

Auto Scaling

FAQ

Issue 01
Date 2025-11-05



Copyright © Huawei 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 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 General	1
1.1 What Are Restrictions on Using AS?	1
1.2 Must I Use AS Together With ELB and Cloud Eye?	2
1.3 Is AS Billed?	2
1.4 Will an Abrupt Change in Monitoring Metric Values Trigger an Unnecessary Scaling Action?	2
1.5 How Many AS Policies and AS Configurations Can I Create and Use?	2
1.6 Can AS Automatically Scale Up or Down vCPUs, Memory, and Bandwidth of ECSs?	2
1.7 How Do I Fix the Error "The key pair does not exist" When I Connect to an Instance?	3
1.8 What Types of EIP and Shared Bandwidths Can Be Adjusted by AS?	3
1.9 What Types of ECS and HECS Instances Can Be Scaled by AS?	3
1.10 If the OS of an ECS Instance in an AS Group Is Upgraded, Can AS Automatically Add Instances with the Newest OS During the Next Scale-out Action?	3
1.11 Can AS Add Yearly/Monthly ECS Instances When It Scales Out an AS Group?	3
1.12 Can AS Increase Bandwidths as Required?	3
2 AS Group	12
2.1 What Can I Do If the AS Group Fails to Be Enabled?	12
2.2 How Can I Handle an AS Group Exception?	15
2.3 What Operations Will Be Suspended If an AS Group Is Disabled?	17
2.4 Can I Use an ECS Instance ID to Learn What AS Group the Instance Is In?	18
3 AS Policy	19
3.1 How Many AS Policies Can I Enable?	19
3.2 What Are the Conditions to Trigger an Alarm-based AS Policy?	19
3.3 What Is a Cooldown Period and Why Is It Required?	19
3.4 Can AS Scale Capacity Based on Custom Monitoring by Cloud Eye?	20
3.5 What Monitoring Metrics for an AS Group Will Be Affected If VMTTools Are Not Installed on the Instances in the Group?	20
3.6 What Can I Do If an AS Policy Fails to Be Enabled?	20
3.7 How Do I Install the Agent on Instances in an AS Group to Use Agent Monitoring Metrics?	21
4 Instance	25
4.1 How Do I Prevent Instances Manually Added to an AS Group from Being Automatically Removed?	25
4.2 If the AS Configuration of an AS Group Uses Multiple Flavors, in What Order Are the Flavors Selected During Scale-out Actions?	26
4.3 When an Instance Is Removed from an AS Group and Deleted, Is the Application Data Saved?	27

4.4 Can I Add Yearly/Monthly ECS Instances to an AS Group?.....	27
4.5 Can AS Automatically Delete ECS Instances Added Based on an AS Policy When They Are Not Required?.....	27
4.6 What Is the Expected Number of Instances?.....	27
4.7 How Do I Delete an ECS Instance Created in a Scaling Action?.....	28
4.8 Will Yearly/Monthly ECS Instances Be Deleted If They Become Unhealthy?.....	29
4.9 How Do I Handle Unhealthy Instances in an AS Group?.....	29
4.10 Why Instances in an AS Group Keep Failing Health Checks and Getting Deleted and Recreated?.....	30
4.11 How Do I Prevent ECS Instances from Being Automatically Removed from an AS Group?.....	30
4.12 Why Is an Instance that Was Removed from an AS Group and Deleted Still Displayed in the ECS List?	30
4.13 Do I Need to Bind an EIP to an Instance Before I Manually Add the Instance to an AS Group?.....	30
4.14 Can I Apply the Configurations of an Existing Instance in an AS Group to Instances to Be Added in the Future?.....	30
4.15 Will the System Bind a New EIP or an Existing One a Newly Added ECS Instance?.....	31
5 Others.....	32
5.1 How Can I Automatically Deploy My Application on an Instance?.....	32
5.2 How Does Cloud-Init Affect the AS Service?.....	32
5.3 Why Can't I Use a Key File to Log In to an ECS?.....	33
5.4 Do I Need to Configure an EIP in an AS Configuration When a Load Balancer Has Been Enabled for an AS Group?.....	34
5.5 How Do I Enable Automatic Initialization of EVS Disks on ECS Instances That Have Been Added to an AS Group During Scaling Actions?.....	34

1 General

1.1 What Are Restrictions on Using AS?

Only applications that are stateless and horizontally scalable can run on ECS instances in an AS group. ECS instances in an AS group can be released automatically by AS, so the ECS instances cannot be used to save application status information (such as session statuses) or 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 quotas.

Table 1-1 Quotas

Item	Description	Default
AS group	Maximum number of AS groups per region per account	10
AS configuration	Maximum number of AS configurations per region per account	100
AS policy	Maximum number of AS policies per AS group	10
Instance	Maximum number of instances per AS group	300
Bandwidth scaling policy	Maximum number of bandwidth scaling policies per region per account	10
Lifecycle hook	Maximum number of lifecycle hooks per AS group	5
Notification	Maximum number of notifications per AS group	5
Tag	Maximum number of tags per AS group	10

1.2 Must I Use AS Together With ELB and Cloud Eye?

AS can work independently or in conjunction with ELB and Cloud Eye.

Cloud Eye does not require additional fees and is enabled by default. ELB is not enabled by default, but you can enable it manually if needed, for example, if distributed clusters are required.

