGroupSubnetId, choisissez le sous-réseau de votre cluster VPC où les instances EFA activées seront lancées.
  - Sous PlacementGroupName, laissez le champ vide pour créer un nouveau groupe de placement de clusters pour le groupe de nœuds. Si vous souhaitez utiliser un groupe de placement existant, entrez son nom ici.
  - Sous ClusterSecurityGroupId, choisissez le groupe de sécurité que vous utilisez pour autoriser l'accès aux autres instances du cluster et au AWS PCSAPI. De nombreux clients choisissent le groupe de sécurité par défaut dans leur clusterVPC.
  - Sous SshSecurityGroupId, indiquez l'ID d'un groupe de sécurité que vous utilisez pour autoriser l'SSHaccès entrant aux nœuds de votre cluster.
  - Pour SshKeyName, sélectionnez la paire de SSH clés pour accéder aux nœuds de votre cluster.
  - Pour LaunchTemplateName, entrez un nom descriptif pour le modèle de lancement, tel que*NODEGROUPNAME*`-efa-lt`. Le nom doit être unique Compte AWS à l' Région AWS endroit où vous allez l'utiliser AWS PCS.
- Sous Capacités
  - Cochez la case « Je reconnais que cela AWS CloudFormation peut créer des IAM ressources ».

Surveillez l'état de la CloudFormation pile. Lorsqu'il atteint, `CREATE_COMPLETE` le modèle de lancement est prêt à être utilisé. Utilisez-le avec un groupe de nœuds de AWS PCS calcul, comme décrit ci-dessus dansCréation ou mise à jour d'un groupe de nœuds de calcul.

Utilisation de systèmes de fichiers réseau avec AWS PCS

Vous pouvez associer des volumes de stockage réseau à des nœuds lancés dans un groupe de nœuds de calcul AWS Parallel Computing Service (AWS PCS) afin de fournir un emplacement permanent où les données et les fichiers peuvent être écrits et accessibles. Vous pouvez utiliser les volumes fournis par les AWS services. Les volumes incluent Amazon Elastic File System (AmazonEFS), Amazon FSx for NetApp ONTAP, Amazon FSx for Open ZFS, Amazon FSx for Lustre et Amazon File Cache. Vous pouvez également utiliser des volumes autogérés, tels que des NFS serveurs.

Cette rubrique présente des considérations et des exemples relatifs à l'utilisation de systèmes de fichiers en réseau avec AWS PCS.

Considérations relatives à l'utilisation de systèmes de fichiers réseau

Les détails de mise en œuvre des différents systèmes de fichiers sont différents, mais il existe des considérations communes.

- Le logiciel du système de fichiers approprié doit être installé sur l'instance. Par exemple, pour utiliser Amazon FSx pour Lustre, le Lustre package approprié doit être présent. Cela peut être accompli en l'incluant dans le groupe de nœuds de calcul AMI ou en utilisant un script qui s'exécute au démarrage de l'instance.
- Il doit exister une route réseau entre le volume de stockage partagé et les instances du groupe de nœuds de calcul.
- Les règles du groupe de sécurité relatives au volume de stockage partagé et aux instances du groupe de nœuds de calcul doivent autoriser les connexions aux ports concernés.
- Vous devez maintenir un espace de noms POSIX d'utilisateur et de groupe cohérent entre les ressources qui accèdent aux systèmes de fichiers. Dans le cas contraire, les tâches et les processus interactifs exécutés sur votre PCS cluster risquent de rencontrer des erreurs d'autorisation.
- Les montages de systèmes de fichiers sont effectués à l'aide de modèles de EC2 lancement. Des erreurs ou des délais d'attente lors du montage d'un système de fichiers réseau peuvent empêcher les instances d'être disponibles pour exécuter des tâches. Ceci, à son tour, peut entraîner des coûts imprévus. Pour plus d'informations sur le débogage des modèles de lancement, consultezUtilisation des modèles EC2 de lancement Amazon avec AWS PCS.

Exemples de montages réseau

Vous pouvez créer des systèmes de fichiers à l'aide d'AmazonEFS, Amazon FSx for Lustre, Amazon FSx for Open ZFS et Amazon File Cache. Développez la section correspondante ci-dessous pour voir un exemple de chaque montage réseau.

Amazon EFS

Configuration du système de fichiers

Créez un système de EFS fichiers Amazon. Assurez-vous qu'il dispose d'une cible de montage dans chaque zone de disponibilité où vous lancerez des instances de groupes de nœuds de PCS calcul. Assurez-vous également que chaque cible de montage est associée à un groupe de sécurité qui autorise l'accès entrant et sortant depuis les instances du groupe de nœuds de PCS calcul. Pour plus d'informations, consultez Mount targets and security groups dans le manuel Amazon Elastic File System User Guide.

Modèle de lancement

Ajoutez le ou les groupes de sécurité de la configuration de votre système de fichiers au modèle de lancement que vous utiliserez pour le groupe de nœuds de calcul.

Incluez les données utilisateur qui utilisent le `cloud-config` mécanisme de montage du système de EFS fichiers Amazon. Remplacez les valeurs suivantes dans ce script par vos propres informations :

- *`mount-point-directory`*— Le chemin de chaque instance sur laquelle vous allez monter Amazon EFS
- *`filesystem-id`*— L'ID du système de fichiers pour le système de EFS fichiers

```
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 pour Lustre

Configuration du système de fichiers

Créez un système de fichiers FSx pour Lustre dans l'VPCendroit où vous allez l'utiliser AWS PCS. Pour minimiser les transferts entre zones, déployez dans un sous-réseau de la même zone de disponibilité où vous lancerez la majorité de vos instances de groupes de nœuds de PCS calcul. Assurez-vous que le système de fichiers est associé à un groupe de sécurité qui autorise l'accès entrant et sortant depuis les instances du groupe de nœuds de PCS calcul. Pour plus d'informations sur les groupes de sécurité, consultez la section Contrôle d'accès au système de fichiers avec Amazon VPC dans le guide de l'utilisateur d'Amazon FSx for Lustre.

Modèle de lancement

Incluez les données utilisateur utilisées `cloud-config` pour monter le système de fichiers FSx for Lustre. Remplacez les valeurs suivantes dans ce script par vos propres informations :

- *mount-point-directory*— Le chemin d'une instance sur laquelle vous souhaitez effectuer le montage FSx pour Lustre
- *filesystem-id*— L'ID du système de fichiers pour le système de fichiers FSx for Lustre
- *mount-name*— Le nom de montage du système de fichiers FSx for Lustre
- *region-code*— L' Région AWS endroit où le système de fichiers FSx for Lustre est déployé (doit être le même que celui de votre AWS PCS système)
- (Facultatif) *latest* — Toute version de Lustre supportée par FSx for Lustre

```
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 pour Open ZFS

Configuration du système de fichiers

Créez un système de ZFS fichiers FSx for Open dans l'VPCendroit où vous allez utiliser AWS PCS. Pour minimiser les transferts entre zones, déployez dans un sous-réseau de la même zone de disponibilité où vous lancerez la majorité de vos instances de groupes de nœuds de AWS PCS calcul. Assurez-vous que le système de fichiers est associé à un groupe de sécurité qui autorise l'accès entrant et sortant depuis les instances du groupe de nœuds de AWS PCS calcul. Pour plus d'informations sur les groupes de sécurité, consultez la section Gestion de l'accès au système de fichiers avec Amazon VPC dans le guide de ZFS l'utilisateur FSx for Open.

Modèle de lancement

Incluez les données utilisateur utilisées `cloud-config` pour monter le volume racine d'un système de ZFS fichiers FSx for Open. Remplacez les valeurs suivantes dans ce script par vos propres informations :

- *`mount-point-directory`*— Le chemin sur une instance où vous souhaitez monter votre FSx pour Open ZFS Share
- *`filesystem-id`*— L'ID du système de fichiers FSx pour le système de ZFS fichiers for Open
- *`region-code`*— L' Région AWS endroit où le système de ZFS fichiers FSx for Open est déployé (doit être le même que celui de votre AWS PCS système)

```
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==
```

Cache de fichiers Amazon

Configuration du système de fichiers

Créez un cache de fichiers Amazon dans l'VPCendroit où vous allez l'utiliser AWS PCS. Pour minimiser les transferts entre zones, choisissez un sous-réseau dans la même zone de disponibilité où vous lancerez la majorité de vos instances de groupes de nœuds de PCS calcul. Assurez-vous que le cache de fichiers est associé à un groupe de sécurité qui autorise le trafic entrant et sortant sur le port 988 entre vos PCS instances et le cache de fichiers. Pour plus d'informations sur les groupes de sécurité, consultez la section Contrôle d'accès au cache avec Amazon VPC dans le guide de l'utilisateur d'Amazon File Cache.

Modèle de lancement

Ajoutez le ou les groupes de sécurité de la configuration de votre système de fichiers au modèle de lancement que vous utiliserez pour le groupe de nœuds de calcul.

Incluez les données utilisateur utilisées `cloud-config` pour monter l'Amazon File Cache. Remplacez les valeurs suivantes dans ce script par vos propres informations :

- *`mount-point-directory`*— Le chemin d'une instance sur laquelle vous souhaitez effectuer le montage FSx pour Lustre
- *`cache-dns-name`*— Le nom du système de noms de domaine (DNS) pour le cache de fichiers
- *`mount-name`*— Le nom de montage pour le cache de fichiers

```
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) pour AWS PCS

AWS PCSfonctionne avec AMIs ce que vous fournissez, offrant une grande flexibilité dans le logiciel et la configuration des nœuds de votre cluster. Si vous essayez AWS PCS, vous pouvez utiliser un échantillon AMI fourni et maintenu par AWS. Si vous l'utilisez AWS PCS en production, nous vous recommandons de créer le vôtreAMIs. Cette rubrique explique comment découvrir et utiliser l'exempleAMIs, ainsi que comment créer et utiliser votre propre version personnaliséeAMIs.

Rubriques

Utilisation d'exemples d'Amazon Machine Images (AMIs) avec AWS PCS

AWSfournit un exemple AMIs que vous pouvez utiliser comme point de départ pour travailler AWS PCS.

> **Important**
>
> AMIsLes échantillons sont fournis à des fins de démonstration et ne sont pas recommandés pour les charges de travail de production.

Trouver l' AWS PCSéchantillon actuel AMIs

AWS Management Console

AWSPCSAMIsles exemples ont la convention de dénomination suivante :

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

Valeurs acceptées

- *OS* – amzn2

- *architecture* — x86_64 ou arm64
- *scheduler* – slurm
- *scheduler-major-version* – 23.11

Pour trouver AWS PCS un échantillon AMIs

1. Ouvrez la EC2console Amazon.
2. Naviguez vers AMIs.
3. Choisissez Images publiques.
4. Dans Rechercher AMI par attribut ou tag, recherchez un AMI en utilisant le nom du modèle.

   Exemples

   - Slurm AMI 23.11 compatible avec Graviton

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

   - Exemple AMI pour les instances x86

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

   > **Note**
   >
   > S'il y en a plusieursAMIs, utilisez le AMI horodatage le plus récent.

5. Utilisez l'AMIID lorsque vous créez ou mettez à jour un groupe de nœuds de calcul.

AWS CLI

Vous pouvez trouver le dernier AWS PCS exemple AMI avec les commandes suivantes. Remplacez *region-code* avec l' Région AWS endroit où vous l'utilisez AWS PCS, par exemple`us-east-1`.

- x86_64

```
aws ec2 describe-images --region region-code --owners amazon 533267220047
 654654292779 654654317195 975050324343 \
```

```
--filters 'Name=name,Values=aws-pcs-sample_ami-amzn2-x86_64-slurm-23.11*' \
            'Name=state,Values=available' \
--query 'sort_by(Images, &CreationDate)[-1].[Name,ImageId]' --output text
```

- Bras 64

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

Utilisez l'AMIID lorsque vous créez ou mettez à jour un groupe de nœuds de calcul.

En savoir plus sur AWS PCS Sample AMIs

Pour consulter le contenu et les détails de configuration des versions actuelles et précédentes de l' AWS PCSexempleAMIs, consultezNotes de publication pour AWS PCS un échantillon AMIs.

Créez le vôtre AMIs compatible avec AWS PCS

Pour savoir comment créer le vôtre AMIs qui fonctionne avec AWS PCS, voirImages Amazon Machine personnalisées (AMIs) pour AWS PCS.

Images Amazon Machine personnalisées (AMIs) pour AWS PCS

AWS PCSest conçu pour fonctionner avec Amazon Machine Images (AMI) que vous apportez au service. Des logiciels et des configurations arbitraires AMIs peuvent être installés sur ceux-ci, à condition que l' AWS PCSagent et une version compatible de Slurm soient installés et configurés correctement. Vous devez utiliser les programmes d'installation AWS fournis pour installer le AWS PCS logiciel sur votre compte personnalisé. AMI Nous vous recommandons d'utiliser les programmes d'installation AWS fournis pour installer Slurm sur votre site personnalisé, AMI mais vous pouvez installer Slurm vous-même si vous préférez (ce n'est pas recommandé).

> **Note**
>
> Si vous souhaitez essayer AWS PCS sans créer de personnalisationAMI, vous pouvez utiliser un exemple AMI fourni par AWS. Pour de plus amples informations, veuillez consulter Utilisation d'exemples d'Amazon Machine Images (AMIs) avec AWS PCS.

Ce didacticiel vous aide à créer un outil AMI qui peut être utilisé avec des groupes de nœuds de PCS calcul pour alimenter vos charges de travail HPC et celles de l'AI/ML.

Rubriques

Étape 1 — Lancer une instance temporaire

Lancez une instance temporaire que vous pouvez utiliser pour installer et configurer le AWS PCS logiciel et le planificateur Slurm. Vous utilisez cette instance pour créer un AMI compatible avec AWS PCS.

Pour lancer une instance temporaire

1. Ouvrez la EC2console Amazon.
2. Dans le volet de navigation, choisissez Instances, puis choisissez Launch instances pour ouvrir le nouvel assistant de lancement d'instances.
3. (Facultatif) Dans la section Nom et balises, saisissez un nom pour l'instance, par exemple`PCS-AMI-instance`. Le nom est attribué à l'instance en tant qu'identification de ressource (`Name=PCS-AMI-instance`).
4. Dans la section Images de l'application et du système d'exploitation, sélectionnez un AMI pour l'un des systèmes d'exploitation pris en charge.

5. Dans la section Instance type (Type d'instance), sélectionnez un type d'instance pris en charge.
6. Dans la section Key pair (Paire de clés), sélectionnez la paire de clés à utiliser pour l'instance.
7. Dans la section Paramètres réseau :
   - Pour Firewall (groupes de sécurité), choisissez Sélectionner un groupe de sécurité existant, puis sélectionnez un groupe de sécurité qui autorise l'SSHaccès entrant à votre instance.
8. Dans la section Storage (Stockage), configurez les volumes selon vos besoins. Assurez-vous de configurer suffisamment d'espace pour installer vos propres applications et bibliothèques.
9. Dans le panneau Summary (Récapitulatif), sélectionnez Launch instance (Lancer l'instance).

Étape 2 — Installation de l' AWS PCSagent

Installez l'agent qui configure les instances lancées par AWS PCS pour une utilisation avec Slurm.

Pour installer l'agent AWS PCS

1. Connectez-vous à l'instance que vous avez lancée. Pour plus d'informations, consultez Connect to your Linux instance.
2. (Facultatif) Pour vous assurer que tous vos packages logiciels sont à jour, effectuez une mise à jour logicielle rapide sur votre instance. Ce processus peut prendre quelques minutes.
   - 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. Redémarrez l'instance et reconnectez-vous à celle-ci.
4. Téléchargez les fichiers d'installation de l' AWS PCSagent. Les fichiers d'installation sont regroupés dans un fichier tarball (`.tar.gz`) compressé. Pour télécharger la version stable la plus récente, utilisez la commande suivante. Substitut *`region`* avec l' Région AWS endroit où vous avez lancé votre instance temporaire, par exemple`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
   ```

Vous pouvez également obtenir la dernière version en remplaçant le numéro de version par `latest` dans la commande précédente (par exemple :`aws-pcs-agent-v1-latest.tar.gz`).

> **Note**
>
> Cela pourrait changer dans les futures versions du logiciel de l' AWS PCSagent.

5. (Facultatif) Vérifiez l'authenticité et l'intégrité de l' AWS PCSarchive du logiciel. Nous vous recommandons de le faire pour vérifier l'identité de l'éditeur du logiciel et pour vérifier que le fichier n'a pas été modifié ou endommagé depuis sa publication.

   a. Téléchargez la GPG clé publique pour AWS PCS et importez-la dans votre trousseau de clés. Substitut *`region`* avec l' Région AWS endroit où vous avez lancé votre instance temporaire. La commande doit renvoyer une valeur clé. Enregistrez la valeur clé ; vous l'utiliserez à l'étape suivante.

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

   b. Exécutez la commande suivante pour vérifier l'empreinte digitale de la GPG clé.

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

   La commande doit renvoyer une empreinte identique à ce qui suit :

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

   > **Important**
   >
   > N'exécutez pas le script d'installation de l' AWS PCSagent si l'empreinte digitale ne correspond pas. Contactez AWS Support.

   c. Téléchargez le fichier de signature et vérifiez la signature du fichier tarball du AWS PCS logiciel. Remplacez *`region`* avec l' Région AWS endroit où vous avez lancé votre instance temporaire, par exemple`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
```

La sortie doit ressembler à ce qui suit :

```
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 le résultat inclut `Good signature` et que l'empreinte correspond à l'empreinte renvoyée à l'étape précédente, passez à l'étape suivante.

> **Important**
>
> N'exécutez pas le script d'installation du AWS PCS logiciel si l'empreinte digitale ne correspond pas. Contactez AWS Support.

