

HPC Solution

User Guide

Issue 01
Date 2021-03-29



Copyright © Huawei Technologies Co., Ltd. 2024. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions



HUAWEI and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process*. For details about this process, visit the following web page:

<https://www.huawei.com/en/psirt/vul-response-process>

For vulnerability information, enterprise customers can visit the following web page:

<https://securitybulletin.huawei.com/enterprise/en/security-advisory>

Contents

1 Overview.....	1
1.1 What's HPC-S ² ?.....	1
1.2 HPC-S ² Application and Usage Restrictions.....	3
1.3 Interaction Between HPC-S ² and Other Services.....	4
2 Getting Started.....	5
2.1 Overview.....	5
2.2 Method 1: Purchasing an HPC Cluster.....	6
2.3 Method 2: Deploying an HPC Cluster.....	11
3 User Guide.....	15
3.1 Viewing the Cluster Deployment Progress.....	15
3.1.1 Viewing the Progress of Purchasing an HPC Cluster.....	15
3.1.2 Viewing the Progress of Deploying an HPC Cluster Using Existing Cloud Servers.....	15
3.1.3 Viewing the Scale-Out Progress of an HPC Cluster.....	16
3.1.4 Viewing the Scale-In Progress of an HPC Cluster.....	16
3.2 Viewing HPC Clusters.....	16
3.2.1 Viewing Details of an HPC Cluster.....	16
3.2.2 Viewing Details About Nodes in an HPC Cluster.....	17
3.2.3 Managing Shared Storage of an HPC Cluster.....	17
3.3 HPC Cluster Management.....	17
3.3.1 Scaling Out an HPC Cluster.....	17
3.3.2 Stopping an HPC Cluster.....	20
3.3.3 Starting an HPC Cluster.....	20
3.3.4 Deleting an HPC Cluster.....	21
3.4 Node Management.....	22
3.4.1 Stopping a Node.....	22
3.4.2 Starting a Node.....	22
3.4.3 Restarting a Node.....	23
3.4.4 Removing a Node.....	23
3.5 Shared Storage Management.....	24
3.5.1 Managing Shared Storage.....	24
4 FAQs.....	26
4.1 Why Am I Still Charged After I Deleted My Cluster?.....	26

4.2 What Can I Do If a Compute Node Cannot Be Found or Fails to Be Added to a Cluster?.....	26
4.3 If I Fail to Deploy a Cloud Server in a Cluster, Can I Deploy It in Another Cluster?.....	27
4.4 What Can I Do If I Receive Error Message "Insufficient EIP quota"?.....	27
4.5 What Can I Do If the Maximum Number of Clusters Has Been Reached?.....	27
4.6 What Can I Do If I Receive Error Message "master node cannot reached" But the Master Node is Running?.....	27
5 Change History.....	28

1 Overview

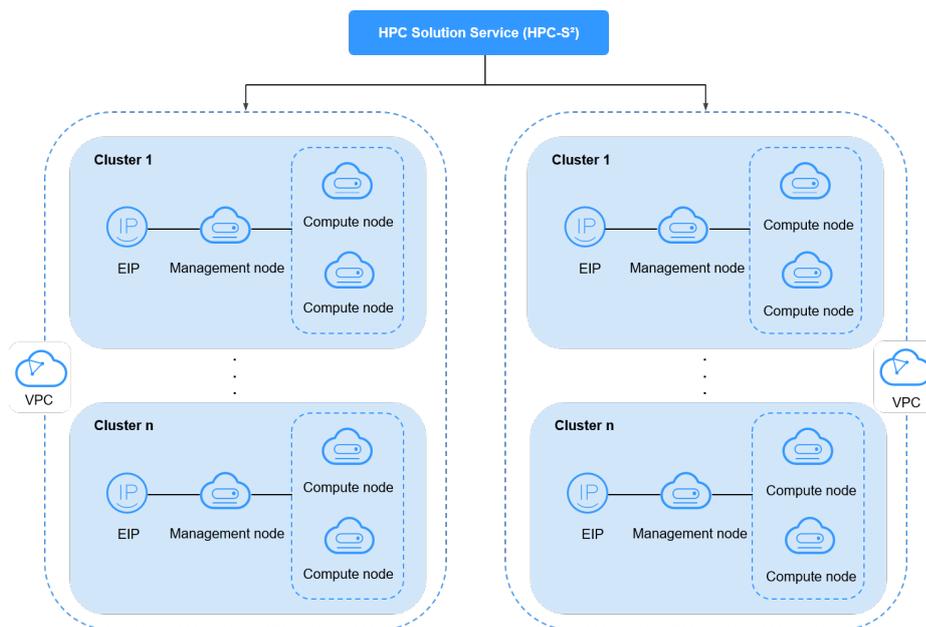
1.1 What's HPC-S²?

