ALIBABA CLOUD

Alibaba Cloud

Apsara Stack Enterprise

Auto Scaling

User Guide

Product Version: v3.16.2

Document Version: 20220727

Alibaba Cloud

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).

5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.

6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	**Danger:** Resetting will result in the loss of user configuration data.
Warning	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	**Warning:** Restarting will cause business interruption. About 10 minutes are required to restart an instance.
Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	**Notice:** If the weight is set to 0, the server no longer receives new requests.
Note	A note indicates supplemental instructions, best practices, tips, and other content.	**Note:** You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click **Settings**> **Network**> **Set network type**.
Bold	Bold formatting is used for buttons , menus, page names, and other UI elements.	Click **OK**.
`Courier font`	Courier font is used for commands	Run the `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`
{} or {a|b}	This format is used for a required value, where only one item can be selected.	`switch {active|stand}`

Table of Contents

1.What is Auto Scaling?

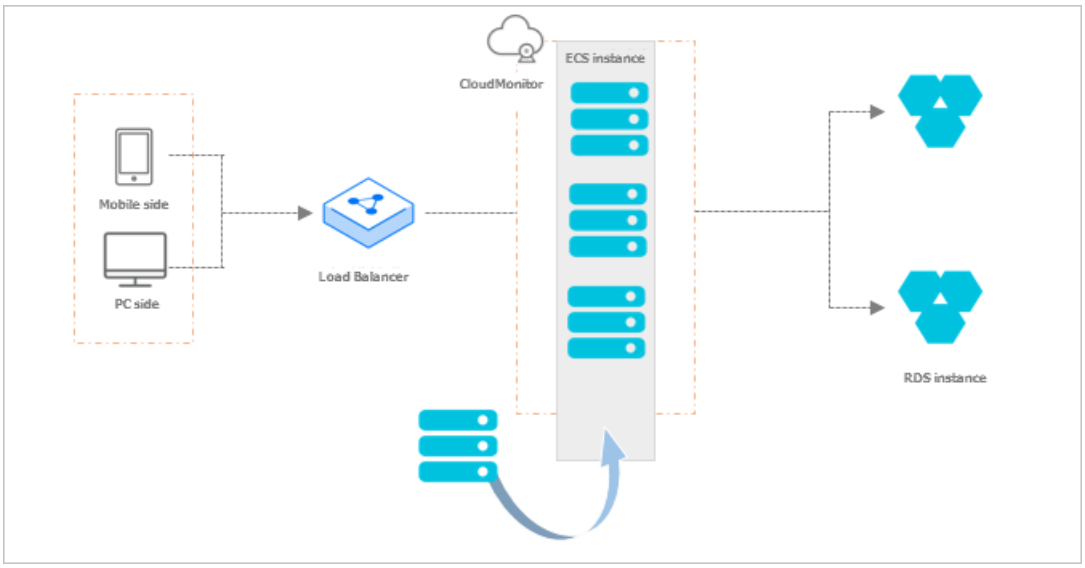
Auto Scaling automatically adjusts your elastic computing resources based on your business requirements and policies that you define.

When demand for services spikes, Auto Scaling automatically scales out Elastic Compute Service (ECS) instances based on your configurations to maintain sufficient computing resources. When demand for services drops, Auto Scaling automatically scales in ECS instances to save costs.

Auto Scaling provides the following features:

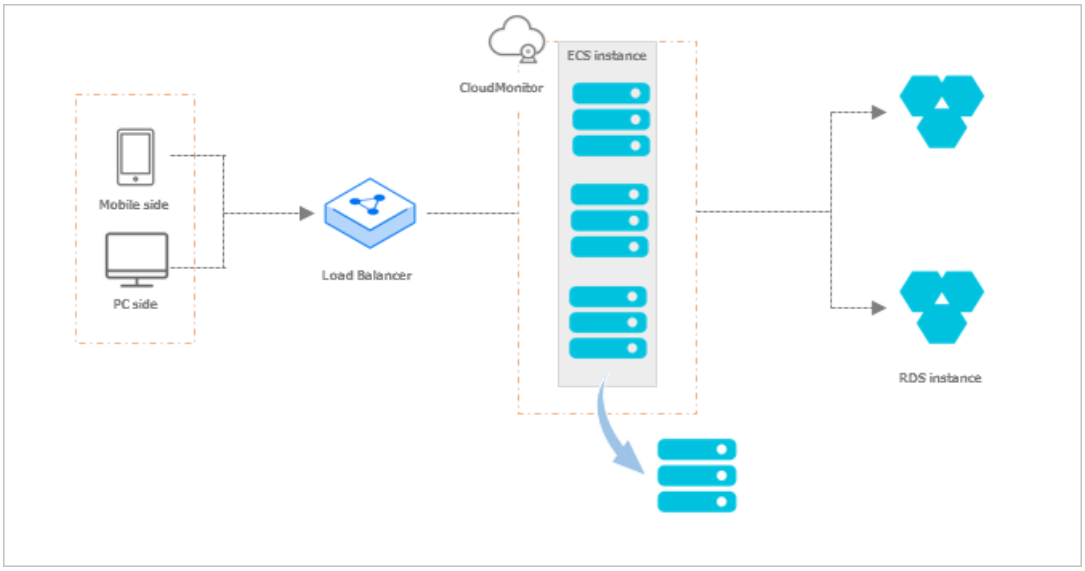
- Scale-out

  When demand for services suddenly grows, Auto Scaling automatically scales out the underlying resources. This ensures that resources are not overloaded and maintains the responsiveness of your servers. For example, if the vCPU utilization of ECS instances exceeds 80%, Auto Scaling scales out ECS resources based on your configurations. During the scale-out event, Auto Scaling automatically creates ECS instances, adds the ECS instances to a scaling group, and then adds the new instances to the backend server groups of the associated Server Load Balancer (SLB) instances and the whitelists of the associated ApsaraDB RDS instances. The following figure shows how a scale-out event is implemented.



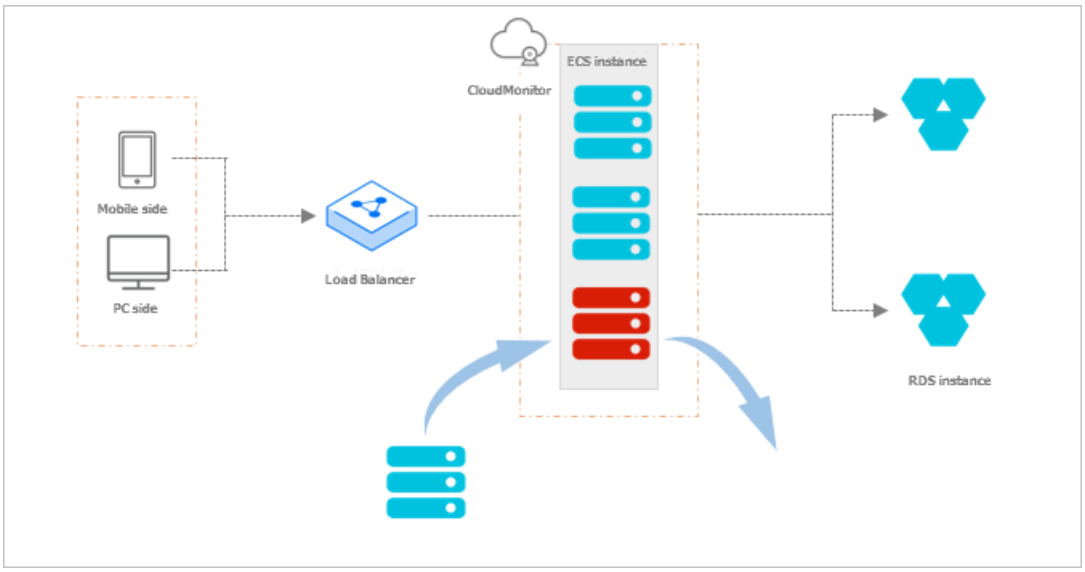
- Scale-in

  When demand for services drops, Auto Scaling automatically releases underlying resources to prevent waste of resources and reduce costs. For example, if the vCPU utilization of ECS instances in a scaling group falls below 30%, Auto Scaling automatically scales in ECS instances based on your configurations. During the scale-in event, Auto Scaling removes ECS instances from the scaling group and also from the backend server groups of the associated SLB instances and the whitelists of the associated ApsaraDB RDS instances. The following figure shows how a scale-in event is implemented.



- Elastic recovery

  If ECS instances in a scaling group are not in the Running state, Auto Scaling considers the instances to be unhealthy. If an ECS instance is considered unhealthy, Auto Scaling automatically releases the instance and creates a new one. This process is called elastic recovery. Elastic recovery ensures that the number of healthy ECS instances in a scaling group does not fall below the minimum number of ECS instances that you specified for the scaling group. The following figure shows how an elastic recovery event is implemented.



2.Notes

2.1. Precautions

This topic describes the precautions when you use Auto Scaling (ESS).

Scaling rules

ESS uses scaling rules to scale ECS instances in a scaling group based on the minimum and maximum numbers of ECS instances specified for the scaling group. Assume that a scaling group can contain up to 45 ECS instances. If you configure a scaling rule to increase the number of ECS instances in the scaling group to 50, ESS only increases the number of ECS instances to 45 at most.

