ALIBABA CLOUD

Alibaba Cloud

Apsara Stack Enterprise

Auto Scaling

Developer Guide

Product Version: v3.16.2

Document Version: 20220916

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.Getting started

1.1. Usage notes on development documentation

This topic describes how to view the desired product development documentation when you use Apsara Stack API (ASAPI).

Prerequisites

The Apsara Stack version is obtained from an O&M personnel or operations administrator.

Procedure

You can go to the Apsara Stack documentation center and view the desired product documentation.

1. Go to the Apsara Stack Documentation Centerpage by using a browser.
2. Click the **Enterprise**, **Agility**, or **All-in-one** tab based on the Apsara Stack version that you use.
3. Select an Apsara Stack version from the drop-down list.
4. In the left-side navigation pane, select the name of the product that you want to view.
5. In the upper-right corner of the product documentation section, click **View Details** or **Download PDF**.

1.2. Make API calls

You can use the Portable OpenAPI Proxy (POP) gateway or the Apsara Stack API (ASAPI) gateway to call ASAPI operations. In Apsara Stack Enterprise V3.16.0 and later, the POP gateway is used by default to ensure that the user experience in Apsara Stack is the same as the user experience in the Alibaba Cloud public cloud. This topic describes how to use the POP gateway to call ASAPI operations.

> **Note** You can continue to use the ASAPI gateway to call API operations in Apsara Stack whose version is earlier than V3.16.0. For more information, see Call the ASAPI gateway, Obtain an endpoint of ASAPI, and Obtain Apsara Stack SDKs.

Flowchart

Procedure

1. Obtain AccessKey credentials.

   Obtain the AccessKey credentials that are required for the API call. For more information, see Obtain AccessKey credentials or Obtain an STS AccessKey 3-tuple.

   > **Note** We recommend that you use the AccessKey credentials of a personal account to call the API operations of the Apsara Uni-manager Management Console and cloud services.

2. Obtain the region ID.

Obtain the region ID that is required for the API call. For more information, see Obtain the endpoint of Auto Scaling.

3. Obtain the organization ID and the resource set ID.

   You can call the GetOrganizationList operation to obtain the organization ID. For more information, see *GetOrganizationList* of *Apsara Uni-manager Management Console Developer Guide*.

   You can call the ListResourceGroup operation to obtain the resource set ID. For more information, see *ListResourceGroup* of *Apsara Uni-manager Management Console Developer Guide*.

   > **Note** When you call an API operation, set the `<x-acs-organizationid>` parameter to the organization ID that you obtained and the `<x-acs-resourcegroupid>` parameter to the resource set ID that you obtained. This way, you can query resources in the specified organization and resource set.
   >
   > - If you do not specify the `<x-acs-organizationid>` parameter, the organization to which the current user belongs is queried.
   > - If you do not specify the `<x-acs-resourcegroupid>` parameter, this parameter is left empty. If you specify a resource set, you must also specify an organization ID.

4. Obtain the API references.

   Refer to the developer guide of the cloud service that you want to use to learn the features of API operations and how to call the API operations. For more information, see Use developer guides.

5. Obtain the endpoint of the cloud service.
6. Obtain the cloud service SDKs and call API operations.

2.Preparations

2.1. Log on to the API Tool console

You can log on to the API Tool console from the Apsara Uni-manager Management Console.

Prerequisites

- Before you log on to the Apsara Uni-manager Management Console, the endpoint of the console is obtained from the deployment staff.
- We recommend that you use Google Chrome.

Procedure

1. Enter the URL of the Apsara Uni-manager Management Console in the address bar and press the Enter key.
2. Enter your username and password.

   Obtain the username and password that you use to log on to the Apsara Uni-manager Management Console from the operations administrator.

   > **Note** The first time that you log on to the Apsara Uni-manager Management Console, you must change the password of your account. For security purposes, your password must meet the minimum complexity requirements. The password must be 10 to 32 characters in length and must contain at least two of the following character types:
   >
   > - Uppercase or lowercase letters
   > - Digits
   > - Special characters, including exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%)

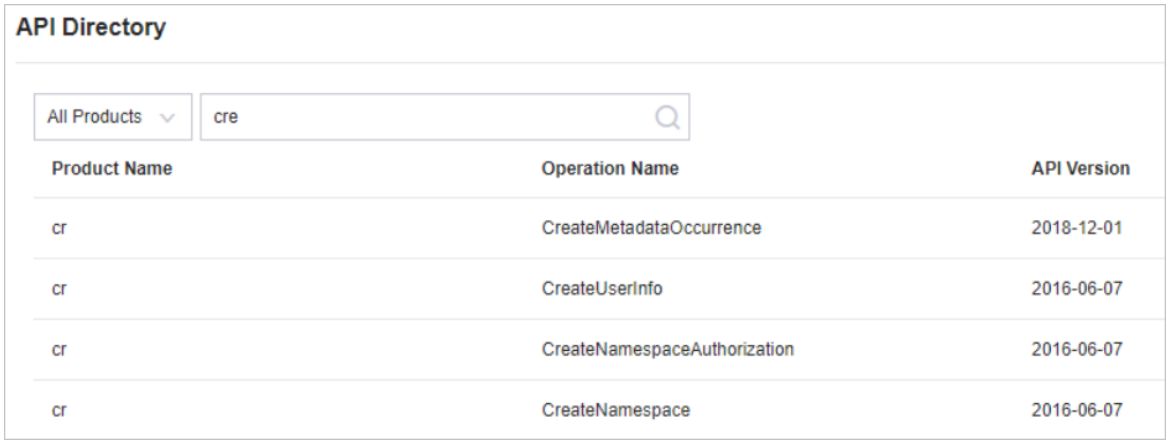
3. Click **Log On**.
4. In the top navigation bar, choose **Products > Others > API Tool**.

2.2. View the information about API operations

You must obtain a service name, API operation name, and API version. All the information is required when you use SDKs.

Procedure

1. Log on to the API Tool console.
2. In the left-side navigation pane, click **API fold**.
3. On the **API Directory** page, select the target service from the drop-down list, enter the API operation name in the search box, and then click the search icon.



4. View and remember the service name, API operation name, and API version.

2.3. Obtain AccessKey credentials

AccessKey credentials are assigned by Resource Access Management (RAM) or Security Token Service (STS). You must use one of the preceding types of AccessKey credentials for authentication when you call API operations. This topic describes how to obtain AccessKey credentials that are assigned by RAM.

Context

The AccessKey ID and AccessKey secret contained in AccessKey credentials are used to perform symmetric encryption and verify the identity of request senders. The AccessKey ID is used to identify a user. The AccessKey secret is a string that is used to encrypt a signature.

You can use RAM or STS to assign AccessKey credentials:

- The AccessKey credentials that are assigned by RAM to a third party consist of an AccessKey ID and an AccessKey secret.
- The AccessKey credentials that are assigned by STS to a third party consist of an AccessKey ID, an AccessKey secret, and a security token. For more information, see Obtain an STS AccessKey 3-tuple.

The Apsara Uni-manager Management Console provides AccessKey credentials for personal accounts and organizations. Only operations administrators and the administrators of level-1 organizations can obtain the AccessKey credentials of organizations. We recommend that you use the AccessKey credentials of personal accounts to call the operations provided by the Apsara Uni-manager Management Console and cloud services. If you use the AccessKey credentials of a personal account, configure the request headers that are described in the following table for access control. Otherwise, an error that indicates insufficient permissions may occur.

Warning The AccessKey credentials of a personal account are managed by the permission system of the Apsara Uni-manager Management Console. If you need to obtain the AccessKey credentials of an organization, make sure that your operation is allowed by an administrator. This ensures security.

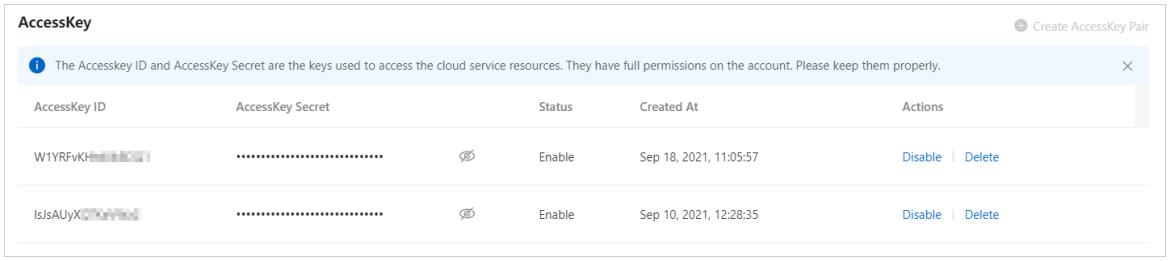
Request header	Description
x-acs-regionid	The ID of the region. Example: cn-hangzhou-****.

Request header	Description
x-acs-organizationid	The ID of the organization in the Apsara Uni-manager Management Console. For more information about how to obtain organization IDs, see *GetOrganizationList* in *Apsara Uni-manager Management Console Developer Guide*.
x-acs-resourcegroupid	The ID of the resource set in the Apsara Uni-manager Management Console. For more information about how to obtain resource set IDs, see *ListResourceGroup* in *Apsara Uni-manager Management Console Developer Guide*.
x-acs-instanceid	The ID of the instance that you want to manage. For example, you can view the ID of an Elastic Compute Service (ECS) instance in the ECS instance list.

Obtain the AccessKey credentials of a personal account

To obtain the AccessKey credentials of a personal account, perform the following steps:

1. Log on to the Apsara Uni-manager Management Console.
2. In the upper-right corner of the page that appears, click the profile picture, and then click **User Information**.
3. In the **AccessKey Pair** section, view the AccessKey credentials of the personal account.



> **Note** The AccessKey credentials consist of an AccessKey ID and an AccessKey secret. The credentials provide full permissions on cloud resources within your Apsara Stack tenant account. You must keep the AccessKey credentials.

Obtain the AccessKey credentials of an organization

To obtain the AccessKey credentials of an organization, perform the following steps:

1. Log on to the Apsara Uni-manager Management Console as an administrator.
2. In the top navigation bar, click **Enterprise**.
3. In the left-side navigation pane, choose **Resources > Organizations**.
4. In the organization list, click the name of the level-1 organization whose AccessKey credentials you want to obtain.
5. Click **Management Accesskey**.
6. In the dialog box that appears, view the AccessKey credentials of the organization.

2.4. Obtain an STS AccessKey 3-tuple

This topic describes how to use Security Token Service (STS) to call API operations in Apsara Stack.

Context

STS allows you to use temporary credentials to access your Apsara Stack resources. You can use STS to call API operations in Apsara Stack.

An STS AccessKey 3-tuple consists of an AccessKey ID, an AccessKey secret, and a security token. When you use STS to call API operations, you must specify an STS AccessKey 3-tuple. For more information about STS, see What is STS?

Procedure

1. Obtain the `RAM role`.
   i. Log on to the Apsara Uni-manager Management Console.
   ii. In the upper-right corner of the homepage, click the profile picture and then click **User Information**.
   iii. Click **View Current Role Policy**.

   In the **View Policies of Current Role (Administrator)** dialog box, view the ARN of your `RAM role`.
2. Obtain the STS AccessKey 3-tuple.

> **Note** In the following sample code, replace the variable parameters with actual values. Delete the angle brackets (<>) in the final code.
>
> Variable parameters:
>
> - <yourRegionID>: Replace the variable parameter with your region ID.
> - <yourAccessKeyID>: Replace the variable parameter with your AccessKey ID.
> - <yourAccessKeySecret>: Replace the variable parameter with your AccessKey secret.
> - <yourRoleSessionName>: Replace the variable parameter with the name of your session.
> - <yourRoleArn>: Replace the variable parameter with the ARN of your RAM role.
> - "sts.aliyuns.com": Replace the variable parameter with the endpoint of STS.

```
public static void main(String[] args) {
    // Create and initialize a DefaultAcsClient instance.
    // Specify the region ID, AccessKey ID, and AccessKey secret.
    DefaultProfile profile = DefaultProfile.getProfile("<yourRegionID>", "<yourAccessKe
yID>", "<yourAccessKeySecret>");
    HttpClientConfig clientConfig = HttpClientConfig.getDefault();
    clientConfig.setIgnoreSSLCerts(true);
    clientConfig.setProtocolType(ProtocolType.HTTPS);
    profile.setHttpClientConfig(clientConfig);
    IAcsClient client = new DefaultAcsClient(profile);
    String assumeRole = callAssumeRole(client);
    if (StringUtils.isEmpty(assumeRole)) return;
    JSONObject assumeRoleJson = JSONObject.parseObject(assumeRole);
    JSONObject credentials = assumeRoleJson.getJSONObject("Credentials");
    // Obtain the STS AccessKey 3-tuple.
    String stsAK = credentials.getString("AccessKeyId");
    String stsSK = credentials.getString("AccessKeySecret");
    String stsToken = credentials.getString("SecurityToken");
}
private static String callAssumeRole(IAcsClient client) {
    CommonRequest request = new CommonRequest();
    // Replace "sts.aliyuns.com" with the endpoint of STS. The format of an STS endpoin
t is sts-vpc.${global:region}.${global:internet-domain}.
    request.setSysDomain("sts.aliyuns.com");
    request.setSysProduct("Sts");
    request.setSysAction("AssumeRole");
    request.setSysVersion("2015-04-01");
    request.setSysMethod(MethodType.POST);
    request.setHttpContentType(FormatType.FORM);
    // Replace <yourRoleSessionName> with the name of your session.
    request.putQueryParameter("RoleSessionName", "<yourRoleSessionName>");
    // Replace <yourRoleArn> with the ARN of the RAM role that you obtain in the preced
ing step.
    request.putQueryParameter("RoleArn", "<yourRoleArn>");
    try {
        CommonResponse response = client.getCommonResponse(request);
        return response.getData();
    } catch (ServerException e) {
        System.out.println(e.getErrCode());
        System.out.println(e.getErrMsg());
    } catch (ClientException e) {
        System.out.println(e.getErrCode());
        System.out.println(e.getErrMsg());
    }
    return null;
}
```

3. Use the STS AccessKey 3-tuple to call an API operation of a cloud service. The following sample code provides an example on how to call the `CreateScalingGroup` operation of Auto Scaling.

Note In the following sample code, replace the variable parameters with actual values. Delete the angle brackets (<>) in the final code.