6. Extrayez les fichiers du `.tar.gz` fichier compressé et accédez au répertoire extrait.

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

7. Installez le logiciel AWS PCS.

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

8. Vérifiez le fichier de version du AWS PCS logiciel pour confirmer la réussite de l'installation.

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

La sortie doit ressembler à ce qui suit :

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

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

Étape 3 — Installation de Slurm

Installez une version de Slurm compatible avec. AWS PCS

Pour installer Slurm

1. Connectez-vous à la même instance temporaire sur laquelle vous avez installé le AWS PCS logiciel.
2. Téléchargez le logiciel d'installation Slurm. Le programme d'installation de Slurm est intégré dans un fichier tarball () `.tar.gz` compressé. Pour télécharger la version stable la plus récente, utilisez la commande suivante. Substitut *`region`* avec le Région AWS de votre instance temporaire, tel que`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
   ```

   Vous pouvez également obtenir la dernière version en remplaçant le numéro de version par `latest` dans la commande précédente (par exemple :`aws-pcs-slurm-23.11-installer-latest.tar.gz`).

   > **Note**
   >
   > Cela pourrait changer dans les futures versions du logiciel d'installation Slurm.

3. (Facultatif) Vérifiez l'authenticité et l'intégrité de l'archive d'installation de Slurm. Nous vous recommandons de le faire pour vérifier l'identité de l'éditeur du logiciel et pour vérifier que le fichier n'a pas été modifié ou endommagé depuis sa publication.

   a. Téléchargez la GPG clé publique pour AWS PCS et importez-la dans votre trousseau de clés. Substitut *`region`* avec l' Région AWS endroit où vous avez lancé votre instance temporaire. La commande doit renvoyer une valeur clé. Enregistrez la valeur clé ; vous l'utiliserez à l'étape suivante.

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

b. Exécutez la commande suivante pour vérifier l'empreinte digitale de la GPG clé.

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

La commande doit renvoyer une empreinte identique à ce qui suit :

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

> **Important**
>
> N'exécutez pas le script d'installation de Slurm si l'empreinte digitale ne correspond pas. Contactez AWS Support.

c. Téléchargez le fichier de signature et vérifiez la signature du fichier tar du programme d'installation de Slurm. Remplacez *region* avec l' Région AWS endroit où vous avez lancé votre instance temporaire, par exemple`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
```