Scaling activities

- Only one scaling activity can be executed at a time in a scaling group.
- An ongoing scaling activity cannot be terminated. For example, if a scaling activity is being executed to create 20 ECS instances but only five have been created, you cannot forcibly terminate the scaling activity.
- If some ECS instances fail to be added to a scaling group during a scaling activity, ESS considers that the scaling activity is complete without trying to add the failed instances to the scaling group. ESS rolls back the ECS instances that fails to be added but not the scaling activity. For example, if ESS has created 20 ECS instances for a scaling group, and 19 of the instances are added to SLB instances, only the one ECS instance that failed to be added is automatically released.

Cooldown period

- During the cooldown period, if you manually execute a scaling task, such as a scaling rule or scheduled task, the task is immediately executed without waiting for the cooldown period to expire.
- The cooldown period starts after the last ECS instance is added to or removed from a scaling group during a scaling activity.

2.2. Manual operations

If you perform a manual operation when an Auto Scaling operation is in progress, the manual operation takes precedence.

Auto Scaling supports manual operations, such as the deletion of automatically created Elastic Compute Service (ECS) instances from the ECS console. The following table describes how Auto Scaling processes manual operations.

Resource	Manual operation type	Processing method

Resource	Manual operation type	Processing method
ECS	A user deletes an ECS instance from a scaling group by using the ECS console or calling API operations.	Auto Scaling performs health checks to determine whether the ECS instance is unhealthy. If the instance is unhealthy, Auto Scaling removes it from the scaling group. The internal IP address of the ECS instance is not automatically deleted from the whitelist of the associated ApsaraDB RDS instance. If the total number of ECS instances in the scaling group falls below the lower limit after the ECS instance is removed, the scaling group automatically creates an ECS instance to ensure that the number of instances is at least equal to the lower limit.
ECS	A user revokes the ECS API permissions granted to Auto Scaling.	Auto Scaling rejects all scaling activity requests.
Server Load Balancer (SLB)	A user manually removes an ECS instance from the associated SLB instance by using the SLB console or calling API operations.	Auto Scaling does not automatically detect this operation or handle this exception. The ECS instance remains in the scaling group. When a scale-in activity is triggered, Auto Scaling releases the ECS instance if the instance meets the conditions specified in the removal policy.
SLB	A user manually deletes an SLB instance or disables the health check feature for an SLB instance by using the SLB console or calling API operations.	Auto Scaling does not add ECS instances to scaling groups that are associated with this SLB instance. Auto Scaling removes ECS instances from the scaling groups if a scaling task triggers a scale-in rule or the ECS instances are considered unhealthy after a health check is performed.
SLB	An SLB instance is unavailable due to a system error.	All scaling activities fail, except for instance removal tasks that are manually executed.
SLB	A user revokes the SLB API permissions granted to Auto Scaling.	Auto Scaling rejects all scaling activity requests for the scaling groups that are associated with SLB instances.
ApsaraDB RDS	A user manually removes the IP address of an ECS instance from the whitelist of the associated ApsaraDB RDS instance by using the ApsaraDB RDS console or calling API operations.	Auto Scaling does not automatically detect this operation or handle this exception. The ECS instance remains in the scaling group. When a scale-in activity is triggered, Auto Scaling releases the ECS instance if the instance meets the conditions specified in the removal policy.

Resource	Manual operation type	Processing method
ApsaraDB RDS	A user manually deletes an ApsaraDB RDS instance by using the ApsaraDB RDS console or calling API operations.	Auto Scaling does not add ECS instances that are associated with the ApsaraDB RDS instance to scaling groups. Auto Scaling removes ECS instances from the scaling groups if a scaling task triggers a scale-in rule or the ECS instances are considered as unhealthy after a health check is performed.
ApsaraDB RDS	An ApsaraDB RDS instance is unavailable due to a system error.	All scaling activities fail, except for instance removal tasks that are manually executed.
ApsaraDB RDS	A user revokes the ApsaraDB RDS API permissions granted to Auto Scaling.	Auto Scaling rejects all scaling activity requests for the scaling groups that are associated with ApsaraDB RDS instances.

2.3. Limits

This topic describes the limits of Auto Scaling.

- Auto Scaling can automatically scale the number of Elastic Compute Service (ECS) instances in a scaling group, but cannot automatically upgrade or downgrade configurations of the ECS instances, such as vCPUs, memory, and bandwidth.
- Applications deployed on the ECS instances in a scaling group must be stateless and horizontally scalable.
- ECS instances in a scaling group can be automatically released. We recommend that you do not store information such as sessions, application data, or logs on the ECS instances in a scaling group. If you need to store data of the applications deployed on the ECS instances, store status information such as sessions on the independent ECS instances, store application data in ApsaraDB RDS, and store logs in Log Service. For more information, see *What is ApsaraDB RDS?* in *ApsaraDB RDS Product Introduction* and *What is Log Service?* in *Log Service Product Introduction*.
- The following table describes the quantity limits that are applied to a scaling group.

Item	Quota
Scaling configuration	You can create a maximum of 10 scaling configurations in a scaling group.
Scaling rule	You can create a maximum of 50 scaling rules in a scaling group.
ECS instance	You can add a maximum of 1,000 ECS instances to a scaling group.

2.4. Scaling group status

This topic describes the states of a scaling group in the console and in an API operation.

State in the console	State in an API operation
Creating	Inactive
Created	Inactive
Enabling	Inactive
Enabled	Active
Disabling	Inactive
Disabled	Inactive
Deleting	Deleting

2.5. Scaling processes

Before you use Auto Scaling, you must understand the processes related to scaling activities.

Automatic scaling of a scaling group

- Automatic scale-out
  i. Check the health status and boundary conditions of the scaling group.
  ii. Assign the activity ID and execute the scaling activity.
  iii. Create ECS instances.
  iv. Modify Total Capacity.
  v. Assign IP addresses to the created ECS instances.
  vi. Add the ECS instances to the whitelist of the associated ApsaraDB RDS instance.
  vii. Start the ECS instances.
  viii. Associate the ECS instances with an SLB instance and set the weight to the SLB weight value that is specified when the scaling configuration is created.
  ix. The cooldown period starts after the scaling activity is complete.
- Automatic scale-in
  i. Check the health status and boundary conditions of the scaling group.
  ii. Assign the activity ID and execute the scaling activity.
  iii. Remove ECS instances from the associated SLB instance.
  iv. Stop the ECS instances.
  v. Remove the ECS instances from the whitelist of the associated ApsaraDB RDS instance.
  vi. Release the ECS instances.
  vii. Modify Total Capacity.
  viii. The cooldown period starts after the scaling activity is complete.

Manually add or remove existing ECS instances

- Manually add instances
  i. Check the health status and boundary conditions of the scaling group, and check the status and type of ECS instances.
  ii. Assign the activity ID and execute the scaling activity.
  iii. Add the ECS instances.
  iv. Modify Total Capacity.
  v. Add the ECS instances to the whitelist of the associated ApsaraDB RDS instance.
  vi. Associate the ECS instances with an SLB instance and set the weight to the SLB weight value that is specified in the active scaling configuration.

  > **Note** If you want to manually add an instance to a scaling group, the instance type of the instance must be the same as that specified in the active scaling configuration of the scaling group. Therefore, you must set the weight to the SLB weight value that is specified in the active scaling configuration.

  vii. The cooldown period starts after the scaling activity is complete.
- Manually remove instances
  i. Check the health status and boundary conditions of the scaling group.
  ii. Assign the activity ID and execute the scaling activity.
  iii. SLB stops forwarding traffic to ECS instances.
  iv. Remove the ECS instances from SLB after 60 seconds.
  v. Remove the ECS instances from the whitelist of the associated ApsaraDB RDS instance.
  vi. Modify Total Capacity.
  vii. Remove the ECS instances from the scaling group.
  viii. After the scaling activity is complete, the cooldown period starts.

2.6. Remove unhealthy ECS instances

Before you use ESS, you must understand information about the removal of unhealthy ECS instances.

After an ECS instance is added to a scaling group, ESS checks the status of the instance on a regular basis. If the ECS instance is not in the Running state, ESS removes the ECS instance from the scaling group. The removal method depends on how the ECS instance is added:

- If an ECS instance is automatically created, ESS immediately removes and releases it.
- If an ECS instance is manually added, ESS immediately removes it, but does not stop or release it.

The removal of unhealthy ECS instances is not limited by the MinSize value. After the unhealthy ECS instances are removed, the number of ECS instances (Total Capacity) may fall below the MinSize value. In this case, ESS automatically creates ECS instances based on the difference between the actual instance number and MinSize value to ensure that the total number of ECS instances is equal to the MinSize value.

2.7. Instance rollback after a failed scaling activity

Before you use ESS, you must understand the mechanism of instance rollback after a failed scaling activity.

If some ECS instances fail to be added to a scaling group during a scaling activity, ESS considers that the scaling activity is complete without trying to add the failed instances to the scaling group. ESS rolls back ECS instances, not the scaling activity.

