Auto Scaling

FAQ

Issue 01

Date 2020-11-27

HUAWEI TECHNOLOGIES CO., LTD.

Trademarks and Permissions


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.

Contents

1 General

1.1 What Are Restrictions on Using AS?

Only applications that are stateless and can be horizontally scaled can run on instances in an AS group. AS automatically releases ECS instances. Therefore, the instances in AS groups cannot be used to save application status information (such as session statuses) and related data (such as database data and logs).

If the application status or related data must be saved, you can store the information on separate servers.

Table 1-1 lists the AS service resource quotas.

Table 1-1 Quota list

Category	Description	Default Value
AS group	Maximum number of AS groups that you can create	10
AS configuration	Maximum number of AS configurations that you can create	100
AS policy	Maximum number of AS policies that can be added to an AS group	10
Instance	Maximum number of instances that can be added to an AS group	300
Bandwidth scaling policy	Maximum number of bandwidth scaling policies that you can create	10

1.2 Are ELB and Cloud Eye Mandatory for AS?

AS can work independently or work together with **ELB** and **Cloud Eye**.

Cloud Eye does not require additional fees and is enabled by default. You can enable the ELB service when required. For example, if distributed clusters are required, you can enable the ELB service.

1.3 Is AS Billed?

AS is free of charge. The pay-per-use instances automatically created in an AS group are billed. For pricing details, see **Billing** EIPs used by the instances are also billed. When the capacity of an AS group is reduced, the automatically created instances will be removed from the AS group and be deleted. After the deletion, these instances are no longer billed. The instances manually added are removed from the AS group but still billed. If you do not need the instances, unsubscribe instances on the ECS console.

For example, two instances are created when an AS group scales out. The AS group scales in after an hour. The two instances are removed from the AS group and are billed for the one-hour usage.

1.4 Does an Abrupt Change on Monitoring Indicator Values Cause an Incorrect Scaling Action?

No. Monitoring data used by **AS** is from **Cloud Eye**. The monitoring interval of Cloud Eye can be set to 5 minutes, 20 minutes, or 1 hour. Therefore, an abrupt change of monitoring indicator values will not cause an incorrect scaling action.

In addition, AS allows you to configure the cooldown period to prevent unnecessary scaling actions caused by frequently reported alarms. You can customize the cooldown period.

1.5 How Many AS Policies and AS Configurations Can I Create and Use?

You can create up to 10 AS groups and 100 AS configurations by default. An AS group supports 1 AS configuration and 10 AS policies at a time.

If the default configurations fail to meet your service requirements, contact the administrator.

1.6 Can AS Automatically Scale Up or Down vCPUs, Memory, and Bandwidth of ECSs?

Currently, AS can only scale up bandwidth.

1.7 What Is the AS Quota?

What Is Quota?

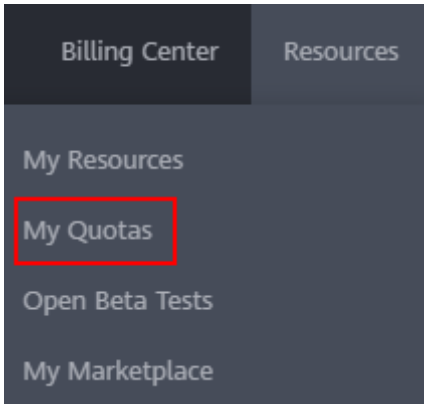
Quotas are enforced for service resources on the platform to prevent unforeseen spikes in resource usage. Quotas can limit the number or amount of resources available to users, for example, how many AS groups you can create. You can apply for increasing quotas if necessary.

This section describes how to view the usage of each type of AS resource and the total quotas in a specified region.

How Do I View My Quotas?

1. Log in to the management console.
2. Click in the upper left corner and select the desired region and project.
3. In the upper right corner of the page, choose **Resources** > **My Quotas**.
   The **Service Quota** page is displayed.

   Figure 1-1 My Quotas

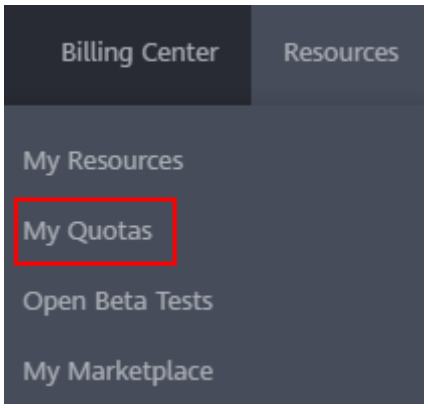
   

4. View the used and total quota of each type of resources on the displayed page.
   If a quota cannot meet service requirements, apply for a higher quota.

How Do I Apply for a Higher Quota?

1. Log in to the management console.
2. In the upper right corner of the page, choose **Resources** > **My Quotas**.
   The **Service Quota** page is displayed.

Figure 1-2 My Quotas



3. Click **Increase Quota**.
4. On the **Create Service Ticket** page, configure parameters as required.

   In **Problem Description** area, fill in the content and reason for adjustment.
5. After all necessary parameters are configured, select **I have read and agree to the Tenant Authorization Letter and Privacy Statement** and click **Submit**.

1.8 Why is a message displayed indicating that the key pair does not exist and the operation is discontinued when several users under the same account operate AS resources?

A key pair cannot be used by multiple users. If the key pair of another user under the same account is configured in the AS configuration, the AS configuration cannot be used to manually provision resources.

If users need to perform operations on others' AS configuration resources without being restricted by the key pair permission, use password authentication for instances.

1.9 What Are the Requirements of AS for the EIP and Shared Bandwidths?

AS can only adjust the bandwidth of pay-per-use EIPs and shared bandwidths, and cannot adjust yearly/monthly bandwidths.

1.10 What Types of ECSs Do AS Support? Does AS Support HECSs?

Supported ECS types are listed on the AS management console. For details about ECS types, see ***Elastic Cloud Server User Guide***.

Currently, AS does not support HECSs.

