

Cloud Container Engine Autopilot

FAQs

Issue 01
Date 2025-01-22



Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2025. 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 Cloud Computing Technologies Co., Ltd.

Trademarks and Permissions



HUAWEI and other Huawei trademarks are the property 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 Cloud 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.

Huawei Cloud Computing Technologies Co., Ltd.

Address: Huawei Cloud Data Center Jiaoxinggong Road
Qianzhong Avenue
Gui'an New District
Gui Zhou 550029
People's Republic of China

Website: <https://www.huaweicloud.com/intl/en-us/>

Contents

1 Billing	1
1.1 How Is a CCE Autopilot Cluster Billed?	1
1.2 How Do I Change the Billing Mode of the vCPU or Memory Required by Pods from Pay-per-Use to Packages?	2
2 Workloads	4
2.1 Workload Exceptions	4
2.1.1 What Do I Do If an Image Can't Be Pulled from SWR During Workload Creation?	4
2.1.2 What Do I Do If a Public Image Can't Be Pulled During Workload Creation?	4
2.1.3 What Do I Do If "Cluster pod max limit exceeded" Is Displayed for a Workload?	6
2.1.4 What Do I Do If Pods Are Repeatedly Created During Workload Creation?	6
2.2 Monitoring Logs	8
2.2.1 Why Is the Reported Container Memory Usage Inconsistent with the Auto Scaling Action?	8
3 Network Management	10
3.1 How Do I Configure Security Group Rules for a Cluster?	10
3.2 How Can I Check Whether a Network Interface Is Used by a Cluster?	13
4 Storage	15
4.1 Can PVs of the EVS Type in a CCE Autopilot Cluster Be Restored After They Are Deleted or Expire?	15
4.2 What Can I Do If a Storage Volume Fails to Be Created?	15
4.3 Can CCE Autopilot PVCs Detect Underlying Storage Faults?	16
4.4 How Can I Delete the Underlying Storage If It Remains After a Dynamically Created PVC is Deleted?	16
5 Permissions	18
5.1 Can I Configure Only Namespace Permissions Without Cluster Management Permissions?	18
5.2 Can I Use APIs If the Cluster Management Permissions Are Not Configured?	18
5.3 Can I Use kubectl If the Cluster Management Permissions Are Not Configured?	19
5.4 Why Can't an IAM User Make API Calls?	19

1 Billing

1.1 How Is a CCE Autopilot Cluster Billed?

Billing Modes

CCE Autopilot supports two billing modes: pay-per-use and packages.

- **Pay-per-use:** You can start using CCE Autopilot resources first and then pay as you go. Their usage durations are calculated by the second but billed every hour. This allows you to flexibly adjust the resources. You neither need to prepare for resources in advance, nor end up with excessive or insufficient preset resources. Pay-per-use billing is a good option for scenarios where there are sudden traffic bursts, such as e-commerce promotions.
- **Packages:** You need to pay for packages before using them. Their usages are settled every hour. You can buy packages to save money. Packages are a good option for long-term, stable services.

NOTE

On the **Overview** page of the cluster console, you can purchase CPU and/or memory packages required by pods. **After you purchase a package in a region, it can be used by all pods in all CCE Autopilot clusters in that region.**

There are no packages for the cluster management and VPC endpoints required by CCE Autopilot clusters. Also, packages are unavailable for other cloud service resources on the CCE console. When a cloud service resource is created on the CCE console, it is billed on a pay-per-use basis by default. You can go to the corresponding cloud service console to purchase yearly/monthly cloud service resources or packages. These resources or packages will be billed based on their standard prices.

Table 1-1 Cluster billing modes

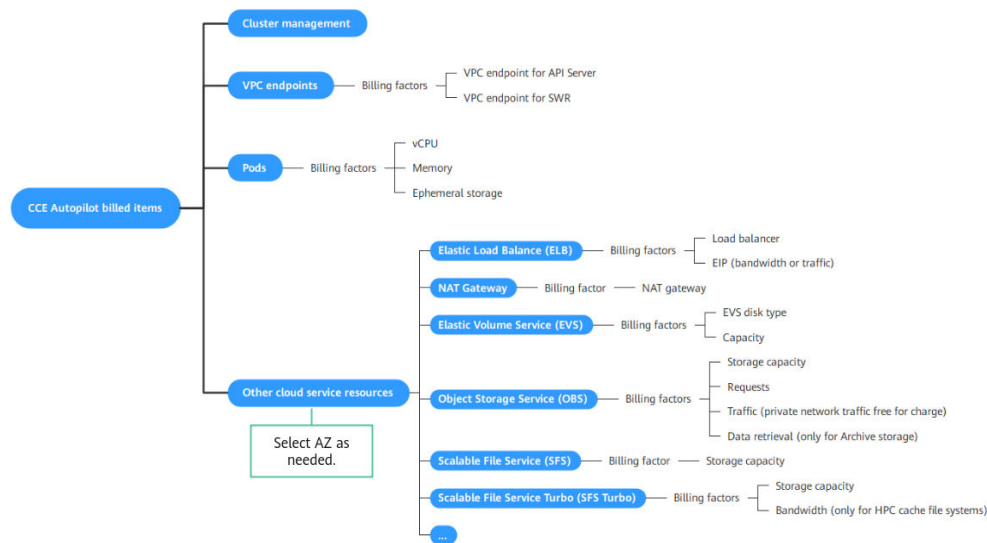
Billing Mode	Pay-per-use	Package
Payment	Postpaid. You are billed based on the actual usage durations.	Prepaid. The packages are used first during the subscription period.