Variable parameters:

- <yourRegionID>: Replace the variable parameter with your region ID.
- <yourStsAccesskeyID>: Replace the variable parameter with your STS AccessKey ID.
- <yourStsAccessKeySecret>: Replace the variable parameter with your STS AccessKey secret.
- <yourStsToken>. Replace the variable parameter with your security token.
- <yourResourceGroupID>: Replace the variable parameter with your resource set ID.
- <yourOrganizationID>: Replace the variable parameter with your organization ID.

```
public static void main(String[] args) {
    // Replace <yourRegionID> with your region ID.
    // Replace <yourStsAccessKeyID> and <yourStsAccessKeySecret> with the AccessKey ID
and AccessKey secret of the STS AccessKey 3-tuple obtained in the previous step.
    DefaultProfile profile = DefaultProfile.getProfile("<yourRegionID>", "<yourStsAcces
skey>", "<yourStsAccessSecret>");
    HttpClientConfig clientConfig = HttpClientConfig.getDefault();
    clientConfig.setIgnoreSSLCerts(true);
    clientConfig.setProtocolType(ProtocolType.HTTPS);
    profile.setHttpClientConfig(clientConfig);
    IAcsClient client = new DefaultAcsClient(profile);
    CommonRequest request = new CommonRequest();
    request.setSysDomain("ess.aliyuncs.com");
    request.setSysProduct("ess");
    request.setSysAction("CreateScalingGroup");
    request.setSysVersion("2014-08-28");
    request.setSysMethod(MethodType.POST);
    request.setHttpContentType(FormatType.FORM);
    // Replace <yourRegionID> with your region ID.
    request.putHeadParameter("x-acs-regionid", "<yourRegionID>");
    // Replace <yourStsToken> with the security token of the STS AccessKey 3-tuple obta
ined in the previous step.
    request.putQueryParameter("SecurityToken", "<yourStsToken>");
    // Optional. Replace the variable parameter with your resource set ID.
    // request.putHeadParameter("x-acs-resourcegroupid", "<yourResourceGroupID>");
    // Required. Replace the variable parameter with your organization ID.
    request.putHeadParameter("x-acs-organizationid", "<yourOrganizationID>");
    try {
        CommonResponse response = client.getCommonResponse(request);
        System.out.println(response.getData());
    } catch (ServerException e) {
        System.out.println(e.getErrCode());
        System.out.println(e.getErrMsg());
    } catch (ClientException e) {
        System.out.println(e.getErrCode());
        System.out.println(e.getErrMsg());
    }
```

2.5. Obtain the values of common header parameters

You must specify multiple header parameters when you call Apsara Stack API operations. This topic describes the common header parameters and how to obtain the values of the parameters. This topic also describes the return values of specific parameters after you call API operations.

Header parameters

Parameter	Description
x-acs-regionid	The ID of the region. For more information about how to obtain region IDs, see Obtain a region ID.
x-acs-organizationid	The ID of the organization in the Apsara Uni-manager Management Console. For more information about how to obtain organization IDs, see Obtain an organization ID.
x-acs-resourcegroupid	The ID of the resource set in the Apsara Uni-manager Management Console. For more information about how to obtain resource set IDs, see Obtain a resource set ID.
x-acs-instanceid	The ID of the instance on which you want to perform operations. For more information about how to obtain instance IDs, see Obtain an instance ID.

Obtain a region ID

1. Log on to the Apsara Uni-manager Management Console as an administrator.
2. In the top navigation bar, click **Enterprise**.
3. In the left-side navigation pane, choose **Resources > Regions**.
4. Click the name of the organization whose region ID you want to view.
5. In the **Regions** section, view the region ID.

Obtain an organization ID

You can call the GetOrganizationList operation to obtain the organization ID. For more information, see *GetOrganizationList* in *Apsara Uni-manager Management Console Developer Guide*.

Obtain a resource set ID

You can call the ListResourceGroup operation to obtain the resource set ID. For more information, see *ListResourceGroup* in *Apsara Uni-manager Management Console Developer Guide*.

Obtain an instance ID

You can obtain the ID of an instance in the cloud service console of the Apsara Uni-manager Management Console. For example, to obtain the ID of an Auto Scaling instance, perform the following steps:

1. Log on to the Apsara Uni-manager Management Console.
2. In the top navigation bar, choose **Products > Elastic Computing > Auto Scaling**.
3. In the left-side navigation pane, choose **Instances**.
4. View the **instance ID** in the instance list.

Response parameters

The following table describes the additional response parameters returned by default when you call operations to query the instances of cloud services.

Parameter	Description
ResourceGroupName	The name of the resource set.
ResourceGroup	The ID of the resource set.
DepartmentName	The name of the organization.
Department	The ID of the organization.

2.6. Obtain the endpoint of Auto Scaling

This topic describes how to obtain the endpoint of Auto Scaling.

Context

When you call Auto Scaling API operations by using the Portable OpenAPI Proxy (POP) gateway, you must obtain the endpoint of Auto Scaling. You can contact an Alibaba Cloud on-site O&M engineer to obtain the endpoint. You can also perform the following steps to obtain the endpoint.

Procedure

1. Log on to Apsara Infrastructure Management.
   i. Log on to the Apsara Uni-manager Operations Console.
   ii. In the top navigation bar, click **Products**.
   iii. Choose **Platforms > Apsara Infrastructure Management Framework**.
2. In the left-side navigation pane, click **Reports**.
3. On the **All Reports** page, enter `Registration Vars of Services` in the search box and click the search icon. In the report list that appears, click the name of the report.

   
4. On the **Registration Vars of Services** page, click the  icon to the right of **Service** and search for `ess`.

5. Find the `ess-coordinator` service, right-click the **Service Registration** column, and then select **Show More** from the shortcut menu.

   On the **Details** page, view the **ess.openapi.endpoint** value. The value is the `endpoint` of Auto Scaling.

2.7. Obtain Auto Scaling SDKs

When you call Auto Scaling API operations by using the Portable OpenAPI Proxy (POP) gateway, you must obtain Auto Scaling SDKs. This topic describes how to obtain Alibaba Cloud SDKs.

Background information

To use Auto Scaling SDKs, you must install Auto Scaling SDKs and Alibaba Cloud core libraries. For example, to use Auto Scaling SDK for Java, you must download `aliyun-java-sdk-ess` and `aliyun-java-sdk-core`.

- Auto Scaling SDKs encapsulate the API that was released on August 28, 2014. The SDKs identify callers based on the AccessKey credentials and provide features such as automatic signing. This way, you can call API operations to create and manage resources. Alibaba Cloud SDKs include code and examples. You can create cloud applications by using an SDK for the programming language that you require.
- We recommend that you use the SDK core library that is provided by Alibaba Cloud. The SDK core library is suitable for all cloud services and is compatible with Apsara Stack.

References

The SDKs for Java, Python, Go, and other programming languages are provided. You can obtain the SDKs from Alibaba Cloud SDKs based on your business requirements.

> **Note** We recommend that you use an SDK for Java, Python, or Go. If you use SDKs for other programming languages, the SDKs may be incompatible with your system.

3.Examples of using Auto Scaling SDK for Java

3.1. Getting started

This topic describes how to obtain and use Auto Scaling SDK for Java.

Prerequisites

- An Apsara Stack tenant account that has the permissions on Auto Scaling SDK for Java and an AccessKey pair are obtained. The AccessKey pair consists of an AccessKey ID and an AccessKey secret. For information about how to obtain an AccessKey pair, see Obtain an AccessKey pair.
- The endpoint of Auto Scaling is obtained. For information about how to obtain the endpoint of Auto Scaling, see Obtain the endpoint of Auto Scaling.
- Java Development Kit (JDK) 1.8 or later is installed. SDKs for Java provided by Alibaba Cloudrequire JDK 1.8 or later.

Install Auto Scaling SDK for Java

If you use Apache Maven to manage a Java project, you need to only import the JAR file to the pom.xml file of the project.

To use an Auto Scaling SDK, you must install the SDK core library and the Auto Scaling SDK. We recommend that you use version 4.5.17 or later for the SDK core library. In this example, Auto Scaling SDK uses version 3.2.0.

Use Auto Scaling SDK for Java

To use Auto Scaling SDK for Java, perform the following steps:

1. Create a DefaultAcsClient instance and initialize the instance.
2. Create an API request and configure the required parameters.
3. Initiate the API request and handle the response or exception.

> **Note**
>
> In the following sample code, replace the variable parameters with actual values. Delete the angle brackets <> in the final code.
>
> Variable parameters:
>
> - <yourRegionID>: Replace the variable parameter with your region ID.
> - <yourAccessKeyID>: Replace the variable parameter with your AccessKey ID.
> - <yourAccessKeySecret>: Replace the variable parameter with your AccessKey secret.
> - <Product>: Replace the variable parameter with the name of your service.
> - <Endpoint>: Replace the variable parameter with the endpoint of your service.
> - <readTimeout>: Replace the variable parameter with a read timeout period based on your business requirements. Unit: milliseconds.
> - <connectionTimeout>: Replace the variable parameter with a connection timeout period based on your business requirements. Unit: milliseconds.
> - <maxRequestsPerHost>: Replace the variable parameter with the maximum number of requests that are supported by each host.
> - <maxRequests>: Replace the variable parameter with the maximum number of requests.
> - <http://192.168.XX.XX:33XX>: Replace the variable parameter with the IP address and port of your proxy.

```
import com.aliyuncs.profile.DefaultProfile;
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.exceptions.ServerException;
import com.aliyuncs.ess.model.v20140828.*;
public class Main {
    public static void main(String[] args) {
        // Create and initialize a DefaultAcsClient instance.
        DefaultProfile profile = DefaultProfile.getProfile(
            "<yourRegionID>",          // Replace the variable parameter with your regi
on ID.
            "<yourAccessKeyID>",      // Replace the variable parameter with your Acces
sKey ID.
            "<yourAccessKeySecret>"); // Replace the variable parameter with your Acces
sKey secret.
        // Specify the endpoint of your service. Replace the variable parameter with th
e endpoint of your service.
        profile.addEndpoint("<yourRegionID>", "<Product>", "<Endpoint>");
        // Configure the request parameters.
        HttpClientConfig clientConfig = HttpClientConfig.getDefault();
        clientConfig.setIgnoreSSLCerts(true);                        // Specify whether
to ignore the certificate.
        clientConfig.setReadTimeoutMillis(<readTimeout>);            // Optional. Speci
fy the read timeout period. Unit: milliseconds.
        clientConfig.setConnectionTimeoutMillis(<connectionTimeout>);// Optional. Speci
fy the connection timeout period. Unit: milliseconds.
```

```
        clientConfig.setMaxRequestsPerHost(<maxRequestsPerHost>);    // Optional. Speci
fy the maximum number of requests per host.
        clientConfig.setMaxRequests(<maxRequests>);                  // Optional. Speci
fy the maximum number of requests.
        clientConfig.setProtocolType(ProtocolType.HTTP);             // Optional. Speci
fy the request protocol. Default value: HTTP.
        /* Optional. Specify the IP address and port of the proxy.
        clientConfig.setHttpProxy("http://192.168.XX:33XX");
        or
        clientConfig.setHttpsProxy("https://192.168.XX:33XX");
        */
        profile.setHttpClientConfig(clientConfig);
        IAcsClient client = new DefaultAcsClient(profile);
        // Create an API request and configure the required parameters.
        DescribeScalingGroupsRequest request = new DescribeScalingGroupsRequest();
        // Configure service-related parameters.
        request.setPageSize(10);
        request.setPageNumber(1);
        // Header parameters. Specify the organization ID and resource set ID to query
specified resources.
        request.putHeadParameter("x-acs-organizationid", "1");
        request.putHeadParameter("x-acs-resourcegroupid", "11");
        // Send the request and handle the response or exception.
        DescribeScalingGroupsResponse response;
        try {
            response = client.getAcsResponse(request);
            }
        } catch (ServerException e) {
            e.printStackTrace();
        } catch (ClientException e) {
            e.printStackTrace();
        }
    }
}
```

3.2. Work with Auto Scaling SDK for Java

3.2.1. Install Auto Scaling SDK for Java

This topic describes how to install and use Auto Scaling SDK for Java.

Auto Scaling SDK for Java supports Java Development Kit (JDK) 1.8 or later. You can install Auto Scaling SDK for Java by adding Maven dependencies or downloading the SDK. For more information, see Alibaba Cloud SDKs.

To use the development tool kit of Auto Scaling, you must install the core library of Alibaba Cloud SDK for Java and Auto Scaling SDK for Java. We recommend that you use the core library version 4.5.17 or later and Auto Scaling SDK for Java version 2.1.1.

To install the core library of Alibaba Cloud SDK for Java and Auto Scaling SDK for Java, you must add the following dependencies to the pom.xml file. For information about the specific SDK version, visit Alibaba Cloud SDKs.

```
<!-- Alibaba Cloud SDK for Java -->
<dependency>
    <groupId>com.aliyun</groupId>
    <artifactId>aliyun-java-sdk-core</artifactId>
    <version>4.5.17</version>
</dependency>
<!-- ess java sdk -->
<dependency>
    <groupId>com.aliyun</groupId>
    <artifactId>aliyun-java-sdk-ess</artifactId>
    <version>2.1.1</version>
</dependency>
```

3.2.2. Configure authentication credentials

When you use Alibaba Cloud SDKs to access Apsara Stack services, you must provide an Apsara Stack tenant account for authentication. This topic describes how to configure authentication credentials.

Auto Scaling SDK for Java supports authentication based on an AccessKey pair. For information about how to obtain an AccessKey pair, see Obtain AccessKey credentials.

If you want to use your AccessKey pair to access Apsara Stack, you must specify the AccessKey pair as the credentials when you initialize the client.

Notice Do not disclose code that contains your AccessKey pair. For example, do not commit the code to public GitHub projects. Otherwise, your Apsara Stack tenant account may be compromised.

```
DefaultProfile profile = DefaultProfile.getProfile(
    "<yourRegionID>",          // Replace the variable parameter with your region ID.
    "<yourAccessKeyID>",      // Replace the variable parameter with your AccessKey ID.
    "<yourAccessKeySecret>"); // Replace the variable parameter with your AccessKey secret.
```

Note In the preceding sample code, replace the variable parameters with your actual information. Delete the angle brackets (<>) in the final code.

3.2.3. Configure connection requests

This topic describes how to configure a connection request by using the HttpClientConfig object.

You can use the following code to configure the HttpClientConfig object. We recommend that you select and configure the following parameters based on your business requirements.

> **Note**
>
> You must use actual values to replace the variables in the following sample code. Do not retain the angle brackets (<>) in the final code.
>
> The following list describes the variables:
>
> - <readTimeout>: Specify the read timeout period based on your business requirements. Unit: milliseconds.
> - <connectionTimeout>: Specify the connection timeout period based on your business requirements. Unit: milliseconds.
> - <maxRequestsPerHost>: Specify the maximum number of requests that are supported by each server based on your business requirements. The supported data type is INTEGER.
> - <maxRequests>: Specify the maximum number of requests based on your business requirements. The supported data type is INTEGER.
> - <http://192.168.XX.XX:33XX>: Replace it with the actual IP address and port of the proxy.

```
// Obtain the HttpClientConfig object.
HttpClientConfig clientConfig = HttpClientConfig.getDefault();
// Specify whether to ignore certificates.
clientConfig.setIgnoreSSLCerts(true);
// (Optional) Specify the read timeout period. Unit: milliseconds.
clientConfig.setReadTimeoutMillis(<readTimeout>);
// (Optional) Specify the connection timeout period. Unit: milliseconds.
clientConfig.setConnectionTimeoutMillis(<connectionTimeout>);
// (Optional) Specify the maximum number of requests that are supported by each server. The
supported data type is INTEGER.
clientConfig.setMaxRequestsPerHost(<maxRequestsPerHost>);
// (Optional) Specify the maximum number of requests. The supported data type is INTEGER.
clientConfig.setMaxRequests(<maxRequests>);
// (Optional) Specify the protocol type.
clientConfig.setProtocolType(ProtocolType.HTTP);
/** (Optional) Specify whether to configure the proxy based on your business requirements.
The following code is available:
clientConfig.setHttpProxy("<http://192.168.XX.XX:33XX>");
Or
clientConfig.setHttpsProxy("<https://192.168.XX.XX:33XX>");
Replace https://192.168.XX.XX:33XX with the actual IP address and port of the proxy.
*/
profile.setHttpClientConfig(clientConfig);
```

3.2.4. Use Auto Scaling SDK for Java to initiate requests

This topic describes how to use Auto Scaling SDK for Java to initiate a request.