La sortie doit ressembler à ce qui suit :

```
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 le résultat inclut `Good signature` et que l'empreinte correspond à l'empreinte renvoyée à l'étape précédente, passez à l'étape suivante.

> **Important**
>
> N'exécutez pas le script d'installation de Slurm si l'empreinte digitale ne correspond pas. Contactez AWS Support.

4. Procédez à l'extraction des fichiers à partir du fichier compressé `.tar.gz` et accédez au répertoire extrait.

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

5. Installez Slurm. Le programme d'installation télécharge, compile et installe Slurm et ses dépendances. Cela prend plusieurs minutes, selon les spécifications de l'instance temporaire que vous avez sélectionnée.

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

6. Consultez le fichier de version du planificateur pour confirmer l'installation.

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

   La sortie doit ressembler à ce qui suit :

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

Étape 4 — (Facultatif) Installation de pilotes, de bibliothèques et de logiciels d'application supplémentaires

Installez des pilotes, des bibliothèques et des logiciels d'application supplémentaires sur l'instance temporaire. Les procédures d'installation varient en fonction des applications et bibliothèques spécifiques. Si vous n'avez encore jamais créé de solution personnaliséeAMI, nous vous recommandons d'abord de créer et de tester une version AMI avec uniquement le AWS PCS logiciel et Slurm installés, puis d'ajouter progressivement vos propres logiciels et configurations une fois que vous aurez confirmé le succès initial. AWS PCS

Exemples

- Logiciel Elastic Fabric Adapter (EFA). Pour plus d'informations, consultez la section Commencer avec EFA et MPI pour les HPC charges de travail sur Amazon EC2 dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.
- Client Amazon Elastic File System (AmazonEFS). Pour plus d'informations, consultez la section Installation manuelle du EFS client Amazon dans le guide de l'utilisateur Amazon Elastic File System.
- Client Lustre, pour utiliser Amazon FSx for Lustre et Amazon File Cache. Pour plus d'informations, consultez la section Installation du client Lustre dans le guide de l'utilisateur de FSx for Lustre.
- CloudWatch Agent Amazon, pour utiliser CloudWatch les journaux et les métriques. Pour plus d'informations, consultez la section Installation de l' CloudWatch agent dans le guide de CloudWatch l'utilisateur Amazon.
- AWS Neuron, pour utiliser les types d'instance trn* et inf*. Pour plus d'informations, consultez la documentation AWS Neuron.
- NVIDIAPiloteCUDA, etDCGM, pour utiliser les types d'instance p* ou g*.

Étape 5 — Créez un AMI compatible avec AWS PCS

Après avoir installé les composants logiciels requis, vous en créez un AMI que vous pouvez réutiliser pour lancer des instances dans des groupes de nœuds de AWS PCS calcul.

Pour créer un fichier AMI à partir de votre instance temporaire

1. Ouvrez la EC2console Amazon.
2. Dans le panneau de navigation, choisissez Instances.
3. Sélectionnez l'instance temporaire que vous avez créée. Choisissez Actions, Image, Créer une image.
4. Pour Créer une image, procédez comme suit :
    a. Dans Nom de l'image, entrez un nom descriptif pourAMI.
    b. (Facultatif) Dans Description de l'image, entrez une brève description de l'objectif duAMI.
    c. Choisissez Create image (Créer une image).
5. Dans le volet de navigation, choisissez AMIs.

6. Localisez le AMI fichier que vous avez créé dans la liste. Attendez que son statut passe de En attente à Disponible, puis utilisez-le avec un groupe de nœuds de AWS PCS calcul.

Étape 6 — Utiliser la personnalisation AMI avec un groupe de nœuds de AWS PCS calcul

Vous pouvez utiliser votre configuration personnalisée AMI avec un groupe de nœuds de AWS PCS calcul nouveau ou existant.

New compute node group

Pour utiliser la personnalisation AMI

1. Ouvrez la AWS PCSconsole.
2. Dans le panneau de navigation, choisissez Clusters.
3. Choisissez le cluster dans lequel vous allez utiliser la personnalisationAMI, puis sélectionnez Compute node groups.
4. Créez un nouveau groupe de nœuds de calcul. Pour de plus amples informations, veuillez consulter Création d'un groupe de nœuds de calcul dans AWS PCS. Sous AMIID, recherchez le nom ou l'ID de la personnalisation que AMI vous souhaitez utiliser. Terminez la configuration du groupe de nœuds de calcul, puis choisissez Créer un groupe de nœuds de calcul.
5. (Facultatif) Vérifiez que l'instance AMI prend en charge le lancement. Lancez une instance dans le groupe de nœuds de calcul. Vous pouvez le faire en configurant le groupe de nœuds de calcul pour qu'il n'ait qu'une seule instance statique, ou vous pouvez soumettre une tâche à une file d'attente qui utilise le groupe de nœuds de calcul.

    a. Vérifiez la EC2 console Amazon jusqu'à ce qu'une instance apparaisse étiquetée avec le nouvel ID de groupe de nœuds de calcul. Pour plus d'informations à ce sujet, voirRecherche d'instances de groupes de nœuds de calcul dans AWS PCS.

    b. Lorsque vous voyez une instance se lancer et terminer son processus d'amorçage, vérifiez qu'elle utilise le protocole attenduAMI. Pour ce faire, sélectionnez l'instance, puis inspectez l'AMIID sous Détails. Il doit correspondre à celui AMI que vous avez configuré dans les paramètres du groupe de nœuds de calcul.

    c. (Facultatif) Mettez à jour la configuration de dimensionnement du groupe de nœuds de calcul selon vos valeurs préférées.

Existing compute node group

Pour utiliser la personnalisation AMI

1. Ouvrez la AWS PCSconsole.
2. Dans le panneau de navigation, choisissez Clusters.
3. Choisissez le cluster dans lequel vous allez utiliser la personnalisationAMI, puis sélectionnez Compute node groups.
4. Sélectionnez le groupe de nœuds que vous souhaitez configurer et choisissez Modifier. Sous AMIID, recherchez le nom ou l'ID de la personnalisation que AMI vous souhaitez utiliser. Terminez la configuration du groupe de nœuds de calcul, puis choisissez Mettre à jour. Les nouvelles instances lancées dans le groupe de nœuds de calcul utiliseront l'AMIID mis à jour. Les instances existantes continueront à utiliser les anciennes AMI jusqu'à ce qu' AWS PCSelles soient remplacées. Pour de plus amples informations, veuillez consulter Mise à jour d'un groupe AWS PCS de nœuds de calcul.
5. (Facultatif) Vérifiez que l'instance AMI prend en charge le lancement. Lancez une instance dans le groupe de nœuds de calcul. Vous pouvez le faire en configurant le groupe de nœuds de calcul pour qu'il n'ait qu'une seule instance statique, ou vous pouvez soumettre une tâche à une file d'attente qui utilise le groupe de nœuds de calcul.

   a. Vérifiez la EC2 console Amazon jusqu'à ce qu'une instance apparaisse étiquetée avec le nouvel ID de groupe de nœuds de calcul. Pour plus d'informations à ce sujet, voirRecherche d'instances de groupes de nœuds de calcul dans AWS PCS.

   b. Lorsque vous voyez une instance se lancer et terminer son processus d'amorçage, vérifiez qu'elle utilise le protocole attenduAMI. Pour ce faire, sélectionnez l'instance, puis inspectez l'AMIID sous Détails. Il doit correspondre à celui AMI que vous avez configuré dans les paramètres du groupe de nœuds de calcul.

   c. (Facultatif) Mettez à jour la configuration de dimensionnement du groupe de nœuds de calcul selon vos valeurs préférées.

Étape 7 — Mettre fin à l'instance temporaire

Une fois que vous avez confirmé que votre AMI appareil fonctionne comme prévu AWS PCS, vous pouvez mettre fin à l'instance temporaire pour ne plus avoir à payer de frais pour celle-ci.

Pour résilier l'instance temporaire

1. Ouvrez la EC2console Amazon.
2. Dans le panneau de navigation, choisissez Instances.
3. Sélectionnez l'instance temporaire que vous avez créée et choisissez Actions, État de l'instance, Terminate instance.
4. Lorsque vous êtes invité à confirmer, choisissez Terminate.

Des installateurs de logiciels pour lesquels créer des logiciels personnalisés AMIs AWS PCS

AWS fournit un fichier téléchargeable qui permet d'installer le AWS PCS logiciel sur une instance. AWS fournit également un logiciel capable de télécharger, de compiler et d'installer les versions pertinentes de Slurm et de ses dépendances. Vous pouvez utiliser ces instructions AMIs pour créer une version personnalisée à utiliser avec AWS PCS ou vous pouvez utiliser vos propres méthodes.

Table des matières

AWS PCSprogramme d'installation du logiciel

Le programme d'installation du AWS PCS logiciel configure une instance à utiliser AWS PCS pendant le processus de démarrage de l'instance. Vous devez utiliser les programmes d'installation AWS fournis pour installer le AWS PCS logiciel sur votre compte personnalisé. AMI

Installateur Slurm

Le programme d'installation de Slurm télécharge, compile et installe les versions pertinentes de Slurm et de ses dépendances. Vous pouvez utiliser le programme d'installation de Slurm pour créer une version personnalisée AMIs pour. AWS PCS Vous pouvez également utiliser vos propres

mécanismes s'ils sont compatibles avec la configuration logicielle fournie par le programme d'installation de Slurm.

Le logiciel AWS fourni installe les éléments suivants :

- Slurm à la version majeure et à la version de maintenance demandées (version actuelle 23.11.8) - Licence 2 GPL
  - Slurm est construit avec `--sysconfdir` un set pour `/etc/slurm`
  - Slurm est conçu avec l'option et `--enable-pam` `--without-munge`
  - Slurm est conçu avec l'option `--sharedstatedir=/run/slurm/`
  - Slurm est conçu avec un support PMIX JWT
  - Slurm est installé sur `/opt/aws/pcs/schedulers/slurm-23.11`
- Open PMIX (version 4.2.6) — Licence
  - Open PMIX est installé en tant que sous-répertoire de `/opt/aws/pcs/scheduler/`
- libjwt (version 1.15.3) — Licence -2.0 MPL
  - libjwt est installé en tant que sous-répertoire de `/opt/aws/pcs/scheduler/`

Le logiciel AWS fourni modifie la configuration du système comme suit :

- Le `systemd` fichier Slurm créé par le build est copié `/etc/systemd/system/` avec le nom du fichier. `slurmd-23.11.service`
- S'ils n'existent pas, un utilisateur et un groupe Slurm (`slurm:slurm`) sont créés avecUID/GIDof. `401`
- Sur Amazon Linux 2 et Rocky Linux 9, l'installation ajoute le EPEL référentiel pour installer le logiciel requis pour créer Slurm ou ses dépendances.
- Lors RHEL9 de l'installation, vous pourrez activer `codeready-builder-for-rhel-9-rhui-rpms` et `epel-release-latest-9` `fedoraproject` installer le logiciel requis pour créer Slurm ou ses dépendances.

Systèmes d'exploitation pris en charge

Le AWS PCS logiciel et les programmes d'installation de Slurm sont compatibles avec les systèmes d'exploitation suivants :

- Amazon Linux 2

- RedHat Linux d'entreprise 9
- Rocky Linux 9
- Ubuntu 22.04

> **Note**
>
> AWS Apprentissage profond (deep learning) AMIs (DLAMI) les versions basées sur Amazon Linux 2 et Ubuntu 22.04 doivent être compatibles avec le AWS PCS logiciel et les installateurs de Slurm. Pour plus d'informations, consultez la section Choosing Your DLAMI dans le guide du AWS Apprentissage profond (deep learning) AMIs développeur.

Types d'instance pris en charge

AWS PCSle logiciel et les installateurs Slurm prennent en charge tout type d'instance x86_64 ou arm64 capable d'exécuter l'un des systèmes d'exploitation pris en charge.

Versions de Slurm prises en charge

Les versions principales suivantes de Slurm sont prises en charge :

- Slurm 23,11

Vérifiez les installateurs à l'aide d'une somme de contrôle

Vous pouvez utiliser des SHA256 checksums pour vérifier les fichiers tarball du programme d'installation (.tar.gz). Nous vous recommandons de le faire pour vérifier l'identité de l'éditeur du logiciel et pour vérifier que l'application n'a pas été modifiée ou endommagée depuis sa publication.

Pour vérifier une archive

Utilisez l'utilitaire sha256sum pour la somme de SHA256 contrôle et spécifiez le nom du fichier tarball. Vous devez exécuter la commande depuis le répertoire dans lequel vous avez enregistré le fichier tarball.

- SHA256

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

La commande doit renvoyer une valeur de somme de contrôle au format suivant.

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

Comparez la valeur de somme de contrôle renvoyée par la commande avec la valeur de somme de contrôle fournie dans le tableau suivant. Si les sommes de contrôle correspondent, vous pouvez exécuter le script d'installation en toute sécurité.

> **Important**
>
> Si les checksums ne correspondent pas, n'exécutez pas le script d'installation. Contactez AWS Support.

Par exemple, la commande suivante génère la SHA256 somme de contrôle pour l'archive Slurm 23.11.9.

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

Exemple de sortie :

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

Le tableau suivant répertorie les sommes de contrôle pour les versions récentes des programmes d'installation. Remplacez *us-east-1* avec l' Région AWS endroit où vous l'utilisez AWS PCS.

Installer	Télécharger URL	SHA256somme de contrôle
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 PCSagent 1.0.0	`https://aws-pcs-repo-us-east-1.s3.amazo`	`d2d3d68d00c685435c38af471d7e2492dde5`

Installer	Télécharger URL	SHA256somme de contrôle
	`naws.com/aws-pcs-agent/aws-pcs-agent-v1.0.0-1.tar.gz`	`ce9eb222d7b6ef0042144b134ce0`

Versions Slurm en AWS PCS

SchedMD améliore continuellement Slurm avec de nouvelles fonctionnalités, optimisations et correctifs de sécurité. SchedMD publie une nouvelle version majeure à intervalles réguliers et prévoit de prendre en charge jusqu'à 3 versions à la fois. AWS PCSsupporte initialement Slurm 23.11. Vous pouvez mettre à jour votre version majeure de Slurm après la publication d'une nouvelle version. AWS PCSest conçu pour mettre à jour automatiquement le contrôleur Slurm avec des versions de patch.

Lorsque SchedMD met fin au support d'une version majeure particulière, il met AWS PCS également fin au support de cette version majeure. AWS PCSenvoie un préavis si une version majeure de Slurm approche de sa fin de vie, afin d'aider les clients à savoir quand mettre à niveau leurs clusters vers une version plus récente prise en charge.

Nous vous recommandons d'utiliser la dernière version prise en charge de Slurm pour déployer votre cluster, afin d'accéder aux avancées et améliorations les plus récentes.

Questions fréquemment posées sur les versions de Slurm

Combien de temps dure une AWS PCS version de Slurm ?

AWS PCSsuit les cycles de support de SchedMD pour les versions majeures. AWS PCSprend en charge jusqu'à 3 versions majeures à la fois. Une fois que SchedMD a publié une nouvelle version majeure, la plus ancienne version AWS PCS prise en charge est supprimée. AWS PCSpublie une nouvelle version majeure de Slurm dès que possible, mais il se peut qu'il y ait un délai entre la sortie de SchedMD et sa disponibilité dans. AWS PCS

Quand m' AWS PCSinformerai-je de la fin de la durée de vie du support (EOSL) pour les versions de Slurm ?

AWS PCSvous avertit plusieurs fois, à une cadence prédéterminée, avant la date. EOSL

Que dois-je faire à l'approche d'une version de Slurm ? EOSL

Vous devez d'abord mettre à jour vos versions de Slurm EOSL afin de maintenir un environnement sécurisé et pris en charge.

Comment puis-je mettre à jour mes clusters pour utiliser une nouvelle version majeure de Slurm ?

Pour mettre à jour la version de Slurm, vous devez créer un nouveau cluster. Vous devez également effectuer une mise à niveau vers le AWS PCS logiciel équivalent AMI et l'utiliser pour créer les groupes de nœuds de calcul pour votre nouveau cluster.

Comment mes clusters recevront-ils les nouvelles versions de patch de Slurm ?

AWS PCSest conçu pour appliquer automatiquement des correctifs afin de remédier aux vulnérabilités et aux expositions courantes de Slurm (). CVEs AWS PCSapplique les correctifs aux contrôleurs de cluster qui s'exécutent sur des comptes appartenant au service interne. Vous devez utiliser les AWS PCS API actions AWS Management Console ou pour installer des correctifs sur EC2 les instances de votre Compte AWS.

Et si je ne mets pas Slurm à jour avant la EOSL date prévue ?

AWS PCSest conçu pour arrêter les clusters dont la version de Slurm n'est pas prise en charge. Vous devez mettre à jour la version majeure de Slurm du contrôleur de cluster et le AWS PCS logiciel installé sur les groupes de nœuds de calcul.

Combien de versions de Slurm sont prises en charge ? AWS PCS

AWS PCSprend en charge jusqu'à 3 versions majeures de Slurm à la fois, y compris la version majeure actuelle et les 2 versions majeures précédentes.

Quelles mises à jour de version de Slurm dois-je appliquer ?

Nous vous recommandons vivement d'utiliser la même version majeure pour tous les composants de votre cluster et d'installer les derniers correctifs dès leur publication. AMIsPour vos groupes de nœuds de calcul, vous devez utiliser une version du logiciel Slurm compatible avec la version Slurm du contrôleur de cluster. La version principale de Slurm AMIs doit se trouver dans les 2 versions de la version principale de Slurm sur le contrôleur de cluster. La version de Slurm installée dans AMI et sur les EC2 instances en cours d'exécution du cluster ne peut pas être plus récente que la version de Slurm installée sur le contrôleur de cluster. Pour maintenir le support de votre cluster, vous AMIs devez utiliser une version AWS PCS logicielle prise en charge.

Et si je mets à jour la version majeure de Slurm mais que j'utilise l'ancien logiciel Slurm dans mes groupes de nœuds de calcul AMI ?

Vous devez mettre à jour le AWS PCS logiciel vers la même version pour utiliser les nouvelles fonctionnalités de Slurm. Pour un AWS PCS support complet, tous les composants de Slurm doivent utiliser des versions prises en charge. Pour résumer :

- Nous sommes en mesure de fournir un support complet lorsque le contrôleur de cluster et tous les composants (AWS PCSpackages) de vos Compte AWS deux utilisateurs utilisent les versions prises en charge.
- AWS PCSest conçu pour arrêter un cluster si la version Slurm de son contrôleur atteint. EOSL
- Si la version Slurm des composants est à votre Compte AWS portéeEOSL, votre cluster ne sera pas pris en charge.

Dans quel ordre dois-je mettre à jour les composants de mon cluster ?

Vous devez mettre à jour la version Slurm de votre contrôleur de cluster avant de l'utiliser AMI avec une version plus récente de Slurm. Vous mettez à jour un groupe de nœuds de calcul pour utiliser leAMI. AWS PCSutilise le AMI pour lancer de nouvelles EC2 instances dans le groupe de nœuds de calcul. AWS PCSne met pas à jour les EC2 instances existantes qui ont des tâches en cours d'exécution ; AWS PCS est conçu pour mettre fin à ces instances une fois leurs tâches terminées.

AWS PCSPropose-t-il un support étendu pour les versions de Slurm ?

Non. Nous communiquerons des informations détaillées sur les options de support étendues, y compris les éventuels coûts supplémentaires et la couverture d'assistance spécifique fournie.

Sécurité dans le service de calcul AWS parallèle

La sécurité du cloud AWS est la priorité absolue. En tant que AWS client, vous bénéficiez de centres de données et d'architectures réseau conçus pour répondre aux exigences des entreprises les plus sensibles en matière de sécurité.

La sécurité est une responsabilité partagée entre vous AWS et vous. Le modèle de responsabilité partagée décrit cela comme la sécurité du cloud et la sécurité dans le cloud :

- Sécurité du cloud : AWS est chargée de protéger l'infrastructure qui exécute les AWS services dans le AWS Cloud. AWS vous fournit également des services que vous pouvez utiliser en toute sécurité. Des auditeurs tiers testent et vérifient régulièrement l'efficacité de notre sécurité dans le cadre des programmes de AWS conformité Programmes de de conformité. Pour en savoir plus sur les programmes de conformité qui s'appliquent au service de calcul AWS parallèle, voir AWS Services concernés par programme de conformitéAWS .
- Sécurité dans le cloud — Votre responsabilité est déterminée par le AWS service que vous utilisez. Vous êtes également responsable d'autres facteurs, y compris de la sensibilité de vos données, des exigences de votre entreprise, ainsi que de la législation et de la réglementation applicables.

Cette documentation vous aide à comprendre comment appliquer le modèle de responsabilité partagée lors de son utilisation AWS PCS. Les rubriques suivantes expliquent comment procéder à la configuration AWS PCS pour atteindre vos objectifs de sécurité et de conformité. Vous apprendrez également à utiliser d'autres AWS services qui vous aident à surveiller et à sécuriser vos AWS PCS ressources.

Rubriques

- Bonnes pratiques de sécurité pour le service de calcul AWS parallèle

Protection des données dans le service de calcul AWS parallèle

Le modèle de responsabilité AWS partagée s'applique à la protection des données dans le service de calcul AWS parallèle. Comme décrit dans ce modèle, AWS est chargé de protéger l'infrastructure mondiale qui gère tous les AWS Cloud. La gestion du contrôle de votre contenu hébergé sur cette infrastructure relève de votre responsabilité. Vous êtes également responsable des tâches de configuration et de gestion de la sécurité des Services AWS que vous utilisez. Pour plus d'informations sur la confidentialité des données, consultez la section Confidentialité des données FAQ. Pour plus d'informations sur la protection des données en Europe, consultez le modèle de responsabilitéAWS partagée et le billet de GDPR blog sur le blog sur la AWS sécurité.

À des fins de protection des données, nous vous recommandons de protéger les Compte AWS informations d'identification et de configurer les utilisateurs individuels avec AWS IAM Identity Center ou AWS Identity and Access Management (IAM). Ainsi, chaque utilisateur se voit attribuer uniquement les autorisations nécessaires pour exécuter ses tâches. Nous vous recommandons également de sécuriser vos données comme indiqué ci-dessous :

- Utilisez l'authentification multifactorielle (MFA) pour chaque compte.
- UtilisezSSL/TLSpour communiquer avec les AWS ressources. Nous avons besoin de la TLS version 1.2 et recommandons la TLS version 1.3.
- Configuration API et journalisation de l'activité des utilisateurs avec AWS CloudTrail.
- Utilisez des solutions de AWS chiffrement, ainsi que tous les contrôles de sécurité par défaut qu'ils contiennent Services AWS.
- Utilisez des services de sécurité gérés avancés tels qu'Amazon Macie, qui contribuent à la découverte et à la sécurisation des données sensibles stockées dans Amazon S3.
- Si vous avez besoin de FIPS 140 à 3 modules cryptographiques validés pour accéder AWS via une interface de ligne de commande ou unAPI, utilisez un point de terminaison. FIPS Pour plus d'informations sur les FIPS points de terminaison disponibles, voir Federal Information Processing Standard (FIPS) 140-3.

Nous vous recommandons fortement de ne jamais placer d'informations confidentielles ou sensibles, telles que les adresses e-mail de vos clients, dans des balises ou des champs de texte libre tels que le champ Name (Nom). Cela inclut lorsque vous travaillez avec AWS PCS ou d'autres Services AWS

utilisateurs de la consoleAPI, AWS CLI, ou AWS SDKs. Toutes les données que vous entrez dans des balises ou des champs de texte de forme libre utilisés pour les noms peuvent être utilisées à des fins de facturation ou dans les journaux de diagnostic. Si vous fournissez un URL à un serveur externe, nous vous recommandons vivement de ne pas inclure d'informations d'identification dans le URL afin de valider votre demande auprès de ce serveur.

Chiffrement au repos

Le chiffrement est activé par défaut pour les données au repos lorsque vous créez un cluster AWS Parallel Computing Service (AWS PCS) avec le AWS Management Console AWS CLI, AWS PCSAPI, ou AWS SDKs. AWS PCSutilise une KMScléAWS détenue pour chiffrer les données au repos. Pour plus d'informations, consultez la section Clés client et AWS clés dans le guide du AWS KMS développeur. Le secret du cluster est stocké AWS Secrets Manager et chiffré avec la KMS clé gérée Secrets Manager. Pour de plus amples informations, veuillez consulter Utilisation des secrets de cluster dans AWS PCS.

Dans un AWS PCS cluster, les données suivantes sont inactives :

- État du planificateur : il inclut des données sur les tâches en cours d'exécution et les nœuds provisionnés dans le cluster. Il s'agit des données que Slurm conserve telles que `StateSaveLocation` définies dans votre. `slurm.conf` Pour plus d'informations, consultez la description StateSaveLocationdans la documentation de Slurm. AWS PCSsupprime les données d'une tâche une fois celle-ci terminée.
- Secret d'authentification du planificateur : l' AWS PCSutilise pour authentifier toutes les communications du planificateur dans le cluster.

Pour les informations sur l'état du planificateur, chiffre AWS PCS automatiquement les données et les métadonnées avant de les écrire dans le système de fichiers. Le système de fichiers chiffré utilise l'algorithme de cryptage standard AES -256 pour les données au repos.

Chiffrement en transit

Vos connexions AWS PCS API utilisent le TLS cryptage avec le processus de signature Signature Version 4, que vous utilisiez le AWS Command Line Interface (AWS CLI) ou AWS SDKs. Pour plus d'informations, consultez la section Signature AWS API des demandes dans le guide de AWS Identity and Access Management l'utilisateur. AWS gère le contrôle d'accès API grâce aux IAM politiques relatives aux informations d'identification de sécurité que vous utilisez pour vous connecter.

AWS PCSutilise TLS pour se connecter à d'autres AWS services.

Au sein d'un cluster Slurm, le planificateur est configuré avec le plug-in `auth/slurm` d'authentification qui fournit une authentification pour toutes les communications du planificateur. Slurm ne fournit pas de chiffrement au niveau de l'application pour ses communications, toutes les données circulant entre les instances de cluster restent locales EC2 VPC et sont donc VPC cryptées si ces instances prennent en charge le chiffrement en transit. Pour plus d'informations, consultez la section Chiffrement en transit dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud. La communication est cryptée entre le contrôleur (fourni dans un compte de service) et les nœuds du cluster de votre compte.

Gestion des clés

AWS PCSutilise une KMScléAWS détenue pour chiffrer les données. Pour plus d'informations, consultez la section Clés client et AWS clés dans le guide du AWS KMS développeur. Le secret du cluster est stocké AWS Secrets Manager et chiffré avec la KMS clé gérée Secrets Manager. Pour de plus amples informations, veuillez consulter Utilisation des secrets de cluster dans AWS PCS.

Confidentialité du trafic inter-réseaux

AWS PCSles ressources de calcul d'un cluster se situent dans la limite de 1 VPC dans le compte du client. Par conséquent, tout le trafic de AWS PCS service interne au sein d'un cluster reste sur le AWS réseau et ne transite pas par Internet. La communication entre l'utilisateur et AWS PCS les nœuds peut se faire via Internet et nous vous recommandons d'utiliser SSH ou Systems Manager pour vous connecter aux nœuds. Pour plus d'informations, voir Qu'est-ce que c'est AWS Systems Manager ? dans le guide de AWS Systems Manager l'utilisateur.

Vous pouvez également utiliser les offres suivantes pour connecter votre réseau local à AWS :

- AWS Site-to-Site VPN. Pour plus d'informations, voir Qu'est-ce que c'est AWS Site-to-Site VPN ? dans le guide de AWS Site-to-Site VPN l'utilisateur.
- Un AWS Direct Connect. Pour plus d'informations, voir Qu'est-ce que c'est AWS Direct Connect ? dans le guide de AWS Direct Connect l'utilisateur.

Vous y accédez AWS PCS API pour effectuer des tâches administratives pour le service. Vous et vos utilisateurs accédez aux ports du point de terminaison Slurm pour interagir directement avec le planificateur.

Chiffrer le trafic API

Pour y accéder AWS PCSAPI, les clients doivent prendre en charge Transport Layer Security (TLS) 1.2 ou version ultérieure. Nous avons besoin de la TLS version 1.2 et recommandons la TLS version 1.3. Les clients doivent également prendre en charge les suites de chiffrement dotées de Perfect Forward Secrecy (PFS), telles que Ephemeral Diffie-Hellman () ou Elliptic Curve Diffie-Hellman Ephemeral (DHE). ECDHE La plupart des systèmes modernes tels que Java 7 et les versions ultérieures prennent en charge ces modes. En outre, les demandes doivent être signées à l'aide d'un identifiant de clé d'accès et d'une clé d'accès secrète associés à un IAM principal. Vous pouvez également utiliser AWS Security Token Service (AWS STS) pour générer des informations de sécurité temporaires afin de signer les demandes.

Chiffrement du trafic de données

Le chiffrement des données en transit est activé à partir des EC2 instances prises en charge accédant au point de terminaison du planificateur et entre les ComputeNodeGroup instances depuis le. AWS Cloud Pour de plus amples informations, veuillez consulter Chiffrement en transit.

Accédez au service de calcul AWS parallèle à l'aide d'un point de terminaison d'interface (AWS PrivateLink)