1.3 Is AS Billed?

AS is free, but you need to pay for pay-per-use instances in your AS groups. 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 deleted. After they are deleted, they no longer generate any fees. Instances manually added, however, are only removed from the AS group. They are not deleted, so they are still billed. If you do not need these instances, unsubscribe from them on the ECS console.

For example, if two instances are added by the system to an AS group but are removed one hour later, you will be billed for that one hour of use.

1.4 Will an Abrupt Change in Monitoring Metric Values Trigger an Unnecessary Scaling Action?

No. Monitoring data used by AS is from Cloud Eye. The monitoring period can be set to 5 minutes, 20 minutes, or 1 hour, so an abrupt change in monitoring metric values will not impact scaling actions.

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

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 can use 1 AS configuration and 10 AS policies at a time.

If the default quotas do not meet your service requirements, contact customer service.

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

Currently, AS can automatically scale up or down bandwidths, and scale in or out ECS instances in an AS group.

1.7 How Do I Fix the Error "The key pair does not exist" When I Connect to an Instance?

A key pair is specific to each user. If the key pair of a user who belongs to the same account as you is configured for an AS configuration, you cannot connect the instances scaled out using that AS configuration.

If you want to connect to these instances without being restricted by the key pair permission, password authentication needs to be configured as the login mode.

1.8 What Types of EIP and Shared Bandwidths Can Be Adjusted by AS?

AS can only adjust bandwidths of EIPs and share bandwidths that are billed on a pay-per-use basis.

1.9 What Types of ECS and HECS Instances Can Be Scaled by AS?

Supported ECS types are listed on the AS management console. For details about ECS specifications, see [Elastic Cloud Server User Guide](#).

Currently, AS cannot scale HECS instances.

1.10 If the OS of an ECS Instance in an AS Group Is Upgraded, Can AS Automatically Add Instances with the Newest OS During the Next Scale-out Action?

No. You need to compile a monitoring plug-in and install it on the target instance. Whenever the OS of this instance is updated, the plug-in will call the relevant APIs automatically to create an image from this instance, create an AS configuration using the image, and then replace the AS configuration used by the AS group with the newly created one.

1.11 Can AS Add Yearly/Monthly ECS Instances When It Scales Out an AS Group?

Only pay-per-use or spot ECS instances can be added during scale-out events.

1.12 Can AS Increase Bandwidths as Required?

Yes. Auto Scaling can increase bandwidths based on specific scenarios.

Scenarios

You can automatically adjust your purchased EIP bandwidth and shared bandwidth using a bandwidth scaling policy. This section describes how to create a bandwidth scaling policy.

When creating a bandwidth scaling policy, you need to configure basic information. The system supports three types of bandwidth scaling policies: alarm-based, scheduled, and periodic.

The basic information for creating a bandwidth scaling policy includes the policy name, resource type, policy type, and trigger condition.

Creating an Alarm-based Bandwidth Scaling Policy

1. Log in to the management console.
2. Under **Compute**, click **Auto Scaling**. In the navigation pane on the left, choose **Bandwidth Scaling**.
3. Click **Create Bandwidth Scaling Policy**.
4. Set parameters, such as the policy name, resource type, policy type, and trigger condition. For details, see [Table 1-2](#).

Table 1-2 Alarm policy parameters

Parameter	Description	Example Value
Region	Select the region where the bandwidth scaling policy is applied.	-
Policy Name	Enter a name for the bandwidth scaling policy. The name consists of only letters, digits, underscores (_), and hyphens (-).	-
Resource Type	Select the type of the bandwidth to be adjusted. You can select EIP or Shared bandwidth .	Shared bandwidth
EIP	This parameter is available only when Resource Type is set to EIP . Select the EIP whose bandwidth needs to be scaled. NOTE Only bandwidths of pay-per-use EIPs can be scaled.	-
Policy Type	Select Alarm .	Alarm

Parameter	Description	Example Value
Alarm Rule	<p>Choose whether to create a new alarm rule or use an existing one.</p> <p>If you select Use existing, you can select an alarm rule drop-down list. Alternatively, click Create Alarm Rule on the right to create an alarm rule on the Cloud Eye console. For details, see Creating an Alarm Rule.</p> <p>If you select Create, configure the following parameters:</p> <ul style="list-style-type: none">• Rule Name Enter a name for the new alarm rule, for example, as-alarm-7o1u.• Trigger Condition Select a monitoring metric and trigger condition. Table 1-3 lists the supported monitoring metrics. An example value is Outbound Traffic Avg. > 100 bit/s.• Monitoring Interval Specifies the monitoring period for the metric, for example, 5 minutes.• Consecutive Occurrences Specifies the number of consecutive times, for example, one time, for triggering a scaling action during a monitoring period.	-
Alarm Policy Type	<p>This parameter is available only when Policy Type is set to Alarm. The value can be Simplified scaling or Refined scaling.</p> <ul style="list-style-type: none">• If you select Simplified scaling, you can set an action to be performed when a monitored metric hits the configured threshold, for example, Monitored metric >, ≥, <, or ≤ xx.• If you select Refined scaling, you can set the action to be performed when the monitored metric falls within a range with both an upper and a lower limit, for example, xx ≤ Monitored metric < xx.	Simplified scaling