2 AS Group

2.1 What Can I Do If the AS Group Fails to Be Enabled?

See section "How Can I Handle an AS Group Exception?"

2.2 How Can I Handle an AS Group Exception?

The handling method varies depending on the possible cause.

- Issue description: Insufficient quota for ECSs, EVS disks, or EIPs.

  Possible cause: insufficient quota

  Handling method: Increase the quota or delete unnecessary resources, and then enable the AS group.
- Issue description: The VPC or subnet does not exist.

  Possible cause: The VPC service encounters an exception or resources have been deleted.

  Handling method: Wait until the VPC service recovers, or modify parameters of the VPC and subnet in the AS group, and then enable the AS group.
- Issue description: The ELB listener or backend ECS group does not exist, and the load balancer is unavailable.

  Possible cause: The ELB service encounters an exception or resources have been deleted.

  Handling method: Wait until the ELB service recovers, or modify load balance parameters in the AS group, and then enable the AS group.
- Issue description: The number of backend ECSs that you add to the ELB listener exceeds the upper limit.

  Possible cause: If classical load balancer is used by an AS group, instances added to the AS group are automatically added to the ELB listener. A maximum of 300 backend ECSs can be added to an ELB listener.

  Handling method: Remove the backend ECSs that are both not required and not in the AS group from the listener. Then enable the AS group.
- Issue description: The image used by the AS configuration, the flavor, or the key pair does not exist.

Possible cause: Resources have been deleted.

Handling method: Change the AS configuration for the AS group and then enable the AS group.

- Issue description: The notification subject configured for your lifecycle hook does not exist.

  Possible cause: The AS group adds a lifecycle hook, while its configured notification subject has been deleted before the scaling action starts. If the notification subject is deleted after the scaling action starts, an AS group exception will occur in the next scaling action.

  Handling method: Change the notification subject used by the lifecycle hook or delete the lifecycle hook. Then enable the AS group.

- Issue description: The subnet you select does not have enough private IP addresses.

  Possible cause: Private IP addresses in the subnet used by the AS group have been used up.

  Handling method: Modify the subnet information and enable the AS group.

- Issue description: The ECS resources of this type in the selected AZ have been sold out.

  Possible cause: ECSs of this type have been sold out or are not supported in the AZ selected for the AS group. ECSs of this type are the ECS flavor selected in the AS configuration.

  Handling method: Change the AS configuration for the AS group and then enable the AS group. If there is no instance in the AS group, you can also change the AZ for the AS group and then enable the AS group.

- Issue description: The selected specifications and the disk do not match.

  Possible cause: The ECS type in the AS configuration does not match the disk type, leading to the ECS creation failure.

  Handling method: Change the AS configuration for the AS group and then enable the AS group.

- Issue description: The selected specifications and the image do not match.

  Possible cause: The ECS type in the AS configuration does not match the image, leading to the ECS creation failure.

  Handling method: Change the AS configuration for the AS group and then enable the AS group.

- Issue description: Storage resources of this type have been sold out in the selected AZ.

  Possible cause: Storage resources of this type have been sold out or are not supported in the AZ selected for the AS group. Storage resources of this type refer to the system and data disk types selected for the AS configuration.

  Handling method: Change the AS configuration for the AS group and then enable the AS group. If there is no instance in the AS group, you can also change the AZ for the AS group and then enable the AS group.

- Issue description: The shared bandwidth defined in the AS configuration does not exist.

  Possible cause: Resources have been deleted.

Handling method: Use a newly purchased or an existing shared bandwidth to create an AS configuration. Then change the AS configuration for the AS group and enable the AS group.

- Issue description: The number of EIPs bound to the shared bandwidth specified in the AS configuration exceeds the limit.

  Possible cause: A maximum of 20 EIPs can be bounded to a shared bandwidth.

  Handling method: Apply for a higher EIP quota, remove unnecessary EIPs from the shared bandwidth, or change another AS configuration for the AS group. Then enable the AS group.
- Issue description: The DeH selected in your AS configuration does not exist. Change the AS configuration.

  Possible cause: Resources have been deleted.

  Handling method: Use a newly purchased or an existing DeH to create an AS configuration. Then change the AS configuration for the AS group and enable the AS group.
- Issue description: No DeH is available. Ensure that there are available DeH resources.

  Handling method: Rectify the DeH fault and restore the DeH to the available state, or enable the automatic placement attribute for the DeH, and then enable the AS group again. You can also use a newly purchased DeH to create an AS configuration, change the AS configuration for the AS group, and enable the AS group.
- Issue description: The DeH selected in your AS configuration does not have sufficient capacity.

  Handling method: You can delete unnecessary ECSs from the DeH and enable the AS group again. You can also use a newly purchased DeH to create an AS configuration, change the AS configuration for the AS group, and enable the AS group.
- Issue description: No DeH is available in the AZ selected for your AS group.

  Handling method: Purchase a DeH in the AZ, use it to create an AS configuration, change the AS configuration for the AS group, and enable the AS group. If there is no instance in the AS group, change the AZ for the AS group and then enable the AS group.
- Issue description: The DeH selected in your AS configuration does not support this type of ECS. Change the AS configuration.

  Handling method: Select the ECS specifications supported by the DeH, create an AS configuration, change the AS configuration for the AS group, and then enable the AS group again.
- Issue description: A system error has occurred.

  Possible cause: An error has occurred in the AS service, peripheral service, or network.

  Handling method: Try again later or contact technical support.
- Issue description: The specification defined in the AS configuration is unavailable.

  Handling method: Change specifications by creating an AS configuration as prompted by the error message and use this AS configuration for the AS group. Then enable the AS group.

- Issue description: The selected AS configuration cannot be used by the AS group.

  Handling method: Create an AS configuration as prompted by the error message and use this AS configuration for the AS group. Then enable the AS group.
- Issue description: AS group expansion fails.

  Possible cause: Your account is in arrears or the balance is insufficient.

  Handling method: Top up your account and enable the AS group.