Vous pouvez l'utiliser AWS PrivateLink pour créer une connexion privée entre votre VPC et AWS Parallel Computing Service (AWS PCS). Vous pouvez y accéder AWS PCS comme s'il se trouvait dans votre ordinateurVPC, sans utiliser de passerelle Internet, d'NATappareil, de VPN connexion ou de AWS Direct Connect connexion. Les instances de votre VPC ordinateur n'ont pas besoin d'adresses IP publiques pour y accéder AWS PCS.

Vous établissez cette connexion privée en créant un point de terminaison d’interface optimisé par AWS PrivateLink. Nous créons une interface réseau de point de terminaison dans chaque sous-réseau que vous activez pour le point de terminaison d’interface. Il s'agit d'interfaces réseau gérées par les demandeurs qui servent de point d'entrée pour le trafic à destination de. AWS PCS

Pour plus d'informations, consultez la section Accès Services AWS par AWS PrivateLink le biais du AWS PrivateLink guide.

Considérations relatives à AWS PCS

Avant de configurer un point de terminaison d'interface pour AWS PCS, consultez la section Accès à un AWS service à l'aide d'un point de VPC terminaison d'interface dans le AWS PrivateLink Guide.

AWS PCSprend en charge les appels à toutes ses API actions via le point de terminaison de l'interface.

Si vous VPC ne disposez pas d'un accès direct à Internet, vous devez configurer un VPC point de terminaison pour permettre aux instances de votre groupe de nœuds de calcul d'appeler l' AWS PCS`RegisterComputeNodeGroupInstance`APIaction.

Créez un point de terminaison d'interface pour AWS PCS

Vous pouvez créer un point de terminaison d'interface pour AWS PCS utiliser la VPC console Amazon ou le AWS Command Line Interface (AWS CLI). Pour plus d'informations, consultez Création d'un point de terminaison d'interface dans le Guide AWS PrivateLink .

Créez un point de terminaison d'interface pour AWS PCS utiliser le nom de service suivant :

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

Remplacez *`region`* avec l'ID du dans lequel Région AWS créer le point de terminaison, tel que `us-east-1`.

Si vous activez le mode privé DNS pour le point de terminaison de l'interface, vous pouvez API envoyer des demandes AWS PCS en utilisant son DNS nom régional par défaut. Par exemple, `pcs.us-east-1.amazonaws.com`.

Création d'une politique de point de terminaison pour votre point de terminaison d'interface

Une politique de point de terminaison est une IAM ressource que vous pouvez associer à un point de terminaison d'interface. La politique de point de terminaison par défaut autorise un accès complet AWS PCS via le point de terminaison de l'interface. Pour contrôler l'accès autorisé à AWS PCS partir de votreVPC, associez une politique de point de terminaison personnalisée au point de terminaison de l'interface.

Une politique de point de terminaison spécifie les informations suivantes :

- Les principaux qui peuvent effectuer des actions (Comptes AWS IAMutilisateurs et IAM rôles).
- Les actions qui peuvent être effectuées.
- La ressource sur laquelle les actions peuvent être effectuées.

Pour plus d'informations, consultez Contrôle de l'accès aux services à l'aide de politiques de point de terminaison dans le Guide AWS PrivateLink .

Exemple : politique des VPC terminaux pour les AWS PCS actions

Voici un exemple de politique de point de terminaison personnalisée. Lorsque vous attachez cette politique au point de terminaison de votre interface, elle accorde l'accès aux AWS PCS actions répertoriées pour tous les principaux du cluster avec le paramètre spécifié *cluster-id*. Remplacer *region* avec l'ID Région AWS du cluster, tel queus-east-1. Remplacez *account-id* avec le Compte AWS numéro du cluster.

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

Identity and Access Management pour le service de calcul AWS parallèle

AWS Identity and Access Management (IAM) est un outil Service AWS qui permet à un administrateur de contrôler en toute sécurité l'accès aux AWS ressources. IAMles administrateurs

contrôlent qui peut être authentifié (connecté) et autorisé (autorisé) à utiliser les AWS PCS ressources. IAMest un Service AWS outil que vous pouvez utiliser sans frais supplémentaires.

Rubriques

Public ciblé

La façon dont vous utilisez AWS Identity and Access Management (IAM) varie en fonction du travail que vous effectuez AWS PCS.

Utilisateur du service : si vous utilisez le AWS PCS service pour effectuer votre travail, votre administrateur vous fournit les informations d'identification et les autorisations dont vous avez besoin. Au fur et à mesure que vous utilisez de nouvelles AWS PCS fonctionnalités pour effectuer votre travail, vous aurez peut-être besoin d'autorisations supplémentaires. En comprenant bien la gestion des accès, vous saurez demander les autorisations appropriées à votre administrateur. Si vous ne pouvez pas accéder à une fonctionnalité dans AWS PCS, consultezRésolution des problèmes d'identité et d'accès au service de calcul AWS parallèle.

Administrateur du service — Si vous êtes responsable des AWS PCS ressources de votre entreprise, vous avez probablement un accès complet à AWS PCS. C'est à vous de déterminer les AWS PCS fonctionnalités et les ressources auxquelles les utilisateurs de votre service doivent accéder. Vous devez ensuite envoyer des demandes à votre IAM administrateur pour modifier les autorisations des utilisateurs de votre service. Consultez les informations de cette page pour comprendre les concepts

de base deIAM. Pour en savoir plus sur la façon dont votre entreprise peut utiliser IAM avec AWS PCS, voirComment fonctionne AWS Parallel Computing Service avec IAM.

IAMadministrateur — Si vous êtes IAM administrateur, vous souhaiterez peut-être en savoir plus sur la manière dont vous pouvez rédiger des politiques pour gérer l'accès à AWS PCS. Pour consulter des exemples de politiques AWS PCS basées sur l'identité que vous pouvez utiliser dansIAM, consultez. Exemples de politiques basées sur l'identité pour le service de calcul AWS parallèle

Authentification par des identités

L'authentification est la façon dont vous vous connectez à AWS l'aide de vos informations d'identification. Vous devez être authentifié (connecté à AWS) en tant que Utilisateur racine d'un compte AWS, en tant qu'IAMutilisateur ou en assumant un IAM rôle.

Vous pouvez vous connecter en AWS tant qu'identité fédérée en utilisant les informations d'identification fournies par le biais d'une source d'identité. AWS IAM Identity Center Les utilisateurs (IAMIdentity Center), l'authentification unique de votre entreprise et vos informations d'identification Google ou Facebook sont des exemples d'identités fédérées. Lorsque vous vous connectez en tant qu'identité fédérée, votre administrateur a préalablement configuré la fédération d'identité à l'aide de IAM rôles. Lorsque vous accédez à AWS l'aide de la fédération, vous assumez indirectement un rôle.

Selon le type d'utilisateur que vous êtes, vous pouvez vous connecter au portail AWS Management Console ou au portail AWS d'accès. Pour plus d'informations sur la connexion à AWS, consultez la section Comment vous connecter à votre compte Compte AWS dans le guide de Connexion à AWS l'utilisateur.

Si vous y accédez AWS par programmation, AWS fournit un kit de développement logiciel (SDK) et une interface de ligne de commande (CLI) pour signer cryptographiquement vos demandes à l'aide de vos informations d'identification. Si vous n'utilisez pas d' AWS outils, vous devez signer vous-même les demandes. Pour plus d'informations sur l'utilisation de la méthode recommandée pour signer vous-même les demandes, consultez la section Signature des AWS API demandes dans le guide de IAM l'utilisateur.

Quelle que soit la méthode d'authentification que vous utilisez, vous devrez peut-être fournir des informations de sécurité supplémentaires. Par exemple, il vous AWS recommande d'utiliser l'authentification multifactorielle (MFA) pour renforcer la sécurité de votre compte. Pour en savoir plus, consultez Authentification multifactorielle dans le guide de AWS IAM Identity Center l'utilisateur et Utilisation de l'authentification multifactorielle (MFA) AWS dans le guide de l'IAMutilisateur.

Compte AWS utilisateur root

Lorsque vous créez un Compte AWS, vous commencez par une identité de connexion unique qui donne un accès complet à toutes Services AWS les ressources du compte. Cette identité est appelée utilisateur Compte AWS root et est accessible en vous connectant avec l'adresse e-mail et le mot de passe que vous avez utilisés pour créer le compte. Il est vivement recommandé de ne pas utiliser l'utilisateur racine pour vos tâches quotidiennes. Protégez vos informations d'identification d'utilisateur racine et utilisez-les pour effectuer les tâches que seul l'utilisateur racine peut effectuer. Pour obtenir la liste complète des tâches qui nécessitent que vous vous connectiez en tant qu'utilisateur root, consultez la section Tâches nécessitant des informations d'identification utilisateur root dans le Guide de IAM l'utilisateur.

Identité fédérée

La meilleure pratique consiste à obliger les utilisateurs humains, y compris ceux qui ont besoin d'un accès administrateur, à utiliser la fédération avec un fournisseur d'identité pour accéder à l'aide Services AWS d'informations d'identification temporaires.

Une identité fédérée est un utilisateur de l'annuaire des utilisateurs de votre entreprise, d'un fournisseur d'identité Web AWS Directory Service, du répertoire Identity Center ou de tout utilisateur qui y accède à l'aide des informations d'identification fournies Services AWS par le biais d'une source d'identité. Lorsque des identités fédérées y accèdent Comptes AWS, elles assument des rôles, qui fournissent des informations d'identification temporaires.

Pour une gestion des accès centralisée, nous vous recommandons d'utiliser AWS IAM Identity Center. Vous pouvez créer des utilisateurs et des groupes dans IAM Identity Center, ou vous pouvez vous connecter et synchroniser avec un ensemble d'utilisateurs et de groupes dans votre propre source d'identité afin de les utiliser dans toutes vos applications Comptes AWS et applications. Pour plus d'informations sur IAM Identity Center, consultez Qu'est-ce qu'IAMIdentity Center ? dans le guide de AWS IAM Identity Center l'utilisateur.

Utilisateurs et groupes IAM

Un IAMutilisateur est une identité au sein de vous Compte AWS qui possède des autorisations spécifiques pour une seule personne ou une seule application. Dans la mesure du possible, nous vous recommandons de vous appuyer sur des informations d'identification temporaires plutôt que de créer des IAM utilisateurs dotés d'informations d'identification à long terme, telles que des mots de passe et des clés d'accès. Toutefois, si vous avez des cas d'utilisation spécifiques qui nécessitent des informations d'identification à long terme auprès des IAM utilisateurs, nous vous recommandons

de faire pivoter les clés d'accès. Pour plus d'informations, voir Rotation régulière des clés d'accès pour les cas d'utilisation nécessitant des informations d'identification à long terme dans le Guide de IAM l'utilisateur.

Un IAMgroupe est une identité qui définit un ensemble d'IAMutilisateurs. Vous ne pouvez pas vous connecter en tant que groupe. Vous pouvez utiliser les groupes pour spécifier des autorisations pour plusieurs utilisateurs à la fois. Les groupes permettent de gérer plus facilement les autorisations pour de grands ensembles d'utilisateurs. Par exemple, vous pouvez nommer un groupe IAMAdminset lui donner les autorisations nécessaires pour administrer IAM des ressources.

Les utilisateurs sont différents des rôles. Un utilisateur est associé de manière unique à une personne ou une application, alors qu'un rôle est conçu pour être endossé par tout utilisateur qui en a besoin. Les utilisateurs disposent d'informations d'identification permanentes, mais les rôles fournissent des informations d'identification temporaires. Pour en savoir plus, voir Quand créer un IAM utilisateur (au lieu d'un rôle) dans le Guide de IAM l'utilisateur.

IAMrôles

Un IAMrôle est une identité au sein de Compte AWS vous dotée d'autorisations spécifiques. Il est similaire à un IAM utilisateur, mais n'est pas associé à une personne en particulier. Vous pouvez assumer temporairement un IAM rôle dans le en AWS Management Console changeant de rôle. Vous pouvez assumer un rôle en appelant une AWS API opération AWS CLI or ou en utilisant une option personnaliséeURL. Pour plus d'informations sur les méthodes d'utilisation des rôles, consultez la section Utilisation IAM des rôles dans le Guide de IAM l'utilisateur.

IAMles rôles dotés d'informations d'identification temporaires sont utiles dans les situations suivantes :

- Accès utilisateur fédéré – Pour attribuer des autorisations à une identité fédérée, vous créez un rôle et définissez des autorisations pour le rôle. Quand une identité externe s'authentifie, l'identité est associée au rôle et reçoit les autorisations qui sont définies par celui-ci. Pour plus d'informations sur les rôles pour la fédération, voir Création d'un rôle pour un fournisseur d'identité tiers dans le guide de IAM l'utilisateur. Si vous utilisez IAM Identity Center, vous configurez un ensemble d'autorisations. Pour contrôler les accès auxquels vos identités peuvent accéder après leur authentification, IAM Identity Center met en corrélation l'ensemble d'autorisations avec un rôle dans. IAM Pour plus d'informations sur les jeux d'autorisations, consultez la rubrique Jeux d'autorisations dans le Guide de l'utilisateur AWS IAM Identity Center .
- Autorisations IAM utilisateur temporaires : un IAM utilisateur ou un rôle peut assumer un IAM rôle afin d'obtenir temporairement différentes autorisations pour une tâche spécifique.

- Accès entre comptes : vous pouvez utiliser un IAM rôle pour autoriser une personne (un mandant fiable) d'un autre compte à accéder aux ressources de votre compte. Les rôles constituent le principal moyen d'accorder l'accès intercompte. Toutefois, dans certains Services AWS cas, vous pouvez associer une politique directement à une ressource (au lieu d'utiliser un rôle comme proxy). Pour connaître la différence entre les rôles et les politiques basées sur les ressources pour l'accès entre comptes, voir Accès aux ressources entre comptes IAM dans le guide de l'IAMutilisateur.
- Accès multiservices — Certains Services AWS utilisent des fonctionnalités dans d'autres Services AWS. Par exemple, lorsque vous effectuez un appel dans un service, il est courant que ce service exécute des applications dans Amazon EC2 ou stocke des objets dans Amazon S3. Un service peut le faire en utilisant les autorisations d'appel du principal, un rôle de service ou un rôle lié au service.
  - Sessions d'accès transmises (FAS) — Lorsque vous utilisez un IAM utilisateur ou un rôle pour effectuer des actions AWS, vous êtes considéré comme un mandant. Lorsque vous utilisez certains services, vous pouvez effectuer une action qui initie une autre action dans un autre service. FASutilise les autorisations du principal appelant an Service AWS, combinées à la demande Service AWS pour adresser des demandes aux services en aval. FASles demandes ne sont effectuées que lorsqu'un service reçoit une demande qui nécessite des interactions avec d'autres personnes Services AWS ou des ressources pour être traitée. Dans ce cas, vous devez disposer d'autorisations nécessaires pour effectuer les deux actions. Pour plus de détails sur les politiques relatives FAS aux demandes, consultez la section Transférer les sessions d'accès.
  - Rôle de service — Un rôle de service est un IAMrôle qu'un service assume pour effectuer des actions en votre nom. Un IAM administrateur peut créer, modifier et supprimer un rôle de service de l'intérieurIAM. Pour plus d'informations, consultez la section Création d'un rôle auquel déléguer des autorisations Service AWS dans le Guide de IAM l'utilisateur.
  - Rôle lié à un service — Un rôle lié à un service est un type de rôle de service lié à un. Service AWS Le service peut endosser le rôle afin d'effectuer une action en votre nom. Les rôles liés à un service apparaissent dans votre Compte AWS répertoire et appartiennent au service. Un IAM administrateur peut consulter, mais pas modifier les autorisations pour les rôles liés à un service.
- Applications exécutées sur Amazon EC2 : vous pouvez utiliser un IAM rôle pour gérer les informations d'identification temporaires pour les applications qui s'exécutent sur une EC2 instance et qui AWS CLI soumettent des AWS API demandes. Cela est préférable au stockage des clés d'accès dans l'EC2instance. Pour attribuer un AWS rôle à une EC2 instance et le rendre disponible pour toutes ses applications, vous devez créer un profil d'instance attaché à l'instance. Un profil d'instance contient le rôle et permet aux programmes exécutés sur l'EC2instance d'obtenir des informations d'identification temporaires. Pour plus d'informations, consultez la section Utilisation

d'un IAM rôle pour accorder des autorisations aux applications exécutées sur des EC2 instances Amazon dans le Guide de IAM l'utilisateur.

Pour savoir s'il faut utiliser IAM des rôles ou des IAM utilisateurs, voir Quand créer un IAM rôle (au lieu d'un utilisateur) dans le guide de IAM l'utilisateur.

Gestion des accès à l'aide de politiques

Vous contrôlez l'accès en AWS créant des politiques et en les associant à AWS des identités ou à des ressources. Une politique est un objet AWS qui, lorsqu'il est associé à une identité ou à une ressource, définit leurs autorisations. AWS évalue ces politiques lorsqu'un principal (utilisateur, utilisateur root ou session de rôle) fait une demande. Les autorisations dans les politiques déterminent si la demande est autorisée ou refusée. La plupart des politiques sont stockées AWS sous forme de JSON documents. Pour plus d'informations sur la structure et le contenu des documents de JSON politique, voir Présentation des JSON politiques dans le guide de IAM l'utilisateur.

Les administrateurs peuvent utiliser AWS JSON des politiques pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

Par défaut, les utilisateurs et les rôles ne disposent d'aucune autorisation. Pour autoriser les utilisateurs à effectuer des actions sur les ressources dont ils ont besoin, un IAM administrateur peut créer des IAM politiques. L'administrateur peut ensuite ajouter les IAM politiques aux rôles, et les utilisateurs peuvent assumer les rôles.

IAMles politiques définissent les autorisations pour une action, quelle que soit la méthode que vous utilisez pour effectuer l'opération. Par exemple, supposons que vous disposiez d'une politique qui autorise l'action `iam:GetRole`. Un utilisateur appliquant cette politique peut obtenir des informations sur le rôle auprès du AWS Management Console AWS CLI, ou du AWS API.

Politiques basées sur l'identité

Les politiques basées sur l'identité sont JSON des documents de politique d'autorisation que vous pouvez joindre à une identité, telle qu'un IAM utilisateur, un groupe d'utilisateurs ou un rôle. Ces politiques contrôlent quel type d'actions des utilisateurs et des rôles peuvent exécuter, sur quelles ressources et dans quelles conditions. Pour savoir comment créer une politique basée sur l'identité, consultez la section Création de IAM politiques dans le Guide de l'IAMutilisateur.

Les politiques basées sur l’identité peuvent être classées comme des politiques en ligne ou des politiques gérées. Les politiques en ligne sont intégrées directement à un utilisateur, groupe ou rôle. Les politiques gérées sont des politiques autonomes que vous pouvez associer à plusieurs utilisateurs, groupes et rôles au sein de votre Compte AWS. Les politiques gérées incluent les politiques AWS gérées et les politiques gérées par le client. Pour savoir comment choisir entre une politique gérée ou une politique intégrée, voir Choisir entre des politiques gérées et des politiques intégrées dans le Guide de l'IAMutilisateur.

Politiques basées sur les ressources

Les politiques basées sur les ressources sont des documents JSON de stratégie que vous attachez à une ressource. Les politiques de confiance dans les IAM rôles et les politiques relatives aux compartiments Amazon S3 sont des exemples de politiques basées sur les ressources. Dans les services qui sont compatibles avec les politiques basées sur les ressources, les administrateurs de service peuvent les utiliser pour contrôler l’accès à une ressource spécifique. Pour la ressource dans laquelle se trouve la politique, cette dernière définit quel type d’actions un principal spécifié peut effectuer sur cette ressource et dans quelles conditions. Vous devez spécifier un principal dans une politique basée sur les ressources. Les principaux peuvent inclure des comptes, des utilisateurs, des rôles, des utilisateurs fédérés ou. Services AWS

Les politiques basées sur les ressources sont des politiques en ligne situées dans ce service. Vous ne pouvez pas utiliser de politiques AWS gérées depuis une IAM stratégie basée sur les ressources.

Listes de contrôle d'accès (ACLs)

Les listes de contrôle d'accès (ACLs) contrôlent les principaux (membres du compte, utilisateurs ou rôles) autorisés à accéder à une ressource. ACLssont similaires aux politiques basées sur les ressources, bien qu'elles n'utilisent pas le format du document JSON de stratégie.

Amazon S3 et Amazon VPC sont des exemples de services compatiblesACLs. AWS WAF Pour en savoir plusACLs, consultez la présentation de la liste de contrôle d'accès (ACL) dans le guide du développeur Amazon Simple Storage Service.

Autres types de politique

AWS prend en charge d'autres types de politiques moins courants. Ces types de politiques peuvent définir le nombre maximum d’autorisations qui vous sont accordées par des types de politiques plus courants.

- Limites d'autorisations — Une limite d'autorisations est une fonctionnalité avancée dans laquelle vous définissez le maximum d'autorisations qu'une politique basée sur l'identité peut accorder à une IAM entité (IAMutilisateur ou rôle). Vous pouvez définir une limite d'autorisations pour une entité. Les autorisations en résultant représentent la combinaison des politiques basées sur l'identité d'une entité et de ses limites d'autorisation. Les politiques basées sur les ressources qui spécifient l'utilisateur ou le rôle dans le champ `Principal` ne sont pas limitées par les limites d'autorisations. Un refus explicite dans l'une de ces politiques remplace l'autorisation. Pour plus d'informations sur les limites d'autorisations, voir Limites d'autorisations pour les IAM entités dans le Guide de IAM l'utilisateur.
- Politiques de contrôle des services (SCPs) : SCPs JSON politiques qui spécifient les autorisations maximales pour une organisation ou une unité organisationnelle (UO) dans AWS Organizations. AWS Organizations est un service permettant de regrouper et de gérer de manière centralisée Comptes AWS les multiples propriétés de votre entreprise. Si vous activez toutes les fonctionnalités d'une organisation, vous pouvez appliquer des politiques de contrôle des services (SCPs) à l'un ou à l'ensemble de vos comptes. Les SCP limites d'autorisations pour les entités présentes dans les comptes des membres, y compris chacune d'entre elles Utilisateur racine d'un compte AWS. Pour plus d'informations sur les OrganizationsSCPs, voir Politiques de contrôle des services dans le Guide de AWS Organizations l'utilisateur.
- Politiques de séance : les politiques de séance sont des politiques avancées que vous utilisez en tant que paramètre lorsque vous créez par programmation une séance temporaire pour un rôle ou un utilisateur fédéré. Les autorisations de séance en résultant sont une combinaison des politiques basées sur l'identité de l'utilisateur ou du rôle et des politiques de séance. Les autorisations peuvent également provenir d'une politique basée sur les ressources. Un refus explicite dans l'une de ces politiques annule l'autorisation. Pour plus d'informations, consultez la section Politiques de session dans le guide de IAM l'utilisateur.

Plusieurs types de politique

Lorsque plusieurs types de politiques s'appliquent à la requête, les autorisations en résultant sont plus compliquées à comprendre. Pour savoir comment AWS déterminer s'il faut autoriser une demande lorsque plusieurs types de politiques sont impliqués, consultez la section Logique d'évaluation des politiques dans le guide de IAM l'utilisateur.

Comment fonctionne AWS Parallel Computing Service avec IAM

Avant d'utiliser IAM pour gérer l'accès à AWS PCS, découvrez quelles IAM fonctionnalités sont disponibles AWS PCS.

IAMfonctionnalités que vous pouvez utiliser avec AWS Parallel Computing Service

IAMfonctionnalité	AWS PCSsoutien
Politiques basées sur l'identité	Oui
Politiques basées sur les ressources	Non
Actions de politique	Oui
Ressources de politique	Oui
Clés de condition de politique (spécifiques au service)	Oui
ACLs	Non
ABAC(balises dans les politiques)	Oui
Informations d'identification temporaires	Oui
Autorisations de principal	Oui
Fonctions du service	Non
Rôles liés à un service	Oui

Pour obtenir une vue d'ensemble de la façon dont AWS PCS les autres AWS services fonctionnent avec la plupart des IAM fonctionnalités, consultez la section AWS Services compatibles IAM dans le Guide de IAM l'utilisateur.

Politiques basées sur l'identité pour AWS PCS

Prend en charge les politiques basées sur l'identité : oui

Les politiques basées sur l'identité sont JSON des documents de politique d'autorisation que vous pouvez joindre à une identité, telle qu'un IAM utilisateur, un groupe d'utilisateurs ou un rôle. Ces politiques contrôlent quel type d'actions des utilisateurs et des rôles peuvent exécuter, sur quelles ressources et dans quelles conditions. Pour savoir comment créer une politique basée sur l'identité, consultez la section Création de IAM politiques dans le Guide de l'IAMutilisateur.

Avec les politiques IAM basées sur l'identité, vous pouvez spécifier les actions et les ressources autorisées ou refusées ainsi que les conditions dans lesquelles les actions sont autorisées ou refusées. Vous ne pouvez pas spécifier le principal dans une politique basée sur une identité car celle-ci s'applique à l'utilisateur ou au rôle auquel elle est attachée. Pour en savoir plus sur tous les éléments que vous pouvez utiliser dans une JSON politique, consultez la référence aux éléments de IAM JSON politique dans le Guide de IAM l'utilisateur.

Exemples de politiques basées sur l'identité pour AWS PCS

Pour consulter des exemples de politiques AWS PCS basées sur l'identité, consultez. Exemples de politiques basées sur l'identité pour le service de calcul AWS parallèle

Politiques basées sur les ressources au sein de AWS PCS

Prend en charge les politiques basées sur les ressources : non

Les politiques basées sur les ressources sont des documents JSON de stratégie que vous attachez à une ressource. Les politiques de confiance dans les IAM rôles et les politiques relatives aux compartiments Amazon S3 sont des exemples de politiques basées sur les ressources. Dans les services qui sont compatibles avec les politiques basées sur les ressources, les administrateurs de service peuvent les utiliser pour contrôler l'accès à une ressource spécifique. Pour la ressource dans laquelle se trouve la politique, cette dernière définit quel type d'actions un principal spécifié peut effectuer sur cette ressource et dans quelles conditions. Vous devez spécifier un principal dans une politique basée sur les ressources. Les principaux peuvent inclure des comptes, des utilisateurs, des rôles, des utilisateurs fédérés ou. Services AWS

Pour activer l'accès entre comptes, vous pouvez spécifier un compte entier ou IAM des entités d'un autre compte comme principal dans une politique basée sur les ressources. L'ajout d'un principal entre comptes à une politique basée sur les ressources ne représente qu'une partie de l'instauration de la relation d'approbation. Lorsque le principal et la ressource sont différents Comptes AWS, un IAM administrateur du compte de confiance doit également accorder à l'entité principale (utilisateur ou rôle) l'autorisation d'accéder à la ressource. Pour ce faire, il attache une politique basée sur une identité à l'entité. Toutefois, si une politique basée sur des ressources accorde l'accès à un principal dans le même compte, aucune autre politique basée sur l'identité n'est requise. Pour plus d'informations, voir Accès aux ressources entre comptes IAM dans le Guide de IAM l'utilisateur.

Actions politiques pour AWS PCS

Prend en charge les actions de politique : oui

Les administrateurs peuvent utiliser AWS JSON des politiques pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

L'`Action`élément d'une JSON politique décrit les actions que vous pouvez utiliser pour autoriser ou refuser l'accès dans une politique. Les actions de stratégie portent généralement le même nom que l' AWS APIopération associée. Il existe certaines exceptions, telles que les actions avec autorisation uniquement qui n'ont pas d'opération correspondante. API Certaines opérations nécessitent également plusieurs actions dans une politique. Ces actions supplémentaires sont nommées actions dépendantes.

Intégration d'actions dans une stratégie afin d'accorder l'autorisation d'exécuter les opérations associées.

Pour consulter la liste des AWS PCS actions, reportez-vous à la section Actions définies par le service de calcul AWS parallèle dans la référence d'autorisation du service.

Les actions de politique en AWS PCS cours utilisent le préfixe suivant avant l'action :