Billing Method	Calculated by the second but billed every hour	Billed by the subscription term you purchase
Billing Items	All billed items	Only vCPU and memory required by the pods
Application Scenario	Recommended when the resource demands are likely to fluctuate and you want more flexibility	A cost-effective option for scenarios where the resource usage duration is predictable. It is recommended for resources expected to be used for a long term.

Billed Items

The total expenditures generated when you use a CCE Autopilot consist of the cluster management, pods, VPC endpoints, and other cloud service resources. For details, see "Billed Items" in "Billing" in the User Guide.

Figure 1-1 Billed items



1.2 How Do I Change the Billing Mode of the vCPU or Memory Required by Pods from Pay-per-Use to Packages?

When a CCE Autopilot cluster is being used, the vCPUs and memory required by pods are billed on a pay-per-use basis by default. If the vCPUs and memory required for creating pods do not meet your requirements, you can purchase packages as needed to enjoy more discounts. For details, see [Procedure](#).

NOTE

When purchasing packages, note the following:

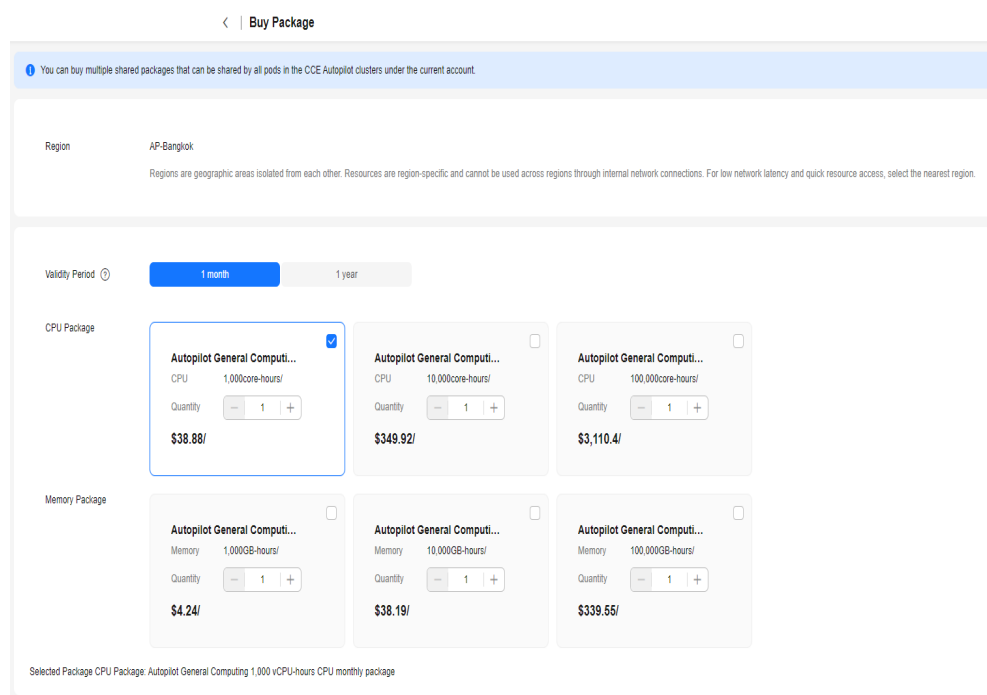
- Packages are billed once and effective immediately upon payment. Currently, you cannot set a future effective date or unsubscribe from the packages.
- After the packages expire, you can still use CCE Autopilot clusters. Ensure that your account balance is sufficient. The system will automatically settle the expenditures on a pay-per-use basis.
- The validity period can be one month or one year. After the validity period expires, the remaining resources cannot be used.