High-performance computing (HPC) is the practice of aggregating a large number of compute nodes to work concurrently to solve large-scale computing problems. It has been widely used in scientific research, weather forecast, simulation experiment, biopharmaceuticals, gene sequencing, and image processing.

The requirements for high-performance computing have been increasingly growing. HUAWEI CLOUD launched the HPC Solution Service (HPC-S²) to provide highly scalable high-performance computing resources as a cloud service. With basic services and software pre-integrated, HPC-S² allows you to provision and deploy HPC resources with just a few clicks. HPC-S² can automatically deploy job scheduling software such as Simple Linux Utility for Resource Management (Slurm), Sun Grid Engine (SGE), and Huawei commercial HPC Donau Scheduler to facilitate task assignments on parallel computers. On the HPC-S² console, you can view and manage clusters, nodes, and shared storage in an intuitive way.

You can also use the HPC-S² console to manage multiple clusters in different VPCs, as shown in [Figure 1-1](#).

Figure 1-1 HPC-S² architecture



HPC-S² Functions

The HPC-S² is mainly used by enterprises that require a large amount of compute resources. These enterprises can use the HPC-S² console to easily manage clusters, nodes, tasks, and shared storage, provision clusters, and deploy applications.

Function	Description
Cluster management	Create HPC-S ² clusters; view cluster creation, deployment, scale-in/out progress; view details about clusters; add, start, stop, and remove clusters; and manage instance topology.
Node management	Query, start, stop, restart, and delete nodes in HPC-S ² clusters.
Shared storage management	Provide storage for HPC-S ² clusters by using Scalable File Service (SFS).
Task management	Manage HPC-S ² cluster-related job queueing, dynamic scheduling (including GPU scheduling), bin packing and workload consolidation, and quotas for CPU, memory, and instances.

HPC-S² Advantages

Compared with traditional HPC, HPC-S² offers the following advantages:

- HPC resources are provisioned on demand and are available instantly upon provisioning.

- Multiple types of cloud servers are available for you to choose the most appropriate compute servers for your workloads.
- HPC-S² can work with security services, such as Anti-DDoS, Web Application Firewall (WAF), and Vulnerability Scan Service (VSS), to provide multi-dimensional protection to improve data security and availability.
- With HPC-S², you do not need to build your own equipment rooms or worry about hardware upgrade or replacement. You can obtain resources for exclusive use from virtual resource pools anytime and these resources can be elastically scaled based on service changes.
- HPC-S² can automatically configure password-free login, network information services (NISs), and schedulers for users.

1.2 HPC-S² Application and Usage Restrictions

HPC-S² Application Scenarios

- HPC-S² allows you to create HPC clusters with a few clicks. During the cluster creation, the schedulers, domain controllers, and runtime dependency are automatically deployed.
- You can manage clusters and nodes and view their status on the HPC-S² console.

HPC-S² Usage Restrictions

- The HPC-S² solution is available only in the AP-Singapore region.
- Only Slurm and SGE schedulers can be used for scheduling.
- The SGE scheduler can be in the Kunpeng or x86 architecture, but the Slurm scheduler can only be in the x86 architecture.
- A single user can create a maximum of three clusters.
- A cluster can contain a maximum of 2,000 compute nodes.
- Data disks cannot be automatically attached to compute nodes. If needed, you can manually attach them for use.
- If nodes fail to be created or added, the cluster cannot be automatically restored.
- Only CentOS 7 series public images can be used during cluster creation or capacity expansion.
- If the EIPs or data disks of the nodes are not deleted when the nodes are deleted, they will still incur fees.
- Do not perform operations on the nodes in the HPC cluster on the ECS console. Otherwise, the cluster may encounter with errors.
- Do not modify or delete files in the `/usr/local/hpccluster` directory. Otherwise, the cluster may encounter errors.

The **authorized_keys** file in the `/usr/local/hpccluster/role/.ssh/` directory is used only for system maintenance.

- The clusters created by users with different member accounts cannot have duplicated names if such users belong to the same master account.

1.3 Interaction Between HPC-S² and Other Services

Service	Interaction with HPC-S ²	Reference
Elastic Cloud Server (ECS)	You can purchase ECSs to create HPC clusters or use existing ECSs to deploy HPC clusters.	Purchasing an ECS
Virtual Private Cloud (VPC)	The cloud servers required for the same high-performance computing task are located in a single VPC, where subnets and security groups are configured for access control.	<ul style="list-style-type: none"> • Creating a VPC and a Subnet • Modifying Details About a VPC
Image Management Service (IMS)	You can use public, private, or shared images to create cloud servers required for high-performance computing. You can also create private images from the cloud servers fulfilling high-performance computing.	Creating a Private Image
Elastic Volume Service (EVS)	EVS disks can be attached to cloud servers to provide additional storage.	Purchasing an EVS Disk
Scalable File Service (SFS)	SFS provides hosted shared file storage for cloud servers.	Sharing Files

2 Getting Started

2.1 Overview

Scenarios

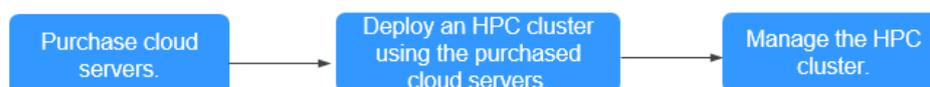
The typical HPC applications include DNA sequencing, industrial simulation, and other scenarios that require high performance computing.