2.3 What Operation Will Be Suspended After An AS Group Is Disabled?

After an AS group is disabled, the group will not automatically any trigger scaling actions, but the on-going scaling action will continue. Scaling policies will not trigger any scaling actions. After you manually change the number of expected instances, no scaling action is triggered although the number of actual instances is not equal to that of expected instances.

The health check continues to check the health status of the instances but does not remove the instances.

2.4 Can I Use an ECS ID to Obtain the AS Group Accommodating the ECS?

No.

To obtain details about an AS group and the ECSs in the group, perform the following operations:

Step 1 Log in to the management console. Choose **Computing** > **Auto Scaling** > **Instance Scaling**.

Step 2 On the **AS Groups** tab page, click the name of the target AS group.

Step 3 Click the **Instances** tab to view the instances in the AS group.

----End

3 AS Policy

3.1 How Many AS Policies Can Be Enabled?

Enable one or more AS policies as required.

3.2 What Are the Conditions to Trigger an Alarm in the AS Policy?

Alarms will be triggered by metrics of CPU Usage, Memory Usage, Inband Incoming Rate, Inband Outgoing Rate, Disk Read Rate, Disk Write Rate, Disk Read Requests, and Disk Write Requests. These alarms will in turn trigger the policy to increase or decrease instances in the AS group.

3.3 What Is a Cooldown Period? Why Is It Required?

A cooldown period is a period of time after each scaling action is complete. During the cooldown period, scaling actions triggered by alarms will be denied. Scheduled and periodic scaling actions are not restricted.

Before an instance is added to the AS group, it requires 2 to 3 minutes to execute the configuration script to install and configure applications. The time varies depending on many factors, such as the instance specifications and startup scripts. Therefore, if an instance is put into use without cooldown after started, the system will continuously increase instances until the load decreases. After the new instances take over services, the system detects that the load is too low and decreases instances in the AS group. A cooldown prevents the AS group from repeatedly triggering unnecessary scaling actions.

The following uses an example to introduce the cooling principles:

When a traffic peak occurs, an alarm policy is triggered. In this case, AS automatically adds an instance to the AS group to help handle the added demands. However, it takes several minutes for the instance to start. After the instance is started, it takes a certain period of time to receive requests from ELB. During this period, alarms may be triggered continuously. As a result, an instance

is added each time an alarm is triggered. If you set a cooldown time, after an instance is started, AS stops adding new instances according to the alarm policy until the specified period of time (300 seconds by default) passes. Therefore, the newly started instance has time to start processing application traffic. If an alarm is triggered again after the cooldown period elapses, AS starts another instance and the cooldown period takes effect again.

3.4 Can AS Scale Capacity Based on Custom Monitoring of Cloud Eye?

Yes. AS can scale capacity based on custom monitoring of Cloud Eye.

3.5 What Will Monitoring Metrics for an AS Group Be Affected If VM Tools Are Not Installed on ECSs?

If VM tools have not been installed on ECSs, Cloud Eye can monitor the Outband Incoming Rate and Outband Outgoing Rate. However, it cannot monitor the Memory Usage, Inband Incoming Rate, and Inband Outgoing Rate, which reduces data accuracy of the CPU usage.

If the ECSs are of I/O-optimized type, Cloud Eye cannot monitor the disk usage, inband incoming rate, and inband outgoing rate metrics of ECSs, regardless of whether VM tools are installed.

If VM tools are not installed on ECSs, AS cannot obtain the memory usage, inband incoming rate, and inband outgoing rate.

3.6 What Can I Do If an AS Policy Fails to Be Enabled?

- Description: The alarm rule does not exist.

  Possible cause: The alarm rule used in the alarm policy is deleted.

  Handling method: Change the alarm rule used in the alarm policy and enable the AS policy again.
- Description: The triggering time of the periodic policy falls outside the effective time range of the policy.

  Possible cause: The effective time of the periodic policy has expired.

  Handling method: Change the start time and end time of the periodic policy and enable the policy again.
- Description: The triggering time of the scheduled policy must be later than the current time.

  Possible causes: The triggering time of the scheduled policy has expired.

  Handling method: Change the triggering time of the scheduled policy and enable the policy again.
- Description: A system error has occurred.

  Handling method: Try again later or contact technical support.

3.7 How Can I Install the Agent Plug-in on the Instances in an AS Group to Use Agent Monitoring Metrics?

Scenarios

When the scaling policy of an AS group is alarm-triggered, you can use Agent monitoring metrics to trigger scaling actions. Compared with basic monitoring, Agent monitoring provides system-level, proactive, and fine-grained monitoring services for instances. Before using Agent monitoring metrics, make sure that the Agent plug-in has been installed on the instances in the AS group. For details, see this section.

Procedure

1. Log in to the management console and click **Elastic Cloud Server** under **Computing**.

   The **Elastic Cloud Server** page is displayed.

2. Create an ECS and install the Agent plug-in.

   To install the plug-in, see **Agent Installation and Configuration**.

3. After installing the Agent plug-in, log in to the Cloud Eye console, choose **Server Monitoring** > **Elastic Cloud Server**, and ensure that the plug-in is running and that the Agent monitoring metrics can be collected.