Procedure

Step 1 Log in to the [CCE console](#) and click the cluster name to go to the **Overview** page. In the **Autopilot Resource Package** area on the right, click **Buy Package**.

Step 2 On the page that is displayed, select the required package specifications as prompted.

Figure 1-2 Buying a package



Step 3 Confirm the specifications, click **Pay Now** in the lower right corner. In the displayed dialog box, click **OK**.

Step 4 On the **Buy Cloud Container Engine (CCE)** page, pay for the order as prompted.

----End

2 Workloads

2.1 Workload Exceptions

2.1.1 What Do I Do If an Image Can't Be Pulled from SWR During Workload Creation?

Symptom

The following information is displayed when a workload is created in a CCE Autopilot cluster:

```
Failed to pull image "swr.cn-north-**.myhuaweicloud.com/**/nginx:latest": rpc error: code = Unknown desc = failed to pull and unpack image "swr.cn-north-7.myhuaweicloud.com/**/nginx:latest": failed to resolve reference "swr.cn-north-7.myhuaweicloud.com/**/nginx:latest": failed to do request: Head "https://swr.cn-north-**.myhuaweicloud.com/v2/**/nginx/manifests/latest": dial tcp 100.79.**.**:443: i/o timeout
```

Fault Location

The error information indicates that the SWR image cannot be pulled during workload creation. Check whether the VPC endpoints for accessing OBS and SWR are running properly.

Solution

[Create VPC endpoints for accessing OBS and SWR.](#)

2.1.2 What Do I Do If a Public Image Can't Be Pulled During Workload Creation?

Symptom

The following information is displayed when a workload is created in a CCE Autopilot cluster:

```
Failed to pull image "100.125.**.**:32334/**/nginx:1.0": rpcerror: code =DeadlineExceeded desc = failed to pull and unpack image "100.125.**.**:32334/**/nginx:1.0": failed to resolve reference "100.125.**.**:32334/**/
```

```
nginx:1.0": failed to do request Head: Head "https://100.125.**.**:32334/v2/**/nginx/manifests/1.0": dial tcp 100.125.**.**:32334: i/o timeout
```

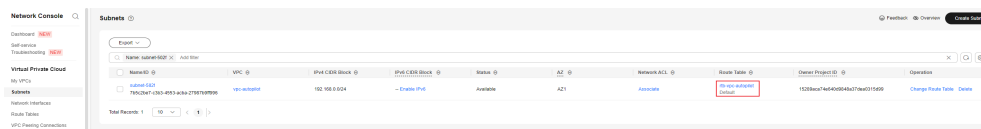
Fault Locating

When the CCE Autopilot cluster pulls the image from the public network, the NAT gateway cannot access the public network because there is no route destined for the NAT gateway in the route table of the subnet.

Solution

Add the route destined for 0.0.0.0/0 over the NAT gateway in the default route table or custom route table of the subnet.

- Step 1** Log in to the CCE console and click the cluster name to access the cluster console.
- Step 2** In the navigation pane, choose **Overview**. In the **Networking Configuration** area, view the cluster subnet.
- Step 3** Switch to the **Network Console**. In the navigation pane, choose **Virtual Private Cloud > Subnets**. Locate the subnet by name and click the route table name.

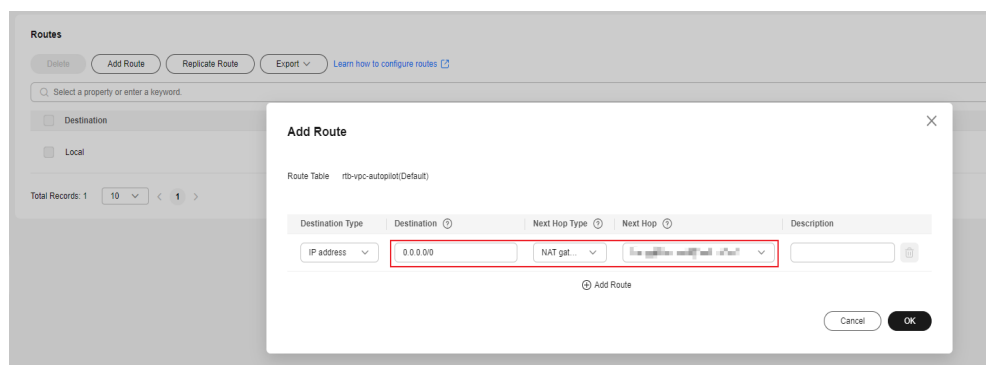


- Step 4** On the **Summary** tab, check whether a route to the NAT gateway exists.

If no, manually add a route. Click **Add Route**.