Parameter	Description	Example Value
Scaling Action	<p>Define the scaling action to be performed. The actions include:</p> <ul style="list-style-type: none"> • Add When a scaling action is triggered, the bandwidth is increased. • Reduce When a scaling action is triggered, the bandwidth is decreased. • Set to The bandwidth is set to a fixed value. <p>NOTE The step (minimum unit for bandwidth adjustment) varies depending on the bandwidth value range. The bandwidth will be automatically adjusted to the nearest value according to the actual step.</p> <ul style="list-style-type: none"> • If the bandwidth is less than or equal to 300 Mbit/s, the default step is 1 Mbit/s. • If the bandwidth ranges from 300 Mbit/s to 1,000 Mbit/s, the default step is 50 Mbit/s. • If the bandwidth is greater than 1,000 Mbit/s, the default step is 500 Mbit/s. 	-
Cooldown Period	<p>A cooldown period (in seconds) is the period of time between two scaling actions. AS recounts the cooldown period after a scaling action is complete. During the cooldown period, AS denies all scaling requests triggered by alarm-based policies. Scaling requests triggered manually or by scheduled or periodic policies are not affected.</p>	300s

Table 1-3 Monitoring metrics supported by the alarm policy

Metric	Description
Inbound Bandwidth	Indicates the network rate of inbound traffic.
Inbound Traffic	Indicates the network traffic going into the cloud platform.
Outbound Bandwidth	Indicates the network rate of outbound traffic.
Outbound Traffic	Indicates the network traffic going out of the cloud platform.
Outbound Bandwidth Usage	Indicates the usage of network rate of outbound traffic in the unit of percentage.

5. After setting the parameters, click **Create Now**.

The newly created bandwidth scaling policy is displayed on the **Bandwidth Scaling** page and is in **Enabled** state by default.

Creating an Alarm Rule

When creating an alarm-based bandwidth scaling policy, you can click **Create Alarm Rule** to the right of **Rule Name** to create an alarm rule. To do so, perform the following operations:

1. Click **Create Alarm Rule** to the right of **Rule Name** to switch to the **Alarm Rules** page of Cloud Eye.
2. On the **Alarm Rules** page, click **Create Alarm Rule** in the upper right corner.
3. Set parameters based on [Figure 1-1](#) and [Table 1-4](#). For more information about how to set alarm rules, see [Cloud Eye User Guide](#).

Figure 1-1 Creating an alarm rule

< | **Create Alarm Rule** ⓘ

* Name

Description 0/256 ↕

* Alarm Type Metric Event

* Cloud Product

* Resource Level ⓘ Cloud product Specific dimension

* Monitoring Scope All resources Resource groups Specific resources

* Instance Selected Resources:1 [Reselect](#)

Name

Table 1-4 Key parameters for creating an alarm rule

Parameter	Description	Example Value
Name	Specifies the name of the alarm rule.	alarm-bandwidth
Description	(Optional) Provides supplementary information about the alarm rule.	-
Alarm Type	Select the alarm type to which the alarm rule will apply. The type can be Metric , Event , Website , WAN , or Availability .	Metric
Cloud Product	Select the cloud product you want to monitor. This parameter is only available if you select Metric for Alarm Type .	Virtual Private Cloud - Bandwidths
Resource Level	This parameter is only available if you select Metric for Alarm Type . You can select Cloud product (recommended) or Specific dimension .	Cloud product
Monitoring Scope	Select the resources to which the alarm rule applies. Set this parameter to Specific resources .	Specific resources
Instance	This parameter is available only if Monitoring Scope is set to Specific resources . Search for the bandwidth using its name or ID. The name and ID can be obtained from the EIP details.	Select the created shared bandwidth.
Method	Select Associate template or Configure manually . NOTE If you select Associate template , after the associated template is modified, the policies contained in this alarm rule to be created will be modified accordingly.	Configure manually
Alarm Policy	Configure the alarm policy for triggering the alarm rule as required. For details about the monitoring metrics, see Table 1-3 .	-

- After setting the parameters, click **Create**.
- On the **Create Bandwidth Scaling Policy** page, click  to the right of **Rule Name**, and select the created alarm rule.

Alternatively, you can create an alarm rule on the Cloud Eye console page before creating a bandwidth scaling policy. Ensure that the EIP bandwidth you want to

adjust is configured for the alarm rule. After the alarm rule is created, you can select the rule when creating a bandwidth scaling policy.

Creating a Scheduled or Periodic Bandwidth Scaling Policy

1. Log in to the management console.
2. Under **Compute**, click **Auto Scaling**. In the navigation pane on the left, choose **Bandwidth Scaling**.
3. Click **Create Bandwidth Scaling Policy**.
4. Set parameters, such as the policy name, resource type, policy type, and trigger condition. For details, see [Table 1-5](#).

Table 1-5 Scheduled or periodic policy parameters

Parameter	Description	Example Value
Region	Select the region where the bandwidth scaling policy is applied.	-
Policy Name	Enter a name for the bandwidth scaling policy. The name consists of only letters, digits, underscores (_), and hyphens (-).	as-policy-p6g5
Resource Type	Select the type of the bandwidth to be adjusted. You can select EIP or Shared bandwidth .	EIP
EIP	Select the EIP whose bandwidth needs to be scaled. This parameter is mandatory when Resource Type is set to EIP . NOTE Only bandwidths of pay-per-use EIPs can be scaled.	-
Shared Bandwidth	Specifies the shared bandwidth to be scaled. This parameter is mandatory when Resource Type is set to Shared bandwidth .	-