   Figure 3-1 Checking the plug-in status

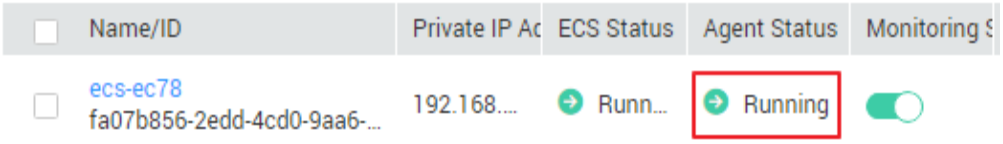
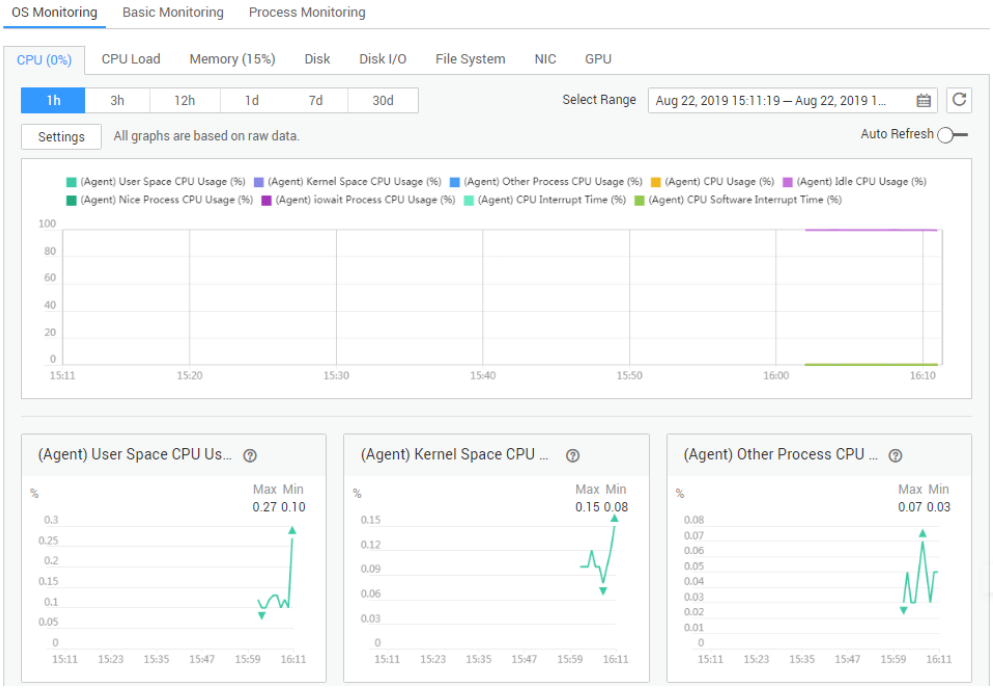


Figure 3-2 Viewing Agent monitoring metrics

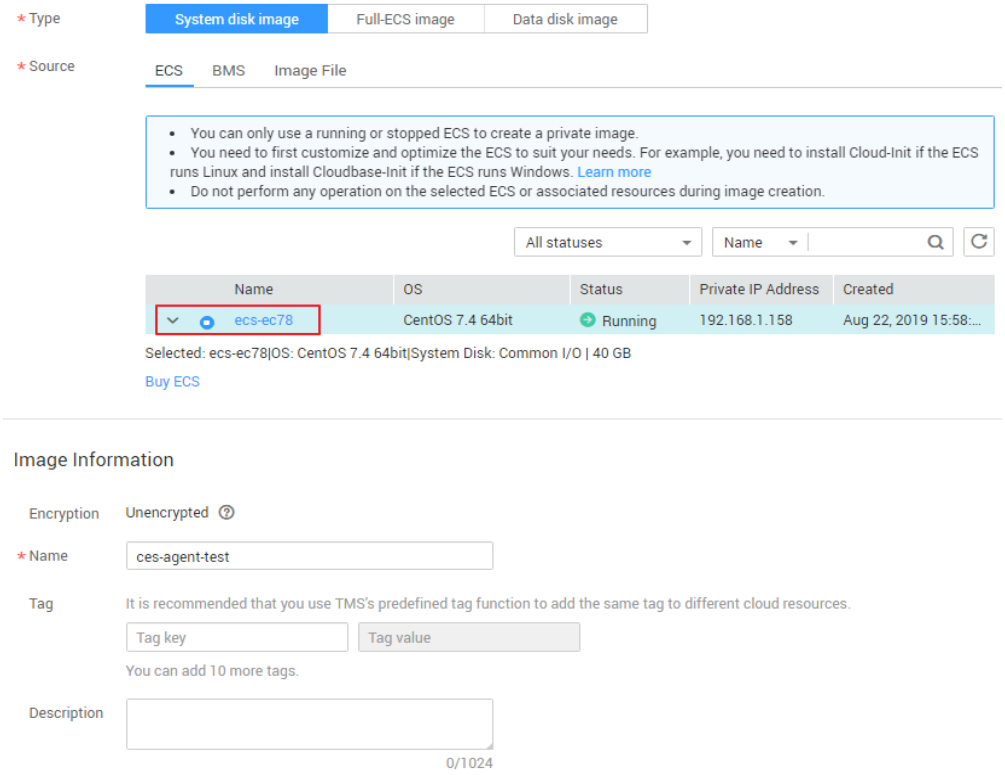


4. Add the AK/SK to the ECS **conf.json** configuration file.
   a. Click the username, choose **My Credentials** > **Access Keys**, obtain the AK/SK.
      - If you have obtained the access key, obtain the AK/SK in the **credentials.csv** file saved when you created **Access Keys**.
      - If **Access Keys** is not available, click **Create Access Key** to create one. Then, save the **credentials.csv** file and obtain the AK/SK in it.
   b. Log in to the ECS and run the **cd /usr/local/telescope/bin** command to go to the Agent installation directory.
   c. Run the **vi conf.json** command to open the configuration file and enter the obtained AK/SK.

```
{
  "InstanceId":"fa07b      4cd0-9aa6-e5c791569e3a",
  "ProjectId": "050b1      572f8cc01f3740bed5",
  "AccessKey": "MK8NR3      7FUMUB",
  "SecretKey": "sPHiTB8                N4wWv3YCNwcUFqj",
  "RegionId": "cn-north-1"
}
```

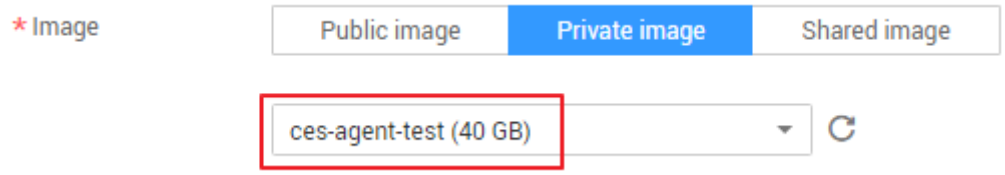
   If the Agent has been installed during ECS creation, the AK/SK has been added during user data injection. You only need to check the AK/SK in this step.
   d. Press **Esc** and enter **:wq** to save and exit the configuration file.
5. Go to the **Image Management Service** page and use this ECS to create a private image. For details, see **Creating a Private Image**.