For example, if a scaling group has created 20 ECS instances, and 19 of the instances are added to SLB instances, only the one ECS instance that failed to be added is automatically released.

2.8. Instance lifecycle management

Before you use Auto Scaling, we recommend that you understand the instance lifecycle.

Automatically created ECS instances

The ECS instances are automatically created by Auto Scaling based on user-defined scaling configurations and rules.

Auto Scaling manages the entire lifecycle of automatically created ECS instances. Auto Scaling creates ECS instances during scale-out activities, and stops and releases ECS instances during scale-in activities.

Manually added ECS instances

The ECS instances are manually added to scaling groups.

Auto Scaling does not manage the entire lifecycle of manually added ECS instances. These instances are not automatically created by Auto Scaling but are manually added by a user to a scaling group. If the ECS instances are manually or automatically removed from the scaling group, Auto Scaling removes the instances but does not stop or release them.

Instance status

An ECS instance in a scaling group can change to the following status during its lifecycle:

- Pending: The ECS instance is being added to the scaling group. The instance is being created, added as the backend server of the associated SLB instance, or added to the whitelist of the associated ApsaraDB RDS instance.
- InService: The ECS instance is added to the scaling group and is providing services as expected.
- Removing: The ECS instance is being removed from the scaling group.

Instance health status

An ECS instance in a scaling group can change to the following health status:

- Healthy
- Unhealthy

If an ECS instance is not in the Running state, Auto Scaling considers the instance unhealthy and automatically removes it from the scaling group.

- Auto Scaling stops and releases automatically created ECS instances.
- Auto Scaling does not stop or release manually added ECS instances.

3.Quick start

3.1. Overview

This topic describes how to get started with Auto Scaling.

You can perform the following steps to get started with Auto Scaling:

1. Create a scaling group

   Configure the parameters for the scaling group, such as Maximum Number of Instances and Minimum Number of Instances.

2. Create a scaling configuration

   Configure the parameters for the scaling configuration, such as Instance Type and Image.

3. Enable a scaling group

   Enable the scaling group for which the scaling configuration is enabled.

4. Create a scaling rule

   Add ECS instances to or remove ECS instances from the scaling group.

5. Create a scheduled task

   Create scheduled tasks to add or remove instances at a specified point in time. Auto Scaling executes the scheduled tasks and scaling rules at the specified time. For example, Auto Scaling can trigger a task to execute a specified scaling rule at 08:00.

3.2. Log on to the Auto Scaling console

This topic describes how to log on to the Auto Scaling console.

Prerequisites

- The URL of the Apsara Uni-manager Management Console is obtained from the deployment personnel before you log on to the Apsara Uni-manager Management Console.
- A browser is available. We recommend that you use the Google Chrome browser.

Procedure

1. In the address bar, enter the URL of the Apsara Uni-manager Management Console. Press the Enter key.
2. Enter your username and password.

   Obtain the username and password that you can use to log on to the console from the operations administrator.

> **Note** When you log on to the Apsara Uni-manager Management Console for the first time, you must change the password of your username. Your password must meet complexity requirements. The password must be 8 to 20 characters in length and must contain at least two of the following character types:
>
> - Uppercase or lowercase letters
> - Digits
> - Special characters, which include ! @ # $ %

3. Click **Login**.
4. If your account has multi-factor authentication (MFA) enabled, perform corresponding operations in the following scenarios:
   - It is the first time that you log on to the console after MFA is forcibly enabled by the administrator.
     a. On the Bind Virtual MFA Device page, bind an MFA device.
     b. Enter the account and password again as in Step 2 and click **Log On**.
     c. Enter a six-digit MFA verification code and click **Authenticate**.
   - You have enabled MFA and bound an MFA device.

     Enter a six-digit MFA authentication code and click **Authenticate**.

     > **Note** For more information, see the *Bind a virtual MFA device to enable MFA* topic in *Apsara Uni-manager Operations Console User Guide*.
5. In the top navigation bar, choose **Products > Elastic Computing > Auto Scaling**.

3.3. Create a scaling group

This topic describes how to create a scaling group. A scaling group is a group of Elastic Compute Service (ECS) instances that can be used in similar business scenarios. When you create a scaling group, you can specify the minimum and maximum numbers of ECS instances that are allowed in the scaling group.

Prerequisites

- A virtual private cloud (VPC) and a vSwitch are created. For more information, see the "Create a VPC" and "Create a vSwitch" topics in *VPC User Guide*.
- To associate your scaling group with a Server Load Balancer (SLB) instance, make sure that the following requirements are met:
  - You have one or more SLB instances in the **Running** state. For information about how to create an SLB instance, see the *Create an SLB instance* topic in *SLB User Guide*.
  - The SLB instance and the scaling group are in the same organization, resource set, and region.
  - If the network type of the SLB instance and the scaling group is VPC, the SLB instance and the scaling group must be in the same VPC.
  - If the network type of the SLB instance is classic network, the network type of the scaling group is VPC, and the backend server group of the SLB instance contains ECS instances of the VPC type, the ECS instances and the scaling group must be in the same VPC.

- At least one listener is configured for the SLB instance. For more information about listeners, see the *Listener overview* topic in *SLB User Guide*.
- The health check feature is enabled for the SLB instance. For more information about how to enable the health check feature, see the *Configure health checks* topic in *SLB User Guide*.

- To associate your scaling group with an ApsaraDB RDS instance, make sure that the following requirements are met:
  - You have one or more ApsaraDB RDS instances that are in the **Running** state. For more information about ApsaraDB RDS, see the *What is ApsaraDB RDS?* topic in *ApsaraDB RDS Product Introduction*.
  - The ApsaraDB RDS instance and the scaling group are in the same organization, resource set, and region.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scaling Groups** page, click **Create Scaling Group**.
5. Configure the parameters for the scaling group. The following table describes the parameters.

Parameter	Required	Description
Scaling Group Name	Yes	The name of the scaling group. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Organization/Resource Group	Yes	The organization and the resource set to which the scaling group belongs.
Region	Yes	The region where you want to create the scaling group.
Maximum Number of Instances (Units)	Yes	The maximum number of instances that the scaling group can contain. To minimize costs, specify a value based on your business requirements. Valid values: 0 to 1000.
Minimum Number of Instances (Units)	Yes	The minimum number of instances that a scaling group must contain. To ensure service availability, specify a value based on your business requirements. When a scaling group is enabled, Auto Scaling automatically creates a specific number of instances in a scaling group to maintain the minimum requirement. Valid values: 0 to 1000.

Parameter	Required	Description
Default Cooldown Time (Seconds)	Yes	The cooldown time after each scaling activity is complete. During the cooldown time, Auto Scaling rejects all scaling requests of event-triggered tasks. However, scaling requests that are triggered by other types of tasks, such as manually triggered tasks and scheduled tasks, can be processed even before the cooldown time expires. The value must be an integer that is greater than or equal to zero. Unit: Seconds.
Removal Policy	No	The policy that is used to automatically remove the ECS instances from the scaling group. This parameter contains the **First Filter** and **Then Filter** fields. You cannot specify the same values for both fields. Valid values of the two fields: ○ **Instances Created From the Earliest Scaling Configuration**: filters instances that are created based on the earliest scaling configuration. ○ **Earliest Added ECS Instances**: filters instances that are added to the scaling group at the earliest point in time. ○ **Most Recently Added Instances**: filters instances that are added to the scaling group at the most recent point in time. ○ **No Policy**: specifies that Auto Scaling does not filter instances from the instances that are obtained based on the First Filter field. This value is available only for the **Then Filter** field. For example, if Auto Scaling filters instances based on the **Earliest Added ECS Instances** value of the First Filter field, you can select only one of the following values for the Then Filter field: ○ **No Policy**: Auto Scaling does not remove instances from the instances that are obtained based on the First Filter field. ○ **Most Recently Added ECS Instances**: filters the most recently added instances from the instances that are obtained based on the First Filter field. ○ **Instances Created From the Earliest Scaling Configuration**: filters instances that are created based on the earliest scaling configuration from the instances that are obtained based on the First Filter field.
VPC ID	Yes	The ID of the VPC in which you want to create the scaling group.

Parameter	Required	Description
vSwitch	Yes	The vSwitch with which you want to associate the scaling group.
Associate SLB Instance	No	After you associate an SLB instance with the scaling group, the ECS instances that are added to the scaling group are automatically added as the backend servers of the associated SLB instance. You can specify one of the following server groups to which you want to add the ECS instances: ◦ Default server group: a group of ECS instances that are used to receive requests. If you do not specify a vServer group or a primary/secondary server group for a listener, requests are forwarded to the ECS instances in the default server group. ◦ vServer group: If you want to forward requests to backend servers that are not in the default server group or configure domain name-based or URL-based routing methods, you can use vServer groups.
Associate ApsaraDB RDS Instance	No	After you associate an ApsaraDB RDS instance with the scaling group, the private IP addresses of the ECS instances that are added to the scaling group are automatically added to the whitelist that manages access to the ApsaraDB RDS instance. This way, the ApsaraDB RDS instance and the ECS instances can communicate with each other over the internal network.