Parameter	Description	Example Value
Policy Type	<p>Specifies the policy type. You can select a scheduled or periodic policy.</p> <p>If you select Periodic, you are required to configure two more parameters:</p> <ul style="list-style-type: none">• Time Range Specifies the time range during which the AS policy can be triggered.• Period<ul style="list-style-type: none">- Day- Week- Month	<p>Day 2023/03/01 00:00:00 - 2023/03/31 23:59:59</p> <p>In this example, the AS policy will trigger a scaling action every day in March, and will become invalid from April 1, 2023 00:00:00.</p>
Triggered At	<p>Specifies the time at which the AS policy is triggered.</p> <p>NOTE The selected triggering time must fall inside the effective time range of the policy.</p>	-
Scaling Action	<p>Specifies the action to be performed.</p> <p>The following scaling action options are available:</p> <ul style="list-style-type: none">• Add When a scaling action is triggered, the bandwidth is increased.• Reduce When a scaling action is triggered, the bandwidth is decreased.• Set to The bandwidth is set to a fixed value. <p>NOTE The step (minimum unit for bandwidth adjustment) varies depending on the bandwidth value range. The bandwidth will be automatically adjusted to the nearest value according to the actual step.</p> <ul style="list-style-type: none">• If the bandwidth is less than or equal to 300 Mbit/s, the default step is 1 Mbit/s.• If the bandwidth ranges from 300 Mbit/s to 1,000 Mbit/s, the default step is 50 Mbit/s.• If the bandwidth is greater than 1,000 Mbit/s, the default step is 500 Mbit/s.	-

Parameter	Description	Example Value
Cooldown Period	A cooldown period (in seconds) is the period of time between two scaling actions. AS recounts the cooldown period after a scaling action is complete. During the cooldown period, AS denies all scaling requests triggered by alarm-based policies. Scaling requests triggered manually or by scheduled or periodic policies are not affected.	300s

5. After setting the parameters, click **Create Now**.

2 AS Group

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

The possible causes and solutions are as follows:

- Issue description: Insufficient ECS, EVS disk, or EIP quota.
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: This type of ECSs has been sold out in the selected AZ.
Possible cause: This type of ECSs has been sold out or is not supported in the AZ selected for the AS group. The ECS flavor selected in the AS configuration is sold out.
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: This type of storage resources has 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 specified 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 bound 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 flavor 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.

2.2 How Can I Handle an AS Group Exception?

The handling method depends on the reported possible cause.

- Issue description: Insufficient ECS, EVS disk, or EIP quota.
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: This type of ECSs has been sold out in the selected AZ.
Possible cause: This type of ECSs has been sold out or is not supported in the AZ selected for the AS group. The ECS flavor selected in the AS configuration is sold out.
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: This type of storage resources has 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 specified 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 bound 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 flavor 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 Operations Will Be Suspended If an AS Group Is Disabled?

If an AS group is disabled, new scaling actions will not happen, but any scaling actions already in progress will continue. Scaling policies will not trigger any scaling actions. Even if you manually change the number of expected instances, no scaling action will be triggered even though the number of actual instances is not equal to that of expected instances.

Health checks continue to be performed but will not remove any instances.

2.4 Can I Use an ECS Instance ID to Learn What AS Group the Instance Is In?

No.

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

- Step 1** Log in to the management console. Choose **Compute > Auto Scaling > Instance Scaling**.
 - Step 2** On the **AS Groups** 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 I Enable?

You can enable one or more AS policies as required.

3.2 What Are the Conditions to Trigger an Alarm-based 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 scale instances in or out in the AS group.

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

A cooldown period is the period of time between two scaling actions. AS recounts the cooldown period after a scaling action is complete. During the cooldown period, AS denies all scaling requests triggered by alarm-based policies. Scaling requests triggered manually or by scheduled or periodic policies are not affected.

Before an instance is put into use after it is added to the AS group, it takes 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. If an instance is put into use without cooldown, the system will keep adding instances until the load decreases. As the new instances take over services, the system will detect that the load is too low and start removing instances from the AS group. A cooldown prevents the AS group from repeatedly triggering unnecessary scaling actions.

For example:

When a traffic peak occurs, an alarm policy is triggered and AS automatically adds an instance to the AS group to help handle the increased load. However, it takes time for the instance to start. After the instance is started, it takes time to receive requests from ELB. During this period, alarms may continue to be triggered and

instances may continue to be added. If you set a cooldown time, after an instance is started, AS stops adding new instances in response to the alarms until the specified period of time (300 seconds by default) passes. That way 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 starts up again.

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

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

3.5 What Monitoring Metrics for an AS Group Will Be Affected If VMTools Are Not Installed on the Instances in the Group?

If VMTools are not installed on ECS instances, Cloud Eye can monitor metrics Outband Incoming Rate and Outband Outgoing Rate but cannot monitor metrics Memory Usage, Inband Incoming Rate, and Inband Outgoing Rate. This will reduce data accuracy of CPU usage.

If the ECS instances are I/O-optimized, Cloud Eye cannot monitor metrics Disk Usage, Inband Incoming Rate, and Inband Outgoing Rate of instances no matter VMTools are installed or not.

If VMTools are not installed on ECS instances, 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 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 Do I Install the Agent on Instances in an AS Group to Use Agent Monitoring Metrics?

Scenarios

If an AS group uses an alarm-based scaling policy, you can install the Agent on the instances to better monitor the AS group. Compared with basic monitoring, Agent monitoring (OS monitoring) provides proactive, fine-grained, system-level monitoring of instances. This section describes how to install the Agent on instances.