Figure 3-3 Creating a private image



6. Go to the **Auto Scaling** page and use the private image created in step **5** to create an AS configuration.

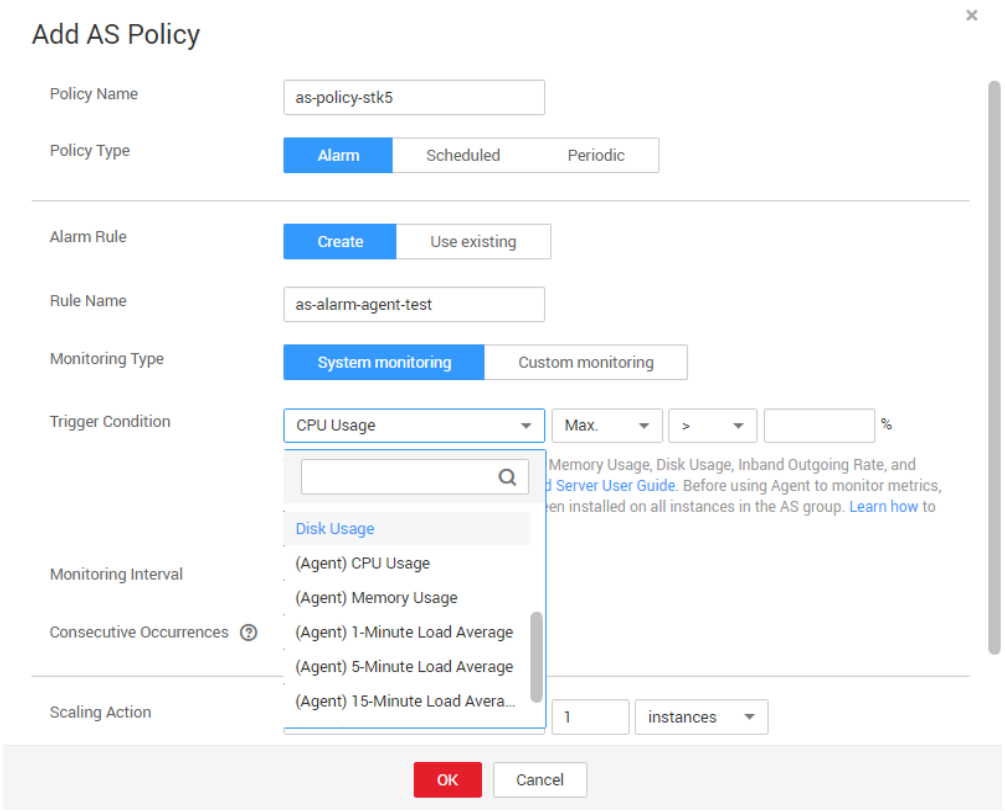
Figure 3-4 Selecting the private image



Click **Private image** for **Image**, select **ces-agent-test** from the drop-down list, and set other parameters as required.

7. Create an AS group and associate the AS configuration created in step **6** with the AS group.
8. Add an AS policy for the AS group: Set **Policy Type** to **Alarm** and **Trigger Condition** to an Agent monitoring metric, such as **(Agent) Memory Usage**.

Figure 3-5 Selecting a trigger condition



9. Manually add the ECS created in step 2 to the AS group.
10. Perform the following operations to check whether the Agent monitoring metric takes effect:
    - Verify that the Agent monitoring metric is displayed on the **Monitoring** tab of the page providing details about the AS group.
    - When the alarm threshold is reached, verify that the alarm policy is triggered on the **Scaling Actions** tab of the page providing details about the AS group and that instances are added for capacity expansion.
    - The Agent monitoring data is available for all instances that are automatically added to the AS group.

4 Instance

4.1 How Do I Prevent Instances Manually Added to an AS Group from Being Removed Automatically?

If you have manually added N instances into an AS group and do not want these instances to be removed automatically, you can use either of the following methods to ensure this:

Method 1

Perform following configurations in the AS group:

- Set the minimum number of instances in the AS group to N or greater than N.
- Set **Instance Removal Policy** to **Oldest instance created from oldest AS configuration** or **Newest instance created from oldest AS configuration**.

Based on the scaling rules, the manually added instances do not correspond to any AS configuration (because they are not created using the AS configuration). Therefore, the instances automatically created using the AS configuration are removed at first. Only when such instances are removed, the instances manually added are removed. Since you have set the minimum number of instances to N or greater than N, the instances manually added are not selected.

Note: If the instances manually added are stopped or they malfunction, they are regarded as unhealthy and removed from the AS group. This is because health check ensures that instances in the AS group must be healthy.

Method 2

Enable instance protection for the N instances. For details, see **Instance Protection**.

You can enable instance protection for the N instances at the same time. When an AS group reduces the capacity, protected instances will not be removed from the AS group. Note: Instances that fail to pass a health check will still be removed from the AS group.

4.2 What Are the Sequence of Selecting Flavors in Multi-Flavor AS Configuration?

When multiple flavors are selected in an AS configuration, the sequence varies according to the extension policies of a single AZ and multiple AZs. This section describes the sequence of selecting flavors in a single AZ and multiple AZs.