- **Destination:** Set this parameter to **0.0.0.0/0**, which means any IP address.
- **Next Hop Type:** Select **NAT gateway**.
- **Next Hop:** Select the NAT gateway configured for the subnet.

Click **OK**.



----End

2.1.3 What Do I Do If "Cluster pod max limit exceeded" Is Displayed for a Workload?

Symptom

The following error occurs when a workload is created:

Cluster pod max limit exceeded(x)

Fault Location

This indicates that the maximum number of pods in the cluster is reached and no more pods can be created. *x* indicates the maximum of pods allowed in the cluster. The default value is **500**.

Solution

Increase the quota by [submitting a service ticket](#).

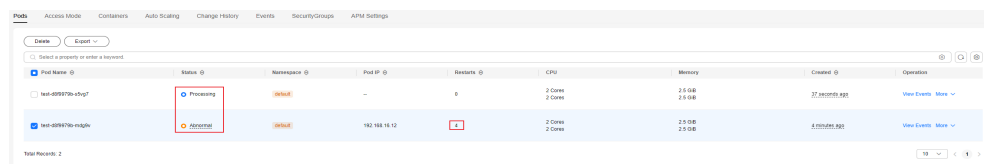
NOTE

The add-on pods installed in the cluster occupy the pod quota. Plan the pod quota properly.

2.1.4 What Do I Do If Pods Are Repeatedly Created During Workload Creation?

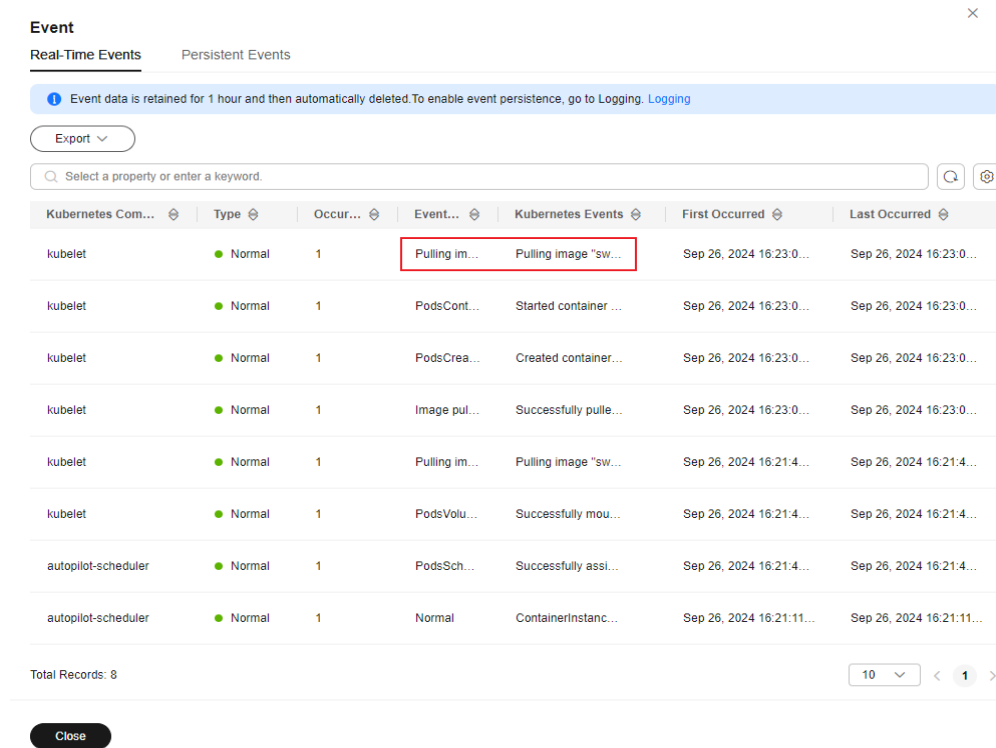
When a workload is being created and in the **Processing** or **Unready** state, the pods for this workload are created repeatedly, and the **Pulling image xx** event lasts for a long time during pod creation. The 30-GiB free ephemeral storage provided for each pod is smaller than the disk capacity required for pulling the image. In this case, you need to expand the storage capacity.