This section describes how to create and configure an HPC cluster and how to complete HPC jobs on the HPC-S² console.

Methods of Creating an HPC Cluster

- **Purchasing an HPC cluster:** Purchase cloud servers with specified specifications as the master and compute nodes.

Figure 2-1 Purchasing an HPC cluster



- **Deploying an HPC cluster:** Select existing cloud servers as the master and compute nodes.

Figure 2-2 Deploying an HPC cluster using existing cloud servers



2.2 Method 1: Purchasing an HPC Cluster

Scenarios

If you want to use new cloud servers to deploy an HPC cluster, perform operations in this section.

Notes

- When you use VPC for the first time, the system automatically creates a VPC for you, including a default security group and NIC. You can also create a VPC by yourself. For more information about VPC, see [Virtual Private Cloud User Guide](#).
- You can use the default security group provided by the system or create a security group by yourself.
To allow secure access to cloud servers in the cluster, ensure that port 22 in the security group rule is open. For details, see [Adding a Security Group Rule](#).
- Scalable File Service (SFS) is used for shared file access spanning cloud servers. For more information about SFS, see [Scalable File Service User Guide](#).

Procedure

1. Log in to the HPC-S² console.
2. Click  in the upper left corner and select your region and project.
3. Click **Buy HPC Cluster**.
The **Buy HPC Cluster** page is displayed.
4. Select a billing mode, **Yearly/Monthly** or **Pay-per-use**.
 - **Yearly/Monthly**: Specify a required duration. The system will deduct the fees incurred from your account based on the service price.

NOTE

Yearly/Monthly cloud servers cannot be deleted. You can only unsubscribe from them. If such a cloud server is no longer needed, go to the corresponding service console to unsubscribe from it.

- **Pay-per-use**: You do not need to set a required duration because the system will deduct the fees incurred from your account based on the service duration.
5. Select a region.
Cloud servers in different regions cannot communicate with each other over an intranet. For low network latency and quick resource access, select the nearest region.
 6. Select an AZ.
An AZ is a physical location that uses an independent power supply and network. AZs in the same region can communicate with each other over an intranet.

- To enhance high availability of applications, create cloud servers in different AZs.
 - To shorten network latency, create cloud servers in the same AZ.
7. Configure the master node.
- **Flavor:** Select a flavor from the drop-down list.
 - **System Disk:** Select a disk type and set the disk size.

 **NOTE**

Before selecting a cloud server flavor, familiarize yourself with the cloud server types. For details, see [ECS Types](#).

8. Configure compute nodes.
- **Flavor:** Select a flavor from the drop-down list.
 - **System Disk:** Select a disk type and set the disk size.

 **NOTE**

Before selecting a cloud server flavor, familiarize yourself with the cloud server types. For details, see [ECS Types](#).

- **Quantity:** You can deploy a maximum of 50 compute nodes in an HPC cluster.
9. Configure the network, including **VPC**, **Security Group**, and **NIC** settings.

 **NOTE**

- The EIP is bound to the master node.
- When you use VPC for the first time, the system automatically creates a VPC for you, including a default security group and NIC.

Table 2-1 Network parameters

Parameter	Description
VPC	<p>A VPC provides a network, including subnets and security groups, for the cloud servers.</p> <p>Select an existing VPC or create one.</p> <p>For more information about VPC, see Virtual Private Cloud User Guide.</p> <p>NOTE DHCP must be enabled in the VPC to which the cloud servers belong.</p>

Parameter	Description
Security Group	<p>A security group controls access to the cloud servers in the group. By default, cloud servers in the same security group can access each other. The access from outside the security group is restricted by security group rules. You can define access rules for a security group to protect the cloud servers that are added to this group.</p> <p>For details, see Adding a Security Group Rule.</p> <p>Enable protocols and ports as needed. Common protocols and ports are as follows:</p> <ul style="list-style-type: none">• Port 80: default port for web page access through HTTP.• Port 443: web page access through HTTPS.• ICMP: checks communication statuses between cloud servers through ping operations.• Port 22: reserved for logging in to a Linux cloud server using SSH.• Port 3389: reserved for logging in to a Windows cloud server using SSH.
NIC	Select a subnet from the drop-down list.
EIP	<p>An EIP is a static public IP address bound to a cloud server in a VPC. Using the EIP, the cloud server provides services externally.</p> <p>The following options are provided:</p> <ul style="list-style-type: none">• Auto assign: The system automatically assigns an EIP with a dedicated bandwidth to the cloud server. The bandwidth is configurable.• Use existing: An existing EIP will be assigned to the cloud server. The EIP is only bound to the master node.
EIP Type	Dynamic BGP provides automatic failover and chooses the optimal path when a network connection fails.
Bandwidth	This parameter is available only if EIP is set to Auto assign .