Procedure

1. Create an AcsClient.

```
IAcsClient client = new DefaultAcsClient(profile);
```

2. Create a request.

   The request name is in the `${apiName}Request` format, where ${apiName} is the API name. Sample API name: *CreateScalingGroup*.

   When the SDKs of multiple services are used, different requests may have the same name. Make sure that you differentiate the requests by SDK package.

```
CreateScalingGroupRequest request = new CreateScalingGroupRequest();
// Service-related parameters.
request.setPageSize(10);
request.setPageNumber(1);
// // Header parameters. Specify the organization ID and the resource set ID to query s
pecified resources.
request.putHeadParameter("x-acs-organizationid", "1");
request.putHeadParameter("x-acs-resourcegroupid", "");
```

3. Initiate the request and handle the response.

```
CreateScalingGroupRequest request = new CreateScalingGroupRequest();
request.setRegionId("cn-hangzhou");
request.setMinSize("0");
request.setMaxSize(1);
try {
    CreateScalingGroupResponse response = client.getCreateScalingGroupResponse(request)
;
    System.out.println(response.getScalingGroups());
} catch (ServerException e) {
    e.printStackTrace();
} catch (ClientException e) {
    e.printStackTrace();
}
```

   In most cases, all response fields are deserialized into the response. You can call `response.getXXX()` to obtain the values of the response fields.

```
System.out.println(response.getScalingGroups());
```

   If an exception occurs or if you require the original HTTP response, call `doAction()` to obtain the original response.

```
HttpResponse response = client.doAction(request);
```

3.2.5. Handle errors

If an error occurs on the server side or SDK side, the Apsara Stack Java SDK returns the corresponding exception message. An exception message contains the detailed error information, namely the error code and the error message.

Generally, you are not required to handle the errors returned by the Apsara Stack Java SDK. You are only required to handle the errors on the server side according to the error message.

- `ServerException` is the error message on the server side.
- `ClientException` is the error message on the SDK side.

For example, if the following error occurs, modify the ID of AccessKey according to the detailed error message.

```
com.aliyuncs.exceptions.ClientException: InvalidAccessKeyId.NotFound : Specified access key
is not found.
```

To handle the errors returned by the Apsara Stack Java SDK, see the following code example for reference:

```
try {
    FooResponse response = client.getAcsResponse(request);
    // Handle the response.
    // ...
}catch (ServerException e){
    // Add your own error handling logic here.
    // For example, print the detailed error message.
    System.out.println("ErrorCode=" + e.getErrCode());
    System.out.println("ErrorMessage=" + e.getErrMsg());
    // If the problem is tricky, open a ticket and provide the RequestId for us.
    System.out.println("ResponseId=" + e.getRequestId());
}catch (ClientException e){
    // Add your own error handling logic here.
    // For example, print the detailed error message.
    System.out.println("ErrorCode=" + e.getErrCode());
    System.out.println("ErrorMessage=" + e.getErrMsg());
}
```

3.2.6. Use CommonRequest to call API operations

If you want to call Auto Scaling API operations for which no SDK is provided, or if the requests and responses of API operations that you call by using an SDK do not meet your business requirements, you can use CommonRequest to call Auto Scaling API operations. You can use CommonRequest to call all API operations.

Benefits

CommonRequest provides the following benefits:

1. Lightweight: You need to only download the core package. You do not need to download and install SDKs of services.
2. Easy to use: You can call the new API operations of a service without the need to update the SDK of the service.
3. Fast iteration.

Use CommonRequest

If you want to send a CommonRequest request, you must obtain the values of the following parameters. You can obtain the values of the parameters from the API references of cloud services.

- Domain: the endpoint of the service.
- Version: the version of the API in the YYYY-MM-DD format. You can obtain the API version from the sections about common parameters in the API references.
- Operation information: the name of the API operation that you want to call. You must obtain the value of the Action parameter and then use the `request.ApiName = "<Action>"` method to specify the name of the operation.

  For example, if the operation that you want to call is DescribeScalingActivities, specify the operation as `request.ApiName = "DescribeScalingActivities"` before you make the request by using CommonRequest.

Example: Call an RPC API operation

In this example, the CreateScalingGroup API operation of Auto Scaling is used to show how to call an RPC API operation.

> **Note** In the following code, replace the variable parameters with your actual information. Delete the angle brackets (<>) in the final code.
>
> Variable parameters:
>
> - <yourRegionID>: Replace the variable parameter with your region ID.
> - <yourAccessKeyID>: Replace the variable parameter with your AccessKey ID.
> - <yourAccessKeySecret>: Replace the variable parameter with your AccessKey secret.
> - <yourSecurityToken>: Replace the variable parameter with your security token.
> - <yourResourceGroupID>: Replace the variable parameter with your resource set ID.
> - <yourOrganizationID>: Replace the variable parameter with your organization ID.
> - <yourInstanceID>: Replace the variable parameter with your instance ID.
> - <http://192.168.XX.XX:33XX>: Replace the variable parameter with the IP address and the port number of your proxy.

```
import com.aliyuncs.CommonRequest;
import com.aliyuncs.CommonResponse;
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.exceptions.ServerException;
import com.aliyuncs.profile.DefaultProfile;
public class CreateScalingGroup {
    public static void main(String[] args) {
        // Create a DefaultAcsClient instance and initialize the instance.
        DefaultProfile profile = DefaultProfile.getProfile(
            "<yourRegionUD>",          // Replace the variable parameter with your region I
D.
            "<yourAccessKeyID>",      // Replace the variable parameter with your AccessKey
ID.
            "<yourAccessKeySecret>"); // Replace the variable parameter with your AccessKey
secret.
        clientConfig.setProtocolType(ProtocolType.HTTP);
        /**
        You can configure the HTTP proxy based on your business requirements.
```

```
        Format:
        clientConfig.setHttpProxy("http://192.168.XX.XX:33XX");
        Specify the IP address and the port number of the proxy.
        */
        profile.setHttpClientConfig(clientConfig);
        IAcsClient client = new DefaultAcsClient(profile);
        // Create an API request and specify the request parameters.
        CommonRequest request = new CommonRequest();
        // Specify the endpoint of your service.
        request.setDomain("ess.aliyuncs.com");
        request.setSysProduct("ess");
        request.setVersion("2014-08-28");
        request.setAction("CreateScalingGroup");
        request.putQueryParameter("PageNumber", "1");
        request.putQueryParameter("PageSize", "30");
        request.putHeadParameter("x-acs-resourcegroupid", "<yourResourceGroupID>");// Optio
nal. Specify the resource set ID.
        request.putHeadParameter("x-acs-organizationid", "<yourOrganizationID>");// Require
d. Specify the organization ID.
        request.putHeadParameter("x-acs-caller-sdk-source","ApiTestDemo");// Required. Spec
ify the source of the API call.
        request.putHeadParameter("x-acs-regionid", "<yourRegionID>");// Optional. Specify t
he region ID. If you use Apsara Stack Enterprise V3.15.0 or earlier, this parameter is requ
ired.
        // request.putHeadParameter("x-acs-instanceid", "<yourInstanceID>");// Optional. Sp
ecify the instance ID. If you want to delete, modify, or query the details of an instance,
this parameter is required.
        /**
        When you use STS AccessKey credentials to call an API operation, you must add the S
ecurityToken field in the request parameters.
        Format:
        request.putQueryParameter("SecurityToken", "<yourSecurityToken>");
        Specify your security token.
        */
        try {
            CommonResponse response = client.getCommonResponse(request);
            System.out.println(response.getData());
        } catch (ServerException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (ClientException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }
}
```

4.Call examples of Auto Scaling SDK for Python

4.1. Getting started

This topic describes how to obtain and use Auto Scaling SDK for Python.

Prerequisites

- An Apsara Stack tenant account and an AccessKey pair are obtained. The AccessKey pair consists of an AccessKey ID and an AccessKey secret. For more information, see Obtain AccessKey credentials.
- The endpoint of Auto Scaling is obtained. For more information, see Obtain the endpoint of Auto Scaling.

Install Auto Scaling SDK for Python

1. Install the SDK core library.
   - If you use Python 2.x.x, run the following command to install the core library:

     ```
     pip install aliyun-python-sdk-core
     ```

   - If you use Python 3.x.x, run the following command to install the core library:

     ```
     pip3 install aliyun-python-sdk-core-v3
     ```

2. Install Auto Scaling SDK for Python.
   - If you use Python 2.x.x, run the following command to install Auto Scaling SDK for Python:

     ```
     pip install aliyun-python-sdk-ess
     ```

   - If you use Python 3.x.x, run the following command to install Auto Scaling SDK for Python:

     ```
     pip3 install aliyun-python-sdk-ess
     ```

Use Auto Scaling SDK for Python

In the following sample code, the CreateScalingGroup operation is used to describe how to use Auto Scaling SDK for Python to call an API operation.

1. Create a client instance.

   When you create a client instance, you must obtain a region ID, AccessKey ID, and AccessKey secret.

2. Create an API request and configure the required parameters.
3. Initiate the API request and handle the response or exception.

> **Note** In the following sample code, replace the variable parameters with actual values. Delete the angle brackets (<>) in the final code.
>
> Variable parameters:
>
> - <yourRegionID>: Replace the variable parameter with your region ID.
> - <yourAccessKeyID>: Replace the variable parameter with your AccessKey ID.
> - <yourAccessKeySecret>: Replace the variable parameter with your AccessKey secret.
> - <ess.aliyuncs.com>: Replace the variable parameter with the endpoint of your service.

```
#!/usr/bin/env python
#coding=utf-8
from aliyunsdkcore import client
from aliyunsdkcore.profile import region_provider
from aliyunsdkess.request.v20140828 import CreateScalingGroupRequest
# global config
region_id = '<yourRegionID>' # Specify your region ID.
access_key = '<yourAccessKeyID>' # Specify your AccessKey ID.
acces_secret = '<yourAccessKeySecret>' # Specify your AccessKey secret.
# product config
product_name = 'ess' # Specify your service name.
endpoint = '<ess.aliyuncs.com>' # Specify the endpoint of your service.
# create client
region_provider.modify_point(product_name, region_id, endpoint)
clt = client.AcsClient(access_key, acces_secret, region_id)
# setup request
request = CreateScalingGroupRequest()
request.set_accept_format('json') # Specify the format of the response.
# get response
response = clt.do_action_with_exception(request)
print response
```

4.2. Use Auto Scaling SDK for Python

4.2.1. Install Auto Scaling SDK for Python

This topic describes how to install and use Auto Scaling SDK for Python.

Context

Apsara Stack SDK for Python supports Python 2.6.x, Python 2.7.x, and Python 3.x.x or later. You can install Auto Scaling SDK for Python by using pip.

> **Note** To use the SDK for Python of a cloud service, you must install the core library of Apsara Stack SDK for Python. For example, if you want to use Auto Scaling SDK for Python, you must install the core library of Apsara Stack SDK for Python and Auto Scaling SDK for Python.

Procedure

1. Install the core library.

- If you use Python 2.x.x, run the following command to install the core library:

```
pip install aliyun-python-sdk-core
```

- If you use Python 3.x.x, run the following command to install the core library:

```
pip3 install aliyun-python-sdk-core-v3
```

2. Install Auto Scaling SDK for Python.

   The following rules apply:

   - If you use Python 2.x.x, run the following command:

```
pip install aliyun-python-sdk-ess
```

   - If you use Python 3.x.x, run the following command:

```
pip3 install aliyun-python-sdk-ess
```

4.2.2. Configure authentication credentials

When you use Apsara Stack SDKs to access Apsara Stack services, you must provide an Apsara Stack tenant account for authentication.

Auto Scaling SDK for Python supports authentication based on AccessKey credentials. For information about how to obtain the AccessKey credentials, see Obtain AccessKey credentials.

If you want to use your AccessKey credentials to access Apsara Stack, you must specify the AccessKey credentials as the authentication credentials when you initialize the client.

Notice Do not disclose code that may contain your AccessKey credentials. For example, do not commit the code to public GitHub projects. Otherwise, your Apsara Stack tenant account may be compromised.

```
clt = client.AcsClient(secret_id, secret_key, region_id)
```

Note In the preceding code, replace the variable parameters with your actual information. Delete the angle brackets (<>) in the final code.

4.2.3. Set HTTPS requests

The Python SDK supports using HTTP and HTTPS to initiate API requests.

When using the Python SDK, you can specify a protocol (HTTP or HTTPS) for a specific request to use, or set a global default protocol for all requests to use.

Note The default protocol (HTTP or HTTPS) of a product takes precedence over the global default protocol you set.

Add OpenSSL dependency

HTTPS of the Python SDK depends on the OpenSSL support in Python. To use Alibaba Cloud SDK to send HTTPS requests, you must add OpenSSL support in Python.

Run the `python -c "import ssl"` command to check if the Python environment supports OpenSSL. If the `ImportError: No module named ssl` error message does not appear, OpenSSL is supported in Python.

If Python does not support OpenSSL, run the following command to install OpenSSL:

```
pip install pyopenssl
```

Set HTTP/HTTPS for a request

See the following code example for reference to set the HTTPS call for an API:

```
request = CreateInstanceRequest.CreateInstanceRequest()
request.set_protocol_type("https")
# Value: "https" or "http"
```

Set global default protocol for all requests

See the following code example for reference to set the global default protocol for all requests:

```
import aliyunsdkcore.request
aliyunsdkcore.request.set_default_protocol_type("https")
# Create a request and call client.do_action_with_exception() to send the request.
```

4.2.4. Use Auto Scaling SDK for Python to initiate requests

This topic describes how to use Auto Scaling SDK for Python to initiate a request.

Procedure

1. Import the Auto Scaling SDK for Python.

   In the following code, the DescribeScalingGroups API operation is used.

   ```
   from aliyunsdkcore import client
   from aliyunsdkcore.profile import region_provider
   from aliyunsdkess.request.v20140828 import DescribeScalingGroupsRequest
   ```

2. Create an AcsClient.

   ```
   client.region_provider.modify_point(product_name, region_id, endpoint)
   clt = client.AcsClient(access_key, access_secret, region_id)
   ```

3. Create a request.

   ```
   request = DescribeScalingGroupsRequest()
   request.set_accept_format('json')
   ```

4. Initiate a call and print the response.

```
response = clt.do_action_with_exception(request)
print response
```

4.2.5. Handle errors

If an error occurs on the server side or SDK side when you use the Python SDK, the Python SDK returns the corresponding exception message. An exception message contains the detailed error information, namely the error code and the error message.