Procedure

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

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

2. Create an ECS and install the Agent.

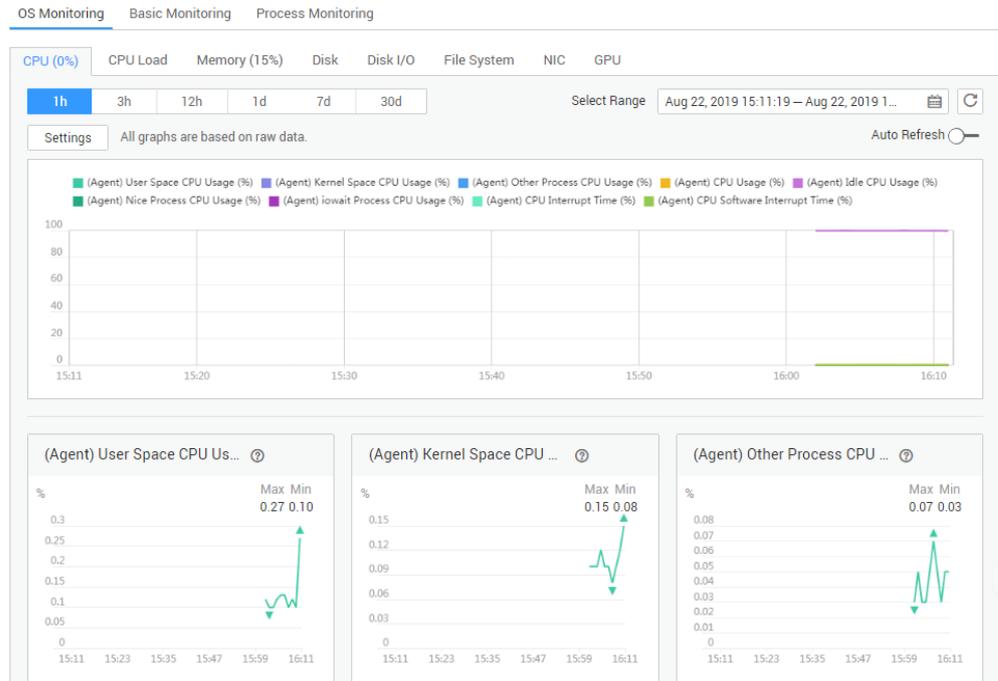
For more information, see [Agent Installation and Configuration](#).

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

Figure 3-1 Checking the Agent status

<input type="checkbox"/>	Name/ID	IP Address	ECS Status	Agent Status ?
<input checked="" type="checkbox"/>	ecs-6e5btest c9976140-3023-490b-85f3-d52cc8...	1.92.158.248 (EIP) 192.168.1.15 (Pri...)	→ Running	→ Running

Figure 3-2 Viewing Agent monitoring metrics



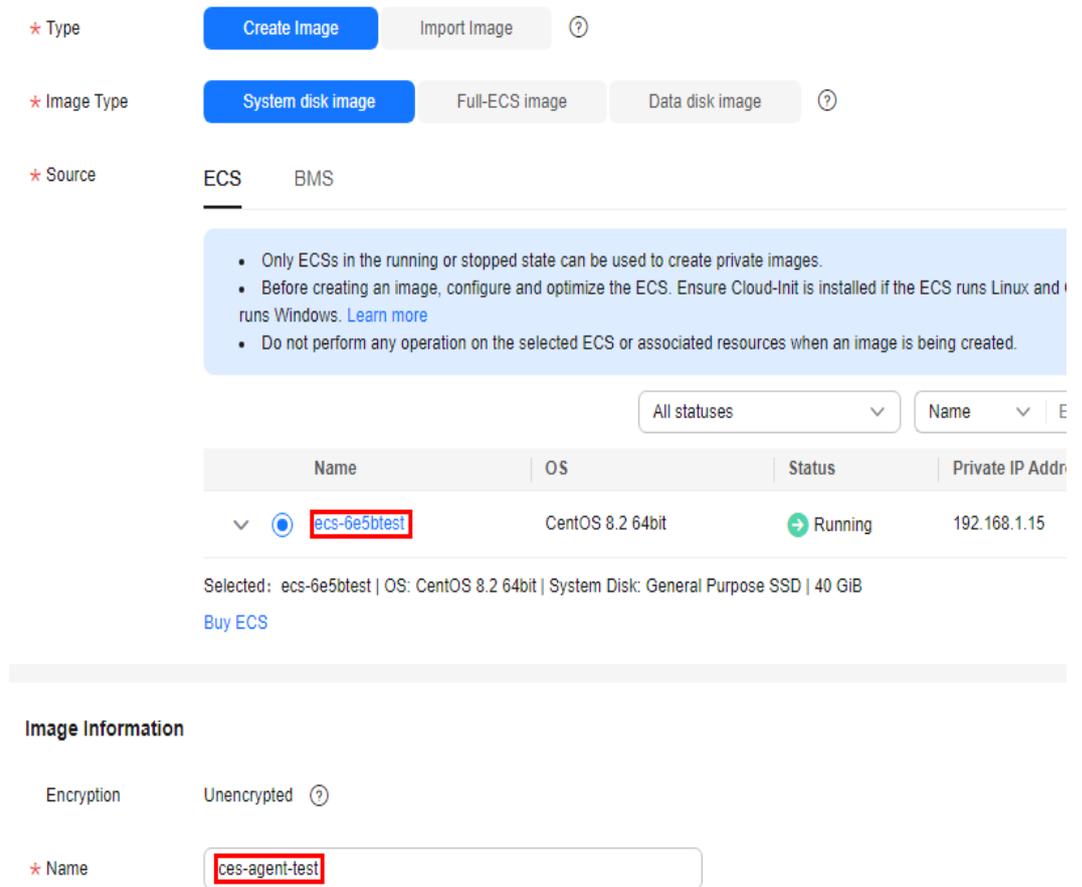
4. Add the AK/SK pair to the **conf.json** configuration file of the ECS.
 - a. Click your username, choose **My Credentials > Access Keys**, and obtain the AK/SK pair.
 - If you have created an access key, obtain the **AccessKey** value and the **SecretKey** value in the **credentials.csv** file you obtained when you create the access key.
 - If no access keys are available, click **Create Access Key** to create one. Save the **credentials.csv** file and obtain the **AccessKey** value and the **SecretKey** value in it.
 - b. Log in to the ECS and use **cd /usr/local/telescope/bin** to go to the Agent installation directory.
 - c. Run **vi conf.json** to open the configuration file and enter the obtained AK/SK pair.

```
{
  "InstanceId": "fa07b[REDACTED]4cd0-9aa6-e5c791569e3a",
  "ProjectId": "050b1[REDACTED]572f8cc01f3740bed5",
  "AccessKey": "MK8NR3[REDACTED]7FUMJB",
  "SecretKey": "sPHiTB8[REDACTED]N4wWw3YCNwcUFqj",
  "RegionId": "cn-north-1"
}
```