Figure 2-1 Repeated creation of a pod



Pod Name	Status	Namespace	Pod IP	Restart	CPU	Memory	Created	Operation
test-659376v-47q7	Processing	default	-	0	2 Cores 2 Cores	25.0GB 25.0GB	37 seconds ago	View Events More
test-659376v-47q7	Aborted	default	192.168.16.12	1	2 Cores 2 Cores	25.0GB 25.0GB	8 minutes ago	View Events More

Figure 2-2 Pod events

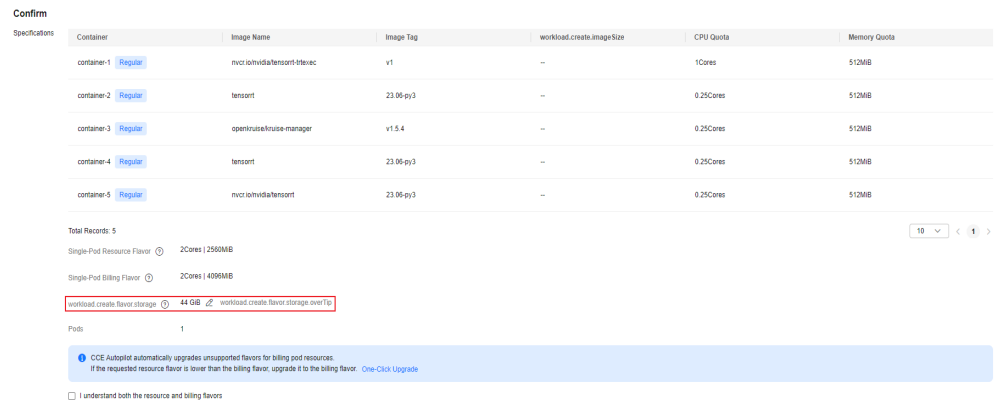


Solutions

You can add ephemeral storage capacity for a workload to meet the storage capacity required for pulling images.

- Step 1** On the **CCE console**, click the cluster name to access the cluster console.
- Step 2** In the navigation pane, choose **Workloads**. On the displayed page, select the workload to be restarted and click **Upgrade** in the **Operation** column. On the **Confirm** page, click the edit icon to modify the ephemeral storage.

Figure 2-3 Modifying the ephemeral storage for a pod



Step 3 Select **I have read and agree to the pricing policy** and click **Upgrade Workload**.

----End

2.2 Monitoring Logs

2.2.1 Why Is the Reported Container Memory Usage Inconsistent with the Auto Scaling Action?

Symptom

The memory usage of a container being monitored does not match the auto scaling requirements. For example, the GUI shows that memory usage of a container is around 40%, but the HPA scale-in threshold is set at 70%. The memory usage displayed on the GUI is below the HPA threshold, but no scale-in action has taken place.

Possible Causes

The way the container memory usage is calculated on the GUI differs from the method used for HPA auto scaling.

- The memory usage of a container displayed on the GUI:
container_memory_rss/Memory limit of the container
container_memory_rss specifies the resident set size (RSS). It includes some memory parts that may not be actively or effectively used.
- The memory usage of a container calculated for HPA auto scaling:
container_memory_working_set_bytes/Memory request of the container
container_memory_working_set_bytes specifies the working set size (WSS) and is calculated by doing as follows:

Run **cat /sys/fs/cgroup/memory/memory.stat** in the pod to obtain the values of **total_cache** (cache memory), **total_rss** (memory used by the current application process), and **total_inactive_file** (memory used by inactive files).

WSS = The value of total_cache + The value of total_rss - The value of total_inactive_file

If an application's memory usage displayed on the GUI is below the HPA scale-in threshold, but no scale-in action has been taken, or if the memory usage is not higher than the HPA scale-out threshold, but a scale-out happened anyways, the HPA scale-out behavior may not work as expected. This can occur when:

- The application cache usage is high, which can cause the WSS to be significantly greater than the RSS, resulting in the container memory usage displayed on the GUI being lower than the memory usage calculated by the HPA.
- The difference between the resource limit and request is large. The request may be significantly lower than the limit, causing the container memory

usage shown on the GUI to be lower than the memory usage calculated by the HPA.

3 Network Management

3.1 How Do I Configure Security Group Rules for a Cluster?

When a CCE Autopilot cluster is created, two security groups are automatically created, one for master nodes, and the other for elastic network interfaces (ENIs). The security group for master nodes is named in the format of *{Cluster name}-cce-control-{Random ID}*, and that for ENIs is in the format of *{Cluster name}-cce-eni-{Random ID}*.

You can modify the security group rules on the VPC console as required. (Log in to the management console, choose **Service List > Networking > Virtual Private Cloud**. On the page displayed, choose **Access Control > Security Groups** in the navigation pane, locate the target security groups, and modify their rules.)

NOTICE

- Modifying or deleting default rules in a security group may affect cluster running. If you need to modify security group rules, do not modify the rules of the port that CCE running depends on.
 - When adding a security group rule, ensure that **this rule does not conflict with the existing rules**. If there is a conflict, existing rules may become invalid, affecting cluster running.
-

Security Group for Master Nodes

The security group automatically created for master nodes is named *{Cluster name}-cce-control-{Random ID}*. [Table 3-1](#) lists the default ports in the security group.