Parameter	Description
Billed By	<p>This parameter is available only if EIP is set to Auto assign.</p> <p>This parameter indicates the bandwidth billing mode of the purchased EIP. Two billing modes are as follows:</p> <ul style="list-style-type: none">• Bandwidth: You will be billed based on the bandwidth size you set.• Traffic: You will be billed based on the total traffic used.

10. Select an image.

– **Public image** (recommended)

A public image is a standard, widely used image. It contains an OS and preinstalled public applications and is available to all users. If a public image does not contain the application environments or software you need, you can use the public image to create a cloud server and then deploy required software as needed.

– **Private image**

A private image is an image that contains an OS or service data, pre-installed public applications, and the user's personal applications. Private images are only available to the users who created them. Using a highly-tailored private image to create cloud servers eliminates the need to manually configure multiple cloud servers repeatedly.

– **Shared image**

A shared image is a private image another user has shared with you.

– **Marketplace image**

The Marketplace is an online store where you can purchase third-party images that have the OS, application environments, and software preinstalled. You can use these images to deploy websites and application development environments in just a few clicks. No additional configuration is required.

11. Set **Login Mode**.

Key pair is recommended because it features higher security than **Password**. If you select **Password**, ensure that the password meets complexity requirements to prevent malicious attacks.

– **Key pair**

A key pair will be used for cloud server login authentication. You can select an existing key pair or create one.

 **NOTE**

If you use an existing key pair, make sure that you have saved the key file locally. Otherwise, logging in to the cloud servers will fail.

– **Password**

A username and its initial password will be used for cloud server login authentication.

The initial password of user **root** will be used for authenticating Linux login, while that of user **Administrator** will be used for authenticating Windows login. The password must meet the following requirements.

Table 2-2 Password complexity requirements

Item	Requirements	Example Value
Password	<ul style="list-style-type: none"> • Consists of 8 to 26 characters. • Contains at least three of the following character types: <ul style="list-style-type: none"> - Uppercase letters - Lowercase letters - Digits - Special characters: \$!@%-_+=[]:./^, {}? • Cannot contain the username or the username spelled backwards. • Cannot contain more than two characters in the same sequence as they appear in the username. (This requirement applies only to Windows cloud servers.) 	YNbUwp! dUc9MClnv NOTE The example password is generated randomly. Do not copy it for use.

 **NOTE**

The system does not periodically change the password. It is recommended that you change it regularly for security.

12. Set a cluster name.

The name must comply with the naming rules in the following table.

Table 2-3 Cluster naming rules

Item	Requirements	Example Value
Cluster name	<ul style="list-style-type: none"> • Consists of 1 to 52 characters. • Contains only letters, digits, hyphens (-), and periods (.). • Cannot start or end with a hyphen (-) or period (.). • Cannot contain consecutive hyphens (-) and periods (.) or the combination of them, for example, --, .., -., or .-. 	hpc-001.p1

13. Select a file system.

Available file systems are filtered and displayed in the drop-down list based on the network settings. If no files systems are available, you can create one and add it to the VPC specified in [Table 2-1](#). For details, see [Create a File System](#).

14. Select a scheduler.

Currently, only the **slurm** and **sgx** schedulers are available. The default one is **slurm**.

15. Confirm the configurations and click **Buy Now**. A page is displayed for you to confirm the order.
 - In **Yearly/Monthly** billing mode, after confirming the specifications and price of the order, agree to the service agreement, click **Submit**, and complete the payment as prompted.
 - In **Pay-per-use** billing mode, after confirming the specifications and price of the order, agree to the service agreement, and click **Submit**.
16. After the order is submitted or paid, the system returns back to the **Clusters** page, on which you can see that the cluster is in the **Creating** state. Wait until the cluster status changes to **Running** which indicates that the cluster is successfully created.

Follow-up Operations

View the cluster details, nodes, and shared storage, as instructed in [Viewing HPC Clusters](#).

2.3 Method 2: Deploying an HPC Cluster

Scenarios

If you want to use existing cloud servers to deploy an HPC cluster, perform operations in this section.

NOTICE

Select a master node and then compute nodes. These nodes must have the same CPU architecture and OS version and in the same VPC and security groups.

Notes

- When you use VPC for the first time, the system automatically creates a VPC for you, including a default security group and NIC. You can also create a VPC by yourself. For more information about VPC, see [Virtual Private Cloud User Guide](#).
- You can use the default security group provided by the system or create a security group by yourself.

To allow secure access to cloud servers in the cluster, ensure that port 22 in the security group rule is open. For details, see [Adding a Security Group Rule](#).