6. Click **OK**.

Result

The scaling group that you created is displayed in the scaling group list but is in the **Disable** state. You must create and enable a scaling configuration for the scaling group. For more information, see Create a scaling configuration.

3.4. Create a scaling configuration

This topic describes how to create a scaling configuration for a scaling group. Auto Scaling uses the instance types that are specified in the scaling configuration to create instances.

Prerequisites

A security group is available in the virtual private cloud (VPC) where your scaling group for which you want to create a scaling configuration resides. For more information, see the *Create a security group* topic in *ECS User Guide*.

Context

You can create only a limited number of scaling configurations for a scaling group. For more information, see the *Limits* topic in *Auto Scaling Product Introduction*.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instance Configuration Sources** tab.
6. Click **Create Scaling Configuration**.
7. Configure the parameters for the scaling configuration. The following table describes the parameters.

Section	Parameter	Required	Description
Region	Region	Yes	The region where you want to create Elastic Compute Service (ECS) instances based on the scaling configuration.
	Zone	Yes	The zone where you want to create ECS instances based on the scaling configuration.
Security Group	Security Group	Yes	The security group with which you want to associate ECS instances that you want to create based on the scaling configuration.
Instance	Instance Type	Yes	The instance type that you want to use to create ECS instances.

Section	Parameter	Required	Description
Image	Image Type	Yes	◦ **Public Image**: You can select public images that are provided by Alibaba Cloud. Public images are licensed to provide a secure and stable operating environment for applications on ECS instances. ◦ **Custom Image**: You can create custom images that you can use to install software or deploy projects that have special requirements. ◦ **Shared Custom Image**: You can select a shared custom image.
Storage	System Disk (GB)	Yes	The category and size of the system disk. The OS is installed on the system disk. You can select **Ultra Disk** or **Standard SSD**.
	Data Disk (GB)	No	The category and size of the data disk. You can select **Ultra Disk** or **Standard SSD**. You can add up to 16 data disks. The maximum capacity of each data disk is 32 TiB. You can select **Release with Instance** and **Encryption** for each data disk.
	Password Setting	Yes	The time to specify a password. Valid values: **Set Now** and **Set After Purchase**. If you set the Password Setting parameter to Set After Purchase, you must reset the password in the console after the scaling configuration is created. For more information, see the *Change the logon password* topic in *ECS User Guide*.

Section	Parameter	Required	Description
Password	Logon Password	No	The password used to log on to the ECS instances that are created based on the scaling configuration. You must configure this parameter if you set the Password Setting parameter to **Set Now**. The logon password is subject to the image type that you select. For more information, see the *Reset the logon password of an instance* topic in *ECS User Guide*. **Note** The password is used to log on to the OS and is not the VNC password.
	Confirm Password	No	You must enter the password again after you configure the Logon Password parameter.
Deployment Set	Deployment Set	No	The deployment set to which the instances that are created based on the scaling configuration belong. You can select a deployment set from the Deployment Set drop-down list. You can also create a deployment set.
Instance Name	Scaling Configuration	No	The name of the scaling configuration that you want to create.
	Instance Name	No	The name of the ECS instance that you want to create based on the scaling configuration.
User Data	User Data	No	The Windows OS supports batch and PowerShell scripts. Before you perform Base64 encoding of user data, make sure that the first line of the data is included in `[bat]` or `[powershell]`. The Linux OS supports shell scripts.

Section	Parameter	Required	Description
Instances	Quantity	No	The number of instances that you want to create based on the scaling configuration.

8. Click **Submit**.

Result

After the scaling configuration is created, the scaling configuration enters the **Disable** state and is displayed in your scaling configuration list. You can click Apply in the Actions column to enable the scaling configuration. Auto Scaling can create ECS instances in a scaling group that has an enabled scaling configuration. For more information, see Apply a scaling configuration.

3.5. Enable a scaling group

This topic describes how to enable a scaling group. Auto Scaling can trigger scaling activities and create instances in a scaling group only after the scaling group is enabled.

Prerequisites

- The scaling group is in the **Disable** state.
- The scaling group has a scaling configuration that is in the **Enable** state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the scaling group that you want to enable and click **Enable** in the **Actions** column.
5. Click **OK**.

Result

The status of the scaling group is changed from **Disable** to **Enable** in the **Status** column.

3.6. Create a scaling rule

This topic describes how to create a scaling rule. You can specify the number of instances that you want to add to or remove from a scaling group. For example, you can create a scaling rule to add one instance to a scaling group.

Context

- You can create only a limited number of scaling rules for a scaling group. For more information, see the *Limits* topic in *Auto Scaling Product Introduction*.
- If the total number of instances in a scaling group is out of the specified range after a scaling rule is executed, Auto Scaling creates or removes a specific number of instances to keep the total number within the specified range.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Scaling Rules** tab.
6. Click **Create Scaling Rule**.
7. Configure the parameters for the scaling rule. The following table describes the parameters.

Parameter	Required	Description
Rule Name	Yes	The name of the scaling rule. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Scaling Activity	Yes	Valid values: ○ **To N Units**: changes the number of instances in the scaling group to N. ○ **Add N Units**: adds N instances to the scaling group. ○ **Remove N Units**: removes N instances from the scaling group. **Note** Up to 500 instances can be added to or removed from a scaling group during a scaling activity.
Cooldown Time	Yes	The period of cooldown time after a scaling activity is complete. If you do not specify a value for this parameter, the default value is used.

8. Click **OK**.

Result

On the Scaling Rules tab, you can view the information about the new scaling rule, such as the scaling rule name or ID, rule type, time when the rule is executed, and action that is specified in the rule.

3.7. Create a scheduled task

This topic describes how to create a scheduled task. If your business workloads are predictable, you can create a scheduled task to prepare sufficient computing resources before the business workloads increase and release excess computing resources after the business workloads decrease.

Context

- A scheduled task allows Auto Scaling to execute a scaling rule at the specified point in time. This way, computing resources can be automatically scaled to meet your business requirements and minimize the costs of resources. You can also specify an interval at which a scheduled task is executed to cope with the changes of business workloads in an efficient manner.
- If multiple scheduled tasks need to be executed within the same minute, Auto Scaling executes the most recently created scheduled task.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Scheduled Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scheduled Tasks** page, click **Create Scheduled Task**.
5. In the dialog box that appears, configure the parameters. The following table describes the parameters.

Parameter	Required	Description
Name	Yes	The name of the scheduled task. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Description	Yes	The description of the scheduled task.
Organization/Resource Set	Yes	The organization and resource set in which you want to create the scheduled task.
Start From	Yes	The point in time at which you want to execute the scheduled task.
Monitoring Resource/Scaling Rule	Yes	The scaling group for which you want to create the scheduled task and the scaling rule that you want to specify in the scheduled task.
Retry Expiry Time (Seconds)	No	The timeout period during which Auto Scaling retries the scheduled task. Unit: Seconds. If Auto Scaling does not execute the scheduled task at the specified point in time, Auto Scaling retries the scheduled task during the timeout period.
Recurrence Interval Settings (Advanced)	No	Specifies whether to execute the scheduled task on a recurring schedule. After you select **Recurrence Interval Settings (Advanced)**, you must configure the **Recurrence Interval** and **Expires At** parameters. ○ Recurrence Interval: You can select **By Day**, **By Week**, or **By Month**. ○ Expires At: You can use the Select Date or Select Time methods.

6. Click **OK**.

Result

On the Scheduled Tasks page, the scheduled task that you created appears. You can view the information about the scheduled task, such as the scaling group for which the scheduled task is created, scaling rule that is specified in the scheduled task, task status, and task description.

3.8. Create an event-triggered task

This topic describes how to create an event-triggered task. If your business workloads are unpredictable or may surge unexpectedly, you can create an event-triggered task. You must specify a metric that is monitored by CloudMonitor in an event-triggered task. After you create and enable an event-triggered task, Auto Scaling collects data for the specified metric in real time and CloudMonitor triggers an alert when the specified condition is met. Then, Auto Scaling executes the scaling rule that is associated with the event-triggered task to scale instances in your scaling group.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Event-Triggered Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Event-triggered Tasks** page, click **Create Event-triggered Task**.
5. In the Create Event-triggered Task dialog box, configure the parameters. The following table describes the parameters.