Table 3-1 Default ports in the security group of the master nodes

Direction	Port	Source	Description	Modifiable	Modification Suggestion
Inbound	All	IP addresses of this security group	Allow traffic from all IP addresses in this security group	No	None
Outbound	All	All IP addresses: 0.0.0.0/0 or ::/0	Allow traffic on all ports by default.	No	None

Security Group for ENIs

When a CCE Autopilot cluster is created, a security group named *{Cluster name}-cce-eni-{Random ID}* is automatically created for ENIs. By default, pods in the cluster are associated with this security group. [Table 3-2](#) lists the default ports in the security group.

Table 3-2 Default ports in the security group for ENIs

Direction	Port	Source	Description	Modifiable	Modification Suggestion
Inbound	All	IP addresses of this security group	Allow traffic from all IP addresses in this security group	No	None
		CIDR block of the master nodes	Allow the master nodes to access kubelet on each worker node, for example, by running <code>kubectl exec {Pod}</code> .	No	None

Direction	Port	Source	Description	Modifiable	Modification Suggestion
Outbound	All	All IP addresses: 0.0.0.0/0 or ::/0	Allow traffic on all ports by default.	Yes	If you want to harden security by allowing traffic over specific ports, you can modify the rule to allow these ports. For details, see Hardening Outbound Rules for the Security Group of ENIs .

Hardening Outbound Rules for the Security Group of ENIs

By default, all ENI security groups created by CCE Autopilot allow all outbound traffic. You are advised to retain this configuration. If you want to harden security by allowing traffic over specific ports, configure the ports listed in the following table.

Table 3-3 Minimum scope for outbound rules in an ENI security group

Port	Allowed CIDR Block	Description
All	IP addresses of this security group	Allow mutual access within the security group so containers can communicate with each other.
TCP port 5443	VPC CIDR block	Allow access from kube-apiserver, which provides lifecycle management for Kubernetes resources.
TCP port 443	100.125.0.0/16	Access the OBS port or SWR port to pull images.
UDP port 53	100.125.0.0/16	Allow traffic over the port for DNS resolution.
TCP port 443	VPC CIDR block	Pull the images through the SWR endpoint.
All	198.19.128.0/17	Allow worker nodes to access the VPC Endpoint (VPCEP) service.
TCP port 9443	VPC CIDR block	Allow the network add-ons of the worker nodes to access master nodes.

3.2 How Can I Check Whether a Network Interface Is Used by a Cluster?

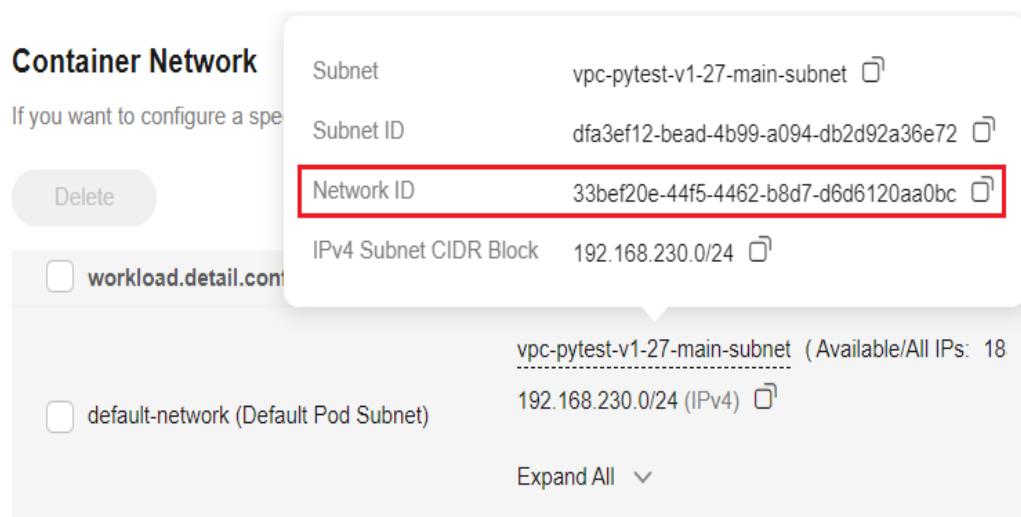
Scenario

For CCE Autopilot v1.27.8-r0, v1.28.6-r0, and later versions, container subnets can be deleted.