- Scalable File Service (SFS) is used for shared file access spanning cloud servers. For more information about SFS, see [Scalable File Service User Guide](#).

Procedure

1. Log in to the HPC-S² console.
2. Click  in the upper left corner and select your region and project.
3. Click **Deploy HPC Cluster**.
The **Deploy HPC Cluster** page is displayed.
4. Select a region.
Cloud servers in different regions cannot communicate with each other over an intranet. For low network latency and quick resource access, select the nearest region.
5. Set a cluster name.
The name must comply with the naming rules in the following table.

Table 2-4 Cluster naming rules

Item	Requirements	Example Value
Cluster name	<ul style="list-style-type: none">• Consists of 1 to 52 characters.• Contains only letters, digits, hyphens (-), and periods (.)• Cannot start or end with a hyphen (-) or period (.)• Cannot contain consecutive hyphens (-) and periods (.) or the combination of them, for example, --, .., -., or .-.	hpc-001.p1

6. Select a cloud server as the master node.
The master node must:
 - Be in the region selected in [2](#).
 - Be in the **Running** state.
 - Have an EIP bound.
 - Not be included in other clusters under the current account.
7. Select security groups.
A security group controls the access to the cloud servers added to this group by defining access rules. You can select multiple security groups.
8. Select compute nodes.
The compute nodes must:
 - Be in the region selected in [2](#).
 - Be in the **Running** state.
 - Use the same CPU architecture as the master node.

- Be in the same VPC and security groups as the master node.
 - Use the same major OS version as the master node.
 - Be no more than 50 in quantity.
9. Select a VPC.
This parameter is automatically set to the VPC of the master node.
10. Set the login mode.

Only **Password** is available. If any of the selected cloud servers is using key pair authentication, go to the corresponding service console to change it to password authentication. In addition, set a strong password to protect yourself from malicious attacks.

The initial password of user **root** will be used for authenticating Linux login, while that of user **Administrator** will be used for authenticating Windows login. The password must meet the following requirements.

Table 2-5 Password complexity requirements

Item	Requirements	Example Value
Password	<ul style="list-style-type: none"> • Consists of 8 to 26 characters. • Contains at least three of the following character types: <ul style="list-style-type: none"> - Uppercase letters - Lowercase letters - Digits - Special characters: \$!@%_-_+[:./^,{}? • Cannot contain the username or the username spelled backwards. • Cannot contain more than two characters in the same sequence as they appear in the username. (This requirement applies only to Windows cloud servers.) 	YNbUwp! dUc9MCl n v NOTE The example password is generated randomly. Do not copy it for use.

11. Select a file system.
Available file systems are filtered and displayed in the drop-down list based on the network settings. If no files systems are available, you can create one and add it to the VPC specified in 9. For details, see [Create a File System](#).
12. Select a scheduler.
Currently, only the **slurm** and **sgs** schedulers are available. The default one is **slurm**.
13. Confirm the configurations and click **Deploy HPC Cluster**.
14. After the order is submitted, the system returns back to the **Clusters** page, where you can see that the cluster is in the **Deploying** state. Wait until the cluster status changes to **Running** which indicates that the cluster is successfully deployed.

Follow-up Operations

View the cluster details, nodes, and shared storage, as instructed in [Viewing HPC Clusters](#).

3 User Guide

3.1 Viewing the Cluster Deployment Progress

3.1.1 Viewing the Progress of Purchasing an HPC Cluster

Scenarios

After you submit the order for purchasing an HPC cluster, you can view the cluster deployment progress on the HPC-S² console.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Clusters**.
3. In the cluster list, locate the row where the cluster you want to view resides and click **View Progress** in the **Operation** column.

On the displayed page, set **Log Type** to **Deployment** and view the deployment progress of the cluster.

3.1.2 Viewing the Progress of Deploying an HPC Cluster Using Existing Cloud Servers

Scenarios

After you submit a request for deploying an HPC cluster using existing cloud servers, you can view the cluster deployment progress on the HPC-S² console.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Clusters**.

3. In the cluster list, locate the row where the cluster you want to view resides and click **View Progress** in the **Operation** column. On the displayed page, set **Log Type** to **Deployment** and view the deployment progress of the cluster.

3.1.3 Viewing the Scale-Out Progress of an HPC Cluster

Scenarios

After you submit a request for scaling out an HPC cluster, you can view the scale-out progress on the HPC-S² console.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Clusters**.
3. In the cluster list, locate the row where the cluster you want to view resides and click **View Progress** in the **Operation** column.

On the displayed page, set **Log Type** to **Manual scale-out** and view the scale-out progress.

3.1.4 Viewing the Scale-In Progress of an HPC Cluster

Scenarios

After you submit a request for scaling in an HPC cluster, you can view the scale-in progress on the HPC-S² console.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Clusters**.
3. In the cluster list, locate the row where the cluster you want to view resides and click **View Progress** in the **Operation** column.