If the Agent was installed during ECS creation, the AK/SK pair was added during user data injection. You only need to check the AK/SK pair in this step.

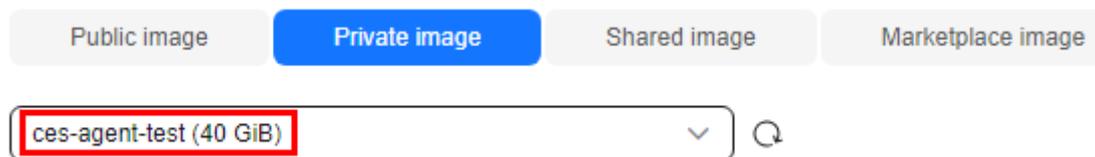
- d. Press **Esc** and enter **:wq** to save the configuration and exit.
5. Go to the **Image Management Service** page and create a private image **ces-agent-test** for this ECS. For details, see [Creating a Private Image](#).

Figure 3-3 Creating a private image



- 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.

- Use the AS configuration created in step 6 to create an AS group.
- 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

Add AS Policy

Policy Name:

Policy Type: **Alarm** | Scheduled | Periodic
Policies of this type are applied only when their associated alarm rules are enabled. [View alarm rules](#)

Alarm Rule: **Create** | Use existing

Rule Name:

Monitoring Type: **System monitoring** | Custom monitoring

Trigger Condition: Max. > %
Dropdown menu items:
Disks Write Requests
(Agent) CPU Usage
(Agent) Memory Usage
(Agent) 1-Minute Load Aver...
(Agent) 5-Minute Load Aver...
(Agent) 15-Minute Load Ave...
(Agent) GPU Usage
(Agent) GPU Memory Usage

Monitoring Interval:

Consecutive Occurrences: ?

ent), the Agent must be installed on all instances in the AS group. [Learn more](#)
ring metrics, see [Monitoring Metrics](#).
ry somewhat by OS. [Learn more](#)

9. Manually add the ECS created in step 2 to the AS group.
10. Perform the following operations to check whether the Agent monitoring has taken effect:
 - Verify that the Agent monitoring metric is displayed on the **Monitoring** tab of the AS group details page.
 - When the alarm threshold is reached, verify that the alarm policy has triggered scaling actions and instance are added or removed.
 - The Agent monitoring data is collected from 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 Automatically Removed?

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 do this:

Method 1

Perform following configurations in the AS group:

- Set the minimum number of instances in the AS group to N or a larger value.
- 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 are not created based on the AS configuration used by the AS group. The instances automatically added using the AS configuration are removed at first. The manually added instances would not be removed until all of the automatically added instances have been removed first. Finally, since you have set the minimum number of instances to N or a larger value, those instances cannot be removed.

Note: If the instances manually added are stopped or if they malfunction, they are marked as unhealthy and removed from the AS group. This is because health checks ensure that all instances in the AS group are healthy.

Method 2

Enable instance protection for the N instances. For details, see [Instance Protection](#).

You can enable instance protection for these instances at the same time. When the AS group scales in, protected instances will not be removed from the AS group as long as they do not fail health checks. Instances that fail health check will be removed even if they are protected.

4.2 If the AS Configuration of an AS Group Uses Multiple Flavors, in What Order Are the Flavors Selected During Scale-out Actions?

The order depends on the AZs and multiple-AZ scaling policy used by the AS group. This section describes how flavors are selected during scale-out events in these the following two different scenarios:

Single AZ

If the AS group uses only one AZ, new instances will be created in the AZ by following the configured flavor selection policy which determines the flavor priority order.