Single AZ

If only one AZ is selected for an AS group, all instances in the AS group are created in the AZ. If multiple flavors are selected in the AS configuration, there are two flavor selection policies:

- Sequenced: During AS group expansion, flavors are used based on the sequence they are selected. For example, you have selected flavors 2, 3, and 1 in sequence. The system selects flavor 2 at first. If flavor 2 is insufficient or an instance fails to be created due to other reasons, the system selects flavor 3. Flavor 1 is used only when flavor 2 and 3 cannot be used.
- Cost-center: During AS group expansion, the flavor with the minimum cost comes first. For example, you select flavors 1, 2, and 3 in sequence. The costs of flavors 1, 3, and 2 are in descending order. The system preferentially selects flavor 2 (with the minimum cost) to create an instance. When flavor 2 fails, select flavor 3. When flavor 3 also fails, flavor 1 is used.

Multiple AZs

When two or more AZs are selected for an AS group, you need to configure the **Multi-AZ Extension Policy** (load-balanced or sequenced). When you select different multi-AZ extension policies, the sequence of creating instances is also different. The sequence is described as follows:

- Load-balanced: When expanding an AS group, preferentially ensure that ECSs are evenly distributed in AZs. If it fails in the target AZ, the system automatically selects another AZ based on the sequenced policy. The following is an example of selecting AZs and flavors:

  You have selected AZ 1, AZ 2, and AZ 3 in sequence and flavors 1, 2, and 3. The priority sequence of the flavors is 2, 3, and 1. AZ 1, AZ 2, and AZ 3 have 3, 2, and 3 instances respectively. According to the load-balanced policy, AZ 2 has fewer instances and is therefore preferentially selected to create instances. Use flavor 2 to create an instance in AZ 2. If the instance is successfully created, the scaling action is successful. If flavor 2 fails, use flavor 3, and so on. If all of them fail, the instance cannot be created in AZ 2. If instances cannot be created using load-balanced policy, select other AZs based on the sequenced policy and try to create instances using flavors 2, 3, and 1 in AZ 1. If ECSs still cannot be created in AZ 1, AZ 3 is selected and the sequence of selecting flavors is also 2, 3, and 1.
- Sequenced: When expanding the ECS capacity, the target AZ is used based on the order in which AZs are selected. The following is an example of selecting AZs and flavors:

  You have selected AZ 1, AZ 2, and AZ 3 in sequence and flavors 1, 2, and 3. The priority sequence of the flavors is 2, 3, and 1. No matter whether

instances in three AZs are evenly distributed, the system creates instances in sequence, that is, AZ 1, AZ 2, and AZ 3. Use flavor 2 to create an instance in AZ 1. If this fails, use flavor 3. Flavor 1 is used when both flavors 2 and 3 fail. If all three flavors fail in AZ 1, AZ 2 is selected. The sequence of flavors is also 2, 3, and 1. Similarly, if AZ 2 also fails, AZ 3 is selected. The sequence of flavors is also 2, 3, and 1.

NOTE

The priority sequence of the flavors is determined by the flavor selection policy in AS configuration. For details, see **Single AZ**.

4.3 Will the Application Data on an Instance Be Retained After the Instance Is Removed from an AS Group and Deleted?

No. AS automatically releases ECS instances. You must ensure that instances in the AS group do not store application status information or important data, such as sessions, databases, and logs. If you want to store your application status, you can store it on an independent server (such as ECS) or database (such as RDS database).

If you want to back up data or download log files before an instance is removed from an AS group, you can add a lifecycle hook of the instance removal type to the AS group. After a lifecycle hook is added to an AS group, when the AS group performs a scaling action, the lifecycle hook suspends the instance that is being removed from the AS group and sets the instance to be in waiting state. During the waiting period, you can perform customized operations on the instance, for example, to back up data or to download log files.

4.4 Can I Add ECSs Charged in Yearly/Monthly Mode?

Yes. Currently, AS automatically creates pay-per-use ECSs by default. In addition, you can manually add ECSs charged in yearly/monthly or pay-per-use mode.

4.5 Can Instances That Have Been Added Based on an AS Policy Be Automatically Deleted When They Are Not Required?

Yes. They can be automatically deleted if one AS policy has been added to trigger scaling actions to delete the ECS.

4.6 What Is the Expected Number of Instances?

The expected number of instances refers to the number of ECSs that are expected to run in an AS group. It is between the minimum number of instances and the maximum number of instances. You can manually change the expected number of instances or change it based on the scheduled, periodic, or alarm policies.

You can set this parameter when creating an AS group. If this value is greater than 0, a scaling action is performed to add the required number of ECSs after the AS group is created. You can also change this value manually or by scaling policies after the AS group is created.

If you manually change this value, the current number of ECSs is not consistent with the expected number. A scaling action is performed to adjust the number of ECSs to the expected number.

If a scaling policy is triggered to add two ECSs to the AS group, the system will add two to the expected number. Then, a scaling action is performed to add two ECSs so that the number of ECSs in the AS group is the same as the expected number.

4.7 How Do I Delete an ECS Created in a Scaling Action?

Handling Methods

Method 1

1. Log in to the management console.
2. Under **Computing**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.
3. Click the target AS group name on the **AS Groups** page.
4. On the AS group details page, click the **Instances** tab.
5. Locate the row that contains the target instance and click **Remove and Delete** in the **Operation** column.

   NOTE

   To delete multiple instances, select the check boxes in front of them and click **Remove and Delete**.

Method 2

1. Log in to the management console.
2. Under **Computing**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.
3. Click the target AS group name on the **AS Groups** page.
4. On the AS group details page, click the **AS Policies** tab.
5. Click **Add AS Policy**. In the displayed **Add AS Policy** dialog box, add an as policy to reduce instances as needed or set the number of instances to a specified value.

Method 3

1. Log in to the management console.
1. Under **Computing**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.

2. Click the target AS group name on the **AS Groups** page.
3. On the AS group details page, click **Modify** in the upper right corner.
4. In the displayed **Modify AS Group** dialog box, change the value of **Expected Instances**.

4.8 Will a Yearly/Monthly ECS Be Deleted When the ECS Becomes Faulty?

No. If a yearly/monthly ECS becomes faulty, it will be removed from the AS group, but will not be deleted.