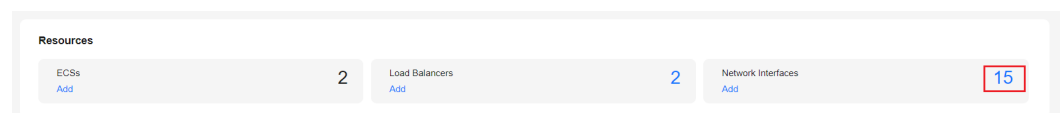
Deleting a cluster container subnet is a high-risk operation. Ensure that the network interfaces used by the cluster are not associated with the subnet.

Procedure

- Step 1** Log in to the [CCE console](#) and click the cluster name in the cluster list.
- Step 2** In the navigation pane, choose **Settings**. Then click the **Network** tab.
- Step 3** In the **Container Network** area, copy the network ID of the subnet. (The **default-network** subnet is used as an example.)



- Step 4** Log in to the [Network Console](#). In the navigation pane, choose **Virtual Private Cloud > Subnets**. Search for the container subnet using the network ID.
- Step 5** Click the subnet name to go to the **Summary** tab. In the **Resources** area, click the number on the **Network Interfaces** tab and view the network interfaces and supplementary network interfaces associated with the subnet.



- Step 6** View the network interface name or description. If the name or description contains the cluster ID, the network interface is used by the cluster. You can view the cluster ID on the **Overview** page of the cluster on the [CCE console](#).

To clear the network interfaces used by the cluster, [submit a service ticket](#).

Network Interfaces		Supplementary Network Interfaces	
Subnet ID: 33bef20e-44f5-4462-b8d7-d6d6120aa0bc X Add filter			
Private IP Address		Name	
192.168.230.221		yangtse_58ea42d5-5880-11ef-825c-0255ac100b0c_109c69f4-b07c-4de4	
192.168.230.87		yangtse_58ea42d5-5880-11ef-825c-0255ac100b0c_93129873-6d16-40d3	

----End

4 Storage

4.1 Can PVs of the EVS Type in a CCE Autopilot Cluster Be Restored After They Are Deleted or Expire?

You need to manually configure backup policies for EVS disks. If an EVS disk is deleted or expires, you can use the VBS backup to restore data.

For details, see [Backing Up EVS Disks](#).

4.2 What Can I Do If a Storage Volume Fails to Be Created?

Symptom

A PV or PVC fails to be created. The following information is displayed in the event:

```
{"message": "Your account is suspended and resources can not be used.", "code": 403}
```

Possible Causes

The event indicates that your account is suspended or permissions are not granted to the account. Check whether your account is normal.

If the account is normal, check whether you have the permissions to access the namespace. You must have one of the development, O&M, and administrator permissions of the namespace, or have the permission to read and write PVCs and PVs. For details, see [Configuring Namespace Permissions \(on the Console\)](#).

4.3 Can CCE Autopilot PVCs Detect Underlying Storage Faults?

No. CCE Autopilot PersistentVolumeClaims (PVCs) are implemented as they are in Kubernetes. A PVC is defined as a storage declaration and is decoupled from underlying storage. It is not responsible for detecting underlying storage details.

Cloud Eye allows you to view cloud service metrics. These metrics are built in based on cloud service attributes. After you enable a cloud service on the cloud platform, Cloud Eye automatically associates its built-in metrics. You can track the cloud service status by monitoring these metrics.

It is recommended that you use Cloud Eye to monitor underlying storage and send alarm notifications when there are storage faults.

4.4 How Can I Delete the Underlying Storage If It Remains After a Dynamically Created PVC is Deleted?

Symptom

After a dynamically created PVC with the reclaim policy in its StorageClass set to **Delete** was deleted from a cluster, the underlying storage volume of the PVC is not deleted simultaneously.

Trigger Conditions

- A PVC and its bound PV are deleted simultaneously.
- The PV bound to a PVC is deleted first and then the PVC is deleted, but the PV deletion fails due to the PVC/PV binding.

Possible Causes

Under normal circumstances, when a dynamically created PVC is deleted in the open-source csi-provisioner module, the PVC is deleted first, followed by a change in the status of the PV bound to the PVC to **Released**. The csi-provisioner module then listens for PV changes, proceeds to delete the underlying storage volume, and deletes the PV, completing the deletion chain.

During abnormal operations, the PV bound to a PVC may be directly deleted without deleting the PVC first. However, the **kubernetes.io/pv-protection** finalizer on the PV prevents immediate deletion. Instead, **deletionTimestamp** is added to the PV. After the PVC is deleted, the PV status is changed to **Released**. Although csi-provisioner listens for PV changes, it skips the process of deleting the underlying storage volume because **deletionTimestamp** has been added to the PV. As a result, csi-provisioner directly deletes the PV, and both the PVC and PV are deleted, but the underlying storage volume remains. For details about the code logic, see [controller](#).