- **Sequenced:** In this policy, the flavors are used in the order they were selected. For example, if you selected flavors 2, 3, and 1, and in that sequence, the system attempts to use flavor 2 to create instances at first. If flavor 2 is not available in the current AZ or an instance fails to be created due to other reasons, the system follows up with flavor 3 and again with flavor 1.
- **Cost-centered:** In this policy, the least expensive flavor is used first. For example, if you select flavors 1, 2, and 3 in sequence, where flavor 1 is the most expensive, followed by flavor 3, and flavor 2 is the least expensive, the system preferentially scales out the AS group by starting with flavor 2, then flavor 3, and then flavor 1.

Multiple AZs

If the AS group spans two or more AZs, you need to configure the multi-AZ scaling policy. The system will scale out the AS group by following the scaling policy you configured using the flavor with the highest priority.

- **Balanced:** When scaling out an AS group, the system preferentially distributes ECS instances evenly across AZs used by the AS group. If it fails in the target AZ, it automatically selects another AZ based on the sequenced policy. For example, if:
 - You have selected AZ 1, AZ 2, and AZ 3 in sequence.
 - You have selected flavors 1, 2, and 3, where the flavor priority order is 2, 3, and 1.
 - AZ 1, AZ 2, and AZ 3 have, respectively, 3, 2, and 3 instances.AZ 2 has fewer instances. The system will try to create instances in AZ 2 first using flavor 2 (the one with the highest priority). If this attempt succeeds, the scaling action is complete. If this attempt fails, the system will try to create instances in AZ 2 using flavor 3, and then flavor 1. If all attempts fail in AZ 2, the system will try to fulfill the required capacity by starting with flavor 2, then flavor 3, and then flavor 1 in the AZ 1 based on the Sequenced policy. If all attempts fail in AZ 1, the system will try in AZ 3.
- **Sequenced:** When scaling out an AS group, the system distributes ECS instances to the AZs based on the order in which AZs are specified. For example, if:

- You have selected AZ 1, AZ 2, and AZ 3 in sequence.
- You have selected flavors 1, 2, and 3, where the flavor priority order is 2, 3, and 1.

Regardless of how instances are distributed in three AZs, the system will try to use flavor 2 to create instances in AZ 1. If this attempt fails, the system will try flavor 3 and then flavor 1 in AZ 1. If all attempts fail in AZ 1, the system will do that again in AZ 2, and then AZ 3.

NOTE

The flavor priority order is determined by the flavor selection policy in the AS configuration. For details, see [Single AZ](#).

4.3 When an Instance Is Removed from an AS Group and Deleted, Is the Application Data Saved?

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

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. When the lifecycle hook is added to the AS group, if the AS group scales in, the lifecycle hook suspends the instance that is being removed from the AS group and puts the instance in a wait state. During the waiting period, you can perform operations on the instance, like backing up data or downloading log files.

4.4 Can I Add Yearly/Monthly ECS Instances to an AS Group?

Yes. Currently, AS automatically creates pay-per-use ECS instances by default, but you can also add yearly/monthly or pay-per-use instances manually if needed.

4.5 Can AS Automatically Delete ECS Instances Added Based on an AS Policy When They Are Not Required?

Yes. AS can do it if an AS policy has been added to trigger scaling actions to delete the ECS instances.

4.6 What Is the Expected Number of Instances?

The expected number of instances refers to the number of ECS instances 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 ECS instances 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 instances will be inconsistent with the expected number, and a scaling action will be performed to bring the number of instances in line with the expected number.

If a scaling policy is triggered to add two instances to an AS group, the system will increase the expected number of instances by 2. Then, a scaling action is performed to add two instances so that the number of instances in the AS group is the same as the expected number.

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

Handling Methods

Method 1

1. Log in to the management console.
2. Under **Compute**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.
3. Click the 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 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 **Compute**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.
3. Click the 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 remove ECS instances as needed or maintain a specified number of instances.

Method 3

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

2. Click the 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 Yearly/Monthly ECS Instances Be Deleted If They Become Unhealthy?

No. If an ECS instance billed on a yearly/monthly basis becomes unhealthy, it will be removed from the AS group, but will not be deleted.

4.9 How Do I Handle Unhealthy Instances in an AS Group?

Normally, you do not need to handle unhealthy instances manually because AS periodically checks the health status of instances in an AS group. When an AS group is enabled, unhealthy instances are removed and new instances are created to ensure that the expected number of instances are running in the AS group. When an AS group is disabled, AS still checks the health of instances, but does not remove instances.

It should be noted that if ELB health check is selected, ELB sends heartbeat messages to ECS instances through an intranet. To ensure that the ELB health check can be performed properly, ensure that your ECS instances can be accessed through that intranet. To check this, perform the following steps:

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 instance to be checked.
 - **Check Path:** If HTTP is used for the health check, ensure that the health check path for the instance is correct.
2. Confirm that there is no software such as firewall on the ECS instance blocking the source IP address used for performing the health check.
3. Confirm that the rules of instance security groups and network ACL allow access from 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 type of health check is used, service ports of the ECS instances must be enabled.
 - If the health check port is different from service ports of the ECS instances, communication between the service ports and health check port must be enabled.
4. If the issue persists, contact technical support.

4.10 Why Instances in an AS Group Keep Failing Health Checks and Getting Deleted and Recreated?

The rules of security group that the instances are in must allow access from the 100.125.0.0/16 network segment over the protocol and port used by ELB for health checks, or the health checks will fail. As a result, the instances will be deleted and created again and again.

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

You can enable instance protection for in-service instances in an AS group. After the configuration, the protected in-service instances will not be removed during scale-in events. You can also modify the minimum number of instances for an AS group and use an instance removal policy to ensure that the AS group always has some in-service instances.