Generally, you are not required to handle the errors returned by the Python SDK. You are only required to handle the errors on the server side according to the error message.

- `ServerException` is the error message on the server side.
- `ClientException` is the error message on the SDK side.

For example, if the following error occurs, modify the ID of AccessKey according to the detailed error message.

```
aliyunsdkcore.acs\_exception.exceptions.ServerException: HTTP Status: 404 Error:InvalidAcce
ssKeyId.NotFound Specified access key is not found.
```

To handle the errors returned by the Python SDK, see the following code example for reference:

```
try:
    response = client.do_action_with_exception(request)
except ServerException as e:
    # Add your own error handling logic here.
    # For example, print the detailed error message.
    print e.get_http_status()
    print e.get_error_code()
    print e.get_error_msg()
```

4.2.6. Use CommonRequest to call API operations

If you want to call the API operations of Auto Scaling that do not provide an SDK or if the requests and responses do not meet your business requirements, you can use CommonRequest. CommonRequest supports all API operations.

Benefits

CommonRequest provides the following benefits:

1. Lightweight: You need to only download the core package. You do not need to download and install SDKs of Apsara Stack services.
2. Easy to use: You can call the new API operations of a service without the need to update the SDK of the service.
3. Fast iteration.

Use CommonRequest

If you want to send a CommonRequest request, you must obtain the values of the following parameters. You can obtain the values of these parameters from Auto Scaling API References.

- Domain: the service endpoint.
- Version: the version of the API in the YYYY-MM-DD format. You can obtain the API version from the common parameters in the API documents.
- Operation information: the name of the API operation that you want to call. You must obtain the value of the Action parameter and then run `request.ApiName = "<Action>"` to specify the name of the operation.

  For example, if you want to call the DescribeScalingActivities operation, you must specify the operation as `request.ApiName = "DescribeScalingActivities"` before you send the request by using CommonRequest.

Example: Call an RPC API operation

In this example, the CreateScalingGroup API operation of Auto Scaling is used to show how to call an RPC API to create a scaling group.

> **Note** In the following code, replace the variable parameters with your actual values. Delete the angle brackets <> in the final code.
>
> Variable parameters:
>
> - <yourRegionID>: Replace the variable parameter with your region ID.
> - <yourAccessKeyID>: Replace the variable parameter with your AccessKey ID.
> - <yourAccessKeySecret>: Replace the variable parameter with your AccessKey secret.
> - <yourSecurityToken>: Replace the variable parameter with your security token.
> - <yourResourceGroupID>: Replace the variable parameter with your resource set ID.
> - <yourOrganizationID>: Replace the variable parameter with your organization ID.
> - <yourInstanceID>: Replace the variable parameter with your instance ID.
> - <your-endpoint>: Replace the variable parameter with the endpoint of your service.

```
from aliyunsdkcore.client import AcsClient
from aliyunsdkcore.request import CommonRequest
# Specify the AccessKey ID, AccessKey secret, the region ID.
client = AcsClient('{<yourAccessKeyID>}', '{<yourAccessKeySecret>}', '{<yourRegionID>}')
request = CommonRequest()
request.set_product('ess')
request.set_action_name('CreateScalingGroup')
request.set_version('2014-08-28')
request.set_domain('<ess.aliyuncs.com>') # Specify the endpoint of the service.
# or:
# request = = CommonRequest(product='ess', action_name='CreateScalingGroup',version='2014-0
8-28',domain='ess.aliyuncs.com')
######
# Specify the request protocol. Valid values: https and http. Default value: http.
# The following example shows how to set the request protocol to https:
# request.set_protocol_type('https')
######
request.set_accept_format('json')
request.set_method('POST')
request.add_query_param('PageNumber', '1')
request.add_query_param('PageSize', '30')
######
# When you use STS credentials to call an API operation, you must configure the SecurityTok
en field.
# Format:
# request.add_query_param("SecurityToken", "<yourSecurityToken>");
# Specify the STS token.
######
# Optional. Specify the resource set ID.
request.add_header("x-acs-resourcegroupid", "<yourResourceGroupID>")
# Required. Specify the organization ID.
request.add_header("x-acs-organizationid", "<yourOrganizationID>")
# Required. Specify the source of the API call.
request.add_header("x-acs-caller-sdk-source","ApiTestDemo")
# Optional. Specify the region ID. If you use Apsara Stack Enterprise V3.15.0 or earlier, t
his parameter is required.
request.add_header("x-acs-regionid", "<yourRegionID>")
# Optional. Specify the instance ID. If you want to modify, delete, or query an instance, t
his parameter is required.
# request.add_header("x-acs-instanceid", "<yourInstanceID>")
response = client.do_action_with_exception(request)
print(str(response, encoding = 'utf-8'))
```

5.Examples of using Auto Scaling SDK for Go

5.1. Getting started

Alibaba Cloud SDK for Go allows you to access services without the need for complex coding. This topic describes how to obtain and use Auto Scaling SDK for Go.

Environment preparation

- An Apsara Stack tenant account and an AccessKey pair are obtained. The account must have the required permissions. The AccessKey pair consists of an AccessKey ID and an AccessKey secret. For information about how to obtain AccessKey pairs, see Obtain AccessKey credentials.
- The endpoint of Auto Scaling is obtained. For more information, see Obtain the endpoint of Auto Scaling.

Install the core package of Apsara Stack SDK for Go

Apsara Stack SDK for Go supports Go 1.10.x or later. You can run the following command to install the core package:

```
go get -u github.com/aliyun/alibaba-cloud-sdk-go/sdk
```

Use Auto Scaling SDK for Go

In the following code, the DescribeScalingGroups is used to describe how to use Auto Scaling SDK for Go to call an API operation.

1. Initialize a client instance.
2. Create an API request and configure the required parameters.
3. Send the API request and handle the response or exception.

> **Note** In the following code, replace the variable parameters with your actual information. Delete the angle brackets (<>) in the final code.
>
> Variable parameters:
>
> - <yourRegionID>: Replace the variable parameter with your region ID.
> - <yourAccessKeyID>: Replace the variable parameter with your AccessKey ID.
> - <yourAccessKeySecret>: Replace the variable parameter with your AccessKey secret.
> - <Endpoint>: Replace the variable parameter with your endpoint.

```
package main
import (
    "github.com/aliyun/alibaba-cloud-sdk-go/services/ess"
    "github.com/aliyun/alibaba-cloud-sdk-go/sdk/requests"
    "fmt"
)
func main() {
    // Create an essClient object.
    essClient, err := ess.NewClientWithAccessKey(
        "<yourRegionID>",             // Replace the variable parameter with your regio
n ID.
        "<yourAccessKeyID>",         // Replace the variable parameter with your Access
Key ID.
        "<yourAccessKeySecret>")        // Replace the variable parameter with your Acc
essKey secret.
    if err != nil {
        // Handle exceptions.
        panic(err)
    }
    // Create a request and configure the required parameters.
    request := ess.CreateDescribeScalingGroupsRequest()
    // Specify the endpoint of Auto Scaling.
    request.Domain = "<Endpoint>"
    // Set the request.PageSize parameter to 10.
    request.PageSize = requests.NewInteger(10)
    // Initiate the request and handle exceptions.
    response, err := Client.DescribeScalingGroups(request)
    if err != nil {
        // Handle exceptions.
        panic(err)
    }
    fmt.Println(response)
}
```

4. (Optional) Switch to HTTP or HTTPS.

```
request.Scheme = "HTTPS"
```

5.2. Work with Auto Scaling SDK for Go

5.2.1. Install Go SDK

This topic describes how to install the Go SDK.

The Alibaba Cloud Go SDK supports Go 1.7 or later versions. Install the Go SDK manually.

1. Download the source code of the Go SDK from the official site of Apsara Stack.
2. Extract the package to the directory where the *GO_PATH* file of the current user resides.

5.2.2. Configure authentication credentials

When you use Alibaba Cloud SDKs to access Apsara Stack services, you must provide an Apsara Stack tenant account for authentication. This topic describes how to configure authentication credentials.

Auto Scaling SDK for Go supports authentication based on an AccessKey pair. For more information, see Obtain AccessKey credentials.

If you want to use your AccessKey pair to access Apsara Stack, you must specify the AccessKey pair as the credentials when you initialize the client.

Notice Do not disclose any code that contains your AccessKey pair. For example, do not commit the code to public GitHub projects. Otherwise, your account may be compromised.

The following sample code provides an example on how to initialize an Auto Scaling client:

```
// Initialize the Auto Scaling client.
essClient, err := ess.NewClientWithAccessKey(
        "<yourRegionID>",                 // Replace the variable parameter with your regio
n ID.
        "<yourAccessKeyID>",             // Replace the variable parameter with your Access
Key ID.
        "<yourAccessKeySecret>")        // Replace the variable parameter with your AccessK
ey secret.
```

Note In the preceding code, replace the variable parameters with your actual information. Delete the angle brackets (<>) in the final code.

5.2.3. Initiate a request

This topic describes how to use Auto Scaling SDK for Go to initiate a request.

To use Auto Scaling SDK for Go to initiate a request, perform the following steps:

1. Import Auto Scaling SDK for Go

   ```
   import github.com/aliyun/alibaba-cloud-sdk-go/services/ess
   ```

2. Create an AcsClient instance that contains all API operations of Auto Scaling.

   Note In the following sample code, replace the variable parameters with actual values. Delete the angle brackets (<>) in the final code.

   ```
   essClient,err := <ess>.NewClientWithAccessKey(  // Replace the variable parameter with
   the name of your service.
       "<yourRegionID>",  // Replace the variable parameter with your region ID.
       "<yourAccessKeyID>",  // Replace the variable parameter with your AccessKey ID.

       "<yourAccessKeySecret>")  // Replace the variable parameter with your AccessKey sec
   ret.
   essClient.Domain = "ess.aliyuncs.com" // Specify the endpoint of your service.
   ```

3. Use the `xxx.CreateXXXRequest` method to create a request.

The naming convention for this method is `${service}.Create${apiName}Request`.

- ${service} specifies the name of the service in lowercase. Example: *ess*.
- ${apiName} specifies the name of the operation that you want to perform. Example: *DescribeScalingGroups*.

```
request := ess.CreateDescribeScalingGroupsAttributeRequest()
request.InstanceId = ""
```

4. Initiate the call.

```
response, err := essClient.DescribeScalingGroupsAttribute(request)
```

5.2.4. Handle errors

If an error occurs on the server or in the SDK when you can use Auto Scaling SDK for Go, a `ServerError` or a `ClientError` error is returned. Both error types are included in `sdk.errors`. The errors conform to the standard Go error interface and can be handled in a similar manner as standard Go errors.

Handle errors

In the following sample code, the DescribeScalingGroups operation is used to describe how to handle errors.

```
response,err  := essClient.DescribeScalingGroups(request)
if err != nil {
    // Obtain the error information.
   err.Error()
}
```

ClientError

If an error occurs in the SDK when you initiate a request and the error cannot be automatically handled, the SDK returns a `ClientError`.

In the following sample example, the DescribeScalingGroups operation is used to describe how to determine and obtain the error information in a `ClientError` that is returned by Auto Scaling SDK for Go.

Notice When you use a specific error struct, you must add the error package `github.com/aliyun/alibaba-cloud-sdk-go/sdk/errors` by using the following code:

```
package main
import (
    "github.com/aliyun/alibaba-cloud-sdk-go/services/ess" // Replace ess with the name
of your service.
    "github.com/aliyun/alibaba-cloud-sdk-go/sdk/requests"
    // Specify the error package that you want to add.
    "github.com/aliyun/alibaba-cloud-sdk-go/sdk/errors"
    "fmt"
)
```

```
response, err := essClient.DescribeScalingGroups(request)
if clientError, ok := err.(*errors.ClientError); ok{
    // Obtain the error code.
    clientError.ErrorCode()
    // Obtain the error information.
    clientError.Message()
    // Obtain the original error. The return value may be nil.
    clientError.OriginError()
}
```

ServerError

When the server returns an error message, the SDK wraps the message into a `ServerError` and returns it to the client. In this case, you can still obtain the original HTTP response from Response.

In the following sample example, the DescribeScalingGroups operation is used to describe how to determine and obtain the error information in a ServerError and a Response that are returned by Auto Scaling SDK for Go.

Notice When you use a specific error struct, you must add the error package `github.com/aliyun/alibaba-cloud-sdk-go/sdk/errors` by using the following code:

```
package main
import (
    "github.com/aliyun/alibaba-cloud-sdk-go/services/ess" // Replace ess with the name
of your service.
    "github.com/aliyun/alibaba-cloud-sdk-go/sdk/requests"
    // Specify the error package that you want to add.
    "github.com/aliyun/alibaba-cloud-sdk-go/sdk/errors"
    "fmt"
)
```

```
response, err := essClient.DescribeScalingGroups(request)
if serverError, ok := err.(*errors.ServerError); ok{
    // Obtain the error code.
    serverError.ErrorCode()
    // Obtain the error information.
    serverError.Message()
    // Obtain the original HTTP response.
    response.GetOriginHttpResponse()
}
```

5.2.5. Use CommonRequest to call API operations

You can use CommonRequest to call API operations in the following scenarios: You want to call Auto Scaling API operations that do not provide SDKs. The requests and responses of the API operations that you called by using an Auto Scaling SDK do not meet your business requirements. You can use CommonRequest to call all API operations.

Benefits

CommonRequest provides the following benefits:

1. Lightweight: You need to only download the core package. You do not need to download and install SDKs of Apsara Stack services.
2. Easy to use: You can call the new API operations of a service without the need to update the SDK of the service.
3. Fast iteration.

Use CommonRequest

If you want to send a CommonRequest request, you must obtain the values of the following parameters. You can obtain the values of these parameters from Auto Scaling API References.

- Domain: the service endpoint.
- Version: the version of the API in the YYYY-MM-DD format. You can obtain the API version from the common parameters in the API documents.
- Operation information: the name of the API operation that you want to call. You must obtain the value of the Action parameter and then run `request.ApiName = "<Action>"` to specify the name of the operation.

  For example, if you want to call the DescribeScalingActivities operation, you must specify the operation as `request.ApiName = "DescribeScalingActivities"` before you send the request by using CommonRequest.

Example: Call an RPC API operation

The following sample code provides an example on how to obtain information about scaling activities by calling the DescribeScalingActivities operation of Auto Scaling.

Note In the following sample code, replace the variable parameters with actual values. Delete the angle brackets (<>) in the final code.