Solution

1. Manually delete the residual underlying storage volumes.
2. Directly delete the dynamically created PVCs that remained after the deletion. The PVs and underlying storage volumes will be deleted automatically.

5 Permissions

5.1 Can I Configure Only Namespace Permissions Without Cluster Management Permissions?

Namespace permissions and cluster management permissions are independent and complementary to each other.

- Namespace permissions: apply to clusters and are used to manage operations on cluster resources (such as creating workloads).
- Cluster management (IAM) permissions: apply to cloud services and used to manage CCE Autopilot clusters and peripheral resources (such as VPC, ELB, and ECS).

Administrators of the IAM Admin user group can grant cluster management permissions (such as CCE Administrator and CCE FullAccess) to IAM sub-users or grant namespace permissions for a cluster on the CCE console. However, the permissions you have on the CCE console are determined by the IAM system policy. If the cluster management permissions are not configured, you do not have the permissions to access the CCE console.

If you only run `kubectl` commands to operate cluster resources, you only need to obtain the `kubeconfig` file with the namespace permission. For details, see [Can I Use APIs If the Cluster Management Permissions Are Not Configured?](#). Note that information leakage may occur when you use the `kubeconfig` file.

5.2 Can I Use APIs If the Cluster Management Permissions Are Not Configured?

CCE Autopilot has cloud service APIs and cluster APIs.

- Cloud service APIs: You can perform operations on the infrastructure (for example, creating nodes) and allow you to perform operations on cluster-level resources (such as creating workloads).

When using cloud service APIs, the cluster management (IAM) permission must be configured.

- Cluster APIs: You can perform operations on cluster-level resources (such as creating workloads) using the native API Server of Kubernetes. You cannot perform operations on the basic infrastructure (such as creating nodes).
When using cluster APIs, you only need to add the cluster certificate. Only the users with the cluster management (IAM) permission can **download** the cluster certificate. Note that information leakage may occur during certificate transmission.

5.3 Can I Use kubectl If the Cluster Management Permissions Are Not Configured?

IAM authentication is not required for running kubectl commands. Therefore, you can run kubectl commands without configuring cluster management (IAM) permissions. The prerequisite is that the kubectl configuration file (kubeconfig) with the namespace permissions needs to be obtained. In the following scenarios, information leakage may occur during file transmission.

- Scenario 1
If an IAM user has been configured with the cluster management permissions and namespace permissions, downloads the kubeconfig authentication file and then deletes the cluster management permissions (reserving the namespace permissions), kubectl can still be used to perform operations on Kubernetes clusters. If you want to permanently delete the user's permissions, you must also delete the cluster management permissions and namespace permissions of the user.
- Scenario 2
An IAM user has certain cluster management and namespace permissions and downloads the kubeconfig authentication file. CCE Autopilot determines which Kubernetes resources can be accessed by kubectl based on the user information. Essentially, the user's authentication information is stored in kubeconfig, which can be used by anyone to access the cluster.

5.4 Why Can't an IAM User Make API Calls?

Symptom

When an IAM user makes an API call, an error message similar to the following is displayed:

```
"code":403,"message": "This user only supports console access, not programmatic access."
```

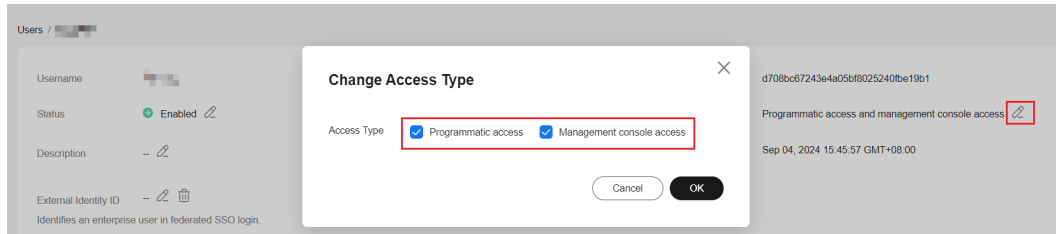
This error message indicates that the IAM user does not have programmatic access permissions.

Solution

- Step 1** Contact the account administrator and log in to the [IAM console](#).
- Step 2** Locate the IAM user to be modified and click the username.

Step 3 Change the access mode and select both **Programmatic access** and **Management console access**.

Figure 5-1 Changing the access mode of an IAM user



Step 4 Confirm the configuration.

----End