On the displayed page, set **Log Type** to **Manual scale-in** and view the scale-in progress.

3.2 Viewing HPC Clusters

3.2.1 Viewing Details of an HPC Cluster

Scenarios

You can view details about an HPC cluster, including the cluster ID, status, region, VPC, security groups, nodes, shared storage, and topology.

Procedure

1. Log in to the HPC-S² console.

2. In the navigation pane on the left, choose **Clusters**.
3. In the cluster list, click the name of the cluster you want to view.
On the cluster details page, you can view the cluster ID, status, region, VPC, security groups, nodes, shared storage, and topology.

3.2.2 Viewing Details About Nodes in an HPC Cluster

Scenarios

You can view node details of an HPC cluster, including the node name/ID, type, status, specifications, IP address, shared storage, AZ, and billing mode.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Nodes**.
3. On the displayed page, select the cluster you want to view from the cluster drop-down list. The nodes in the cluster are displayed.
4. Locate the node you want to view and click its name to go to the ECS console for its details.

3.2.3 Managing Shared Storage of an HPC Cluster

Scenarios

You can view details about the shared storage used by an HPC cluster, including the name and ID of the SFS file system, protocol, shared path, and mount point.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Shared Storage**.
3. On the displayed page, select the cluster you want to view from the cluster drop-down list, and view details about the shared storage of the cluster.

3.3 HPC Cluster Management

3.3.1 Scaling Out an HPC Cluster

Scenarios

If the HPC cluster does not meet your requirements, you can scale out the cluster.

NOTE

Only clusters in the **Running** state can be scaled out.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Clusters**.
3. On the displayed page, locate the row where the cluster you want to scale out resides and click **Scale** in the **Operation** column. The scale-out details page is displayed.
4. Select a billing mode.

NOTE

- During scale-out, you can select a different billing mode for scaled nodes. If the cluster contains both the pay-per-use nodes and yearly/monthly billing nodes, the cluster billing mode is displayed as **Yearly/Monthly and Pay-per-use**.
- **Yearly/Monthly**: Specify a required duration. The system will deduct the fees incurred from your account based on the service price.
 - **Pay-per-use**: You do not need to set a required duration because the system will deduct the fees incurred from your account based on the service duration.
5. Specify **Region** and **AZ**.
 - By default, the region is unconfigurable and set to the region where the cluster is created.
 - You can select a different AZ.
 6. Configure the specifications for new nodes.
 - **Flavor**: Select a flavor from the drop-down list.
 - **System Disk**: Select a disk type and set the disk size.
 7. Configure the network.

VPC, **Security Group**, and **NIC** are automatically set to those configured when the cluster was created.
 8. Select an image.

NOTE

- You are advised to select a public image for new nodes. Otherwise, the scale-out may fail.
- **Public image** (recommended)

A public image is a standard, widely used image that is visible to all users. It contains an OS and preinstalled public applications. Currently, only CentOS 7 images are available.
 - **Private image**

A private image is only available to the user who created it. It contains an OS, preinstalled public applications, and the user's personal applications. Using a highly-tailored private image to create cloud servers eliminates the need to manually configure multiple cloud servers repeatedly.

You can also select an encrypted image. For details, see *Image Management Service User Guide*.
 - **Shared image**

A shared image is a private image another user has shared with you.

- **Marketplace image**

The Marketplace is an online store where you can purchase third-party images that have the OS, application environments, and software preinstalled.

9. Set **Login Mode**.

 **NOTE**

The login mode must be the same as that configured when the cluster was created.

- **Key pair**

A key pair will be used for cloud server login authentication. You can select an existing key pair or create one.

 **NOTE**

If you use an existing key pair, make sure that you have saved the key file locally. Otherwise, logging in to the cloud servers will fail.

- **Password**

A username and its initial password will be used for cloud server login authentication.

The password must meet the following requirements.

Item	Requirement	Example Value
Password	<ul style="list-style-type: none"> Consists of 8 to 26 characters. Contains at least three of the following character types: <ul style="list-style-type: none"> Uppercase letters Lowercase letters Digits Special characters: \$!@%-_+=[]:./^,{}? Cannot contain the username or the username spelled backwards. 	YNbUwp! dUc9MCl nv NOTE The example password is generated randomly. Do not copy it for use.

10. Select the amount of time you want to use the nodes for.

If you set **Billing Mode** to **Yearly/Monthly**, you need to select the required duration and determine whether to select **Auto-renew**. If you select **Auto-renew**, the node will be automatically renewed when the subscription period expires.

11. Specify **Quantity**.

The total number of existing and new compute nodes in the cluster cannot exceed 50.