```
pcs
```

Pour indiquer plusieurs actions dans une seule déclaration, séparez-les par des virgules.

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

Pour consulter des exemples de politiques AWS PCS basées sur l'identité, consultez. Exemples de politiques basées sur l'identité pour le service de calcul AWS parallèle

Ressources politiques pour AWS PCS

Prend en charge les ressources de politique : oui

Les administrateurs peuvent utiliser AWS JSON des politiques pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

L'élément `Resource` JSON de stratégie indique le ou les objets auxquels s'applique l'action. Les instructions doivent inclure un élément `Resource` ou `NotResource`. Il est recommandé de spécifier une ressource en utilisant son Amazon Resource Name (ARN). Vous pouvez le faire pour des actions qui prennent en charge un type de ressource spécifique, connu sous la dénomination autorisations de niveau ressource.

Pour les actions qui ne sont pas compatibles avec les autorisations de niveau ressource, telles que les opérations de liste, utilisez un caractère générique (*) afin d'indiquer que l'instruction s'applique à toutes les ressources.

```
"Resource": "*"
```

Pour consulter la liste des types de AWS PCS ressources et leurs caractéristiquesARNs, consultez la section Ressources définies par le service de calcul AWS parallèle dans la référence d'autorisation du service. Pour savoir avec quelles actions vous pouvez spécifier pour chaque ressource, voir Actions définies par le service de calcul AWS parallèle. ARN

Pour consulter des exemples de politiques AWS PCS basées sur l'identité, consultez. Exemples de politiques basées sur l'identité pour le service de calcul AWS parallèle

Clés de conditions de politique pour AWS PCS

Prend en charge les clés de condition de politique spécifiques au service : oui

Les administrateurs peuvent utiliser AWS JSON des politiques pour spécifier qui a accès à quoi. C'est-à-dire, quel principal peut effectuer des actions sur quelles ressources et dans quelles conditions.

L'élément `Condition` (ou le bloc `Condition`) vous permet de spécifier des conditions lorsqu'une instruction est appliquée. L'élément `Condition` est facultatif. Vous pouvez créer des expressions conditionnelles qui utilisent des opérateurs de condition, tels que les signes égal ou inférieur à, pour faire correspondre la condition de la politique aux valeurs de la demande.

Si vous spécifiez plusieurs éléments `Condition` dans une instruction, ou plusieurs clés dans un seul élément `Condition`, AWS les évalue à l'aide d'une opération `AND` logique. Si vous spécifiez plusieurs valeurs pour une seule clé de condition, AWS évalue la condition à l'aide d'une `OR` opération logique. Toutes les conditions doivent être remplies avant que les autorisations associées à l'instruction ne soient accordées.

Vous pouvez aussi utiliser des variables d'espace réservé quand vous spécifiez des conditions. Par exemple, vous pouvez autoriser un IAM utilisateur à accéder à une ressource uniquement si celle-ci est étiquetée avec son nom IAM d'utilisateur. Pour plus d'informations, consultez IAMla section Éléments de politique : variables et balises dans le Guide de IAM l'utilisateur.

AWS prend en charge les clés de condition globales et les clés de condition spécifiques au service. Pour voir toutes les clés de condition AWS globales, voir les clés contextuelles de condition AWS globales dans le guide de IAM l'utilisateur.

Pour consulter la liste des clés de AWS PCS condition, voir Clés de condition pour le service de calcul AWS parallèle dans la référence d'autorisation de service. Pour savoir avec quelles actions et ressources vous pouvez utiliser une clé de condition, voir Actions définies par AWS Parallel Computing Service.

Pour consulter des exemples de politiques AWS PCS basées sur l'identité, consultez. Exemples de politiques basées sur l'identité pour le service de calcul AWS parallèle

ACLsdans AWS PCS

Supports ACLs : Non

Les listes de contrôle d'accès (ACLs) contrôlent les principaux (membres du compte, utilisateurs ou rôles) autorisés à accéder à une ressource. ACLssont similaires aux politiques basées sur les ressources, bien qu'elles n'utilisent pas le format du document JSON de stratégie.

ABACavec AWS PCS

Supports ABAC (balises dans les politiques) : Oui

Le contrôle d'accès basé sur les attributs (ABAC) est une stratégie d'autorisation qui définit les autorisations en fonction des attributs. Dans AWS, ces attributs sont appelés balises. Vous pouvez associer des balises à IAM des entités (utilisateurs ou rôles) et à de nombreuses AWS ressources. Le balisage des entités et des ressources est la première étape deABAC. Vous concevez ensuite des ABAC politiques pour autoriser les opérations lorsque le tag du principal correspond à celui de la ressource à laquelle il essaie d'accéder.

ABACest utile dans les environnements qui se développent rapidement et aide dans les situations où la gestion des politiques devient fastidieuse.

Pour contrôler l'accès basé sur des étiquettes, vous devez fournir les informations d'étiquette dans l'élément de condition d'une politique utilisant les clés de condition `aws:ResourceTag/key-name`, `aws:RequestTag/key-name` ou `aws:TagKeys`.

Si un service prend en charge les trois clés de condition pour tous les types de ressources, alors la valeur pour ce service est Oui. Si un service prend en charge les trois clés de condition pour certains types de ressources uniquement, la valeur est Partielle.

Pour plus d'informationsABAC, voir Qu'est-ce que c'est ABAC ? dans le guide de IAM l'utilisateur. Pour consulter un didacticiel présentant les étapes de configurationABAC, voir Utiliser le contrôle d'accès basé sur les attributs (ABAC) dans le guide de l'IAMutilisateur.

Utilisation d'informations d'identification temporaires avec AWS PCS

Prend en charge les informations d'identification temporaires : oui

Certains Services AWS ne fonctionnent pas lorsque vous vous connectez à l'aide d'informations d'identification temporaires. Pour plus d'informations, y compris celles qui Services AWS fonctionnent avec des informations d'identification temporaires, consultez Services AWS la section relative à l'utilisation IAM dans le Guide de IAM l'utilisateur.

Vous utilisez des informations d'identification temporaires si vous vous connectez à l' AWS Management Console aide d'une méthode autre qu'un nom d'utilisateur et un mot de passe. Par exemple, lorsque vous accédez à AWS l'aide du lien d'authentification unique (SSO) de votre entreprise, ce processus crée automatiquement des informations d'identification temporaires. Vous créez également automatiquement des informations d'identification temporaires lorsque vous vous connectez à la console en tant qu'utilisateur, puis changez de rôle. Pour plus d'informations sur le changement de rôle, consultez la section Passage à un rôle (console) dans le guide de IAM l'utilisateur.

Vous pouvez créer manuellement des informations d'identification temporaires à l'aide du AWS CLI ou AWS API. Vous pouvez ensuite utiliser ces informations d'identification temporaires pour y accéder AWS. AWS recommande de générer dynamiquement des informations d'identification temporaires au lieu d'utiliser des clés d'accès à long terme. Pour plus d'informations, consultez la section Informations d'identification de sécurité temporaires dans IAM.

Autorisations principales interservices pour AWS PCS

Prend en charge les sessions d'accès transféré (FAS) : Oui

Lorsque vous utilisez un IAM utilisateur ou un rôle pour effectuer des actions AWS, vous êtes considéré comme un mandant. Lorsque vous utilisez certains services, vous pouvez effectuer une action qui initie une autre action dans un autre service. FASutilise les autorisations du principal appelant an Service AWS, combinées à la demande Service AWS pour adresser des demandes aux services en aval. FASles demandes ne sont effectuées que lorsqu'un service reçoit une demande qui nécessite des interactions avec d'autres personnes Services AWS ou des ressources pour être traitée. Dans ce cas, vous devez disposer d'autorisations nécessaires pour effectuer les deux actions. Pour plus de détails sur les politiques relatives FAS aux demandes, consultez la section Transférer les sessions d'accès.

Rôles de service pour AWS PCS

Supporte les rôles de service : Non

Un rôle de service est un IAMrôle qu'un service assume pour effectuer des actions en votre nom. Un IAM administrateur peut créer, modifier et supprimer un rôle de service de l'intérieurIAM. Pour plus d'informations, consultez la section Création d'un rôle auquel déléguer des autorisations Service AWS dans le Guide de IAM l'utilisateur.

> **Warning**
>
> La modification des autorisations associées à un rôle de service peut perturber AWS PCS les fonctionnalités. Modifiez les rôles de service uniquement lorsque AWS PCS vous recevez des instructions à cet effet.

Rôles liés à un service pour AWS PCS

Prend en charge les rôles liés aux services : Oui

Un rôle lié à un service est un type de rôle de service lié à un. Service AWS Le service peut endosser le rôle afin d'effectuer une action en votre nom. Les rôles liés à un service apparaissent dans votre Compte AWS répertoire et appartiennent au service. Un IAM administrateur peut consulter, mais pas modifier les autorisations pour les rôles liés à un service.

Pour plus de détails sur la création ou la gestion des rôles liés à un service, consultez la section AWS Services compatibles avec. IAM Recherchez un service dans le tableau qui inclut un `Yes` dans la colonne Rôle lié à un service. Choisissez le lien Oui pour consulter la documentation du rôle lié à ce service.

Exemples de politiques basées sur l'identité pour le service de calcul AWS parallèle

Par défaut, les utilisateurs et les rôles ne sont pas autorisés à créer ou à modifier AWS PCS des ressources. Ils ne peuvent pas non plus effectuer de tâches en utilisant le AWS Management Console, AWS Command Line Interface (AWS CLI) ou AWS API. Pour autoriser les utilisateurs à effectuer des actions sur les ressources dont ils ont besoin, un IAM administrateur peut créer des IAM politiques. L'administrateur peut ensuite ajouter les IAM politiques aux rôles, et les utilisateurs peuvent assumer les rôles.

Pour savoir comment créer une politique IAM basée sur l'identité à l'aide de ces exemples de documents de JSON stratégie, consultez la section Création de IAM politiques dans le guide de l'IAMutilisateur.

Pour plus de détails sur les actions et les types de ressources définis par AWS PCS, y compris le format de ARNs pour chacun des types de ressources, voir Actions, ressources et clés de condition pour le service de calcul AWS parallèle dans la référence d'autorisation du service.

Rubriques

- Bonnes pratiques en matière de politiques
- Utilisation de la AWS PCS console
- Autorisation accordée aux utilisateurs pour afficher leurs propres autorisations

Bonnes pratiques en matière de politiques

Les politiques basées sur l'identité déterminent si quelqu'un peut créer, accéder ou supprimer AWS PCS des ressources dans votre compte. Ces actions peuvent entraîner des frais pour votre Compte AWS. Lorsque vous créez ou modifiez des politiques basées sur l'identité, suivez ces instructions et recommandations :

- Commencez AWS par les politiques gérées et passez aux autorisations du moindre privilège : pour commencer à accorder des autorisations à vos utilisateurs et à vos charges de travail, utilisez les politiques AWS gérées qui accordent des autorisations pour de nombreux cas d'utilisation courants. Ils sont disponibles dans votre Compte AWS. Nous vous recommandons de réduire davantage les autorisations en définissant des politiques gérées par les AWS clients spécifiques à vos cas d'utilisation. Pour plus d'informations, consultez les politiques AWS gérées ou les politiques AWS gérées pour les fonctions professionnelles dans le Guide de IAM l'utilisateur.

- Appliquer les autorisations du moindre privilège : lorsque vous définissez des autorisations à IAM l'aide de politiques, accordez uniquement les autorisations nécessaires à l'exécution d'une tâche. Pour ce faire, vous définissez les actions qui peuvent être entreprises sur des ressources spécifiques dans des conditions spécifiques, également appelées autorisations de moindre privilège. Pour plus d'informations sur l'utilisation IAM pour appliquer des autorisations, consultez la section Politiques et autorisations du Guide de IAM l'utilisateur. IAM
- Utilisez des conditions dans IAM les politiques pour restreindre davantage l'accès : vous pouvez ajouter une condition à vos politiques pour limiter l'accès aux actions et aux ressources. Par exemple, vous pouvez rédiger une condition de politique pour spécifier que toutes les demandes doivent être envoyées en utilisantSSL. Vous pouvez également utiliser des conditions pour accorder l'accès aux actions de service si elles sont utilisées par le biais d'un service spécifique Service AWS, tel que AWS CloudFormation. Pour plus d'informations, voir Éléments IAM JSON de politique : Condition dans le guide de IAM l'utilisateur.
- Utilisez IAM Access Analyzer pour valider vos IAM politiques afin de garantir des autorisations sécurisées et fonctionnelles. IAM Access Analyzer valide les politiques nouvelles et existantes afin qu'elles soient conformes au langage des IAM politiques (JSON) et IAM aux meilleures pratiques. IAMAccess Analyzer fournit plus de 100 vérifications des politiques et des recommandations exploitables pour vous aider à créer des politiques sécurisées et fonctionnelles. Pour plus d'informations, consultez la section Validation des politiques d'IAMAccess Analyzer dans le guide de IAM l'utilisateur.
- Exiger l'authentification multifactorielle (MFA) : si vous avez un scénario qui nécessite des IAM utilisateurs ou un utilisateur root Compte AWS, activez-le MFA pour une sécurité supplémentaire. Pour exiger le MFA moment où les API opérations sont appelées, ajoutez MFA des conditions à vos politiques. Pour plus d'informations, consultez la section Configuration de l'APIaccès MFA protégé dans le Guide de l'IAMutilisateur.

Pour plus d'informations sur les meilleures pratiques en matière de sécuritéIAM, consultez la section Bonnes pratiques en matière de sécurité IAM dans le Guide de IAM l'utilisateur.

Utilisation de la AWS PCS console

Pour accéder à la console AWS Parallel Computing Service, vous devez disposer d'un ensemble minimal d'autorisations. Ces autorisations doivent vous permettre de répertorier et d'afficher les détails AWS PCS des ressources de votre Compte AWS. Si vous créez une stratégie basée sur l'identité qui est plus restrictive que l'ensemble minimum d'autorisations requis, la console ne fonctionnera pas comme prévu pour les entités (utilisateurs ou rôles) tributaires de cette stratégie.

Il n'est pas nécessaire d'accorder des autorisations de console minimales aux utilisateurs qui appellent uniquement le AWS CLI ou le AWS API. Au lieu de cela, autorisez uniquement l'accès aux actions correspondant à l'APIopération qu'ils tentent d'effectuer.

Pour plus d'informations sur les autorisations minimales requises pour utiliser la AWS PCS console, consultezAutorisations minimales pour AWS PCS.

Autorisation accordée aux utilisateurs pour afficher leurs propres autorisations

Cet exemple montre comment créer une politique qui permet aux IAM utilisateurs de consulter les politiques intégrées et gérées associées à leur identité d'utilisateur. Cette politique inclut les autorisations permettant d'effectuer cette action sur la console ou par programmation à l'aide du AWS CLI ou. 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 politiques gérées pour le service de calcul AWS parallèle

Une politique AWS gérée est une politique autonome créée et administrée par AWS. AWS les politiques gérées sont conçues pour fournir des autorisations pour de nombreux cas d'utilisation courants afin que vous puissiez commencer à attribuer des autorisations aux utilisateurs, aux groupes et aux rôles.

N'oubliez pas que les politiques AWS gérées peuvent ne pas accorder d'autorisations de moindre privilège pour vos cas d'utilisation spécifiques, car elles sont accessibles à tous les AWS clients. Nous vous recommandons de réduire encore les autorisations en définissant des politiques gérées par le client qui sont propres à vos cas d'utilisation.

Vous ne pouvez pas modifier les autorisations définies dans les politiques AWS gérées. Si les autorisations définies dans une politique AWS gérée sont AWS mises à jour, la mise à jour affecte toutes les identités principales (utilisateurs, groupes et rôles) auxquelles la politique est attachée. AWS est le plus susceptible de mettre à jour une politique AWS gérée lorsqu'une nouvelle Service AWS est lancée ou que de nouvelles API opérations sont disponibles pour les services existants.

Pour plus d'informations, consultez la section Politiques AWS gérées dans le Guide de IAM l'utilisateur.

AWS politique gérée : AWSPCSServiceRolePolicy

Vous ne pouvez pas vous attacher AWSPCSServiceRolePolicy à vos IAM entités. Cette politique est associée à un rôle lié à un service qui permet d' AWS PCSeffectuer des actions en votre nom. Pour de plus amples informations, veuillez consulter Rôles liés à un service pour AWS PCS.

Détails de l'autorisation

Cette politique inclut les autorisations suivantes.

- `ec2`— Permet AWS PCS de créer et de gérer les EC2 ressources Amazon.
- `iam`— Permet AWS PCS de créer un rôle lié à un service pour la EC2 flotte Amazon et de le transmettre à Amazon. EC2

- `cloudwatch`— Permet AWS PCS de publier des statistiques de service sur Amazon CloudWatch.
- `secretsmanager`— Permet AWS PCS de gérer les secrets des ressources AWS PCS du cluster.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PermissionsToCreatePCSNetworkInterfaces",
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface"
            ],
            "Resource": "arn:aws:ec2:*:*:network-interface/*",
            "Condition": {
                "Null": {
                    "aws:RequestTag/AWSPCSManaged": "false"
                }
            }
        },
        {
            "Sid": "PermissionsToCreatePCSNetworkInterfacesInSubnet",
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface"
            ],
            "Resource": [
                "arn:aws:ec2:*:*:subnet/*",
                "arn:aws:ec2:*:*:security-group/*"
            ]
        },
        {
            "Sid": "PermissionsToManagePCSNetworkInterfaces",
            "Effect": "Allow",
            "Action": [
                "ec2:DeleteNetworkInterface",
                "ec2:CreateNetworkInterfacePermission"
            ],
            "Resource": "arn:aws:ec2:*:*:network-interface/*",
            "Condition": {
                "Null": {
                        "aws:ResourceTag/AWSPCSManaged": "false"
                }
```

```
            }
        },
        {
            "Sid": "PermissionsToDescribePCSResources",
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "ec2:DescribeNetworkInterfaces",
                "ec2:DescribeLaunchTemplates",
                "ec2:DescribeLaunchTemplateVersions",
                "ec2:DescribeInstances",
                "ec2:DescribeInstanceTypes",
                "ec2:DescribeInstanceStatus",
                "ec2:DescribeInstanceAttribute",
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeKeyPairs",
                "ec2:DescribeImages",
                "ec2:DescribeImageAttribute"
            ],
            "Resource": "*"
        },
        {
            "Sid": "PermissionsToCreatePCSLaunchTemplates",
            "Effect": "Allow",
            "Action": [
                "ec2:CreateLaunchTemplate"
            ],
            "Resource": "arn:aws:ec2:*:*:launch-template/*",
            "Condition": {
                "Null": {
                    "aws:RequestTag/AWSPCSManaged": "false"
                }
            }
        },
        {
            "Sid": "PermissionsToManagePCSLaunchTemplates",
            "Effect": "Allow",
            "Action": [
                "ec2:DeleteLaunchTemplate",
                "ec2:DeleteLaunchTemplateVersions",
                "ec2:CreateLaunchTemplateVersion"
            ],
            "Resource": "arn:aws:ec2:*:*:launch-template/*",
```

```
            "Condition": {
                "Null": {
                    "aws:ResourceTag/AWSPCSManaged": "false"
                }
            }
        },
        {
            "Sid": "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 PCSmises à jour des politiques AWS gérées

Consultez les détails des mises à jour des politiques AWS gérées AWS PCS depuis que ce service a commencé à suivre ces modifications. Pour recevoir des alertes automatiques concernant les modifications apportées à cette page, abonnez-vous au RSS fil sur la page Historique du AWS PCS document.

Modification	Description	Date
AWS PCS a démarré le suivi des modifications	AWS PCSa commencé à suivre les modifications apportées AWS à ses politique s gérées.	28 août 2024

Rôles liés à un service pour AWS PCS

AWS Le service de calcul parallèle utilise AWS Identity and Access Management (IAM) des rôles liés au service. Un rôle lié à un service est un type unique de IAM rôle directement lié à. AWS PCS Les rôles liés au service sont prédéfinis par AWS PCS et incluent toutes les autorisations dont le service a besoin pour appeler d'autres AWS services en votre nom.

Un rôle lié à un service facilite la configuration AWS PCS car vous n'avez pas à ajouter manuellement les autorisations nécessaires. AWS PCSdéfinit les autorisations associées à ses rôles liés aux services et, sauf indication contraire, seul AWS PCS peut assumer ses rôles. Les autorisations définies incluent la politique de confiance et la politique d'autorisations, et cette politique d'autorisations ne peut être attachée à aucune autre IAM entité.

Vous pouvez supprimer un rôle lié à un service uniquement après la suppression préalable de ses ressources connexes. Cela protège vos AWS PCS ressources car vous ne pouvez pas supprimer accidentellement l'autorisation d'accès aux ressources.

Pour plus d'informations sur les autres services qui prennent en charge les rôles liés à un service, consultez la section AWS Services qui fonctionnent avec IAM et recherchez les services dont la valeur est Oui dans la colonne Rôles liés à un service. Choisissez un Oui ayant un lien permettant de consulter les détails du rôle pour ce service.

Autorisations de rôle liées à un service pour AWS PCS

AWS PCSutilise le rôle lié au service nommé AWSServiceRoleForPCS— Autoriser AWS PCS pour gérer les ressources AmazonEC2.

Le rôle AWSServiceRoleForPCS lié à un service fait confiance aux services suivants pour assumer le rôle :

- `pcs.amazonaws.com`

La politique d'autorisations de rôle nommée AWSPCSServiceRolePolicypermet AWS PCS d'effectuer des actions sur des ressources spécifiques.

Vous devez configurer les autorisations de manière à permettre à vos utilisateurs, groupes ou rôles de créer, modifier ou supprimer un rôle lié à un service. Pour plus d'informations, consultez la section Autorisations relatives aux rôles liés à un service dans le Guide de l'IAMutilisateur.

Création d'un rôle lié à un service pour AWS PCS

Il n'est pas nécessaire de créer manuellement un rôle lié à un service. AWS PCScrée un rôle lié à un service pour vous lorsque vous créez un cluster.

Modification d'un rôle lié à un service pour AWS PCS

AWS PCSne vous permet pas de modifier le rôle AWSServiceRoleForPCS lié au service. Une fois que vous avez créé un rôle lié à un service, vous ne pouvez pas changer le nom du rôle, car plusieurs entités peuvent faire référence à ce rôle. Vous pouvez toutefois modifier la description du rôle à l'aide deIAM. Pour plus d'informations, consultez la section Modification d'un rôle lié à un service dans le Guide de l'IAMutilisateur.

Supprimer un rôle lié à un service pour AWS PCS

Si vous n'avez plus besoin d'utiliser une fonctionnalité ou un service qui nécessite un rôle lié à un service, nous vous recommandons de supprimer ce rôle. De cette façon, vous n'avez aucune entité inutilisée qui n'est pas surveillée ou gérée activement. Cependant, vous devez nettoyer les ressources de votre rôle lié à un service avant de pouvoir les supprimer manuellement.

> **Note**
>
> Si le AWS PCS service utilise le rôle lorsque vous essayez de supprimer les ressources, la suppression risque d'échouer. Si cela se produit, patientez quelques minutes et réessayez.

Pour supprimer AWS PCS les ressources utilisées par AWSServiceRoleForPCS

Vous devez supprimer tous vos clusters pour supprimer le rôle AWSServiceRoleForPCS lié au service. Pour plus d'informations, voir Supprimer un cluster.

Pour supprimer manuellement le rôle lié à un service à l'aide de IAM

Utilisez la IAM console AWS CLI, le ou le AWS API pour supprimer le rôle AWSServiceRoleForPCS lié au service. Pour plus d'informations, consultez la section Suppression d'un rôle lié à un service dans le Guide de l'IAMutilisateur.

Régions prises en charge pour les rôles liés à un service AWS PCS

AWS PCSprend en charge l'utilisation de rôles liés au service dans toutes les régions où le service est disponible. Pour de plus amples informations, veuillez consulter AWS Régions et points de terminaison.

Rôle Amazon EC2 Spot pour AWS PCS

Si vous souhaitez créer un groupe de nœuds de AWS PCS calcul utilisant Spot comme option d'achat, vous devez également avoir le rôle AWSServiceRoleForEC2Spotlié au service dans votre. Compte AWS Vous pouvez utiliser la AWS CLI commande suivante pour créer le rôle. Pour plus d'informations, voir Créer un rôle lié à un service et Créer un rôle pour déléguer des autorisations à un AWS service dans le Guide de l'AWS Identity and Access Management utilisateur.