Variable parameters:

- <yourRegionID>: Replace the variable parameter with your region ID.
- <yourAccessKeyID>: Replace the variable parameter with your AccessKey ID.
- <yourAccessKeySecret>: Replace the variable parameter with your AccessKey secret.
- <vpc.aliyuncs.com>: Replace the variable parameter with the endpoint of your service.
- <yourSecurityToken>: Replace the variable parameter with your security token.
- <yourResourceGroupID>: Replace the variable parameter with your resource set ID.
- <yourOrganizationID>: Replace the variable parameter with your organization ID.
- <yourInstanceID>: Replace the variable parameter with your instance ID.

```
package main
import (
    "github.com/aliyun/alibaba-cloud-sdk-go/sdk/requests"
    "github.com/aliyun/alibaba-cloud-sdk-go/sdk"
    "fmt"
)
func main() {
    client, err := sdk.NewClientWithAccessKey(
        "<yourRegionID>",        // Replace the variable parameter with your region ID.
        "<yourAccessKeyID>",     //Replace the variable parameter with your AccessKey ID.
        "<yourAccessKeySecret>") //Replace the variable parameter with your AccessKey secre
t.
    if err != nil {
        panic(err)
    }
    request := requests.NewCommonRequest()
    request.Method = "POST"
    request.Domain = "<ess.aliyuncs.com>" // Replace the variable parameter with the endpoi
nt of your service.
    request.Product = "ess"
    request.Version = "2014-08-28"
    // This is an RPC API operation. Therefore, the ApiName(Action) parameter must be confi
gured.
    request.ApiName = "DescribeScalingGroups"
    // Configure service-related parameters.
    request.QueryParams["PageNumber"] = "1"
    request.QueryParams["PageSize"] = "30"
    /*
     When you use STS AccessKey credentials to call an RPC API operation, you must add the
SecurityToken field to the request parameters.
     Format:
     request.QueryParams["SecurityToken"]="<yourSecurityToken>"
     Specify your security token.
     */
    request.Headers["x-acs-caller-sdk-source"] = "ApiTestDemo" // Required. Specify the sou
rce of the call.
    request.Headers["x-acs-regionid"] = "<yourRegionID>" // Optional. Specify your region I
D. If you use Apsara Stack Enterprise V3.15.0 or earlier, you must configure this parameter
.
    request.Headers["x-acs-resourcegroupid"] = "<yourResourceGroupID>" // Optional. Specify
the resource set ID.
    request.Headers["x-acs-organizationid"] = "<yourOrganizationID>" // Required. Specify t
he organization ID.
    // request.Headers["x-acs-instanceid"] = "<yourInstanceID>" // Optional. Specify the in
stance ID. If you want to modify, delete, or query an instance, you must configure this par
ameter.
    response, err := client.ProcessCommonRequest(request)
    if err != nil {
        panic(err)
    }
    fmt.Print(response.GetHttpContentString())
}
```

6.API reference

6.1. API Reference

6.1.1. API usage instructions

Before you can call Auto Scaling API operations, you must be granted the required permissions on Auto Scaling API in the Auto Scaling console.

The following table describes the errors that occur if you are not granted the required permissions.

Description

HTTP status code	Error code	Error message	Description
403	Forbidden.Unsubscribed	Do not have permission to access this API.	The error message returned because Auto Scaling is not activated and you do not have the permissions to call Auto Scaling API operations.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.

6.1.2. Make API requests

Call an Auto Scaling API operation, send an HTTP GET request to the Auto Scaling API server, and then add corresponding parameters to the request based on the description of the corresponding API operation. After the request is processed, the system returns a result.

6.1.3. Request structure

This topic describes the structure of an API request.

- Endpoints

  For information about how to obtain an endpoint, see Generate the endpoint of ASAPI.

- Protocols

  You can send requests over HTTP or HTTPS. For a higher level of security, we recommend that you send requests over HTTPS.

- Methods

  You can use the HTTP GET method to send the request for calling an API operation. If you use this method, the request parameters must be included in the URL of the request.

- Parameters

For each request, you must specify the operation that you want to perform, such as CreateScalingGroup. Additionally, each request must contain the common request parameters and operation-specific request parameters.

- Encoding

  All requests and responses are encoded in UTF-8.

6.1.4. Common parameters

Common request parameters

Common request parameters must be included in all API requests. The following table describes the common request parameters.

Parameter	Type	Required	Description
Format	String	No	The format of the response. Set the value to JSON.
Version	String	Yes	The version number of the API in the YYYY-MM-DD format. Set the value to 2014-08-28.
AccessKeyId	String	Yes	The AccessKey ID that is provided by Apsara Stack.
AccessKeySecret	String	Yes	The AccessKey secret that is provided by Apsara Stack.
RegionId	String	Yes	The region ID that is provided by Apsara Stack.
Action	String	Yes	The operation that you want to perform.
Product	String	Yes	The name of the service to which the API operation belongs. Set the value to Ess.

You can obtain the values of the `Product` and `Version` parameters in the Apsara Uni-manager Management Console. To obtain the values, perform the following steps:

1. Log on to the Apsara Uni-manager Management Console.
2. In the top navigation bar, choose **Products > Others > API Tool**.
3. In the left-side navigation pane of the API Tool page, click **API Directory**.
4. In the **Operation Name** search box, enter the name of the operation whose information you want to query.

Obtain the values of common header parameters

Parameter	Description
x-acs-organizationId	The ID of the organization in the Apsara Uni-manager Management Console. For information about how to obtain organization IDs, see Obtain an organization ID.

Parameter	Description
x-acs-resourcegroupId	The ID of the resource set in the Apsara Uni-manager Management Console. For information about how to obtain resource set IDs, see Obtain a resource set ID.
x-acs-instanceId	The ID of the instance on which you want to perform operations. For information about how to obtain instance IDs, see Obtain an instance ID.

Note When you call Auto Scaling API operations, you must configure the x-acs-organizationid, x-acs-resourcegroupid, and x-acs-instanceid header parameters to specify the organization ID, resource set ID, and instance ID. You cannot configure the x-acs-instanceid header parameter to call API operations to create instances and query instances. If you do not configure the preceding header parameters, the call is directed to the root resource set and organization.

Common response parameters

API responses use the HTTP response format:

- A 2xx status code indicates a successful call.
- A 4xx or 5xx status code indicates a failed call.

Responses are returned in the JSON format. Every response returns a unique RequestId regardless of whether the call is successful.

6.1.5. Responses

6.1.5.1. Successful response examples

API responses use the HTTP response format where a 2xx status code indicates a successful call and a 4xx or 5xx status code indicates a failed call. Responses can be returned in either the JSON or XML format. You can specify the response format in the request. The default response format is XML.

Note Sample responses in our API documents are formatted in a way that is easier for you to read. The actual responses are not formatted with line breaks or indentation.

Responses in the XML format include information about whether the request succeeds and specific service data.

```
<?xml version="1.0" encoding="UTF-8"? >
<!--Result Root Node-->
<Operation Name+Response>
    <!--Return Request Tag-->
    <RequestId>4C467B38-3910-447D-87BC-AC049166F216</RequestId>
    <!--Return Result Data-->
</Operation Name+Response>
```

JSON format

```
{
    "RequestId": "4C467B38-3910-447D-87BC-AC049166F216",
    /* Return Result Data */
}
```

6.1.5.2. Sample error responses

When an error occurs during an API call, no result data is returned. You can locate the cause of an error based on the error codes defined by each API operation, the error code table in the appendix Client errors, and Server errors.

When an error occurs on an HTTP request for an API operation, a 4xx or 5xx HTTP status code is returned. The response body includes the error code and error message. The response body also contains a globally unique request ID (RequestId) and the ID of the site you accessed with this request (HostId).

XML format

```
<?xml version="1.0" encoding="UTF-8"? >
<Error>
   <RequestId>8906582E-6722-409A-A6C4-0E7863B733A5</RequestId>
   <HostId>ess.example.com</HostId>
   <Code>UnsupportedOperation</Code>
   <Message>The specified action is not supported. </Message>
</Error>
```

JSON format

```
{
    "RequestId": "8906582E-6722-409A-A6C4-0E7863B733A5",
    "HostId": "ess.example.com",
    "Code": "UnsupportedOperation",
    "Message": "The specified action is not supported."
}
```

6.1.6. Request signatures

This topic describes how to use the request signature to verify the identity of the API caller.

Auto Scaling authenticates each access request. Therefore, each request must contain signature information, regardless of whether the request is sent over HTTP or HTTPS. Auto Scaling implements symmetric encryption by using an AccessKey pair to verify the identity of the request sender. You can obtain the AccessKey ID and the AccessKey secret in the Apsara Uni-manager Management Console. For more information, see Obtain an AccessKey pair. The AccessKey ID is used to verify the identity of a user, while the AccessKey secret is used to encrypt and verify the signature string. You must keep your AccessKey secret confidential.

Perform the following operations to sign an access request:

1. Create a canonicalized query string by using request parameters.

i. Arrange all request parameters in alphabetical order. These parameters include all common request parameters and operation-specific parameters except the Signature parameter in the common request parameters.

   If you use the GET method to make a request, the request parameters are included as a part of the request URL. The first parameter follows a question mark (`?`) in the URL, and the other parameters follow an ampersand (`&`).

ii. Encode the name and value of each request parameter.

   The parameter names and values must be URL-encoded in UTF-8 based on the following encoding rules:

   - Uppercase letters, lowercase letters, digits, and some special characters such as ampersands (&), hyphens (-), underscores (_), periods (.), and tildes (~) do not need to be encoded.
   - Other characters must be percent encoded in the `%XY` format. `XY` represents the ASCII code of the characters in hexadecimal notation. For example, double quotation marks (") are encoded as `%22` .
   - Extended UTF-8 characters are encoded in the `%XY%ZA...` format.
   - Spaces must be encoded as `%20` . Do not encode spaces as plus signs `(+)` .

   > **Note** Most libraries that support URL encoding, such as java.net.URLEncoder, comply with the Multipurpose Internet Mail Extensions (MIME) encoding rules of `application/x-www-form-urlencoded` . If this encoding method is used, replace the plus signs (+) with `%20` , the asterisks (*) with `%2A` , and `%7E` with tildes (~) in the encoded strings to obtain the required strings.

iii. Connect the encoded parameter names and values with equal signs (=).

iv. Arrange the parameter name and value pairs connected by equal signs in alphabetical order, and connect them with ampersands (&) to obtain a canonicalized query string.

2. Create a string-to-sign from the encoded canonicalized query string based on the following rules:

```
StringToSign=
HTTPMethod + "&" +
percentEncode("/") + "&" +
percentEncode(CanonicalizedQueryString)
```

3. Use the signature string to calculate the HMAC value of the string-to-sign as defined in RFC 2104.

   The key used for signature calculation is the combination of your AccessKey secret and an ampersand (&) (ASCII 38). Secure Hash Algorithm 1 (SHA1) is used as the hash algorithm.

4. Encode the HMAC value in Base64 to obtain the signature string.
5. Add the signature string to the request as the Signature parameter to sign the request.

   When the signature string is submitted to the ECS instance as the final request parameter value, the string must be URL-encoded as other parameters based on the rules defined in RFC 3986.

DescribeScalingGroups is used in the example. The following section shows the request URL to be signed:

```
http://ess.example.com/?TimeStamp=2014-08-15T11%3A10%3A07Z&Format=xml&AccessKeyId=testid&Ac
tion=DescribeScalingGroups&SignatureMethod=HMAC-SHA1&RegionId=cn-qingdao&SignatureNonce=132
4fd0e-e2bb-4bb1-917c-bd6e437f****&SignatureVersion=1.0&Version=2014-08-28
```

The following section shows the corresponding StringToSign:

```
GET&%2F&AccessKeyId%3Dtestid&Action%3DDescribeScalingGroups&Format%3Dxml&RegionId%3Dcn-qing
dao&SignatureMethod%3DHMAC-SHA1&SignatureNonce%3D1324fd0e-e2bb-4bb1-917c-bd6e437f***&Signat
ureVersion%3D1.0&TimeStamp%3D2014-08-15T11%253A10%253A07Z&Version%3D2014-08-28
```

If the Access Key ID is testid and the AccessKey secret is testsecret, the key used for HMAC calculation is testsecret&. Then, the calculated signature value is SmhZuLUnXmqxSEZ%2FGqyiwGqmf%2BM=.

The following section shows the URL of the signed request:

```
http://ess.example.com/?TimeStamp=2014-08-15T11%3A10%3A07Z&Format=xml&AccessKeyId=testid&Ac
tion=DescribeScalingGroups&SignatureMethod=HMAC-SHA1&RegionId=cn-qingdao&SignatureNonce=132
4fd0e-e2bb-4bb1-917c-bd6e437f****&SignatureVersion=1.0&Version=2014-08-28&Signature=SmhZuLU
nXmqxSEZ%2FGqyiwGqmf%****
```

The signature string is used as the value of the Signature parameter, and the parameter is added to the request URL.

6.1.7. Scaling groups

6.1.7.1. CreateScalingGroup

Creates a scaling group.

Description

A scaling group is a group of Elastic Compute Service (ECS) instances that can be used in similar business scenarios.

- You can create up to 50 scaling groups in one region.
- A scaling group does not immediately take effect after you create it. You must call the EnableScalingGroup operation to enable the scaling group. After the scaling group is enabled, Auto Scaling can execute scaling rules to trigger scaling activities in the scaling group.
- Server Load Balancer (SLB) instances and ApsaraDB RDS instances that are associated with a scaling group must reside in the same region as the scaling group. For more information, see the *Regions and zones* topic in ECS documentation.
- If you specify an SLB instance when you create a scaling group, Auto Scaling automatically adds ECS instances in the scaling group to the backend server group of the specified SLB instance. The default weight of an ECS instance added to the backend server group is 50. The specified SLB instance must meet the following requirements:
  - The SLB instance must be in the Active state. You can call the DescribeLoadBalancers operation to query the status of the specified SLB instance.
  - The health check feature is enabled on all listener ports configured for the SLB instance. Otherwise, the scaling group cannot be created.

- If you specify an ApsaraDB RDS instance when you create a scaling group, Auto Scaling automatically adds the private IP addresses of the ECS instances in the scaling group to the whitelist that manages access to the specified ApsaraDB RDS instance. The specified ApsaraDB RDS instance must meet the following requirements:
  - The ApsaraDB RDS instance must be in the Running state. You can call the DescribeDBInstances operation to query the status of the specified ApsaraDB RDS instance.
  - The number of IP addresses in the whitelist that manages access to the ApsaraDB RDS instance cannot exceed the upper limit. For more information, see the *Configure a whitelist* topic in ApsaraDB RDS documentation.
- If the MultiAZPolicy parameter of the scaling group is set to COST_OPTIMIZED, the following rules apply:
  - You can use the OnDemandBaseCapacity, OnDemandPercentageAboveBaseCapacity, and SpotInstancePools parameters to specify the percentages of pay-as-you-go instances and preemptible instances based on the cost optimization policy. This instance allocation method is prioritized during scaling.
  - If OnDemandBaseCapacity, OnDemandPercentageAboveBaseCapacity, or SpotInstancePools is not specified, the instance types provided at the lowest cost are used to create instances based on the cost optimization policy.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	CreateScalingGroup	The operation that you want to perform. Set the value to CreateScalingGroup.
MaxSize	Integer	Yes	20	The maximum number of ECS instances in the scaling group. Valid values: 0 to 1000. If the number of ECS instances in the scaling group is greater than the value of the MaxSize parameter, Auto Scaling automatically removes ECS instances from the scaling group until the number of instances is equal to the value of this parameter.
MinSize	Integer	Yes	2	The minimum number of ECS instances in the scaling group. Valid values: 0 to 1000. If the number of ECS instances in the scaling group is less than the value of the MinSize parameter, Auto Scaling automatically creates ECS instances until the number of instances is equal to the value of this parameter.