Unhealthy instances are removed from an AS group and new instances are created automatically. Do not stop or delete instances that have been added to an AS group on the ECS console as they will be marked as unhealthy and automatically removed from the AS group. Even when an AS group is disabled, AS still checks the health of instances in the AS group, but does not remove unhealthy instances.

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

If an automatically added instance is protected, it is 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 SDRS.

4.13 Do I Need to Bind an EIP to an Instance Before I Manually Add the Instance to an AS Group?

No. You can manually add the instance to an AS group regardless of if it has an EIP or not.

4.14 Can I Apply the Configurations of an Existing Instance in an AS Group to Instances to Be Added in the Future?

Yes. You can create an AS configuration using the existing instance, and change the AS configuration used by the AS group to the one you just created. Then

scaled-out instances will have the same specifications as the instance that you used to create the AS configuration.

4.15 Will the System Bind a New EIP or an Existing One a Newly Added ECS Instance?

During scale-out actions, the system automatically buys and assigns an EIP to each newly added instance.

5 Others

5.1 How Can I Automatically Deploy My Application on an Instance?

To enable automatic application deployment on instances automatically added to an AS group, create a private image with the application preinstalled and automatic startup settings preconfigured. Create an AS configuration with the private image, and then change the AS configuration used by the AS group to the one you created. Your application will be automatically deployed on instances that are automatically added to the AS group. The procedure is as follows:

1. Install the application on the ECS instance you will use to create a private image, and configure the application to automatically start at boot.
2. Create a private image using the instance. For details, see *Image Management Service User Guide*.
3. Create an AS configuration. For details, see [Creating an AS Configuration from a New Specifications Template](#). During the creation, select the private image created in 2.
4. Go to the page that shows the details about your AS group.
5. Click **Change Configuration** to 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 your application has been installed on the instances. If you encounter any problems, 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 custom configurations, such as the hostname, key pair, and user data, of a newly created ECS. When creating an AS configuration, you can choose an image with Cloud-Init or Cloudbase-Init preinstalled for ECS initialization.

If Cloud-Init or Cloudbase-Init is not installed in the private image specified in the AS configuration of an AS group, the following issues may occur on the ECS instances created in a scaling action:

- If an ECS is created from a Windows image, the system will display a message indicating that the password for logging in to the ECS could not be found. In such a case, you can only log in to the using the image password. If you forgot the image password, use password resetting on the **Elastic Cloud Server** page to reset the ECS password.
- If an ECS is created from a Linux image, the ECS cannot be logged in using the password or key pair configured during ECS creation. In such a case, you can only log in to the ECS using the image password or key pair. If you forgot the image password or key, use password resetting on the **Elastic Cloud Server** page to reset the ECS password.
- User data injection into the ECS instances will fail.
To avoid these issues, confirm that the private image specified in the AS configuration has Cloud-Init or Cloudbase-Init installed. If the program was not installed, use a private image with the program installed to create an AS configuration, and replace the AS configuration of the AS group with the newly created one. The procedure is as follows:
 - a. Log in to the management console.
 - b. Under **Compute**, click **Auto Scaling**. In the navigation pane on the left, choose **Instance Scaling**.
 - c. Click the **AS Configurations** tab.
 - 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 of the AS group to the newly created one.

5.3 Why Can't I Use a Key File to Log In to an ECS?

Issue Description

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

Possible Causes

The image specified in the AS configuration of the AS group is a private image, on which Cloud-Init has not been installed.

In this case, user data cannot be injected into the ECS. You can only log in to the ECS using the password or key pair specified in the image.

Handling Method

1. Check whether the ECS needs to be logged in to.
 - If yes, use the 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. Change the AS configuration for the AS group. For details, see [Changing the AS Configuration for an AS Group](#).

 NOTE

Make sure that Cloud-Init or Cloudbase-Init has been installed on the image specified in the new AS configuration. For how to install Cloud-Init or Cloudbase-Init, see *Image Management Service User Guide*.

After the AS configuration is changed, you can use the key file to log in to the new ECS instances that are added to the AS group during scaling actions. You do not need to use the image password or key pair to log in to these new ECS instances anymore.

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

No. If you have enabled a load balancer for an AS group, you do not have to configure an EIP in the AS configuration. The system automatically associates 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 Do I Enable Automatic Initialization of EVS Disks on ECS Instances That Have Been Added to an AS Group During Scaling Actions?

Scenarios

After an ECS instance is created, you need to manually initialize EVS disks attached to the instance before using them. If multiple ECS instances are added to the AS group, you must initialize the EVS disks on each instance, which takes a while.

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

This section uses CentOS 6.5 as an example. For how to configure automatic initialization of EVS disks on other OSs, see the relevant OS documentation.

Procedure

1. Log in to an ECS instance as user **root**.
2. Run a command to switch to the directory where the script will be stored:
cd /script directory
For example:
cd /home
3. Run the following command to create the script:
vi script name
For example:

vi fdisk_mount.sh

4. Press **i** to enter editing mode.

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 does 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 enter editing mode.

The drive letter, size, and mount 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 **rc.local**:

```
/home/fdisk_mount.sh
```

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

11. Press **Esc**, enter **:wq**, and press **Enter** to save the changes and exit.
12. Create a private image from the .
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 instances added to this AS group in scaling actions will be automatically initialized.