Parameter	Required	Description
Task Name	Yes	The name of the event-triggered task. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Description	No	The description of the event-triggered task.
Organization/Resource Set	Yes	The organization and the resource set in which you want to create the event-triggered task.
Monitoring Resource/Scaling Rule	Yes	The scaling group for which you want to create the event-triggered task and the scaling rule that you want to specify in the event-triggered task.
Monitor Type	Yes	Valid value: **System Monitoring**.
Metric	Yes	The metric that you want to specify in the event-triggered task. Valid values: ○ **Average CPU Usage** ○ **Memory Usage** ○ **Outbound Traffic over Internal Network** ○ **Inbound Traffic over Internal Network** ○ **Average System Load**

Parameter	Required	Description
Statistical Period	Yes	The period during which data is collected, aggregated, and analyzed. The shorter the period, the more frequently the alert is triggered. Unit: Minutes. Valid values: ○ 1 ○ 2 ○ 5 ○ 15
Statistical Method	Yes	The rule that determines whether to trigger an alert. Select **Average**, **Max Capacity**, or **Min Capacity**, and specify a threshold. For example, you can specify one of the following rules to trigger an alert when the CPU utilization exceeds 80%: ○ Average: triggers an alert when the average CPU utilization of all instances in the scaling group exceeds 80%. ○ Max Capacity: triggers an alert when the highest CPU utilization among all instances in the scaling group exceeds 80%. ○ Min Capacity: triggers an alert when the lowest CPU utilization among all instances in the scaling group exceeds 80%.
Trigger Alert After {1} Times	Yes	The number of consecutive times that the threshold must be exceeded before the alert is triggered. Valid values: ○ 1 ○ 2 ○ 3 ○ 5

6. Click **OK**.

Result

On the Event-triggered Tasks page, the event-triggered task that you created appears. You can view the information about the event-triggered task, such as the monitoring type, statistical period, alert condition, and task description.

4.Scaling groups

4.1. Create a scaling group

This topic describes how to create a scaling group. A scaling group is a group of Elastic Compute Service (ECS) instances that can be used in similar business scenarios. When you create a scaling group, you can specify the minimum and maximum numbers of ECS instances that are allowed in the scaling group.

Prerequisites

- A virtual private cloud (VPC) and a vSwitch are created. For more information, see the "Create a VPC" and "Create a vSwitch" topics in *VPC User Guide*.
- To associate your scaling group with a Server Load Balancer (SLB) instance, make sure that the following requirements are met:
  - You have one or more SLB instances in the **Running** state. For information about how to create an SLB instance, see the *Create an SLB instance* topic in *SLB User Guide*.
  - The SLB instance and the scaling group are in the same organization, resource set, and region.
  - If the network type of the SLB instance and the scaling group is VPC, the SLB instance and the scaling group must be in the same VPC.
  - If the network type of the SLB instance is classic network, the network type of the scaling group is VPC, and the backend server group of the SLB instance contains ECS instances of the VPC type, the ECS instances and the scaling group must be in the same VPC.
  - At least one listener is configured for the SLB instance. For more information about listeners, see the *Listener overview* topic in *SLB User Guide*.
  - The health check feature is enabled for the SLB instance. For more information about how to enable the health check feature, see the *Configure health checks* topic in *SLB User Guide*.
- To associate your scaling group with an ApsaraDB RDS instance, make sure that the following requirements are met:
  - You have one or more ApsaraDB RDS instances that are in the **Running** state. For more information about ApsaraDB RDS, see the *What is ApsaraDB RDS?* topic in *ApsaraDB RDS Product Introduction*.
  - The ApsaraDB RDS instance and the scaling group are in the same organization, resource set, and region.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scaling Groups** page, click **Create Scaling Group**.
5. Configure the parameters for the scaling group. The following table describes the parameters.

Parameter	Required	Description

Parameter	Required	Description
Scaling Group Name	Yes	The name of the scaling group. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Organization/Resource Group	Yes	The organization and the resource set to which the scaling group belongs.
Region	Yes	The region where you want to create the scaling group.
Maximum Number of Instances (Units)	Yes	The maximum number of instances that the scaling group can contain. To minimize costs, specify a value based on your business requirements. Valid values: 0 to 1000.
Minimum Number of Instances (Units)	Yes	The minimum number of instances that a scaling group must contain. To ensure service availability, specify a value based on your business requirements. When a scaling group is enabled, Auto Scaling automatically creates a specific number of instances in a scaling group to maintain the minimum requirement. Valid values: 0 to 1000.
Default Cooldown Time (Seconds)	Yes	The cooldown time after each scaling activity is complete. During the cooldown time, Auto Scaling rejects all scaling requests of event-triggered tasks. However, scaling requests that are triggered by other types of tasks, such as manually triggered tasks and scheduled tasks, can be processed even before the cooldown time expires. The value must be an integer that is greater than or equal to zero. Unit: Seconds.

Parameter	Required	Description
Removal Policy	No	The policy that is used to automatically remove the ECS instances from the scaling group. This parameter contains the **First Filter** and **Then Filter** fields. You cannot specify the same values for both fields. Valid values of the two fields: ○ **Instances Created From the Earliest Scaling Configuration**: filters instances that are created based on the earliest scaling configuration. ○ **Earliest Added ECS Instances**: filters instances that are added to the scaling group at the earliest point in time. ○ **Most Recently Added Instances**: filters instances that are added to the scaling group at the most recent point in time. ○ **No Policy**: specifies that Auto Scaling does not filter instances from the instances that are obtained based on the First Filter field. This value is available only for the **Then Filter** field. For example, if Auto Scaling filters instances based on the **Earliest Added ECS Instances** value of the First Filter field, you can select only one of the following values for the Then Filter field: ○ **No Policy**: Auto Scaling does not remove instances from the instances that are obtained based on the First Filter field. ○ **Most Recently Added ECS Instances**: filters the most recently added instances from the instances that are obtained based on the First Filter field. ○ **Instances Created From the Earliest Scaling Configuration**: filters instances that are created based on the earliest scaling configuration from the instances that are obtained based on the First Filter field.
VPC ID	Yes	The ID of the VPC in which you want to create the scaling group.
vSwitch	Yes	The vSwitch with which you want to associate the scaling group.

Parameter	Required	Description
Associate SLB Instance	No	After you associate an SLB instance with the scaling group, the ECS instances that are added to the scaling group are automatically added as the backend servers of the associated SLB instance. You can specify one of the following server groups to which you want to add the ECS instances: ○ Default server group: a group of ECS instances that are used to receive requests. If you do not specify a vServer group or a primary/secondary server group for a listener, requests are forwarded to the ECS instances in the default server group. ○ vServer group: If you want to forward requests to backend servers that are not in the default server group or configure domain name-based or URL-based routing methods, you can use vServer groups.
Associate ApsaraDB RDS Instance	No	After you associate an ApsaraDB RDS instance with the scaling group, the private IP addresses of the ECS instances that are added to the scaling group are automatically added to the whitelist that manages access to the ApsaraDB RDS instance. This way, the ApsaraDB RDS instance and the ECS instances can communicate with each other over the internal network.

6. Click **OK**.

Result

The scaling group that you created is displayed in the scaling group list but is in the **Disable** state. You must create and enable a scaling configuration for the scaling group. For more information, see Create a scaling configuration.

4.2. Enable a scaling group

This topic describes how to enable a scaling group. Auto Scaling can trigger scaling activities and create instances in a scaling group only after the scaling group is enabled.

Prerequisites

- The scaling group is in the **Disable** state.
- The scaling group has a scaling configuration that is in the **Enable** state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. Find the scaling group that you want to enable and click **Enable** in the **Actions** column.
5. Click **OK**.

Result

The status of the scaling group is changed from **Disable** to **Enable** in the **Status** column.

4.3. View scaling groups

This topic describes how to view the list of scaling groups and the information about a specific scaling group.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Enter a scaling group name or ID in the search box and click **Search**.

   You can specify multiple filter options to search for scaling groups.

Filter option	Description
Scaling Group Name	You can search for a scaling group by name.
Scaling Group ID	You can search for a scaling group by ID. Separate multiple scaling group IDs with commas (,).

5. Click the name of the desired scaling group in the **Scaling Group Name/ID** column.
6. On the page that appears, view the information about the scaling group.

Tab	Description
Basic Information	◦ The Basic Information section provides the basic information about the scaling group, such as the scaling group ID, scaling group name, point in time when the scaling group was created, and point in time when the scaling group was last modified. ◦ The Auto Scaling Architecture: ECS Instances section provides information about the instances in the scaling group, such as the maximum number of instances, the minimum number of instances, cooldown time (seconds), and instance status. ◦ The Scaling Group Information section provides information about the configurations of the scaling group, such as the total number of instances, removal policy, network configurations, instance configuration source, and instance reclaim mode.
Instances	The Instances tab provides information about the instances in the scaling group, such as the list of automatically created instances, the list of manually added instances, the number of instances that are in the In Service state, and the number of instances that are in the Disabled state.

Tab	Description
Instance Configuration Sources	The Instance Configuration Sources tab provides information about all scaling configurations in the scaling group. On the Instance Configuration Sources tab, you can create, modify, apply, or delete scaling configurations.
Scaling Activity	The Scaling Activity tab provides information about all scaling activities that are executed in the scaling group, such as the total number of instances after scaling, start time, end time, description or reason, status information, and status.
Scaling Rules	The Scaling Rules tab provides information about all scaling rules that are applied to the scaling group. On the Scaling Rules tab, you can create, execute, modify, or delete scaling rules.

4.4. Modify a scaling group

This topic describes how to modify a scaling group. For example, you can modify the Maximum Number of Instances (Units) and Minimum Number of Instances (Units) settings of a specific scaling group.