4.9 How Should I Handle Abnormal Instances in an AS Group?

In normal cases, you do not need to handle instances in abnormal state because the AS service periodically checks the health status of instances in an AS group. When an AS group is enabled, abnormal instances are removed and new instances are created to ensure that the number of expected instances is the same as current instances. When an AS group is disabled, checking instance health status continues. However, AS will not remove instances.

It should be noted that if the ELB health check mode is selected, ELB sends heartbeat messages to backend ECSs through an intranet. Therefore, to ensure that the ELB health check can run properly, ensure that your ECS can be accessed through an intranet. Perform the following steps to check this:

1. In the **Listener** area, locate the row containing the target listener and click **View** in the **Health Check** column. A dialog box is displayed.
   - **Health Check Protocol**: Ensure that the protocol has been configured and port has been enabled for the ECS to be checked.
   - **Check Path**: If HTTP is used for the health check, ensure that the health check path for backend ECSs is correct.
2. Check whether software (such as firewall) on the ECS masks the source IP address performing the health check.
3. Check whether the rules of backend ECS security groups and network ACL allow access by 100.125.0.0/16, and configure the protocol and port used for health check. Obtain the health check protocol and port from the dialog box displayed in step **1**.
   - If the default health check mode is used, service ports of the backend ECSs must be enabled.
   - If the health check port is different from service ports of the ECSs, communication between the service ports of the ECSs and health check port must be enabled.
4. If the issue persists, contact technical support.

4.10 What Can I Do If Instances in an AS Group Frequently Fail in Health Checks and Are Deleted and Then Created Repeatedly?

The security group rule of the instance must allow communication with the 100.125.0.0/16 network, and the protocol and port number must be the same as those used by ELB for health checks. Otherwise, the health check will fail. As a result, the instances will be deleted and created again and again.

4.11 How Do I Prevent ECSs from Being Removed from an AS Group Automatically?

You can enable instance protection for in-service ECSs in an AS group. After the configuration, when AS automatically reduces the number of ECSs in an AS group, the in-service ECSs with instance protection enabled will not be removed. You can also set the minimum number of instances for an AS group and the instance removal policy to ensure that an AS group has some in-service ECSs.

Unhealthy ECSs are removed from an AS group and new ECSs are created. Therefore, do not stop or delete ECSs that have been added to an AS group on the ECS console. The stopped or deleted ECSs are considered unhealthy and automatically removed from the AS group. Even when an AS group is disabled, AS checks the status of ECSs in the AS group. In this case, however, unhealthy ECSs will not be removed from the AS group.

4.12 Why Is an Instance Removed and Deleted from an AS Group Still Displayed in the ECS List?

If an instance automatically added to an AS group is protected, it is only removed out of the AS group, but not deleted, so that it can still be used by other services.

An instance that is being used by other services are protected generally. For example, an instance is used by IMS for creating a private image, or used by storage DR.

4.13 Do I Need to Bind an EIP to an ECS Before Manually Adding the ECS to an AS Group?

No. No matter whether an ECS has an EIP, you can manually add the ECS to an AS group.

4.14 Can I Apply the Specifications of an Existing ECS to ECSs to Be Scaled in an AS Group?

Yes. You can create an AS configuration using the specifications of an existing ECS, and select the AS configuration when creating an AS group. The scaled ECSs in the AS group will have the same specifications as your existing ECS that you use to create the AS configuration.

5 Others

5.1 What Can I Do to Enable My Application to Be Automatically Deployed on an Instance?

To enable automatic application deployment on instances automatically added to an AS group, you need to create a private image which contains application software and automatic startup settings. When creating an AS group, select the private image you have created for the AS configuration. In this way, applications will be automatically deployed on instances that are added to the AS group. The procedure is as follows:

1. Before creating a private image, install the application and set it to automatically start upon system startup on the ECS which you will use to create the private image.
2. Create a private image. For details, see ***Image Management Service User Guide***.
3. Create an AS configuration. For details, see ***Auto Scaling User Guide***. Ensure that the image in the AS configuration is the private image created in **2**.
4. Click the **AS Groups** tab and then click the name of the target AS group.
5. Click **Change Configuration** on the right of **Configuration Name**. In the displayed dialog box, select the AS configuration created in **3** and click **OK**.

   After new instances are added to the AS group in the next scaling action, you can check whether the required application has been installed on the instances. If any issue occurs, contact technical support.

5.2 How Does Cloud-Init Affect the AS Service?

Cloud-Init is an open-source cloud initialization program, which initializes specified customized configurations, such as the hostname, key pair, and user data, of a newly created ECS. When creating an AS configuration, you can use Cloud-Init to initialize the ECS.

If Cloud-Init or Cloudbase-Init is not installed in the private image specified in an AS configuration for an AS group, the following cases occur on the instance created in a scaling action:

- If an ECS created using a Windows private image without Cloudbase-Init installed is used, the system will display a message indicating that the password for logging in to the ECS cannot be obtained when you obtain the password. In such a case, log in to the ECS using only the image password. If you forget the image password, use the password resetting function available on the **Elastic Cloud Server** page to reset the password.
- If an ECS created using a Linux private image without Cloud-Init installed is used, the ECS cannot be logged in using the password or key pair configured during ECS creation. In such a case, log in to the ECS using only the image password or key pair. If you forget the image password or key, use the password resetting function available on the **Elastic Cloud Server** page to reset the password.
- If an ECS created using a private image without Cloud-Init or Cloudbase-Init installed is used, user data injection fails.

  To prevent the preceding issues from occurring, check whether the private image in the AS configuration has Cloud-Init or Cloudbase-Init installed before using the AS service. Delete the AS configurations that use the private images without Cloud-Init or Cloudbase-Init installed. Use the private images with Cloud-Init or Cloudbase-Init installed to create new AS configurations. The procedure is as follows:

  a. Log in to the management console.
  b. Under **Computing**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.
  c. Click the **AS Configurations** tab and query the AS configuration list.
  d. Click **Create AS Configuration** and select a private image with Cloud-Init or Cloudbase-Init installed to create a desired AS configuration.
  e. Change the AS configuration in the AS group to the newly created one.