12. Confirm the configurations and click **Buy Now**. A page is displayed for you to confirm the order.

- In **Yearly/Monthly** billing mode, after confirming the specifications and price of the order, agree to the service agreement, click **Submit**, and complete the payment as prompted.
 - In **Pay-per-use** billing mode, after confirming the specifications and price of the order, agree to the service agreement, and click **Submit**.
13. After the order is submitted or paid, the system returns back to the **Clusters** page where you can see that the cluster is in the **Creating** state. Wait until the cluster status changes to **Running**.

3.3.2 Stopping an HPC Cluster

Scenarios

If there are no jobs to run on a pay-per-use HPC cluster, you can stop the cluster to save money.

NOTE

- Yearly/monthly cloud servers will continue to be billed after the cluster is stopped. You need to manually unsubscribe from the cloud servers to stop billing.
- Only clusters in the **Running** state can be stopped.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Clusters**.
3. In the cluster list, locate the row where the cluster you want to stop resides, and choose **More > Stop** in the **Operation** column.
4. Click **OK**.
5. Wait until the cluster status changes to **Stopped**.

3.3.3 Starting an HPC Cluster

Scenarios

If you need to execute jobs on a stopped cluster, you can start the cluster.

NOTE

Only clusters in the **Stopped** state can be started.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Clusters**.
3. In the cluster list, locate the row where the cluster you want to start resides, and choose **More > Start** in the **Operation** column.
4. Click **Next**.
5. Wait until the cluster status changes to **Running**.

3.3.4 Deleting an HPC Cluster

Scenarios

- If a cluster is no longer needed, you can delete it. If you do not select **Delete cloud server** when you delete the cluster, these servers will be retained and continue to be billed after the cluster is deleted. To stop billing for these cloud servers, manually delete pay-per-use servers and unsubscribe from yearly/monthly servers.
- If a cluster fails to be created, you can delete the cluster and do not select **Delete cloud server**. Then, use these servers to deploy a cluster again.
- If an exception occurs in a cluster and the cluster cannot be manually restored, you can delete the cluster and do not select **Delete cloud server**. Then, use these servers to deploy a cluster again.

Precautions

- The EIP bound to the cluster will continue to be billed after the cluster is deleted. You need to manually delete the EIP to stop billing.
- Data disks attached to nodes in the cluster will continue to be billed after the cluster is deleted, even if you select **Delete cloud server** when deleting the cluster. You need to manually delete the data disks to stop billing.

Procedure

1. Log in to the HPC-S² console.
1. In the navigation pane on the left, choose **Clusters**.
2. In the cluster list, locate the row where the cluster you want to delete, and choose **More > Delete** in the **Operation** column. On the confirmation page, whether option **Delete cloud server** is available depends on the billing mode of the cluster you want to delete.
 - Pay-per-use
Delete cloud server is available on the confirmation page.
 - Option selected
The cloud servers will be deleted together with the cluster.
 - Option not selected
The cloud servers are retained. You can view the cloud servers on the corresponding service console.
 - Yearly/Monthly
Delete cloud server is unavailable on the confirmation page.
You need to unsubscribe from the yearly/monthly cloud servers in the cluster.
 - Yearly/Monthly and Pay-per-use
Delete cloud server is available on the confirmation page.
 - Option selected

Pay-per-use cloud servers will be deleted together with the cluster. You need to manually unsubscribe from the yearly/monthly cloud servers in the cluster.

- Option not selected
The cloud servers are retained. You can view the cloud servers on the corresponding service console.

3.4 Node Management

3.4.1 Stopping a Node

Scenarios

If a pay-per-use node is faulty or no longer needed, you can stop the node to save money.

NOTE

- A yearly/monthly node will continue to be billed after being stopped. You need to manually unsubscribe from the node to stop billing.
- Only nodes in the **Running** state can be stopped.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Nodes**.
3. Select the target cluster from the cluster drop-down list.
4. Locate the row where the node you want to stop resides, and choose **More > Stop** in the **Operation** column.
5. Click **Yes**.
6. Wait until the node status changes to **Stopped**.

3.4.2 Starting a Node

Scenarios

You can start nodes as needed.

Only nodes in the **Stopped** state can be started.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Nodes**.
3. Select the target cluster from the cluster drop-down list.
4. Locate the row where the node you want to start resides and choose **More > Start** in the **Operation** column.

5. Click **Yes**.
6. Wait until the node status changes to **Running**.

3.4.3 Restarting a Node

Scenarios

You can restart a faulty node to restore it.

Only nodes in the **Running** state can be restarted.

Procedure

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Nodes**.
3. Select the target cluster from the cluster drop-down list.
4. Locate the row where the node you want to restart resides and choose **More > Restart** in the **Operation** column.
5. Click **Yes**.
6. Wait until the node status changes to **Running**.

3.4.4 Removing a Node

Scenarios

- If there are too many idle nodes in the cluster, you can remove some nodes to save money.
- If a faulty node in the cluster cannot be manually restored, you can remove it and then add a new node to the cluster as needed.