Context

After you modify the minimum or maximum number of instances that a scaling group can contain, if the number of instances in the scaling group is outside the specified range, Auto Scaling automatically creates or removes instances until the number of instances is within the range.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the scaling group that you want to modify and click **Edit** in the **Actions** column.
5. Modify the parameter settings of the scaling group.

   You can modify only the scaling configuration and other parameters. You cannot change the organization and resource set. For more information about other parameters, see Create a scaling group.
6. Click **OK**.

Result

After you modify a scaling group, you can view the new settings of the scaling group.

4.5. Disable a scaling group

This topic describes how to disable a scaling group that you no longer need.

Prerequisites

- Make sure that the scaling group that you want to disable does not have scaling activities in progress.
- The scaling group is in the **Enable** state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the scaling group that you want to disable and click ••• in the **Actions** column.
5. In the message that appears, click **OK**.

Result

The status of the scaling group is changed from **Enable** to **Disable** in the **Status** column.

4.6. Delete a scaling group

This topic describes how to permanently delete a scaling group. When you delete a scaling group, Auto Scaling removes and releases instances that are automatically created, removes instances that are manually added, and deletes the scaling configurations and scaling rules. However, the scheduled tasks and event-triggered tasks that are associated with the scaling group are not deleted.

Prerequisites

The scaling group is no longer needed.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Find the scaling group that you want to delete and click **Delete** in the **Actions** column.
5. In the message that appears, click **OK**.

Result

After you delete the scaling group and refresh the Scaling Groups page, the scaling group does not appear.

5.Instances

5.1. Query instances

This topic describes how to query all instances in a scaling group and the status of the instances.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. View the information about the instances.

Information	Description
Auto Create	The instances on the Auto Create tab are created by Auto Scaling. Auto Scaling executes the specified scaling rule and creates instances by using the scaling configuration that you enabled for the scaling group. On the **Auto Create** tab, you can manage the automatically created instances. For example, you can put an instance into the Standby or Protected state. You can remove an instance from the scaling group. You can also remove and delete an instance.
Add Manually	The instances on the Add Manually tab are not created by Auto Scaling. These instances are the instances that you manually add to the scaling group. On the **Add Manually** tab, you can manually manage the instances. For example, you can put an instance into the Standby or Protected state. You can remove an instance from the scaling group. You can also remove and delete an instance.
Number of instances in different states	○ Total: the total number of instances in the scaling group. ○ In Service: the number of instances that are in the In Service state. ○ Standby: the number of instances that are in the Standby state. ○ Protected: the number of instances that are in the Protected state. ○ Adding: the number of instances that are being added to the scaling group. ○ Removing: the number of instances that are being removed from the scaling group. **Note** The Stopped, Pending Add, and Pending Remove states are unavailable.

5.2. Manually add instances to a scaling group

This topic describes how to manually add instances to a scaling group. You can add existing instances to a scaling group to utilize computing resources.

Prerequisites

- The instances that you want to add to a scaling group must meet the following requirements:
  - The instances that you want to add and the scaling group to which you want to add the instances are in the same region, organization, and resource set.
  - The instances are in the **Running** state.
  - The instances do not belong to other scaling groups.
  - The instances and the scaling group are in the same virtual private cloud (VPC).
- The scaling group to which you want to add instances must meet the following requirements:
  - The scaling group is in the **Enable** state.
  - The scaling group has no scaling activities in progress.

Context

- You can manually add instances to a scaling group even before the cooldown time expires.
- If the number of instances in the scaling group is greater than the maximum number of instances after you manually add instances to the scaling group, you cannot add the instances.
- The instance types of the instances that you want to manually add to a scaling group can be different from the instance types specified in the scaling configuration that is in the **Enable** state of the scaling group.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. On the **Add Manually** tab, click **Add Existing Instance**.
7. In the Manually Add ECS Instance dialog box, select the instances that you want to add and click **OK**.

Result

After you manually add instances to the scaling group, the instances appear on the **Add Manually** tab.

5.3. Put an instance into the Standby state

This topic describes how to put an instance into the Standby state. Auto Scaling does not release or perform health checks on instances that are in the Standby state.

Context

After you put an instance into the Standby state, the following rules apply:

- The instance remains in the Standby state until you manually move it out of the Standby state.
- Auto Scaling stops managing the lifecycle of the instance. You must manually manage the lifecycle of the instance.
- If a scale-in activity is triggered, Auto Scaling does not remove the instance from the scaling group.
- When the instance is stopped or restarted, the health check status of the instance is not updated.
- To release the instance, you must first remove the instance from the scaling group.
- If you delete the scaling group, the instance is automatically removed from the Standby state and subsequently released.
- You can perform the following operations on the instance: stop the instance, restart the instance, change the instance type, and change the OS.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. Select the source of an instance.
   - To select an automatically created instance, click the **Auto Created** tab.
   - To select a manually added instance, click the **Manually Added** tab.
7. Find the instance that you want to put into the Standby state and click **Switch to Standby** in the **Actions** column.
8. Click **OK**.

5.4. Move an instance out of the Standby state

You can move an instance out of the Standby state to reuse the instance.

Context

After you move an instance out of the Standby state, the following rules apply:

- The instance enters the In Service state.

- When the instance is stopped or restarted, the health status of the instance is updated.
- Auto Scaling continues to manage the lifecycle of the instance, and can remove the instance from the scaling group during a scale-in activity.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. Specify the source of instances.
   - To select an automatically created instance, click the **Auto Create** tab.
   - To select a manually added instance, click the **Add Manually** tab.
7. Find the instance that you want to manage and click **Remove from Standby** in the **Actions** column.
8. Click **OK**.

5.5. Put an instance into the Protected state

If you do not want an instance to be removed from your scaling group, you can put the instance into the Protected state. Auto Scaling does not check the health status of instances that are in the Protected state or release the instances.

Context

After you put an instance into the Protected state, the following rules apply:

- The instance remains in the Protected state until you manually move it out of the Protected state.
- If a scale-in activity is triggered, Auto Scaling does not remove the instance from the scaling group. The instance can be released only after you move it out of the Protected state.
- When the instance is stopped or restarted, the health check status of the instance is not updated.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. Select the source of an instance.
   - To select an automatically created instance, click the **Auto Created** tab.
   - To select a manually added instance, click the **Manually Added** tab.
7. Find the instance that you want to manage and click **Switch to Protected** in the **Actions**

column.

8. Click **OK**.

5.6. Move an instance out of the Protected state

You can move an instance out of the Protected state. This way, Auto Scaling manages the lifecycle of the instance.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. Specify the source of instances.
   - To select an automatically created instance, click the **Auto Create** tab.
   - To select a manually added instance, click the **Add Manually** tab.
7. Find the instance that you want to manage and click **Remove from Protected** in the **Actions** column.
8. Click **OK**.

5.7. Manually remove an instance from a scaling group

If you no longer need an instance, you can manually remove the instance from your scaling group. However, the instance is not released.

Prerequisites

Before you manually remove an instance from a scaling group, make sure that the following conditions are met:

- The scaling group is in the **Enable** state.
- No scaling activity is in progress in the scaling group.

Context

- You can manually remove an instance from a scaling group without the need to wait for the cooldown time to expire.
- After you remove an instance from a scaling group, the number of instances in the scaling group must be greater than or equal to the minimum number of instances allowed. Otherwise, you cannot remove the instance from the scaling group.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. Select the source of an instance.
   - To select an automatically created instance, click the **Auto Created** tab.
   - To select a manually added instance, click the **Manually Added** tab.
7. Find the instance that you want to remove from the scaling group and click **Remove from Scaling Group** in the **Actions** column.
8. Click **OK**.

5.8. Manually delete an instance

If you no longer need an instance, you can manually remove the instance from your scaling group and permanently delete the instance.

Prerequisites

Before you delete an instance in a scaling group, make sure that the following conditions are met:

- The scaling group is in the **Enable** state.
- No scaling activity is in progress in the scaling group.

Context

- You can manually remove an instance from a scaling group and delete the instance without the need to wait for the cooldown time to expire.
- After you remove an instance from a scaling group and delete the instance, the number of instances in the scaling group must be greater than or equal to the minimum number of instances allowed. Otherwise, you cannot delete the instance.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instances** tab.
6. Select the source of an instance.
   - To select an automatically created instance, click the **Auto Created** tab.
   - To select a manually added instance, click the **Manually Added** tab.
7. Find the instance that you want to delete and click **Remove and Delete** in the **Actions** column.
8. Click **OK**.

6.Scaling configurations

6.1. Create a scaling configuration

This topic describes how to create a scaling configuration for a scaling group. Auto Scaling uses the instance types that are specified in the scaling configuration to create instances.

Prerequisites

A security group is available in the virtual private cloud (VPC) where your scaling group for which you want to create a scaling configuration resides. For more information, see the *Create a security group* topic in *ECS User Guide*.

Context

You can create only a limited number of scaling configurations for a scaling group. For more information, see the *Limits* topic in *Auto Scaling Product Introduction*.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instance Configuration Sources** tab.
6. Click **Create Scaling Configuration**.
7. Configure the parameters for the scaling configuration. The following table describes the parameters.