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao-env17-d01	The ID of the region where the scaling group resides. For more information, see the *Regions and zones* topic in ECS documentation.
ScalingGroupName	String	No	scalinggroup****	The name of the scaling group. The name of a scaling group must be unique in a region. 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. If you do not specify this parameter, the scaling group ID is used.
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the launch template that is used by Auto Scaling to create ECS instances.
LaunchTemplateVersion	String	No	Default	The version of the instance launch template. Valid values: • A fixed template version number. • Default: the default template version. • Latest: the latest template version.
InstanceId	String	No	i-28wt4****	The ID of the ECS instance from which Auto Scaling obtains instance configuration information.
DefaultCooldown	Integer	No	300	The cooldown time after a scaling activity is executed. Valid values: 0 to 86400. Unit: seconds. Default value: 300. **Note** During the cooldown time, Auto Scaling executes only the scaling activities that are triggered by event-triggered tasks monitored by CloudMonitor in the scaling group.

Parameter	Type	Required	Example	Description
LoadBalancerIds	String	No	["lb-bp1u7etiogg38yvwz****", "lb-bp168cqrux9ai9l7f****", "lb-bp1jv3m9zvj22ufxp****"]	The ID of the SLB instance. This value can be a JSON array that consists of up to five SLB instance IDs. Separate multiple IDs with commas (,).
DBInstanceIds	String	No	["rm-bp142f86de0t7*****", "rm-bp18l1z42ar4o****", "rm-bp1lqr97h4aqk****"]	The ID of the ApsaraDB RDS instance. This value can be a JSON array that contains up to eight ApsaraDB RDS instance IDs. Separate multiple IDs with commas (,).
RemovalPolicy.1	String	No	OldestScalingConfiguration	The first step for the instance removal policy. For more information, see the "Removal policies" topic. Valid values: • OldestInstance: removes ECS instances that are added to the scaling group at the earliest point in time. • NewestInstance: removes ECS instances that are most recently added to the scaling group. • OldestScalingConfiguration: removes ECS instances that are created based on the earliest scaling configuration. The default value of RemovalPolicy.1 is OldestScalingConfiguration if both RemovalPolicy.1 and RemovalPolicy.2 are not specified.

Parameter	Type	Required	Example	Description
RemovalPolicy.2	String	No	OldestInstance	The second step for the instance removal policy. This parameter value cannot be the same as the RemovalPolicy.1 value. For more information, see the "Removal policies" topic. Valid values: • OldestInstance: removes ECS instances that are added to the scaling group at the earliest point in time. • NewestInstance: removes ECS instances that are most recently added to the scaling group. • OldestScalingConfiguration: removes ECS instances that are created based on the earliest scaling configuration. The default value of RemovalPolicy.2 is OldestInstance if both RemovalPolicy.1 and RemovalPolicy.2 are not specified.
VSwitchId	String	No	vsw-5le3tkkyc8b7jhr2g****	The ID of the vSwitch. This parameter takes effect only when the network type of the scaling group is Virtual Private Cloud(VPC).

Parameter	Type	Required	Example	Description
VSwitchIds.N	RepeatList	No	vsw-5le3tkkyc8b7jhr2g****	The ID of vSwitch N. Valid values of N: 1 to 5. • If you specify the VSwitchIds.N parameter, the VSwitchId parameter is ignored. • This parameter takes effect only when the network type of the scaling group is VPC. The specified vSwitches and the scaling group must reside in the same VPC. • The vSwitches can reside in different zones. The vSwitches are sorted in ascending order based on the value of N. 1 specifies the highest priority. A greater value of N specifies a lower priority. If Auto Scaling fails to create ECS instances in the zone where the vSwitch that has the highest priority resides, Auto Scaling creates ECS instances in the zone where the vSwitch that has the next highest priority resides.

Parameter	Type	Required	Example	Description
MultiAZPolicy	String	No	PRIORITY	The scaling policy for the multi-zone scaling group that contains ECS instances. Valid values: • PRIORITY: ECS instances are scaled based on the VSwitchIds.N parameter. If Auto Scaling fails to create ECS instances in the zone where the vSwitch that has the highest priority resides, Auto Scaling creates ECS instances in the zone where the vSwitch that has the next highest priority resides. • COST_OPTIMIZED: ECS instances are created based on the unit prices of their vCPUs. ECS instances that have vCPUs of the lowest price are preferentially created. Preemptible instances are created first when preemptible instance types are specified for the scaling configuration. Pay-as-you-go instances are created if no preemptible instances are available due to issues such as insufficient resources. Note The COST_OPTIMIZED setting takes effect only when multiple instance types are specified or preemptible instances are used. • BALANCE: ECS instances are evenly distributed across multiple zones that are specified in the scaling group. If ECS instances are unevenly distributed across multiple zones due to issues such as insufficient resources, you can call the RebalanceInstance operation to evenly distribute the instances across the zones. Default value: PRIORITY

Parameter	Type	Required	Example	Description
HealthCheckType	String	No	ECS	The health check mode of the scaling group. Valid values: • NONE: Auto Scaling does not perform health checks on instances in the scaling group. • ECS: Auto Scaling performs health checks on ECS instances in the scaling group.
LifecycleHook.N.LifecycleHookName	String	No	lifecycle_hook	The name of the lifecycle hook. The name of a lifecycle hook cannot be modified after you specify it.
LifecycleHook.N.LifecycleTransition	String	No	SCALE_OUT	The type of scaling activity to which the lifecycle hook applies. Valid values: • SCALE_OUT • SCALE_IN

Parameter	Type	Required	Example	Description
LifecycleHook.N.DefaultResult	String	No	CONTINUE	The action that you want Auto Scaling to perform when the lifecycle hook times out. Valid values: • CONTINUE: Auto Scaling continues to respond to scale-in or scale-out requests. • ABANDON: Auto Scaling releases ECS instances that are created during scale-out activities. Auto Scaling removes ECS instances during scale-in activities. Default value: CONTINUE. If multiple lifecycle hooks in a scaling group are triggered during scale-in activities and you set the DefaultResult parameter to ABANDON for the lifecycle hook that you want to modify, Auto Scaling immediately starts to perform the action after the lifecycle hook that you want to modify times out. As a result, other lifecycle hooks time out ahead of schedule. In other cases, Auto Scaling performs the action only after all lifecycle hooks time out.

Parameter	Type	Required	Example	Description
LifecycleHook.N.HeartbeatTimeout	Integer	No	600	The timeout period of the lifecycle hook. When the lifecycle hook times out, Auto Scaling performs the default action. Unit: seconds. Valid values: 30 to 21600. Default value: 600. After you create a lifecycle hook, you can call the RecordLifecycleActionHeartbeat operation to extend the timeout period and keep the instances in the Pending Add or Pending Remove state. You can also call the CompleteLifecycleAction operation to end the timeout period in advance.
LifecycleHook.N.NotificationMetadata	String	No	Test	The fixed string that you want to include in the notification that Auto Scaling sends, which indicates that the scaling activity is pending. The parameter value cannot exceed 128 characters in length. Auto Scaling sends the value that you specified for the NotificationMetadata parameter together with the notification. This helps you categorize your notifications. The NotificationMetadata parameter takes effect only after you specify the NotificationArn parameter.
LifecycleHook.N.NotificationArn	String	No	acs:ess:cn-hangzhou:1111111111:queue/queue2	The Alibaba Cloud Resource Name (ARN) of the notification method that Auto Scaling uses when a lifecycle hook takes effect. This method can be a Message Service (MNS) queue or an MNS topic. The parameter value is in the acs:ess:{region}:{account-id}:{resource-relative-id} format. • region: the region where the scaling group resides • account-id: the ID of the Apsara Stack tenant account Example: • MNS queue: acs:ess:{region}:{account-id}:queue/{queuename} • MNS topic: acs:ess:{region}:{account-id}:topic/{topicname}

Parameter	Type	Required	Example	Description
VServerGroup.N.LoadBalancerId	String	No	lb-bp1u7etiogg38yvwz****	The ID of the SLB instance to which vServer group N belongs. You can call the AttachVServerGroups operation to attach vServer groups to SLB instances.
VServerGroup.N.VServerGroupAttribute.N.VServerGroupId	String	No	rsp-bp1443g77****	The ID of vServer group N. You can call the AttachVServerGroups operation to attach vServer groups to SLB instances.
VServerGroup.N.VServerGroupAttribute.N.Port	Integer	No	22	The port number that is used by Auto Scaling to add the ECS instances to vServer group N. Valid values: 1 to 65535. You can call the AttachVServerGroups operation to attach vServer groups to SLB instances.
VServerGroup.N.VServerGroupAttribute.N.Weight	Integer	No	100	The weight that you want to specify for the ECS instance that is added to vServer group N. Valid values: 0 to 100. Default value: 50. You can call the AttachVServerGroups operation to attach vServer groups to SLB instances.
ScalingPolicy	String	No	recycle	The instance reclaim mode of the scaling group. Valid values: • recycle • release You can call the RemoveInstances operation to remove instances from the scaling group.

Parameter	Type	Required	Example	Description
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can contain only ASCII characters and cannot exceed 64 characters in length. For more information, see the *How to ensure idempotence* topic in ECS documentation.
OnDemandBaseCapacity	Integer	No	30	The minimum number of pay-as-you-go instances that must be included in the scaling group. Valid values: 0 to 1000. If the number of pay-as-you-go instances is less than the value of this parameter, Auto Scaling preferentially creates pay-as-you-go instances.
OnDemandPercentageAboveBaseCapacity	Integer	No	20	The percentage of pay-as-you-go instances in the excess instances when the minimum number of pay-as-you-go instances reaches the requirement. Valid values: 0 to 100.
SpotInstanceRemedy	Boolean	No	true	Specifies whether to supplement preemptible instances. If this parameter is set to true, Auto Scaling creates an instance to replace a preemptible instance when Auto Scaling receives a system message indicating that the preemptible instance is to be reclaimed.
CompensateWithOnDemand	Boolean	No	true	Specifies whether to automatically create pay-as-you-go instances to meet the requirements on the number of ECS instances when the expected capacity of preemptible instances cannot be fulfilled due to reasons such as costs or insufficient resources. This parameter takes effect only when MultiAZPolicy is set to COST_OPTIMIZED. Valid values: • true • false Default value: true.

Parameter	Type	Required	Example	Description
SpotInstancePools	Integer	No	5	The number of instance types that you specify. Auto Scaling creates preemptible instances of multiple instance types that are provided at the lowest cost. Valid values: 1 to 10.
DesiredCapacity	Integer	No	5	The expected number of ECS instances in the scaling group. Auto Scaling automatically maintains the expected number of ECS instances that you specify. The expected number cannot be greater than the value of the MaxSize parameter and cannot be less than the value of the MinSize parameter.
GroupDeletionProtection	Boolean	No	true	Specifies whether to enable deletion protection for the scaling group. Valid values: • true: enables deletion protection for the scaling group. This way, the scaling group cannot be deleted. • false: disables deletion protection for the scaling group. Default value: false.
Tag.N.Key	String	No	Department	The key of tag N of the scaling group.
Tag.N.Value	String	No	Finance	The value of tag N of the scaling group.

Parameter	Type	Required	Example	Description
LaunchTemplateOverride.N.InstanceType	String	No	ecs.c5.xlarge	If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.InstanceType and LaunchTemplateOverride.N.WeightedCapacity parameters at the same time. This parameter specifies instance type N, which overwrites the instance type that is specified in the launch template. You can specify N instance types in the extended configurations of the launch template. Valid values of N: 1 to 10. **Note** This parameter takes effect only when you specify the LaunchTemplateId parameter. For information about the valid values of InstanceType in LaunchTemplateOverride.N.InstanceType, see the *Instance type families* topic in ECS documentation.

Parameter	Type	Required	Example	Description
LaunchTemplateOverride.N.WeightedCapacity	Integer	No	4	If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.WeightedCapacity parameter after you specify the LaunchTemplateOverride.N.InstanceType parameter. The values of N in the two parameters must be the same. This parameter specifies the weight of the instance type. The weight specifies the capacity of an instance of the specified instance type in the scaling group. A higher weight specifies that a smaller number of instances of the specified instance type are required to meet the expected capacity. Performance metrics such as the number of vCPUs and the memory size of each instance type may vary. You can specify different weights for different instance types based on your business requirements. Example: • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 To meet the expected capacity, Auto Scaling must create two ecs.c5.xlarge instances. **Note** The capacity of the scaling group cannot exceed the sum of the maximum capacity that is specified by the MaxSize parameter and the maximum weight of the instance type. Valid values of WeightedCapacity in LaunchTemplateOverride.N.WeightedCapacity: 1 to 500.

Parameter	Type	Required	Example	Description
AlbServerGroup.N.AlbServerGroupId	String	No	sgp-ddwb0y0g6y9bjm****	The ID of Application Load Balancer (ALB) server group N. N indicates the serial number of the ALB server group. You can associate only a limited number of ALB server groups with a scaling group. To view the available quota or manually request a quota increase, go to the Quota Center.
AlbServerGroup.N.Port	Integer	No	22	The port number used by the ECS instance after the instance is added to ALB server group N. Valid values: 1 to 65535. N indicates the serial number of the ALB server group. **Note** If the N values are the same but the port numbers are different, the system associates multiple ports in the scaling group with ALB server group N.
AlbServerGroup.N.Weight	Integer	No	100	The weight of the ECS instance as a backend server after the instance is added to ALB server group N. If you increase the weight of an ECS instance in a server group, the number of access requests that are forwarded to the ECS instance increases. If the weight is set to 0, no access requests are forwarded to the ECS instance. Valid values: 0 to 100. N indicates the serial number of the ALB server group.
GroupType	String	No	ECS	The type of instances that are managed by the scaling group. Valid values: • ECS: ECS instances • ECI: elastic container instances Default value: ECS.

Parameter	Type	Required	Example	Description
ContainerGroupId	String	No	eci-bp13fm93jggsvjnuad3z	The ID of the container group.

Response parameters

Parameter	Type	Example	Description
ScalingGroupId	String	asg-bp14wlu85wrpchm0****	The ID of the scaling group. The ID is a globally unique identifier (GUID) that is generated by the system.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=CreateScalingGroup
&RegionId=cn-qingdao-env17-d01
&MaxSize=20
&MinSize=2
&LoadBalancerId=147b46d767c-cn-qingdao-env17-d01-cm5****
&DBInstanceId.1=rm-rdszzzyyuny****
&DBInstanceId.2=rm-rdsia3u3yia****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<CreateScalingGroupResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
      <ScalingGroupId>asg-bp14wlu85wrpchm0****</ScalingGroupId>
</CreateScalingGroupResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "ScalingGroupId": "asg-bp14wlu85wrpchm0****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance is not in the Running state.
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned because the health check feature is not enabled for the specified SLB instance.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified SLB instance is not in the Active state.
400	IncorrectVSwitchStatus	The current status of virtual switch does not support this operation.	The error message returned because the vSwitch is unavailable. Therefore, ECS instances cannot be created.
400	InvalidDBInstanceId.RegionMismatch	DB instance "XXX" and the specified scaling group are not in the same Region.	The error message returned because the specified ApsaraDB RDS instance and the scaling group do not reside in the same region.
400	InvalidLoadBalancerId.IncorrectAddressType	The current address type of specified load balancer does not support this action.	The error message returned because the network type of the specified SLB instance does not match the network type of the specified vSwitch.