```
aws iam create-service-linked-role --aws-service-name spot.amazonaws.com
```

> **Note**
>
> Le message d'erreur suivant s'affiche si vous avez Compte AWS déjà un `AWSServiceRoleForEC2Spot` IAM rôle.
>
> ```
> An error occurred (InvalidInput) when calling the CreateServiceLinkedRole
> operation: Service role name AWSServiceRoleForEC2Spot has been taken in this
> account, please try a different suffix.
> ```

Autorisations minimales pour AWS PCS

Cette section décrit les IAM autorisations minimales requises pour qu'une IAM identité (utilisateur, groupe ou rôle) puisse utiliser le service.

Table des matières

Autorisations minimales pour utiliser API des actions

APIaction	Autorisations minimales	Autorisations supplémentaires pour la console
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`

APIaction	Autorisations minimales	Autorisations supplémentaires pour la console
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`

APIaction	Autorisations minimales	Autorisations supplémentaires pour la console
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`

APIaction	Autorisations minimales	Autorisations supplémentaires pour la console
DeleteQueue	`pcs:DeleteQueue`	

Autorisations minimales requises pour utiliser les balises

Les autorisations suivantes sont requises pour utiliser des balises contenant vos ressources AWS PCS.

```
pcs:ListTagsForResource
pcs:TagResource
pcs:UntagResource
```

Autorisations minimales requises pour prendre en charge les journaux

AWS PCSenvoie les données du journal à Amazon CloudWatch Logs (CloudWatch Logs). Vous devez vous assurer que votre identité dispose des autorisations minimales pour utiliser les CloudWatch journaux. Pour plus d'informations, consultez la section Présentation de la gestion des autorisations d'accès à vos ressources CloudWatch Logs dans le guide de l'utilisateur Amazon CloudWatch Logs.

Pour plus d'informations sur les autorisations requises pour qu'un service envoie des CloudWatch journaux à Logs, consultez la section Activation de la journalisation à partir AWS des services dans le guide de l'utilisateur Amazon CloudWatch Logs.

Autorisations minimales pour un administrateur de service

La IAM politique suivante spécifie les autorisations minimales requises pour une IAM identité (utilisateur, groupe ou rôle) afin de configurer et de gérer le AWS PCS service.

> **Note**
>
> Les utilisateurs qui ne configurent ni ne gèrent le service n'ont pas besoin de ces autorisations. Les utilisateurs qui exécutent uniquement des tâches utilisent Secure Shell (SSH) pour se connecter au cluster. AWS Identity and Access Management (IAM) ne gère pas l'authentification ou l'autorisation pourSSH.

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

Vous pouvez exclure les autorisations suivantes de la politique et utiliser à la place la stratégie gérée correspondante dans IAM :

- `"firehose:*"`

  ```
  AmazonKinesisFirehoseFullAccess
  ```

- `"kms:*"`

  ```
  AWSKeyManagementServicePowerUser
  ```

- `"logs:*"`

```
CloudWatchLogsFullAccess
```

- `"s3:*"`

```
AmazonS3FullAccess
```

IAMprofils d'instance pour AWS Parallel Computing Service

Les applications qui s'exécutent sur une EC2 instance doivent inclure des AWS informations d'identification dans toutes les AWS API demandes qu'elles effectuent. Nous vous recommandons d'utiliser un IAM rôle pour gérer les informations d'identification temporaires sur l'EC2instance. Pour ce faire, vous pouvez définir un profil d'instance et l'associer à vos instances. Pour plus d'informations, consultez les IAMrôles d'Amazon EC2 dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.

> **Note**
>
> Lorsque vous utilisez le AWS Management Console pour créer un IAM rôle pour AmazonEC2, la console crée automatiquement un profil d'instance et lui donne le même nom que le IAM rôle. Si vous utilisez AWS CLI les AWS API actions ou an AWS SDK pour créer le IAM rôle, vous créez le profil d'instance en tant qu'action distincte. Pour plus d'informations, consultez la section Profils d'instance dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.

Vous devez spécifier le profil ARN d'une instance lorsque vous créez un groupe de nœuds de calcul. Vous pouvez choisir différents profils d'instance pour certains ou pour tous les groupes de nœuds de calcul.

Exigences relatives au profil d'instance

Nom du profil d'instance

Le profil d'IAMinstance ARN doit commencer par `AWSPCS` ou contenir `/aws-pcs/` dans son chemin.

Example

- `arn:aws:iam::*:instance-profile/AWSPCS-example-role-1` et

- `arn:aws:iam::*:instance-profile/aws-pcs/example-role-2`.

Autorisations

Le profil d'instance pour AWS PCS doit au minimum inclure la politique suivante. Il permet aux nœuds de calcul d'avertir le AWS PCS service lorsqu'ils deviennent opérationnels.

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

Politiques supplémentaires

Vous pouvez envisager d'ajouter des politiques gérées au profil d'instance. Par exemple :

- AmazonS3 ReadOnlyAccess fournit un accès en lecture seule à tous les compartiments S3.
- A mazonSSMManaged InstanceCore active AWS les fonctionnalités de base du service Systems Manager, telles que l'accès à distance directement depuis Amazon Management Console.
- CloudWatchAgentServerPolicycontient les autorisations requises pour une utilisation AmazonCloudWatchAgent sur les serveurs.

Vous pouvez également inclure vos propres IAM politiques adaptées à votre cas d'utilisation spécifique.

Création d'un profil d'instance

Vous pouvez créer un profil d'instance directement depuis la EC2 console Amazon. Pour plus d'informations, consultez la section Utilisation des profils d'instance dans le Guide de AWS Identity and Access Management l'utilisateur.

Résolution des problèmes d'identité et d'accès au service de calcul AWS parallèle

Utilisez les informations suivantes pour vous aider à diagnostiquer et à résoudre les problèmes courants que vous pouvez rencontrer lorsque vous travaillez avec AWS PCS etIAM.

Rubriques

Je ne suis pas autorisé à effectuer une action dans AWS PCS

Si vous recevez une erreur qui indique que vous n'êtes pas autorisé à effectuer une action, vos politiques doivent être mises à jour afin de vous permettre d'effectuer l'action.

L'exemple d'erreur suivant se produit lorsque l'`mateojackson`IAMutilisateur essaie d'utiliser la console pour afficher les détails d'une *`my-example-widget`* ressource fictive mais ne dispose pas des `pcs:`*`GetWidget`* autorisations fictives.

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 pcs:GetWidget on resource: my-example-widget
```

Dans ce cas, la politique qui s'applique à l'utilisateur `mateojackson` doit être mise à jour pour autoriser l'accès à la ressource *`my-example-widget`* à l'aide de l'action `pcs:`*`GetWidget`*.

Si vous avez besoin d'aide, contactez votre AWS administrateur. Votre administrateur vous a fourni vos informations d'identification de connexion.

Je ne suis pas autorisé à effectuer iam : PassRole

Si vous recevez un message d'erreur indiquant que vous n'êtes pas autorisé à effectuer l'`iam:PassRole`action, vos politiques doivent être mises à jour pour vous permettre de transmettre un rôle AWS PCS.

Certains vous Services AWS permettent de transmettre un rôle existant à ce service au lieu de créer un nouveau rôle de service ou un rôle lié à un service. Pour ce faire, un utilisateur doit disposer des autorisations nécessaires pour transmettre le rôle au service.

L'exemple d'erreur suivant se produit lorsqu'un IAM utilisateur nommé `marymajor` essaie d'utiliser la console pour effectuer une action dans AWS PCS. Toutefois, l'action nécessite que le service ait des autorisations accordées par un rôle de service. Mary ne dispose pas des autorisations nécessaires pour transférer le rôle au service.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

Dans ce cas, les politiques de Mary doivent être mises à jour pour lui permettre d'exécuter l'action `iam:PassRole`.

Si vous avez besoin d'aide, contactez votre AWS administrateur. Votre administrateur vous a fourni vos informations d'identification de connexion.

Je souhaite permettre à des personnes extérieures Compte AWS à moi d'accéder à mes AWS PCS ressources

Vous pouvez créer un rôle que les utilisateurs provenant d'autres comptes ou les personnes extérieures à votre organisation pourront utiliser pour accéder à vos ressources. Vous pouvez spécifier qui est autorisé à assumer le rôle. Pour les services qui prennent en charge les politiques basées sur les ressources ou les listes de contrôle d'accès (ACLs), vous pouvez utiliser ces politiques pour autoriser les utilisateurs à accéder à vos ressources.

Pour en savoir plus, consultez les éléments suivants :

- Pour savoir si ces fonctionnalités sont prises AWS PCS en charge, consultezComment fonctionne AWS Parallel Computing Service avec IAM.
- Pour savoir comment donner accès à vos ressources sur un site Comptes AWS qui vous appartient, consultez la section Fournir l'accès à un IAM utilisateur dans un autre site Compte AWS que vous possédez dans le Guide de IAM l'utilisateur.
- Pour savoir comment fournir l'accès à vos ressources à des tiers Comptes AWS, consultez la section Fournir un accès à des ressources Comptes AWS détenues par des tiers dans le Guide de IAM l'utilisateur.
- Pour savoir comment fournir un accès via la fédération d'identité, consultez la section Fournir un accès aux utilisateurs authentifiés de manière externe (fédération d'identité) dans le guide de l'IAMutilisateur.

- Pour connaître la différence entre l'utilisation de rôles et l'utilisation de politiques basées sur les ressources pour l'accès entre comptes, voir Accès aux ressources entre comptes IAM dans le guide de l'IAMutilisateur.

Validation de conformité pour le service de calcul AWS parallèle

Pour savoir si un programme Services AWS de conformité Service AWS s'inscrit dans le champ d'application de programmes de conformité spécifiques, consultez Services AWS la section de conformité et sélectionnez le programme de conformité qui vous intéresse. Pour des informations générales, voir Programmes de AWS conformité Programmes AWS de .

Vous pouvez télécharger des rapports d'audit tiers à l'aide de AWS Artifact. Pour plus d'informations, voir Téléchargement de rapports dans AWS Artifact .

Votre responsabilité en matière de conformité lors de l'utilisation Services AWS est déterminée par la sensibilité de vos données, les objectifs de conformité de votre entreprise et les lois et réglementations applicables. AWS fournit les ressources suivantes pour faciliter la mise en conformité :

- Guides de démarrage rapide sur la sécurité et la conformité : ces guides de déploiement abordent les considérations architecturales et indiquent les étapes à suivre pour déployer des environnements de base axés sur AWS la sécurité et la conformité.
- Architecture axée sur la HIPAA sécurité et la conformité sur Amazon Web Services : ce livre blanc décrit comment les entreprises peuvent AWS créer HIPAA des applications éligibles.

> **Note**
>
> Tous ne Services AWS sont pas HIPAA éligibles. Pour plus d'informations, consultez la référence des services HIPAA éligibles.

- AWS Ressources de https://aws.amazon.com/compliance/resources/ de conformité — Cette collection de classeurs et de guides peut s'appliquer à votre secteur d'activité et à votre région.
- AWS Guides de conformité destinés aux clients — Comprenez le modèle de responsabilité partagée sous l'angle de la conformité. Les guides résument les meilleures pratiques en matière de sécurisation Services AWS et reprennent les directives relatives aux contrôles de sécurité dans de nombreux cadres (notamment le National Institute of Standards and Technology (NIST),

le Payment Card Industry Security Standards Council (PCI) et l'Organisation internationale de normalisation (ISO)).

- Évaluation des ressources à l'aide des règles du guide du AWS Config développeur : le AWS Config service évalue dans quelle mesure les configurations de vos ressources sont conformes aux pratiques internes, aux directives du secteur et aux réglementations.
- AWS Security Hub— Cela Service AWS fournit une vue complète de votre état de sécurité interne AWS. Security Hub utilise des contrôles de sécurité pour évaluer vos ressources AWS et vérifier votre conformité par rapport aux normes et aux bonnes pratiques du secteur de la sécurité. Pour obtenir la liste des services et des contrôles pris en charge, consultez Référence des contrôles Security Hub.
- Amazon GuardDuty — Cela Service AWS détecte les menaces potentielles qui pèsent sur vos charges de travail Comptes AWS, vos conteneurs et vos données en surveillant votre environnement pour détecter toute activité suspecte et malveillante. GuardDuty peut vous aider à répondre à diverses exigences de conformité PCIDSS, par exemple en répondant aux exigences de détection des intrusions imposées par certains cadres de conformité.
- AWS Audit Manager— Cela vous Service AWS permet d'auditer en permanence votre AWS utilisation afin de simplifier la gestion des risques et la conformité aux réglementations et aux normes du secteur.

Résilience dans les services de calcul AWS parallèle

L'infrastructure AWS mondiale est construite autour Régions AWS de zones de disponibilité. Régions AWS fournissent plusieurs zones de disponibilité physiquement séparées et isolées, connectées par un réseau à faible latence, à haut débit et hautement redondant. Avec les zones de disponibilité, vous pouvez concevoir et exploiter des applications et des bases de données qui basculent automatiquement d'une zone à l'autre sans interruption. Les zones de disponibilité sont davantage disponibles, tolérantes aux pannes et ont une plus grande capacité de mise à l'échelle que les infrastructures traditionnelles à un ou plusieurs centres de données.

Pour plus d'informations sur les zones de disponibilité Régions AWS et les zones de disponibilité, consultez la section Infrastructure AWS globale.

Sécurité de l'infrastructure dans un service de calcul AWS parallèle

En tant que service géré, AWS Parallel Computing Service est protégé par la sécurité du réseau AWS mondial. Pour plus d'informations sur les services AWS de sécurité et sur la manière dont

AWS l'infrastructure est protégée, consultez la section Sécurité du AWS cloud. Pour concevoir votre AWS environnement en utilisant les meilleures pratiques en matière de sécurité de l'infrastructure, consultez la section Protection de l'infrastructure dans le cadre AWS bien architecturé du pilier de sécurité.

Vous utilisez API les appels AWS publiés pour accéder AWS PCS via le réseau. Les clients doivent prendre en charge les éléments suivants :

- Sécurité de la couche de transport (TLS). Nous avons besoin de la TLS version 1.2 et recommandons la TLS version 1.3.
- Des suites de chiffrement parfaitement confidentielles (PFS) telles que (Ephemeral Diffie-Hellman) ou DHE ECDHE (Elliptic Curve Ephemeral Diffie-Hellman). La plupart des systèmes modernes tels que Java 7 et les versions ultérieures prennent en charge ces modes.

En outre, les demandes doivent être signées à l'aide d'un identifiant de clé d'accès et d'une clé d'accès secrète associés à un IAM principal. Vous pouvez également utiliser AWS Security Token Service (AWS STS) pour générer des informations d'identification de sécurité temporaires et signer les demandes.

Lors de AWS PCS la création d'un cluster, le service lance le contrôleur Slurm dans un compte appartenant au service, distinct des nœuds de calcul de votre compte. Pour établir une passerelle entre le contrôleur et les nœuds de calcul, AWS PCS créez une interface réseau élastique (ENI) entre comptes dans votreVPC. Le contrôleur Slurm utilise le ENI pour gérer et communiquer avec les nœuds de calcul à travers différents nœuds Comptes AWS, en maintenant la sécurité et l'isolation des ressources tout en facilitant l'efficacité des opérations d'intelligence artificielle HPC et d'apprentissage automatique.