Section	Parameter	Required	Description
Region	Region	Yes	The region where you want to create Elastic Compute Service (ECS) instances based on the scaling configuration.
	Zone	Yes	The zone where you want to create ECS instances based on the scaling configuration.
Security Group	Security Group	Yes	The security group with which you want to associate ECS instances that you want to create based on the scaling configuration.
Instance	Instance Type	Yes	The instance type that you want to use to create ECS instances.

Section	Parameter	Required	Description
Image	Image Type	Yes	◦ **Public Image**: You can select public images that are provided by Alibaba Cloud. Public images are licensed to provide a secure and stable operating environment for applications on ECS instances. ◦ **Custom Image**: You can create custom images that you can use to install software or deploy projects that have special requirements. ◦ **Shared Custom Image**: You can select a shared custom image.
Storage	System Disk (GB)	Yes	The category and size of the system disk. The OS is installed on the system disk. You can select **Ultra Disk** or **Standard SSD**.
	Data Disk (GB)	No	The category and size of the data disk. You can select **Ultra Disk** or **Standard SSD**. You can add up to 16 data disks. The maximum capacity of each data disk is 32 TiB. You can select **Release with Instance** and **Encryption** for each data disk.
	Password Setting	Yes	The time to specify a password. Valid values: **Set Now** and **Set After Purchase**. If you set the Password Setting parameter to Set After Purchase, you must reset the password in the console after the scaling configuration is created. For more information, see the *Change the logon password* topic in *ECS User Guide*.

Section	Parameter	Required	Description
Password	Logon Password	No	The password used to log on to the ECS instances that are created based on the scaling configuration. You must configure this parameter if you set the Password Setting parameter to **Set Now**. The logon password is subject to the image type that you select. For more information, see the *Reset the logon password of an instance* topic in *ECS User Guide*. **Note** The password is used to log on to the OS and is not the VNC password.
	Confirm Password	No	You must enter the password again after you configure the Logon Password parameter.
Deployment Set	Deployment Set	No	The deployment set to which the instances that are created based on the scaling configuration belong. You can select a deployment set from the Deployment Set drop-down list. You can also create a deployment set.
Instance Name	Scaling Configuration	No	The name of the scaling configuration that you want to create.
	Instance Name	No	The name of the ECS instance that you want to create based on the scaling configuration.
User Data	User Data	No	The Windows OS supports batch and PowerShell scripts. Before you perform Base64 encoding of user data, make sure that the first line of the data is included in `[bat]` or `[powershell]`. The Linux OS supports shell scripts.

Section	Parameter	Required	Description
Instances	Quantity	No	The number of instances that you want to create based on the scaling configuration.

8. Click **Submit**.

Result

After the scaling configuration is created, the scaling configuration enters the **Disable** state and is displayed in your scaling configuration list. You can click Apply in the Actions column to enable the scaling configuration. Auto Scaling can create ECS instances in a scaling group that has an enabled scaling configuration. For more information, see Apply a scaling configuration.

6.2. View scaling configurations

This topic describes how to view scaling configurations.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
   The scaling groups that correspond to the specified organization, resource set, and region are displayed.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instance Configuration Sources** tab.
6. View scaling configurations.

   On the Scaling Configuration Sources tab, you can view information about the scaling configurations such as instance families, images, system disk categories, data disks, and status.

6.3. Modify a scaling configuration

This topic describes how to modify a scaling configuration. You can modify the information about a scaling configuration based on your business requirements.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instance Configuration Sources** tab.
6. Find the scaling configuration that you want to modify and click its name in the **Scaling Configuration Name/ID** column.
7. Modify the information about the scaling configuration.

For more information about the parameters of the scaling configuration, see Create a scaling configuration.

8. Click **Submit**.

6.4. Apply a scaling configuration

This topic describes how to apply a scaling configuration. You can create multiple scaling configurations for a scaling group and apply one of the scaling configurations based on your business requirements.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instance Configuration Sources** tab.
6. Find the scaling configuration that you want to apply and click **Apply** in the **Actions** column.

   You can put only one scaling configuration into the **Enable** state in a scaling group. After you apply a scaling configuration, other scaling configurations in the scaling group enter the **Disable** state.
7. Click **OK**.

Result

The status of the scaling configuration changes from **Disable** to **Enable** in the **Status** column.

6.5. Delete a scaling configuration

This topic describes how to delete a scaling configuration that you no longer require. After you delete a scaling configuration, Elastic Compute Service (ECS) instances that are created based on the scaling configuration are retained in the scaling group.

Prerequisites

The scaling configuration is in the **Disable** state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Instance Configuration Sources** tab.
6. Find the scaling configuration that you want to delete and click **Delete** in the **Actions** column.
7. Click **OK**.

7.Scaling rules
7.1. Create a scaling rule

This topic describes how to create a scaling rule. You can specify the number of instances that you want to add to or remove from a scaling group. For example, you can create a scaling rule to add one instance to a scaling group.

Context

- You can create only a limited number of scaling rules for a scaling group. For more information, see the *Limits* topic in *Auto Scaling Product Introduction*.
- If the total number of instances in a scaling group is out of the specified range after a scaling rule is executed, Auto Scaling creates or removes a specific number of instances to keep the total number within the specified range.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Scaling Rules** tab.
6. Click **Create Scaling Rule**.
7. Configure the parameters for the scaling rule. The following table describes the parameters.

Parameter	Required	Description
Rule Name	Yes	The name of the scaling rule. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Scaling Activity	Yes	Valid values: ◦ **To N Units**: changes the number of instances in the scaling group to N. ◦ **Add N Units**: adds N instances to the scaling group. ◦ **Remove N Units**: removes N instances from the scaling group. **Note** Up to 500 instances can be added to or removed from a scaling group during a scaling activity.

Parameter	Required	Description
Cooldown Time	Yes	The period of cooldown time after a scaling activity is complete. If you do not specify a value for this parameter, the default value is used.

8. Click **OK**.

Result

On the Scaling Rules tab, you can view the information about the new scaling rule, such as the scaling rule name or ID, rule type, time when the rule is executed, and action that is specified in the rule.

7.2. View scaling rules

This topic describes how to view scaling rules.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
   The scaling groups that belong to the specified organization, resource set, and region are displayed.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Scaling Rules** tab.
6. View scaling rules.
   On the Scaling Rules tab, you can view information about the scaling rules, such as the scaling rule IDs, rule types, and actions that are specified in the rules.

What's next

On the Scaling Rules tab, you can manage the scaling rules. For example, you can click Edit, Execute, or Delete in the **Actions** column. For more information, see the following topics:

- Modify a scaling rule
- Execute a scaling rule
- Delete a scaling rule

7.3. Modify a scaling rule

This topic describes how to modify a scaling rule. You can modify the information about a scaling rule, such as the rule name, action that is specified in the rule, and cooldown time.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.

4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Scaling Rules** tab.
6. Find the scaling rule that you want to modify and click **Edit** in the **Actions** column.
7. Modify the rule name and change the action that you want to perform and the cooldown time based on your business requirements.
8. Click **OK**.

7.4. Execute a scaling rule

This topic describes how to manually execute a scaling rule.

Prerequisites

- The scaling group for which the scaling rule is created is in the **Enable** state.
- No scaling activity is in progress in the scaling group for which the scaling rule is created.

Context

If the total number of instances in a scaling group is out of the specified range after a scaling rule is executed, Auto Scaling creates or removes a specific number of instances to keep the total number within the specified range.

Auto Scaling allows you to manually execute scaling rules. You can also associate scaling rules with scheduled tasks or event-triggered tasks to enable auto scaling. For information about how to create scheduled tasks and event-triggered tasks, see Create a scheduled task and Create an event-triggered task.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Scaling Rules** tab.
6. Find the scaling rule that you want to execute and click **Execute** in the **Actions** column.
7. Click **OK**.

Result

On the **Scaling Activity** tab, a scaling activity is generated. You can view the execution status of the scaling activity.

7.5. Delete a scaling rule

This topic describes how to delete a scaling rule that you no longer require.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, click **Scaling Groups**.

3. In the top navigation bar, select an organization, a resource set, and a region.
4. Click the name of the scaling group in the **Scaling Group Name/ID** column.
5. On the scaling group details page, click the **Scaling Rules** tab.
6. Find the scaling rule that you want to delete and click **Delete** in the **Actions** column.
7. Click **OK**.

8.Scheduled tasks

8.1. Create a scheduled task

This topic describes how to create a scheduled task. If your business workloads are predictable, you can create a scheduled task to prepare sufficient computing resources before the business workloads increase and release excess computing resources after the business workloads decrease.

Context

- A scheduled task allows Auto Scaling to execute a scaling rule at the specified point in time. This way, computing resources can be automatically scaled to meet your business requirements and minimize the costs of resources. You can also specify an interval at which a scheduled task is executed to cope with the changes of business workloads in an efficient manner.
- If multiple scheduled tasks need to be executed within the same minute, Auto Scaling executes the most recently created scheduled task.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Scheduled Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scheduled Tasks** page, click **Create Scheduled Task**.
5. In the dialog box that appears, configure the parameters. The following table describes the parameters.