HTTP status code	Error code	Error message	Description
400	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified Load Balancer does not support this action.	The error message returned because the network type of the ECS instance added to the backend server group is different from that of the scaling group.
400	InvalidLoadBalancerId.RegionMismatch	The specified Load Balancer and the specified scaling group are not in the same Region.	The error message returned because the specified SLB instance and the scaling group do not reside in the same region.
400	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in specified Load Balancer are not in the same VPC.	The error message returned because the vSwitch and the ECS instance added to the backend server group of the specified SLB instance do not reside in the same VPC.
400	InvalidParameter	The specified value of parameter "ScalingPolicy" is not valid.	The error message returned because the specified ScalingPolicy parameter is invalid.
400	InvalidParameter.Conflict	The value of parameter <parameter name> and parameter <parameter name> are conflict.	The error message returned because the specified MinSize value is greater than the MaxSize value.
400	InvalidScalingGroupName.Duplicate	The specified value of parameter <parameter name> is duplicated.	The error message returned because the name that you specified for the scaling group is already used by an existing scaling group.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because the maximum number of IP addresses in the whitelist that manages access to the specified ApsaraDB RDS instance has been reached.
400	QuotaExceeded.PrivateIpAddress	Private IP address quota exceeded in the specified virtual switch.	The error message returned because no more private IP addresses are available in the CIDR block of the specified vSwitch.
400	QuotaExceeded.ScalingGroup	Scaling group quota exceeded.	The error message returned because the maximum number of scaling groups has been reached.
400	QuotaExceeded.VPCInstance	Instance quota exceeded in the specified VPC.	The error message returned because the maximum number of instances in the VPC has been reached.
404	InvalidDBInstanceId.NotFound	DB instance "XXX" does not exist.	The error message returned because the specified ApsaraDB RDS instance does not exist.
404	InvalidLoadBalancerId.NotFound	The specified Load Balancer does not exist.	The error message returned because the specified SLB instance does not exist.
404	InvalidRegionId.NotFound	The specified region does not exist.	The error message returned because the specified region does not exist.

HTTP status code	Error code	Error message	Description
404	InvalidVSwitchId.NotFound	The specified virtual switch does not exist.	The error message returned because the specified vSwitch does not exist.
400	LaunchTemplateVersionSet.NotFound	The specific version of launch template is not exist.	The error message returned because the specified version of the launch template does not exist.
400	LaunchTemplateSet.NotFound	The specified launch template set is not found.	The error message returned because the specified launch template does not exist.
400	TemplateMissingParameter.ImageId	The input parameter "ImageId" that is mandatory for processing this request is not supplied.	The error message returned because the ImageId parameter that is required for the specified launch template is not specified.
400	TemplateMissingParameter.InstanceTypes	The input parameter "InstanceTypes" that is mandatory for processing this request is not supplied.	The error message returned because the InstanceTypes parameter that is required for the specified launch template is not specified.
400	TemplateMissingParameter.SecurityGroup	The input parameter "SecurityGroup" that is mandatory for processing this request is not supplied.	The error message returned because the SecurityGroup parameter that is required for the specified launch template is not specified.

HTTP status code	Error code	Error message	Description
400	TemplateVersion.NotNumber	The input parameter "LaunchTemplateVersion" is supposed to be a string representing the version number.	The error message returned because the fixed version number that is specified for the LaunchTemplateVersion parameter contains non-digit characters.

6.1.7.2. ModifyScalingGroup

Modifies a scaling group.

Description

- You cannot modify the following parameters:
  - RegionId
  - LoadBalancerId

    > **Note** If you want to associate a Server Load Balancer (SLB) instance with a scaling group, call the AttachLoadBalancers operation. If you want to disassociate an SLB instance from a scaling group, call the DetachLoadBalancers operation.

  - DBInstanceId

    > **Note** If you want to associate an ApsaraDB RDS instance with a scaling group, call the AttachDBInstances operation. If you want to disassociate an ApsaraDB RDS instance from a scaling group, call the DetachDBInstances operation.

- You can call this operation to modify a scaling group only when the scaling group is in the Active or Inactive state.
- If you enable a new scaling configuration, Elastic Compute Service (ECS) instances that are created based on the previous scaling configuration still run as expected in the scaling group.
- If the total number of instances in the scaling group is greater than the allowed maximum number after you change the value of the MaxSize parameter, Auto Scaling automatically removes instances from the scaling group to ensure that the number of instances is within the new range.
- If the total number of instances in the scaling group is less than the allowed minimum number after you change the value of the MinSize parameter, Auto Scaling automatically adds instances to the scaling group to ensure that the number of instances is within the new range.

Request parameters

Parameter	Type	Required	Example	Description

Parameter	Type	Required	Example	Description
Action	String	Yes	ModifyScalingGroup	The operation that you want to perform. Set the value to ModifyScalingGroup.
ScalingGroupId	String	Yes	asg-bp1ffogfdauy0jw0****	The ID of the scaling group that you want to modify.
ScalingGroupName	String	No	Scalinggroup****	The name of the scaling group. The name must be unique in a region. 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.
MinSize	Integer	No	1	The minimum number of ECS instances in the scaling group. Valid values: 0 to 1000. If the number of ECS instances in the scaling group is less than the value of the MinSize parameter, Auto Scaling automatically creates ECS instances until the number of instances is equal to the value of the MinSize parameter.
MaxSize	Integer	No	99	The maximum number of ECS instances in the scaling group. Valid values: 0 to 1000. If the number of ECS instances in the scaling group is greater than the value of the MaxSize parameter, Auto Scaling automatically removes ECS instances from the scaling group until the number of instances is equal to the value of the MaxSize parameter.

Parameter	Type	Required	Example	Description
VSwitchIds.N	RepeatList	No	vsw-bp1oo2a7isyrb8igf****	The ID of vSwitch N. Valid values of N: 1 to 5. • This parameter takes effect only when the network type of the scaling group is virtual private cloud (VPC). The specified vSwitches and the scaling group must reside in the same VPC. • The vSwitches can reside in different zones. The vSwitches are sorted in ascending order based on the value of N. A value of 1 specifies the highest priority. As N increases, the priority decreases. When an ECS instance cannot be created in the zone where the vSwitch that has the highest priority resides, Auto Scaling creates the ECS instance in the zone where the vSwitch that has the next highest priority resides.
DefaultCooldown	Integer	No	600	The cooldown time after a scaling activity is executed. Valid values: 0 to 86400. Unit: seconds. During the cooldown time, Auto Scaling executes only scaling activities that are triggered by event-triggered tasks monitored by CloudMonitor.
RemovalPolicy.1	String	No	OldestScalingConfiguration	The first step of the instance removal policy. • OldestInstance: removes ECS instances that are added to the scaling group at the earliest point in time. • NewestInstance: removes ECS instances that are most recently added to the scaling group. • OldestScalingConfiguration: removes ECS instances that are created based on the earliest scaling configuration.

Parameter	Type	Required	Example	Description
RemovalPolicy.2	String	No	NewestInstance	The second step of the instance removal policy. • OldestInstance: removes ECS instances that are added to the scaling group at the earliest point in time. • NewestInstance: removes ECS instances that are most recently added to the scaling group. • OldestScalingConfiguration: removes ECS instances that are created based on the earliest scaling configuration.
ActiveScalingConfigurationId	String	No	asc-bp17pelvl720x5ub****	The ID of the active scaling configuration in the scaling group.
HealthCheckType	String	No	ECS	The health check mode of the scaling group. Valid values: • NONE: does not perform health checks. • ECS: performs health checks on ECS instances in the scaling group.
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the launch template that is used by Auto Scaling to create ECS instances.
LaunchTemplateVersion	String	No	Default	The version of the instance launch template. Valid values: • A fixed template version number. • Default: the default template version. • Latest: the latest template version.
OnDemandBaseCapacity	Integer	No	30	The minimum number of pay-as-you-go instances that are required in the scaling group. Valid values: 0 to 1000. If the number of pay-as-you-go instances is less than the value of this parameter, Auto Scaling creates pay-as-you-go instances.

Parameter	Type	Required	Example	Description
OnDemandPercentageAboveBaseCapacity	Integer	No	20	The percentage of pay-as-you-go instances that can be created when instances are added to the scaling group. This parameter takes effect when the number of instances reaches the value of the OnDemandBaseCapacity parameter. Valid values: 0 to 100.
SpotInstanceRemedy	Boolean	No	true	Specifies whether to supplement preemptible instances. If you set this parameter to true, Auto Scaling creates an instance to replace a preemptible instance when Auto Scaling receives the system message that the preemptible instance is to be reclaimed.
CompensateWithOnDemand	Boolean	No	true	Specifies whether to automatically create pay-as-you-go instances to meet the requirement for the number of ECS instances in the scaling group when the number of preemptible instances cannot be reached due to reasons such as costs or insufficient resources. This parameter takes effect if you set the MultiAZPolicy parameter to COST_OPTIMIZED when you call the CreateScalingGroup operation. Valid values: • true • false
SpotInstancePools	Integer	No	5	The number of instance types that are supported. Auto Scaling creates preemptible instances of multiple instance types that are provided at the lowest cost. Valid values: 0 to 10.

Parameter	Type	Required	Example	Description
DesiredCapacity	Integer	No	5	The expected number of ECS instances in the scaling group. Auto Scaling automatically maintains the expected number of ECS instances that you specify. The expected number cannot be greater than the value of the MaxSize parameter and cannot be less than the value of the MinSize parameter.
GroupDeletionProtection	Boolean	No	true	Specifies whether to enable deletion protection for the scaling group. Valid values: • true: enables deletion protection for the scaling group. The scaling group cannot be deleted. • false: disables deletion protection for the scaling group.
LaunchTemplateOverride.N.InstanceType	String	No	ecs.c5.xlarge	If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.InstanceType and LaunchTemplateOverride.N.WeightedCapacity parameters at the same time. This parameter specifies instance type N. Instance type N overrides the instance type that is specified in the launch template. You can specify N instance types in the extended configurations of the launch template. Valid values of N: 1 to 10. **Note** This parameter takes effect only when you specify the LaunchTemplateId parameter. For information about the valid values of InstanceType in the LaunchTemplateOverride.N.InstanceType parameter, see the *Instance families* topic in ECS documentation.

Parameter	Type	Required	Example	Description
LaunchTemplateOverride.N.WeightedCapacity	Integer	No	4	If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.WeightedCapacity parameter after you specify the LaunchTemplateOverride.N.InstanceType parameter. The values of N in the two parameters must be the same. This parameter specifies the weight of the instance type. The weight specifies the capacity of an instance of the specified instance type in the scaling group. A higher weight specifies that a smaller number of instances of the specified instance type are required to meet the expected capacity. Performance metrics, such as the number of vCPUs and the memory size of each instance type, may vary. You can specify different weights for different instance types based on your business requirements. Example: • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 To meet the expected capacity, Auto Scaling creates two ecs.c5.xlarge instances. **Note** Note: The capacity of the scaling group cannot exceed the sum of the maximum capacity that is specified by the MaxSize parameter and the maximum weight of the instance type. Valid values of WeightedCapacity in LaunchTemplateOverride.N.WeightedCapacity: 1 to 500.

Parameter	Type	Required	Example	Description
MultiAZPolicy	String	No	PRIORITY	The scaling policy of the multi-zone scaling group that contains ECS instances. Valid values: • PRIORITY: The first vSwitch that is specified by using the VSwitchIds.N parameter has the highest priority. Auto Scaling scales instances in the zone where the vSwitch that has the highest priority resides. If Auto Scaling fails to scale the instances, Auto Scaling scales instances in the zone where the vSwitch that has the next highest priority resides. • COST_OPTIMIZED: During a scale-out activity, Auto Scaling creates ECS instances that have vCPUs of the lowest price. During a scale-in activity, Auto Scaling removes ECS instances that have vCPUs of the highest price. Preemptible instances are preferentially created when preemptible instance types are specified in the active scaling configuration. You can use the CompensateWithOnDemand parameter to specify whether to automatically create pay-as-you-go instances when Auto Scaling fails to create preemptible instances. Note Note: The COST_OPTIMIZED parameter takes effect only when multiple instance types are specified or at least one instance type is specified for preemptible instances. • BALANCE: ECS instances are evenly distributed across zones that are specified in the scaling group. If ECS instances are unevenly distributed among multiple zones due to insufficient resources, you can call the RebalanceInstance operation to evenly distribute the instances among multiple zones.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=ModifyScalingGroup
&ScalingGroupId=asg-bp1ffogfdauy0jw0****
&ScalingGroupName=Scaling****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<ModifyScalingGroupResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</ModifyScalingGroupResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Apsara Stack tenant account.

HTTP status code	Error code	Error message	Description
400	InvalidScalingGroupName.Duplicate	The specified value of parameter <parameter name> is duplicated.	The error message returned because the name that you specified for the scaling group already exists.
404	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	The error message returned because the specified scaling configuration does not exist in the scaling group.
400	InvalidScalingConfigurationId.InstanceTypeMismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	The error message returned because the instance type in the specified scaling configuration is different from the instance type in the active scaling configuration.
400	InvalidParameter.Conflict	The value of parameter <parameter name> and parameter <parameter name> are confilict.	The error message returned because the value of the MinSize parameter is greater than the value of the MaxSize parameter.
400	LaunchTemplateVersionSet.NotFound	The specific version of launch template is not exist.	The error message returned because the specified version of the launch template does not exist.

HTTP status code	Error code	Error message	Description
400	LaunchTemplateSet.NotFound	The specified launch template set is not found.	The error message returned because the specified launch template does not exist.
400	TemplateMissingParameter.ImageId	The input parameter "ImageId" that is mandatory for processing this request is not supplied.	The error message returned because the ImageId parameter that is required for the specified launch template is not specified.
400	TemplateMissingParameter.InstanceTypes	The input parameter "InstanceTypes" that is mandatory for processing this request is not supplied.	The error message returned because the InstanceTypes parameter that is required for the specified launch template is not specified.
400	TemplateMissingParameter.SecurityGroup	The input parameter "SecurityGroup" that is mandatory for processing this request is not supplied.	The error message returned because the SecurityGroup parameter that is required for the specified launch template is not specified.

HTTP status code	Error code	Error message	Description
400	TemplateVersion.NotNumber	The input parameter "LaunchTemplateVersion" is supposed to be a string representing the version number.	The error message returned because the fixed version number that is specified for the LaunchTemplateVersion parameter of the launch template contains non-digit characters.

6.1.7.3. EnableScalingGroup

Enables a scaling group.