Analyse et gestion des vulnérabilités dans le service de calcul AWS parallèle

La configuration et les contrôles informatiques sont une responsabilité partagée entre vous AWS et vous. Pour plus d'informations, consultez le modèle de responsabilitéAWS partagée. AWS gère les tâches de sécurité de base pour l'infrastructure sous-jacente du compte de service, telles que l'application de correctifs au système d'exploitation sur les instances de contrôleur, la configuration du pare-feu et la reprise après sinistre de AWS l'infrastructure. Ces procédures ont été vérifiées et certifiées par les tiers appropriés. Pour plus de détails, consultez les meilleures pratiques en matière de sécurité, d'identité et de conformité.

Vous êtes responsable de la sécurité de l'infrastructure sous-jacente dans votre Compte AWS :

- Maintenez votre code, y compris les mises à jour et les correctifs de sécurité.
- Corrigez et mettez à jour le système d'exploitation sur les instances de groupes de nœuds.
- Mettez à jour le planificateur pour qu'il reste dans les versions prises en charge.
- Authentifiez et chiffrez les communications entre les clients utilisateurs et les nœuds auxquels ils se connectent.

Prévention du cas de figure de l'adjoint désorienté entre services

Le problème de député confus est un problème de sécurité dans lequel une entité qui n'est pas autorisée à effectuer une action peut contraindre une entité plus privilégiée à le faire. En AWS, l'usurpation d'identité interservices peut entraîner la confusion des adjoints. L'usurpation d'identité entre services peut se produire lorsqu'un service (le service appelant) appelle un autre service (le service appelé). Le service appelant peut être manipulé et ses autorisations utilisées pour agir sur les ressources d'un autre client auxquelles on ne serait pas autorisé d'accéder autrement. Pour éviter cela, AWS fournit des outils qui vous aident à protéger vos données pour tous les services avec des principaux de service qui ont eu accès aux ressources de votre compte.

Nous recommandons d'utiliser les clés contextuelles de condition `aws:SourceAccount`globale `aws:SourceArn`et les clés contextuelles dans les politiques de ressources afin de limiter les autorisations que AWS Parallel Computing Service (AWS PCS) accorde à un autre service à la ressource. Utilisez `aws:SourceArn` si vous souhaitez qu'une seule ressource soit associée à l'accès entre services. Utilisez `aws:SourceAccount` si vous souhaitez autoriser l'association d'une ressource de ce compte à l'utilisation interservices.

Le moyen le plus efficace de se protéger contre le problème de confusion des adjoints consiste à utiliser la clé de contexte de la condition `aws:SourceArn` globale avec l'intégralité ARN de la ressource. Si vous ne connaissez pas l'intégralité ARN de la ressource ou si vous spécifiez plusieurs ressources, utilisez la clé de condition contextuelle `aws:SourceArn` globale avec des caractères génériques (*) pour les parties inconnues duARN. Par exemple, `arn:aws:`*`servicename`*`:*:`*`123456789012`*`:*`.

Si la `aws:SourceArn` valeur ne contient pas l'ID de compte, tel qu'un compartiment Amazon S3ARN, vous devez utiliser les deux clés de contexte de condition globale pour limiter les autorisations.

La valeur de `aws:SourceArn` doit être un clusterARN.

L'exemple suivant montre comment utiliser les touches de contexte de condition `aws:SourceAccount` globale `aws:SourceArn` et globale AWS PCS pour éviter le problème de confusion des adjoints.

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

IAMrôle pour les EC2 instances Amazon mises en service dans le cadre d'un groupe de nœuds de calcul

AWS PCSorchestre automatiquement la EC2 capacité Amazon pour chacun des groupes de nœuds de calcul configurés dans un cluster. Lors de la création d'un groupe de nœuds de calcul, les utilisateurs doivent fournir un profil d'IAMinstance via le `iamInstanceProfileArn` champ. Le profil d'instance spécifie les autorisations associées aux EC2 instances provisionnées. AWS PCSaccepte tout rôle ayant `AWSPCS` comme préfixe de nom de `/aws-pcs/` rôle ou faisant partie du chemin du rôle. L'`iam:PassRole`autorisation est requise pour l'IAMidentité (utilisateur ou rôle) qui crée ou met à jour un groupe de nœuds de calcul. Lorsqu'un utilisateur lance l'`UpdateComputeNodeGroup`APIaction `CreateComputeNodeGroup` ou, AWS PCS vérifie s'il est autorisé à effectuer l'`iam:PassRole`action.

L'exemple de politique suivant accorde des autorisations pour transmettre uniquement IAM les rôles dont le nom commence par`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"
                    ]
                }
            }
        }
    ]
}
```

Bonnes pratiques de sécurité pour le service de calcul AWS parallèle

Cette section décrit les meilleures pratiques de sécurité spécifiques à AWS Parallel Computing Service (AWS PCS). Pour en savoir plus sur les meilleures pratiques en matière de sécurité dans AWS, consultez la section Meilleures pratiques en matière de sécurité, d'identité et de conformité.

AMIsécurité associée

- N'utilisez pas d' AWS PCSéchantillon AMIs pour les charges de travail de production. Les échantillons AMIs ne sont pas pris en charge et sont uniquement destinés à être testés.
- Mettez régulièrement à jour le système d'exploitation et le logiciel des AWS PCS instances pour atténuer les vulnérabilités.
- Utilisez-le AWS Systems Manager pour automatiser l'application des correctifs et garantir la conformité avec vos politiques de sécurité.
- N'utilisez que des AWS PCS packages officiels authentifiés téléchargés à partir de AWS sources officielles.

- Mettez régulièrement à jour les AWS PCS packages sur les nœuds de calcul pour recevoir des correctifs de sécurité et des améliorations. Envisagez d'automatiser ce processus afin de minimiser les vulnérabilités.

Sécurité de Slurm Workload Manager

- Mettez en œuvre des contrôles d'accès et des restrictions réseau pour sécuriser les nœuds de contrôle et de calcul de Slurm. Autorisez uniquement les utilisateurs et les systèmes fiables à soumettre des tâches et à accéder aux commandes de gestion de Slurm.
- Utilisez les fonctionnalités de sécurité intégrées de Slurm, telles que l'authentification Slurm, pour vous assurer que les soumissions de tâches et les communications sont authentifiées.
- Mettez à jour les versions de Slurm pour garantir le bon fonctionnement des opérations et la prise en charge des clusters.

> **Important**
>
> Tout cluster qui utilise une version de Slurm ayant atteint la fin de la durée de vie du support (EOSL) est immédiatement arrêté. Utilisez le lien en haut des pages du guide de l'utilisateur pour vous abonner au RSS flux de AWS PCS documentation afin de recevoir une notification lorsqu'une version de Slurm approche. EOSL

Surveillance et journalisation

- Utilisez Amazon CloudWatch Logs et AWS CloudTrail pour surveiller et enregistrer les actions dans vos clusters et Compte AWS. Utilisez les données pour le dépannage et l'audit.

Sécurité du réseau

- Déployez vos AWS PCS clusters séparément VPC pour isoler votre HPC environnement du reste du trafic réseau.
- Utilisez des groupes de sécurité et des listes de contrôle d'accès réseau (ACLs) pour contrôler le trafic entrant et sortant vers les AWS PCS instances et les sous-réseaux.
- Utilisez AWS PrivateLink nos VPC points de terminaison pour maintenir le trafic réseau entre vos clusters et les autres AWS services du AWS réseau.

Enregistrement et surveillance pour AWS PCS

La surveillance joue un rôle important dans le maintien de la fiabilité, de la disponibilité AWS PCS et des performances de vos autres AWS ressources. AWSfournit les outils de surveillance suivants pour surveiller AWS PCS, signaler tout problème et prendre des mesures automatiques le cas échéant :

- Amazon CloudWatch surveille vos AWS ressources et les applications que vous utilisez AWS en temps réel. Vous pouvez collecter et suivre les métriques, créer des tableaux de bord personnalisés, et définir des alarmes qui vous informent ou prennent des mesures lorsqu'une métrique spécifique atteint un seuil que vous spécifiez. Par exemple, vous pouvez CloudWatch suivre CPU l'utilisation ou d'autres indicateurs de vos EC2 instances Amazon et lancer automatiquement de nouvelles instances en cas de besoin. Pour plus d'informations, consultez le guide de CloudWatch l'utilisateur Amazon.
- Amazon CloudWatch Logs vous permet de surveiller, de stocker et d'accéder à vos fichiers journaux à partir d'EC2instances Amazon et d'autres sources. CloudTrail CloudWatch Les journaux peuvent surveiller les informations contenues dans les fichiers journaux et vous avertir lorsque certains seuils sont atteints. Vous pouvez également archiver vos données de journaux dans une solution de stockage hautement durable. Pour plus d'informations, consultez le guide de l'utilisateur d'Amazon CloudWatch Logs.
- AWS CloudTrailcapture API les appels et les événements connexes effectués par ou pour le compte de votre AWS compte et envoie les fichiers journaux dans un compartiment Amazon S3 que vous spécifiez. Vous pouvez identifier les utilisateurs et les comptes appelés AWS, l'adresse IP source à partir de laquelle les appels ont été effectués et la date des appels. Pour plus d'informations, consultez le AWS CloudTrail Guide de l'utilisateur .

AWS PCSjournaux du planificateur

Vous pouvez configurer AWS PCS pour envoyer des données de journalisation détaillées depuis votre planificateur de cluster à Amazon CloudWatch Logs, Amazon Simple Storage Service (Amazon S3) et Amazon Data Firehose. Cela peut faciliter la surveillance et le dépannage. Vous pouvez configurer les journaux du AWS PCS planificateur à l'aide de la AWS PCS console ou par programmation à l'aide du ou. AWS CLI SDK

Table des matières

Prérequis

Le IAM principal utilisé pour gérer le AWS PCS cluster doit l'autoriser`pcs:AllowVendedLogDeliveryForResource`. Voici un exemple de AWS IAM politique qui l'active.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PcsAllowVendedLogsDelivery",
            "Effect": "Allow",
            "Action": ["pcs:AllowVendedLogDeliveryForResource"],
            "Resource": [
                "arn:aws:pcs:::cluster/*"
            ]
        }
    ]
}
```

Configuration des journaux du planificateur à l'aide de la console AWS PCS

Pour configurer les journaux du AWS PCS planificateur dans la console, procédez comme suit :

1. Ouvrez la AWS PCSconsole.
2. Choisissez Clusters et accédez à la page détaillée du AWS PCS cluster où vous allez activer la journalisation.
3. Choisissez Logs (Journaux).
4. Dans le cadre des livraisons de journaux — Scheduler Logs — facultatif

a. Ajoutez jusqu'à trois destinations de livraison de journaux. Les choix incluent CloudWatch Logs, Amazon S3 ou Firehose.

b. Choisissez Mettre à jour les livraisons de journaux.

Vous pouvez reconfigurer, ajouter ou supprimer des livraisons de journaux en revisitant cette page.

Configuration des journaux du planificateur à l'aide du AWS CLI

Pour ce faire, vous avez besoin d'au moins une destination de livraison, une source de livraison (le PCS cluster) et une livraison, qui est une relation qui connecte une source à une destination.

Création d'une destination de livraison

Vous avez besoin d'au moins une destination de livraison pour recevoir les journaux du planificateur d'un AWS PCS cluster. Vous pouvez en savoir plus sur ce sujet dans la PutDeliveryDestination section du guide de l' CloudWatch APIutilisateur.

Pour créer une destination de livraison à l'aide du AWS CLI

- Créez une destination à l'aide de la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :
  - Remplacez *region-code* avec l' Région AWS endroit où vous allez créer votre destination. Il s'agira généralement de la même région que celle où le AWS PCS cluster est déployé.
  - Remplacez *pcs-logs-destination* avec votre nom préféré. Il doit être unique pour toutes les destinations de livraison de votre compte.
  - Remplacez *resource-arn* avec le ARN pour un groupe de CloudWatch journaux existant dans Logs, un compartiment S3 ou un flux de diffusion dans Firehose. En voici quelques exemples :
    - CloudWatch Groupe de journaux

      ```
      arn:aws:logs:region-code:account-id:log-group:/log-group-name:*
      ```

    - Compartiment S3

      ```
      arn:aws:s3:::bucket-name
      ```

    - Flux de livraison 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
```

Prenez note de la ARN nouvelle destination de livraison, car vous en aurez besoin pour configurer les livraisons.

Activer le AWS PCS cluster en tant que source de diffusion

Pour collecter les journaux du planificateur AWSPCS, configurez le cluster comme source de diffusion. Pour plus d'informations, consultez PutDeliverySourcele manuel Amazon CloudWatch Logs API Reference.

Pour configurer un cluster en tant que source de livraison à l'aide du AWS CLI

- Activez la livraison des journaux depuis votre cluster à l'aide de la commande suivante. Avant d'exécuter la commande, effectuez les remplacements suivants :
  - Remplacez *region-code* avec l' Région AWS endroit où votre cluster est déployé.
  - Remplacez *cluster-logs-source-name* avec un nom pour cette source. Il doit être unique pour toutes les sources de livraison de votre Compte AWS. Envisagez d'incorporer le nom ou l'ID du AWS PCS cluster.
  - Remplacez *cluster-arn* avec le ARN pour votre AWS PCS cluster

```
aws logs put-delivery-source \
    --region region-code \
    --name cluster-logs-source-name \
    --resource-arn cluster-arn \
    --log-type PCS_SCHEDULER_LOGS
```

Connectez la source de diffusion du cluster à la destination de livraison

Pour que les données du journal du planificateur circulent du cluster vers la destination, vous devez configurer une diffusion qui les connecte. Pour plus d'informations, consultez CreateDeliveryle manuel Amazon CloudWatch Logs API Reference.

Pour créer une livraison à l'aide du AWS CLI

- Créez une livraison à l'aide de la commande ci-dessous. Avant d'exécuter la commande, effectuez les remplacements suivants :

- Remplacez *region-code* avec l' Région AWS endroit où se trouvent votre source et votre destination.
- Remplacez *cluster-logs-source-name* avec le nom de votre source de livraison indiqué ci-dessus.
- Remplacez *destination-arn* avec le ARN depuis une destination de livraison où vous souhaitez que les journaux soient livrés.

```
aws logs create-delivery \
    --region region-code \
    --delivery-source-name cluster-logs-source \
    --delivery-destination-arn destination-arn
```

Le planificateur enregistre les chemins et les noms des flux

Le chemin et le nom des journaux du AWS PCS planificateur dépendent du type de destination.

- CloudWatch Journaux
  - Un flux CloudWatch Logs suit cette convention de dénomination.

    ```
    AWSLogs/PCS/${cluster_id}/${log_name}_${scheduler_major_version}.log
    ```

    Example

    ```
    AWSLogs/PCS/abcdef0123/slurmctld_24.05.log
    ```
- Compartiment S3
  - Le chemin de sortie d'un compartiment S3 suit cette convention de dénomination :

    ```
    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.
    ```
  - Le nom d'un objet S3 suit cette convention :

```
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
```

Exemple d'enregistrement du AWS PCS journal du planificateur

AWSPCSles journaux du planificateur sont structurés. Ils incluent des champs tels que l'identifiant du cluster, le type de planificateur, les versions majeures et de correctif, en plus du message de journal émis par le processus du contrôleur Slurm. Voici un exemple.

```
{
    "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"
}
```

Surveillance d'un service de calcul AWS parallèle avec Amazon CloudWatch

Amazon CloudWatch assure la surveillance de l'état et des performances de votre cluster AWS Parallel Computing Service (AWS PCS) en collectant des métriques à partir du cluster à intervalles réguliers. Ces indicateurs sont conservés, ce qui vous permet d'accéder aux données historiques et de mieux comprendre les performances de votre cluster au fil du temps.

CloudWatch vous permet également de surveiller les EC2 instances lancées par AWS PCS afin de répondre à vos exigences de dimensionnement. Bien que vous puissiez inspecter les journaux des instances en cours d'exécution, CloudWatch les métriques et les données de journalisation

sont généralement supprimées une fois les instances fermées. Toutefois, vous pouvez configurer l' CloudWatch agent sur les instances à l'aide d'un modèle de EC2 lancement afin de conserver les métriques et les journaux même après la fermeture de l'instance, ce qui permet une surveillance et une analyse à long terme.

Explorez les rubriques de cette section pour en savoir plus sur la surveillance à AWS PCS l'aide de CloudWatch.

Rubriques

- Surveillance des AWS PCS métriques à l'aide CloudWatch
- Surveillance des AWS PCS instances à l'aide d'Amazon CloudWatch

Surveillance des AWS PCS métriques à l'aide CloudWatch

Vous pouvez surveiller l'état AWS PCS du cluster à l'aide d'Amazon CloudWatch, qui collecte les données de votre cluster et les transforme en métriques en temps quasi réel. Ces statistiques sont conservées pendant une période de 15 mois, afin que vous puissiez accéder aux informations historiques et avoir une meilleure idée des performances de votre cluster. Les métriques du cluster sont envoyées à des CloudWatch intervalles d'une minute. Pour plus d'informations CloudWatch, consultez Qu'est-ce qu'Amazon CloudWatch ? dans le guide de CloudWatch l'utilisateur Amazon.

AWS PCSpublie les métriques suivantes dans l'espace de PCS noms AWS/de CloudWatch. Ils ont une seule dimension,`ClusterId`.

Name (Nom)	Description	Unités
ActualCapacity	IdleCapacity + UtilizedCapacity	Nombre
CapacityUtilization	UtilizedCapacity / ActualCapacity	Nombre
DesiredCapacity	ActualCapacity + PendingCapacity	Nombre
IdleCapacity	Nombre d'instances en cours d'exécution mais non allouées aux tâches	Nombre

Name (Nom)	Description	Unités
UtilizedCapacity	Nombre d'instances en cours d'exécution et allouées aux tâches	Nombre

Surveillance des AWS PCS instances à l'aide d'Amazon CloudWatch

AWSPCSlance EC2 les instances Amazon selon les besoins pour répondre aux exigences de dimensionnement définies dans vos groupes de nœuds de PCS calcul. Vous pouvez surveiller ces instances lorsqu'elles sont en cours d'exécution à l'aide d'Amazon CloudWatch. Vous pouvez consulter les journaux des instances en cours d'exécution en vous y connectant et en utilisant des outils de ligne de commande interactifs. Toutefois, par défaut, CloudWatch les données des métriques ne sont conservées que pendant une période limitée une fois qu'une instance est résiliée, et les journaux d'instance sont généralement supprimés en même temps que les EBS volumes qui soutiennent l'instance. Pour conserver les métriques ou les données de journalisation des instances lancées PCS après leur résiliation, vous pouvez configurer l' CloudWatch agent sur vos instances à l'aide d'un modèle de EC2 lancement. Cette rubrique fournit une vue d'ensemble de la surveillance des instances en cours d'exécution et fournit des exemples de configuration des métriques et des journaux d'instance persistants.

Surveillance des instances en cours d'exécution

Trouver des AWS PCS instances

Pour surveiller les instances lancées parPCS, recherchez les instances en cours d'exécution associées à un cluster ou à un groupe de nœuds de calcul. Ensuite, dans la EC2 console d'une instance donnée, inspectez les sections État, alarmes et surveillance. Si l'accès de connexion est configuré pour ces instances, vous pouvez vous y connecter et inspecter les différents fichiers journaux des instances. Pour plus d'informations sur l'identification des instances géréesPCS, consultezRecherche d'instances de groupes de nœuds de calcul dans AWS PCS.

Permettre des métriques détaillées

Par défaut, les métriques d'instance sont collectées à intervalles de 5 minutes. Pour collecter des métriques à intervalles d'une minute, activez la CloudWatch surveillance détaillée dans votre modèle de lancement de groupe de nœuds de calcul. Pour de plus amples informations, veuillez consulter Activez la CloudWatch surveillance détaillée.

Configuration des métriques et des journaux d'instance persistants

Vous pouvez conserver les métriques et les journaux de vos instances en installant et en configurant l' CloudWatch agent Amazon sur celles-ci. Cela comprend trois étapes principales :

1. Créez une configuration d' CloudWatch agent.
2. Stockez la configuration dans un endroit où elle peut être récupérée par PCS les instances.
3. Rédigez un modèle de EC2 lancement qui installe le logiciel de l' CloudWatch agent, récupère votre configuration et démarre l' CloudWatch agent à l'aide de cette configuration.

Pour plus d'informations, consultez Collecter des métriques, des journaux et des traces avec l' CloudWatch agent dans le guide de CloudWatch l'utilisateur Amazon, etUtilisation des modèles EC2 de lancement Amazon avec AWS PCS.

Création d'une configuration CloudWatch d'agent

Avant de déployer l' CloudWatch agent sur vos instances, vous devez générer un fichier de JSON configuration qui spécifie les métriques, les journaux et les traces à collecter. Les fichiers de configuration peuvent être créés à l'aide d'un assistant ou manuellement à l'aide d'un éditeur de texte. Le fichier de configuration sera créé manuellement pour cette démonstration.

Sur un ordinateur sur lequel vous l'avez AWS CLI installé, créez un fichier CloudWatch de configuration nommé config.json avec le contenu ci-dessous. Vous pouvez également utiliser ce qui suit URL pour télécharger une copie du fichier.