Parameter	Required	Description
Name	Yes	The name of the scheduled task. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Description	Yes	The description of the scheduled task.
Organization/Resource Set	Yes	The organization and resource set in which you want to create the scheduled task.
Start From	Yes	The point in time at which you want to execute the scheduled task.
Monitoring Resource/Scaling Rule	Yes	The scaling group for which you want to create the scheduled task and the scaling rule that you want to specify in the scheduled task.
Retry Expiry Time (Seconds)	No	The timeout period during which Auto Scaling retries the scheduled task. Unit: Seconds. If Auto Scaling does not execute the scheduled task at the specified point in time, Auto Scaling retries the scheduled task during the timeout period.

Parameter	Required	Description
Recurrence Interval Settings (Advanced)	No	Specifies whether to execute the scheduled task on a recurring schedule. After you select **Recurrence Interval Settings (Advanced)**, you must configure the **Recurrence Interval** and **Expires At** parameters. ○ Recurrence Interval: You can select **By Day**, **By Week**, or **By Month**. ○ Expires At: You can use the Select Date or Select Time methods.

6. Click **OK**.

Result

On the Scheduled Tasks page, the scheduled task that you created appears. You can view the information about the scheduled task, such as the scaling group for which the scheduled task is created, scaling rule that is specified in the scheduled task, task status, and task description.

8.2. View scheduled tasks

This topic describes how to view scheduled tasks.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Scheduled Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
   The scheduled tasks that correspond to the specified organization, resource set, and region are displayed.
4. Enter a scheduled task name or ID and click **Search**.

   You can select multiple filter options to narrow down the search results.

Filter option	Description
Scheduled Task Name	You can search for a scheduled task by name.
Scheduled Task ID	You can search for a scheduled task by ID.

5. View scheduled tasks.
   On the Scheduled Tasks page, you can view information about scheduled tasks, such as scheduled task IDs, organizations, resource sets, scaling rules that are specified in scheduled tasks, and task status.

What's next

On the Scheduled Tasks page, you can manage a specific scheduled task. You can click Edit, Disable, Enable, or Delete in the **Actions** column that corresponds to the scheduled task that you want to manage.

- Modify a scheduled task

- Disable a scheduled task
- Enable a scheduled task
- Delete a scheduled task

8.3. Modify a scheduled task

This topic describes how to modify a scheduled task. You can modify the information about a scheduled task, such as the scheduled task name and description, point in time at which the scheduled task is executed, resource for which the scheduled task is created, scaling rule that is specified in the scheduled task, and timeout period during which Auto Scaling can retry the scheduled task.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Scheduled Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scheduled Tasks** page, find the scheduled task that you want to modify and click **Edit** in the **Actions** column.
5. Modify the information about the scheduled task.

   If you select Recurrence Interval Settings (Advanced) when you create the scheduled task, you can modify the Recurrence Interval and Expires At settings. However, you cannot clear Recurrence Interval Settings (Advanced). For more information about other parameter settings, see Create a scheduled task.
6. Click **OK**.

8.4. Disable a scheduled task

This topic describes how to disable a scheduled task that you no longer require.

Prerequisites

The scheduled task is in the **Running** state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Scheduled Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scheduled Tasks** page, find the scheduled task that you want to disable and click **Disable** in the **Actions** column.
5. In the message that appears, click **OK**.

Result

The status of the scheduled task changes from **Running** to **Stop** in the **Status** column.

8.5. Enable a scheduled task

This topic describes how to enable a scheduled task. You can enable a scheduled task that is in the Stop state. This way, Auto Scaling automatically executes the scheduled task at the specified point in time.

Prerequisites

The scheduled task is in the **Stop** state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Scheduled Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scheduled Tasks** page, find the scheduled task that you want to enable and click **Enable** in the **Actions** column.
5. In the message that appears, click **OK**.

Result

The status of the scheduled task changes from **Stop** to **Running** in the **Status** column.

8.6. Delete a scheduled task

This topic describes how to permanently delete a scheduled task that you no longer require.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Scheduled Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Scheduled Tasks** page, find the scheduled task that you want to delete and click **Delete** in the **Actions** column.
5. In the message that appears, click **OK**.

9.Event-triggered tasks

9.1. Create an event-triggered task

This topic describes how to create an event-triggered task. If your business workloads are unpredictable or may surge unexpectedly, you can create an event-triggered task. You must specify a metric that is monitored by CloudMonitor in an event-triggered task. After you create and enable an event-triggered task, Auto Scaling collects data for the specified metric in real time and CloudMonitor triggers an alert when the specified condition is met. Then, Auto Scaling executes the scaling rule that is associated with the event-triggered task to scale instances in your scaling group.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Event-Triggered Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Event-triggered Tasks** page, click **Create Event-triggered Task**.
5. In the Create Event-triggered Task dialog box, configure the parameters. The following table describes the parameters.

Parameter	Required	Description
Task Name	Yes	The name of the event-triggered task. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.
Description	No	The description of the event-triggered task.
Organization/Resource Set	Yes	The organization and the resource set in which you want to create the event-triggered task.
Monitoring Resource/Scaling Rule	Yes	The scaling group for which you want to create the event-triggered task and the scaling rule that you want to specify in the event-triggered task.
Monitor Type	Yes	Valid value: **System Monitoring**.
Metric	Yes	The metric that you want to specify in the event-triggered task. Valid values: ○ **Average CPU Usage** ○ **Memory Usage** ○ **Outbound Traffic over Internal Network** ○ **Inbound Traffic over Internal Network** ○ **Average System Load**

Parameter	Required	Description
Statistical Period	Yes	The period during which data is collected, aggregated, and analyzed. The shorter the period, the more frequently the alert is triggered. Unit: Minutes. Valid values: ○ 1 ○ 2 ○ 5 ○ 15
Statistical Method	Yes	The rule that determines whether to trigger an alert. Select **Average**, **Max Capacity**, or **Min Capacity**, and specify a threshold. For example, you can specify one of the following rules to trigger an alert when the CPU utilization exceeds 80%: ○ Average: triggers an alert when the average CPU utilization of all instances in the scaling group exceeds 80%. ○ Max Capacity: triggers an alert when the highest CPU utilization among all instances in the scaling group exceeds 80%. ○ Min Capacity: triggers an alert when the lowest CPU utilization among all instances in the scaling group exceeds 80%.
Trigger Alert After {1} Times	Yes	The number of consecutive times that the threshold must be exceeded before the alert is triggered. Valid values: ○ 1 ○ 2 ○ 3 ○ 5

6. Click **OK**.

Result

On the Event-triggered Tasks page, the event-triggered task that you created appears. You can view the information about the event-triggered task, such as the monitoring type, statistical period, alert condition, and task description.

9.2. View event-triggered tasks

This topic describes how to view event-triggered tasks.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Event-Triggered Tasks**.

3. In the top navigation bar, select an organization, a resource set, and a region.
   The event-triggered tasks that correspond to the specified organization, resource set, and region are displayed.
4. Enter an event-triggered task name or ID and click **Search**.

   You can select multiple filter options to narrow down the search results.

Filter option	Description
Event-triggered Task Name	You can search for an event-triggered task by name.
Scaling Group ID	You can search for an event-triggered task by scaling group ID.

9.3. Modify an event-triggered task

This topic describes how to modify an event-triggered task. You can modify the information about an event-triggered task, such as the task name and description, metric that is specified in the task, and statistical method.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Event-Triggered Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Event-triggered Tasks** page, find the event-triggered task that you want to modify and click **Edit** in the **Actions** column.
5. Modify the information about the event-triggered task.

   For more information about other parameters of the event-triggered task, see Create an event-triggered task. You cannot modify the following information about the event-triggered task:

   - Organization/Resource Set
   - Monitoring Resource/Scaling Rule
   - Monitoring Type
6. Click **OK**.

9.4. Disable an event-triggered task

This topic describes how to disable an event-triggered task. You can disable an event-triggered task that you no longer require.

Prerequisites

The event-triggered task is in the **Normal** or **Insufficient Data** state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Event-Triggered Tasks**.

3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Event-triggered Tasks** page, find the event-triggered task that you want to disable and click **Disable** in the **Actions** column.
5. In the message that appears, click **OK**.

9.5. Enable an event-triggered task

This topic describes how to enable an event-triggered task. You can enable an event-triggered task that is in the Disable state. This way, Auto Scaling continues to collect data of the metric that is specified in the event-triggered task and the scaling activity that is based on the event-triggered task can be triggered when the specified conditions are met.

Prerequisites

The event-triggered task is in the Disable state.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Event-Triggered Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Event-triggered Tasks** page, find the event-triggered task that you want to enable and click **Enable** in the **Actions** column.
5. In the message that appears, click **OK**.

9.6. Delete an event-triggered task

This topic describes how to permanently delete an event-triggered task that you no longer require.

Procedure

1. Log on to the Auto Scaling console.
2. In the left-side navigation pane, choose **Scaling Tasks > Event-Triggered Tasks**.
3. In the top navigation bar, select an organization, a resource set, and a region.
4. On the **Event-triggered Tasks** page, find the event-triggered task that you want to delete and click **Delete** in the **Actions** column.
5. In the message that appears, click **OK**.