5.3 Why Cannot I Use a Key File to Log In to an ECS?

Issue Description

When I used a key file to attempt to log in to an instance in an AS group, the login failed.

Possible Causes

The image in the AS configuration of the AS group is your private one, and the Cloud-Init tool had not been installed when you created the private image.

If the Cloud-Init tool had not been installed when you created a private image, you would fail to customize the ECS configuration. In such a case, you can log in to the ECS only using the original image password or key pair.

Handling Method

1. Check whether the ECS must be logged in to.
   - If yes, use the original image password or key pair to log in to this ECS.

     The original image password or key pair is the OS password or key pair configured when the private image was created.

- If no, go to step 2.

2. Modify the AS configuration of the AS group. For details, see **Replacing an AS Configuration in an AS Group**.

NOTE

Make sure that the Cloud-Init or Cloudbase-Init tool has been installed in the image of the modified AS configuration. For instructions about how to install the Cloud-Init or Cloudbase-Init tool, see ***Image Management Service User Guide***.

After the AS configuration is modified, you can use the key file to log in to the new ECSs that are added when the AS action is performed in the AS group. In such a case, you do not need to use the original image password or key pair to log in to the new ECSs any more.

5.4 Do I Need to Configure an EIP in an AS Configuration When A Load Balancer Has Been Enabled in an AS Group?

No. If you have enabled a load balancer in an AS group, you do not have to configure an EIP in the AS configuration. The system automatically adds instances in the AS group to the load balancer. These instances will provide services via the EIP bound to the load balancer.

5.5 How Can I Enable Automatic Initialization of EVS Disks of Instances That Have Been Added in a Scaling Action to an AS Group?

Scenarios

After an ECS is created, EVS disks attached to the ECS must be initialized. If multiple ECSs are added to the AS group in scaling actions, you must manually initialize the EVS disks of each ECS, which requires a long period of time.

This section describes how to configure scripts to enable automatic initialization of EVS disks, including disk partitioning and attaching specified directories. The scripts can only be used to initialize one EVS disk.

This section uses CentOS 6.5 as an example. For details about how to configure DHCP on other OSs, see the relevant OS documentation.

Procedure

1. Log in to the instance as user **root**.
2. Run a command to switch to the directory storing the script:

   cd /*script directory*

   An example is as follows:

   cd /home

3. Run the following command to create a script:

   vi *script name*

   An example is as follows:

   vi fdisk_mount.sh

4. Press **i** to go to the script editing page.

   The following script is used as an example to show how to implement automatic initialization of one data disk:

```
#!/bin/bash
bash_scripts_name=fdisk_mount.sh
ini_path=/home/fdisk.ini
disk=
size=
mount=
partition=

function get_disk_from_ini()
{
disk=`cat $ini_path|grep disk| awk -F '=' '{print $2}'`
if [ $disk = "" ]
then
    echo "disk is null in file,exit"
    exit
fi
result=`fdisk -l $disk | grep $disk`
if [ $result = 1 ]
then
    echo "disk path is not exist in linux,exit"
    exit
fi

}

function get_size()
{
size=`cat $ini_path| grep size|awk -F '=' '{print $2}'`
if [ $size = "" ]
then
    echo "size is null,exit"
    exit
fi
}

function make_fs_mount()
{
mkfs.ext4 -T largefile $partition
if [ $? -ne 0 ]
then
    echo "mkfs disk failed,exit"
    exit
fi

dir=`cat $ini_path|grep mount |awk -F '=' '{print $2}'`
if [ $dir = "" ]
then
    echo "mount dir is null in file,exit"
    exit
fi

if [ ! -d $dir ]
then
    mkdir -p $dir
fi

mount $partition $dir
if [ $? -ne 0 ]
```



```
then
    echo "mount disk failed,exit"
    exit
fi

echo "$partition $dir ext3 defaults 0 0" >> /etc/fstab
}

function remove_rc()
{
cat /etc/rc.local | grep $bash_scripts_name
if [ $? ne 0 ]
then
    sed -i '/'$bash_scripts_name'/d' /etc/rc.local
fi
}

################### start ########################
##1. Check whether the configuration file exists.
if [ ! -f $ini_path ]
then
    echo "ini file not exist,exit"
    exit
fi

##2. Obtain the device path for the specified disk from the configuration file.
get_disk_from_ini

##3. Obtain the size of the size partition from the configuration file.
get_size

##4. Partition the disk.
fdisk $disk  <<EOF
n
p
1
1
$size
w
EOF
partition=`fdisk -l $disk 2>/dev/null| grep "^/dev/[xsh].*d" | awk '{print $1}'`

##5. Format the partition and attach the partition to the specified directory.
make_fs_mount

##6. Change startup items to prevent re-execution of the scripts.
remove_rc

echo 'SUCESS'
```

5. Press **Esc**, enter **:wq**, and press **Enter** to save the changes and exit.
6. Run the following command to create the configuration file:

   vi fdisk.ini
7. Press **i** to go to the file editing page.

   The drive letter, size, and directory of the EVS disk are configured in the configuration file. You can change the settings based on the following displayed information.

   ```
   disk=/dev/xdev
   size=+100G
   mount=/opt/test
   ```
8. Press **Esc**, enter **:wq**, and press **Enter** to save the changes and exit.
9. Run the following command to open configuration file **rc.local**:

   vi /etc/rc.local

10. Press **i** to add the following content to the **rc.local** file:

    /home/fdisk_mount.sh

    After the **rc.local** file is configured, the EVS disk initialization script will be automatically executed when the ECS starts.

11. Press **Esc**, enter **:wq**, and press **Enter** to save the changes and exit.
12. Create a private image using an ECS.
13. Create an AS configuration.

    When you specify the AS configuration information, select the private image created in the preceding step and select an EVS disk.

14. Create an AS group.

    When you configure the AS group, select the AS configuration created in the preceding step.

    After the AS group is created, EVS disks of new ECSs added in scaling actions to this AS group will be automatically initialized based on the configuration in the private image.