```
https://aws-hpc-recipes.s3.amazonaws.com/main/recipes/pcs/cloudwatch/assets/config.json
```

Remarques

- Les chemins de journal figurant dans le fichier d'exemple concernent Amazon Linux 2. Si vos instances doivent utiliser un système d'exploitation de base différent, modifiez les chemins comme il convient.
- Pour capturer d'autres journaux, ajoutez des entrées supplémentaires sous`collect_list`.
- Les valeurs saisies `{brackets}` sont des variables modélisées. Pour obtenir la liste complète des variables prises en charge, voir Création ou modification manuelle du fichier de configuration de l' CloudWatch agent dans le guide de CloudWatch l'utilisateur Amazon.
- Vous pouvez choisir d'omettre `logs` ou `metrics` de ne pas collecter ces types d'informations.

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

Ce fichier demande à l' CloudWatch agent de surveiller plusieurs fichiers qui peuvent être utiles pour diagnostiquer les erreurs liées au démarrage des instances, à l'authentification et à la connexion, ainsi que dans d'autres domaines de résolution des problèmes. Il s'agit des licences suivantes :

- `/var/log/cloud-init.log`— Résultat de la phase initiale de configuration de l'instance
- `/var/log/cloud-init-output.log`— Résultat des commandes exécutées lors de la configuration de l'instance
- `/var/log/amazon/pcs/bootstrap.log`— Résultat d'opérations PCS spécifiques exécutées lors de la configuration de l'instance
- `/var/log/slurmd.log`— Résultat du daemon slurmd du gestionnaire de charge de travail Slurm
- `/var/log/messages`— Messages système provenant du noyau, des services système et des applications

- `/var/log/secure`— Journaux liés aux tentatives d'authentificationSSH, tels que sudo et autres événements de sécurité

Les fichiers journaux sont envoyés à un groupe de CloudWatch journaux nommé`/PCSLogs/instances`. Les flux de journaux sont une combinaison de l'ID d'instance et du nom de base du fichier journal. Le groupe de journaux a une durée de conservation de 30 jours.

En outre, le fichier demande à l' CloudWatch agent de collecter plusieurs métriques communes, en les agrégeant par ID d'instance.

Stocker la configuration

Le fichier de configuration de l' CloudWatch agent doit être stocké de manière à être accessible aux instances du nœud de PCS calcul. Il existe deux méthodes courantes pour ce faire. Vous pouvez le télécharger dans un compartiment Amazon S3 auquel les instances de votre groupe de nœuds de calcul auront accès via leur profil d'instance. Vous pouvez également le stocker en tant que SSM paramètre dans Amazon Systems Manager Parameter Store.

Téléchargement vers un compartiment S3

Pour stocker votre fichier dans S3, utilisez les AWS CLI commandes ci-dessous. Avant d'exécuter la commande, effectuez les remplacements suivants :

- Remplacez *DOC-EXAMPLE-BUCKET* avec votre propre nom de compartiment S3

Tout d'abord (cela est facultatif si vous avez un bucket existant), créez un bucket pour contenir vos fichiers de configuration.

```
aws s3 mb s3://DOC-EXAMPLE-BUCKET
```

Ensuite, chargez le fichier dans le compartiment.

```
aws s3 cp ./config.json s3://DOC-EXAMPLE-BUCKET/
```

Stocker en tant que SSM paramètre

Pour enregistrer votre fichier en tant que SSM paramètre, utilisez la commande ci-dessous. Avant d'exécuter la commande, effectuez les remplacements suivants :

- Remplacez *region-code* avec la AWS région avec laquelle vous travaillez AWSPCS.

- (Facultatif) Remplacer *AmazonCloudWatch-PCS* avec votre propre nom pour le paramètre. Notez que si vous modifiez le préfixe du nom de, `AmazonCloudWatch-` vous devrez spécifiquement ajouter un accès en lecture au SSM paramètre dans le profil d'instance de votre groupe de nœuds.

```
aws ssm put-parameter \
   --region region-code \
   --name "AmazonCloudWatch-PCS" \
   --type String \
   --value file://config.json
```

Rédiger un modèle de EC2 lancement

Les détails spécifiques du modèle de lancement varient selon que votre fichier de configuration est stocké dans S3 ouSSM.

Utiliser une configuration stockée dans S3

Ce script installe CloudWatch l'agent, importe un fichier de configuration depuis un compartiment S3 et lance l' CloudWatch agent avec celui-ci. Remplacez les valeurs suivantes dans ce script par vos propres informations :

- *DOC-EXAMPLE-BUCKET* — Le nom d'un compartiment S3 que votre compte peut lire
- */config.json* — Chemin relatif à la racine du compartiment S3 où la configuration est stockée

```
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==--
```

Le profil d'IAMinstance du groupe de nœuds doit avoir accès au bucket. Voici un exemple de IAM politique pour le bucket dans le script de données utilisateur ci-dessus.

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

Notez également que les instances doivent autoriser le trafic sortant vers le S3 et les points de CloudWatch terminaison. Cela peut être réalisé à l'aide de groupes de sécurité ou de VPC points de terminaison, en fonction de l'architecture de votre cluster.

Utiliser une configuration stockée dans SSM

Ce script installe CloudWatch l'agent, importe un fichier de configuration à partir d'un SSM paramètre et lance l' CloudWatch agent avec celui-ci. Remplacez les valeurs suivantes dans ce script par vos propres informations :

- (Facultatif) Remplacer *AmazonCloudWatch-PCS* avec votre propre nom pour le paramètre.

```
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 politique d'IAMinstance du groupe de nœuds doit être CloudWatchAgentServerPolicyassociée à celle-ci.

Si le nom de votre paramètre ne commence pas par, `AmazonCloudWatch-` vous devrez spécifiquement ajouter un accès en lecture au SSM paramètre dans le profil d'instance de votre groupe de nœuds. Voici un exemple de IAM politique qui illustre cela pour le préfixe *DOC-EXAMPLE-PREFIX*.

```
{
  "Version" : "2012-10-17",
  "Statement" : [
    {
      "Sid" : "CustomCwSsmMParamReadOnly",
      "Effect" : "Allow",
      "Action" : [
        "ssm:GetParameter"
      ],
      "Resource" : "arn:aws:ssm:*:*:parameter/DOC-EXAMPLE-PREFIX*"
    }
  ]
}
```

Notez également que les instances doivent autoriser le trafic sortant vers les points de CloudWatch terminaison SSM et. Cela peut être réalisé à l'aide de groupes de sécurité ou de VPC points de terminaison, en fonction de l'architecture de votre cluster.

Enregistrement des API appels du service de calcul AWS parallèle à l'aide de AWS CloudTrail

AWS PCSest intégré à AWS CloudTrail un service qui fournit un enregistrement des actions entreprises par un utilisateur, un rôle ou un AWS service dans AWS PCS. CloudTrail capture tous les API appels AWS PCS sous forme d'événements. Les appels capturés incluent des appels provenant de la AWS PCS console et des appels de code vers les AWS PCS API opérations. Si vous créez un suivi, vous pouvez activer la diffusion continue d' CloudTrail événements vers un

compartiment Amazon S3, y compris les événements pour AWS PCS. Si vous ne configurez pas de suivi, vous pouvez toujours consulter les événements les plus récents dans la CloudTrail console dans Historique des événements. À l'aide des informations collectées par CloudTrail, vous pouvez déterminer la demande qui a été faite AWS PCS, l'adresse IP à partir de laquelle la demande a été faite, qui a fait la demande, quand elle a été faite et des détails supplémentaires.

Pour en savoir plus CloudTrail, consultez le guide de AWS CloudTrail l'utilisateur.

AWS PCSinformations dans CloudTrail

CloudTrail est activé sur votre compte Compte AWS lorsque vous créez le compte. Lorsqu'une activité se produit dans AWS PCS, cette activité est enregistrée dans un CloudTrail événement avec d'autres événements de AWS service dans l'historique des événements. Vous pouvez consulter, rechercher et télécharger les événements récents dans votre Compte AWS. Pour plus d'informations, consultez la section Affichage des événements à l'aide de l'historique des CloudTrail événements.

Pour un enregistrement continu des événements de votre région Compte AWS, y compris des événements pour AWS PCS, créez un parcours. Un suivi permet CloudTrail de fournir des fichiers journaux à un compartiment Amazon S3. Par défaut, lorsque vous créez un journal d'activité dans la console, il s'applique à toutes les régions Régions AWS. Le journal enregistre les événements de toutes les régions de la AWS partition et transmet les fichiers journaux au compartiment Amazon S3 que vous spécifiez. En outre, vous pouvez configurer d'autres AWS services pour analyser plus en détail les données d'événements collectées dans les CloudTrail journaux et agir en conséquence. Pour plus d'informations, consultez les ressources suivantes :

- Présentation de la création d'un journal de suivi
- CloudTrail services et intégrations pris en charge
- Configuration des SNS notifications Amazon pour CloudTrail
- Réception de fichiers CloudTrail journaux de plusieurs régions et réception de fichiers CloudTrail journaux de plusieurs comptes

Toutes les AWS PCS actions sont enregistrées CloudTrail et documentées dans le AWS Parallel Computing Service API Reference. Par exemple, les appels aux `CreateComputeNodeGroupUpdateQueue`, et `DeleteCluster` les actions génèrent des entrées dans les fichiers CloudTrail journaux.

Chaque événement ou entrée de journal contient des informations sur la personne ayant initié la demande. Les informations relatives à l'identité permettent de déterminer les éléments suivants :

- Si la demande a été faite avec les informations d'identification de l'utilisateur root ou AWS Identity and Access Management (IAM).
- Si la demande a été effectuée avec les informations d'identification de sécurité temporaires d'un rôle ou d'un utilisateur fédéré.
- Si la demande a été faite par un autre AWS service.

Pour plus d'informations, consultez l'CloudTrail userIdentityélément.

Comprendre les entrées du fichier CloudTrail journal provenant de AWS PCS

Un suivi est une configuration qui permet de transmettre des événements sous forme de fichiers journaux à un compartiment S3 que vous spécifiez. CloudTrail les fichiers journaux contiennent une ou plusieurs entrées de journal. Un événement représente une demande unique provenant de n'importe quelle source et inclut des informations sur l'action demandée, la date et l'heure de l'action, les paramètres de la demande, etc. CloudTrail les fichiers journaux ne constituent pas une trace ordonnée des API appels publics, ils n'apparaissent donc pas dans un ordre spécifique.

L'exemple suivant montre une entrée de CloudTrail journal pour une `CreateQueue` action.


```
{
    "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"
}
```

Points de terminaison et quotas de service pour AWS PCS

Les sections suivantes décrivent les points de terminaison et les quotas de service pour AWS Parallel Computing Service (AWS PCS). Les quotas de service, anciennement appelés limites, sont le nombre maximum de ressources de service ou d'opérations pour votre Compte AWS.

Vous Compte AWS avez des quotas par défaut pour chaque AWS service. Sauf indication contraire, chaque quota est spécifique à la région. Vous pouvez demander des augmentations pour certains quotas, et d'autres quotas ne peuvent pas être augmentés.

Pour plus d'informations, veuillez consulter la rubrique Quotas du service AWS dans les Références générales AWS .

Table des matières

Points de terminaison de service

Nom de la région	Région	Point de terminaison	Protocole
US East (Virginie du Nord)	us-east-1	pcs.us-east-1.amazonaws.com	HTTPS
USA Est (Ohio)	us-east-2	pcs.us-east-2.amazonaws.com	HTTPS
USA Ouest (Oregon)	us-west-2	pcs.us-west-2.amazonaws.com	HTTPS
Asie-Pacifique (Singapour)	ap-southeast-1	pcs.ap-southeast-1.amazonaws.com	HTTPS

Nom de la région	Région	Point de terminaison	Protocole
Asie-Pacifique (Sydney)	ap-southeast-2	pcs.ap-southeast-2.amazonaws.com	HTTPS
Asie-Pacifique (Tokyo)	ap-northeast-1	pcs.ap-northeast-1.amazonaws.com	HTTPS
Europe (Francfort)	eu-central-1	pcs.eu-central-1.amazonaws.com	HTTPS
Europe (Irlande)	eu-west-1	pcs.eu-west-1.amazonaws.com	HTTPS
Europe (Stockholm)	eu-north-1	pcs.eu-north-1.amazonaws.com	HTTPS

Quotas de service

Nom	Par défaut	Ajustable	Description
Clusters	5	Oui	Le nombre maximum de clusters par Région AWS.

> **Note**
>
> Les valeurs par défaut sont les quotas initiaux définis par AWS. Ces valeurs par défaut sont distincts des valeurs réelles de quotas appliqués et des quotas de service maximaux possible. Pour plus d'informations, veuillez consulter la rubrique Terminologie des Service Quotas dans le Guide de l'utilisateur Service Quotas.

Ces quotas de service sont répertoriés sous AWS Parallel Computing Service (PCS) dans le AWS Management Console. Pour demander une augmentation de quota pour les valeurs indiquées

comme ajustables, consultez la section Demander une augmentation de quota dans le Guide de l'utilisateur du Service Quotas.

> **Important**
>
> N'oubliez pas de vérifier le Région AWS réglage actuel dans le AWS Management Console.

Quotas internes

Les quotas suivants sont internes et non ajustables.

Nom	Par défaut	Ajustable	Description
Création simultanée de clusters	1	Non	Le nombre maximum de clusters dans l'`Creating`état par Région AWS.

Quotas pertinents pour les autres AWS services

AWS PCSutilise d'autres AWS services. Vos quotas de service pour ces services ont un impact sur votre utilisation de AWS PCS.

Quotas EC2 de service Amazon qui ont un impact AWS PCS

- Demandes d'instance ponctuelles
- Exécution d'instances à la demande
- Modèles de lancement
- Versions du modèle de lancement
- EC2APIDemandes Amazon

Pour plus d'informations, consultez les quotas de EC2 service Amazon dans le guide de l'utilisateur d'Amazon Elastic Compute Cloud.

Notes de publication pour AWS PCS un échantillon AMIs

AWS PCSsample AMIs ont une cadence de publication nocturne pour les correctifs de sécurité. Ces correctifs de sécurité incrémentiels ne sont pas inclus dans les notes de publication officielles.

> **Important**
>
> AMIsLes échantillons sont fournis à des fins de démonstration et ne sont pas recommandés pour les charges de travail de production.

Table des matières

AWS PCSexemple x86_64 AMI pour Slurm 23.11 (Amazon Linux 2)

Ce document décrit les derniers changements, ajouts, problèmes connus et correctifs pour AWS PCS Sample x86_64 (AMIAmazon Linux 2).

- Date de création : 15 juillet 2024
- Date de parution : 22 août 2024
- Dernière mise à jour : 22 août 2024

AMInom

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

EC2Instances prises en charge

- Toutes les instances dotées d'un processeur x86 64 bits. Pour trouver des instances compatibles, accédez à la EC2console Amazon. Choisissez Types d'instances, puis recherchez`Architectures=x86_64`.

AMIcontenu

- AWS Service pris en charge : AWS PCS
- Système d'exploitation : Amazon Linux 2
- Architecture de calcul : x86_64
- Noyau Linux : 5.10.220-209.867.amzn2.x86_64
- EBStype de volume : gp2
- AWS PCSInstallateur de Slurm 23.11 : 23.11.9-1
- AWS PCSinstallateur du logiciel : 1.0.0-1
- EFAInstallateur : 1.33.0
- GDRCopy: 2,4
- NVIDIAPilote : 535.154.05
- NVIDIACUDA: 12,2.2_535,04,05

Avis

- Aucun

Date de sortie : 2024-08-22

Mis à jour

- Aucune. Première version.

Ajouté

- Aucune. Première version.

Supprimé

- Aucune. Première version.

AWS PCSexemple d'Arm64 AMI pour Slurm 23.11 (Amazon Linux 2)

Ce document décrit les derniers changements, ajouts, problèmes connus et correctifs pour AWS PCS Sample Arm64 AMI (Amazon Linux 2).

- Date de création : 15 juillet 2024
- Date de parution : 22 août 2024
- Dernière mise à jour : 22 août 2024

AMInom

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

EC2Instances prises en charge

- Toutes les instances dotées d'un processeur Arm 64 bits. Pour trouver des instances compatibles, accédez à la EC2console Amazon. Choisissez Types d'instances, puis recherchez`Architectures=arm64`.

AMIcontenu

- AWS Service pris en charge : AWS PCS
- Système d'exploitation : Amazon Linux 2
- Architecture informatique : arm64
- Noyau Linux : 5.10.220-209.867.amzn2.aarch64
- EBStype de volume : gp2
- AWS PCSInstallateur de Slurm 23.11 : 23.11.9-1
- AWS PCSinstallateur du logiciel : 1.0.0-1
- EFAInstallateur : 1.33.0
- GDRCopy: 2,4
- NVIDIAPilote : 535.154.05
- NVIDIACUDA: 12,2.2_535,04,05

Avis

- Aucun

Date de sortie : 2024-08-22

Mis à jour

- Aucune. Première version.

Ajouté

- Aucune. Première version.

Supprimé

- Aucune. Première version.

Historique du Guide de l'utilisateur AWS PCS

Le tableau suivant décrit les versions de documentation pour AWS PCS.

Date	Modification	Mises à jour de la documentation	APIversions mises à jour
28 août 2024	Page de politiques gérées ajoutée	Pour de plus amples informations, veuillez consulter AWS politiques gérées pour le service de calcul AWS parallèle.	N/A
28 août 2024	AWS PCSlibération	Première publication du guide de AWS PCS l'utilisateur.	AWS SDK: 28/08/2024

AWS Glossaire

Pour la AWS terminologie la plus récente, consultez le AWS glossaire dans la Glossaire AWS référence.

Les traductions sont fournies par des outils de traduction automatique. En cas de conflit entre le contenu d'une traduction et celui de la version originale en anglais, la version anglaise prévaudra.