Precautions

- Master nodes cannot be removed.
- A compute node can be removed only when its cluster is in the **Running** or **Abnormal** state.
- A compute node can be removed only when other nodes in the cluster are in the **Running** state.
- Before removing a compute node, you need to ensure that no jobs are running on the node. Otherwise, the running jobs will be abnormal.
- Attached data disks will be retained and continue to be billed after the node is removed. To stop billing for these data disks, delete them manually.

Removing a Single Node

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Nodes**.
3. Select the target cluster from the cluster drop-down list.
Locate the row where the node you want to remove resides and choose **More > Remove** in the **Operation** column.

- Pay-per-use
If you select **Delete cloud server**, the cloud server will be deleted. Otherwise, the cloud server will be retained. You can view the cloud server on the corresponding service console.
- Yearly/Monthly
You need to manually unsubscribe from yearly/monthly cloud servers.

Removing Nodes in Batches

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Nodes**.
3. Select the target cluster from the cluster drop-down list.
Select the nodes to be removed and click **Remove** on the left of the cluster name.
 - Pay-per-use
If you select **Delete cloud server**, the cloud servers will be deleted. Otherwise, the cloud servers will be retained. You can view the cloud servers on the corresponding service console.
 - Yearly/Monthly
You need to manually unsubscribe from yearly/monthly cloud servers.
 - Yearly/Monthly and Pay-per-use
If you select **Delete cloud server**, pay-per-use cloud servers will be deleted and yearly/monthly cloud servers will be retained and continue to be billed. To stop billing for these yearly/monthly cloud servers, unsubscribe from them manually. If you do not select **Delete cloud server**, all cloud servers will be retained.

3.5 Shared Storage Management

3.5.1 Managing Shared Storage

Scenarios

The HPC-S² console enables you to mount SFS file systems to and unmount them from running clusters. SFS provides shared storage for clusters.

Mounting

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Shared Storage**.
3. On the **Shared Storage** page, select the target cluster from the drop-down list in the upper left corner.
4. In the upper right corner of the displayed page, click **Mount**.
5. On the **Mount Shared Storage** page, enter the mount point for the shared storage and select a file system from the drop-down list.

6. Click **Yes**.

Unmounting

1. Log in to the HPC-S² console.
2. In the navigation pane on the left, choose **Shared Storage**.
3. On the **Shared Storage** page, select the target cluster from the drop-down list in the upper left corner.
4. In the list, view the file system attached to the current cluster and click **Unmount** in the **Operation** column.
5. Click **Yes**.

4 FAQs

4.1 Why Am I Still Charged After I Deleted My Cluster?

- You deselected **Delete cloud server** when you delete a pay-per-use cluster. Therefore, only the cluster is deleted, but cloud servers in the cluster are still being billed.
- When you delete a yearly/monthly cluster, only the cluster is deleted, but cloud servers in the cluster are still being billed. To stop billing, manually unsubscribe from the cloud servers.
- You have purchased an EIP for the master node in the cluster, but the EIP is not deleted even if you select **Delete cloud server** when deleting the cluster. To stop billing, manually delete the EIP.
- If an EVS disk is attached to a node in the cluster, the disk is not deleted and is still being billed even if you select **Delete cloud server** when you delete the cluster. To stop billing, manually delete the EVS disk.

4.2 What Can I Do If a Compute Node Cannot Be Found or Fails to Be Added to a Cluster?

Symptom

If a compute node cannot be found or fails to be added to a cluster, the cluster cannot be deployed or created.

Handling Methods

- Delete the cluster. Manually bind an EIP to a compute node that was successfully created and then use it together with other compute nodes to create a new cluster.
- Delete the abnormal compute node to bring the cluster back to work.

4.3 If I Fail to Deploy a Cloud Server in a Cluster, Can I Deploy It in Another Cluster?

Deselect **Delete cloud server** and delete this cluster. Use the cloud server to deploy another cluster.

4.4 What Can I Do If I Receive Error Message "Insufficient EIP quota"?

Symptom

Error message **Insufficient EIP quota** is displayed after I submit a request for creating a cluster.

Handling Methods

Take one of the following actions:

- Set **EIP** to **Use existing**.
- Submit a service ticket to increase your EIP quota.

4.5 What Can I Do If the Maximum Number of Clusters Has Been Reached?

Each user can create up to three clusters. You can delete the clusters that are no longer needed and create new ones again.

4.6 What Can I Do If I Receive Error Message "master node cannot reached" But the Master Node is Running?

Symptom

The master node is **Running**, but error message **master node cannot reached** is displayed.

Handling Methods

1. Check that `/usr/local/.hpccluster/role/.ssh/authorized_keys` has been added to **AuthorizedKeysFile** in `/etc/ssh/sshd_config`.
2. Check that **authorized_keys** exists in `/usr/local/.hpccluster/role/.ssh/`.

5 Change History

Released On	Description
2021-03-29	This issue is the first official release.