Description

- You can call this operation to enable a scaling group that is in the Inactive state and has an instance configuration source. The instance configuration source can be a scaling configuration, a launch template, or the Elastic Compute Service (ECS) instances that you specified when you created the scaling group. If a scaling group is not in the Inactive state or does not have an instance configuration source, you cannot call this operation to enable the scaling group.

> **Note** A scaling group can have only one active instance configuration source at a time. When you call this operation to enable a scaling group, you can specify a scaling configuration or a launch template for the scaling group. If an instance configuration source is configured for the scaling group before you call this operation, the scaling configuration or launch template that you specify in the request overwrites the original scaling configuration or launch template.

- If you specify a value for the InstanceId.N parameter when you call the operation, Auto Scaling checks whether the total number of ECS instances is within the range allowed for the scaling group after you call the operation.
  - If the total number of ECS instances is less than the minimum number of instances allowed in the scaling group, Auto Scaling automatically creates the required number of pay-as-you-go ECS instances to maintain the minimum number. For example, if the minimum number of instances allowed in your scaling group is five, and you specify the InstanceId.N parameter to add two created ECS instances to the scaling group, Auto Scaling automatically creates three instances in the scaling group after the two instances are added.
  - If the total number of instances is greater than the maximum number of instances allowed in the scaling group, the operation fails.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	EnableScalingGroup	The operation that you want to perform. Set the value to EnableScalingGroup.
ScalingGroupId	String	Yes	asg-bp14wlu85wrpchm0****	The ID of the scaling group.
ActiveScalingConfigurationId	String	No	asc-bp1ffogfdauy0nu5****	The ID of the scaling configuration that you want to put into the Active state.
InstanceId.N	RepeatList	No	i-283vv****	The ID of ECS instance N that you want to add to the scaling group after the scaling group is enabled. Valid values of N: 1 to 20. Before you add ECS instances to a scaling group, make sure that the instances meet the following requirements: • The instances and the scaling group must reside in the same region. • The instances are in the Running state. • The instances cannot belong to another scaling group. • The instances are billed on a subscription or pay-as-you-go basis, or the instances are preemptible instances. • If you specify the VswitchID parameter for the scaling group, the instances must reside in the same virtual private cloud (VPC) as the specified vSwitch. You cannot add the instances that reside in the classic network or other VPCs to the scaling group. • If you do not specify the VswitchID parameter for the scaling group, you cannot add VPC-type ECS instances to the scaling group.

Parameter	Type	Required	Example	Description
LoadBalancerWeight.N	RepeatList	No	50	The weight of ECS instance N that acts as a backend server of the associated Server Load Balancer (SLB) instance. Valid values of N: 1 to 20. Valid values of this parameter: 1 to 100. Default value: 50.
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the launch template that is used by Auto Scaling to create ECS instances.
LaunchTemplateVersion	String	No	Default	The version of the launch template. Valid values: • A fixed template version number. • Default: the default template version. • Latest: the latest template version.
LaunchTemplateOverride.N.InstanceType	String	No	ecs.c5.xlarge	If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.InstanceType and LaunchTemplateOverride.N.WeightedCapacity parameters. This parameter specifies instance type N, which overwrites the instance type that is specified in the launch template. You can specify N instance types in the extended configurations of the launch template. Valid values of N: 1 to 10. **Note** This parameter takes effect only if you specify the LaunchTemplateId parameter. For information about the valid values of InstanceType in InstanceTypeOverride.N.InstanceType, see the *Instance families* topic in ECS documentation.

Parameter	Type	Required	Example	Description
LaunchTemplateOverride.N.WeightedCapacity	Integer	No	4	If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.WeightedCapacity parameter after you specify the LaunchTemplateOverride.N.InstanceType parameter. The values of N in the two parameters must be the same. This parameter specifies the weight of the instance type. The weight specifies the capacity of an instance of the specified instance type in the scaling group. A higher weight specifies that a smaller number of instances of the specified instance type are required to meet the expected capacity. Performance metrics such as the number of vCPUs and the memory size of each instance type may vary. You can specify different weights for different instance types based on your business requirements. Example: • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 To meet the expected capacity, Auto Scaling must create two ecs.c5.xlarge instances. **Note** The capacity of the scaling group cannot exceed the sum of the maximum capacity that is specified by the MaxSize parameter and the maximum weight of the instance type. Valid values of WeightedCapacity in LaunchTemplateOverride.N.WeightedCapacity: 1 to 500.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=EnableScalingGroup
&ScalingGroupId=asg-bp1ffogfdauy0jw0****
&InstanceId.1=i-283vv****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<EnableScalingGroupResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</EnableScalingGroupResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within your Apsara Stack tenant account.

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the specified scaling group is being deleted.
404	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	The error message returned because the specified scaling configuration does not exist in the specified scaling group.
400	InvalidScalingConfigurationId.InstanceTypeMismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	The error message returned because the instance type of the specified scaling configuration is different from the instance type of the active scaling configuration.
400	MissingActiveScalingConfiguration	An active scaling configuration for the specified scaling group is not supplied.	The error message returned because no active scaling configuration is specified for the scaling group.
404	InvalidInstanceId.NotFound	Instance “XXX” does not exist.	The error message returned because the specified ECS instance does not exist within your Apsara Stack tenant account.

HTTP status code	Error code	Error message	Description
400	InvalidInstanceId.RegionMismatch	Instance “XXX” and the specified scaling group are not in the same Region.	The error message returned because the specified ECS instance and the scaling group do not reside in the same region.
400	InvalidInstanceId.InstanceTypeMismatch	Instance "XXX" and existing active scaling configuration have different instance type.	The error message returned because the instance type of the specified ECS instance is different from the instance type of the active scaling configuration.
400	IncorrectInstanceStatus	The current status of instance "XXX" does not support this action.	The error message returned because the specified ECS instance is not in the Running state.
400	InvalidInstanceId.NetworkTypeMismatch	The network type of instance “XXX” does not support this action.	The error message returned because the network type of the specified ECS instance is different from the network type of the scaling group.
400	InvalidInstanceId.VPCMismatch	Instance "XXX" and the specified scaling group are not in the same VPC.	The error message returned because the added ECS instance and the specified scaling group do not reside in the same VPC.
400	InvalidInstanceId.InUse	Instance "XXX" is already attached to another scaling group.	The error message returned because the specified ECS instance has been added to another scaling group.

HTTP status code	Error code	Error message	Description
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified SLB instance is not in the Active state.
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned because health check is not enabled for the specified SLB instance.
400	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified Load Balancer does not support this action.	The error message returned because the network type of an ECS instance that is attached to the specified SLB instance is different from the network type of the scaling group.
400	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in specified Load Balancer are not in the same VPC.	The error message returned because an ECS instance that is attached to the specified SLB instance does not reside in the same VPC as the vSwitch that is specified by the VswitchID parameter.
400	IncorrectDBInstanceStatus	The current status of DB instance “XXX” does not support this action.	The error message returned because the specified ApsaraDB RDS instance is not in the Running state.

HTTP status code	Error code	Error message	Description
400	IncorrectCapacity.MaxSize	To attach the instances, the total capacity will be greater than the max size.	The error message returned because the total number of ECS instances exceeds the specified value of the MaxSize parameter after ECS instances are added to the specified scaling group.
400	LaunchTemplateVersionSet.NotFound	The specific version of launch template is not exist.	The error message returned because the specified version of the launch template does not exist.
400	LaunchTemplateSet.NotFound	The specified launch template set is not found.	The error message returned because the specified launch template does not exist.
400	TemplateMissingParameter.ImageId	The input parameter "ImageId" that is mandatory for processing this request is not supplied.	The error message returned because the ImageId parameter that is required for the specified launch template is not specified.
400	TemplateMissingParameter.InstanceTypes	The input parameter "InstanceTypes" that is mandatory for processing this request is not supplied.	The error message returned because the InstanceTypes parameter that is required for the specified launch template is not specified.

HTTP status code	Error code	Error message	Description
400	TemplateMissingParameter.SecurityGroup	The input parameter "SecurityGroup" that is mandatory for processing this request is not supplied.	The error message returned because the SecurityGroup parameter that is required for the specified launch template is not specified.
400	TemplateVersion.NotNumber	The input parameter "LaunchTemplateVersion" is supposed to be a string representing the version number.	The error message returned because the fixed version number that is specified for the LaunchTemplateVersion parameter contains non-digit characters.

6.1.7.4. DisableScalingGroup

Disables a scaling group.

Description

- The scaling activity that is being executed before this operation is called continues until the scaling activity is complete. Scaling activity requests that are issued after this operation is called are rejected.
- You can call this operation only if the scaling group is in the Active state.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DisableScalingGroup	The operation that you want to perform. Set the value to DisableScalingGroup.
ScalingGroupId	String	Yes	dmIDKNcyWfzncX9MWX1****	The ID of the scaling group.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=DisableScalingGroup
&ScalingGroupId=dmIDKNcyWfzncX9MWX1****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DisableScalingGroupResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DisableScalingGroupResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within your Apsara Stack tenant account.

6.1.7.5. DeleteScalingGroup

Deletes a scaling group.

Description

- When a scaling group is deleted, the scaling configurations, rules, activities, and scaling requests that are associated with the scaling group are also deleted.
- The following tasks and instances are not deleted after a scaling group is deleted: scheduled tasks, event-triggered tasks, Server Load Balancer (SLB) instances, and ApsaraDB RDS instances.

Request parameters

Parameter	Type	Required	Example	Description

Parameter	Type	Required	Example	Description
Action	String	Yes	DeleteScalingGroup	The operation that you want to perform. Set the value to DeleteScalingGroup.
ScalingGroupId	String	Yes	dmIDKNcyWfzncX9MWX1****	The ID of the scaling group.

ForceDelete	Boolean	No	false	Specifies whether to forcibly delete the scaling group, and remove and release Elastic Compute Service (ECS) instances if the scaling group has ECS instances or ongoing scaling activities. Valid values:
				• true: The scaling group is disabled and new scaling requests are rejected. After all existing scaling requests are processed, the ECS instances in the scaling group are removed. Then, the scaling group is deleted. If ECS instances are manually added to the scaling group, the ECS instances are removed. If ECS instances are automatically created in the scaling group, the ECS instances are removed and released.
				• false: You can set this parameter to false if the following conditions are met:
				  ○ No ongoing scaling activities exist in the scaling group.
				  ○ The TotalCapacity parameter is set to 0. A value of 0 specifies that no ECS instances exist in the scaling group.
				Default value: false.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=DeleteScalingGroup
&ScalingGroupId=dmIDKNcyWfzncX9MWX1****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DeleteScalingGroupResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DeleteScalingGroupResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within your Apsara Stack tenant account.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
400	InstanceInUse	You cannot delete a scaling configuration or scaling group while there is an instance associated with it.	The error message returned because the specified scaling group contains ECS instances.

6.1.7.6. DescribeScalingGroups

Queries scaling groups.

Description

A scaling group can have the following states during its lifecycle:

- Active: The scaling group is active. Scaling groups that are in the Active state can receive requests to execute scaling rules and trigger scaling activities.
- Inactive: The scaling group is inactive. Scaling groups that are in the Inactive state cannot receive any requests to execute scaling rules.
- Deleting: The scaling group is being deleted. Scaling groups that are in the Deleting state cannot receive any requests to execute scaling rules.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DescribeScalingGroups	The operation that you want to perform. Set the value to DescribeScalingGroups.
RegionId	String	Yes	cn-qingdao-env17-d01	The ID of the region to which the scaling group belongs.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	10	The number of entries to return on each page. Maximum value: 50. Default value: 10
ScalingGroupName	String	No	dyrSuvBOtO1dEdIlIbp****	The name of the scaling group that you want to query. You can query 1 to 20 scaling groups in each call. The names of inactive scaling groups are not displayed in the query results and no error is reported.
GroupType	String	No	ECS	The type of instances that are managed by the scaling group. Valid values: • ECS: ECS instances • ECI: elastic container instances Default value: ECS.

Parameter	Type	Required	Example	Description
ScalingGroupId.N	RepeatList	No	asg-bp14wlu85wrpchm0****	The ID of scaling group N that you want to query. Valid values of N: 1 to 20. The IDs of inactive scaling groups are not displayed in the query results and no error is reported.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	10	The number of entries returned per page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingGroups	Array of ScalingGroup		Details of the scaling groups.
ScalingGroup			
ActiveCapacity	Integer	1	The number of ECS instances that are added to the scaling group and can provide services.
ActiveScalingConfigurationId	String	asc-bp1et2qekq3ojr33****	The ID of the scaling configuration that is in the Active state in the scaling group.
AlbServerGroups	Array of AlbServerGroup		Details of Application Load Balancer (ALB) server groups.
AlbServerGroup			
AlbServerGroupId	String	sgp-ddwb0y0g6y9bjm*****	The ID of the ALB server group.
Port	Integer	80	The port number used by the ECS instance after the instance is added to the ALB server group.

Parameter	Type	Example	Description
Weight	Integer	100	The weight of the ECS instance as a backend server after the instance is added to the ALB server group.
CompensateWithOnDemand	Boolean	true	Indicates whether pay-as-you-go instances are automatically created to meet the requirements on the number of ECS instances when the expected capacity of preemptible instances cannot be fulfilled due to reasons such as costs or insufficient resources. This parameter takes effect only when the MultiAZPolicy parameter is set to COST_OPTIMIZED. Valid values: • true: Pay-as-you-go instances can be automatically created. • false: Pay-as-you-go instances cannot be automatically created.
CreationTime	String	2014-08-14T10:58Z	The time at which the scaling group was created.
DBInstanceIds	List	rm-bp15556qzebg1****	The IDs of the ApsaraDB RDS instances that are associated with the scaling group.
DefaultCooldown	Integer	60	The default cooldown time of the scaling group. During the cooldown time, Auto Scaling executes only the scaling activities that are triggered by event-triggered tasks monitored by CloudMonitor in the scaling group.
DesiredCapacity	Integer	5	The expected number of ECS instances in the scaling group. Auto Scaling automatically maintains the expected number of ECS instances that you specify.
GroupDeletionProtection	Boolean	true	Indicates whether deletion protection is enabled for the scaling group. Valid values: • true: Deletion protection is enabled for the scaling group. The scaling group cannot be deleted. • false: Deletion protection is disabled for the scaling group.

Parameter	Type	Example	Description
GroupType	String	ECS	The type of instances that are managed by the scaling group.
HealthCheckType	String	ECS	The health check mode of the scaling group. Valid values: • NONE: No health checks are performed. • ECS: Auto Scaling performs health checks on ECS instances in the scaling group.
LaunchTemplateId	String	lt-m5e3ofjr1zn1aw7****	The ID of the launch template used by the scaling group.
LaunchTemplateOverrides	Array of LaunchTemplateOverride		Details of the instance type specified in the extended configurations.
LaunchTemplateOverride			
InstanceType	String	ecs.c5.xlarge	The instance type. The specified instance type overwrites the instance type specified in the launch template.
WeightedCapacity	Integer	4	The weight of the instance type. The value of this parameter indicates the capacity of an instance of this instance type in the scaling group. A greater weight indicates that a smaller number of instances of the specified instance type are required to meet the expected capacity.
LaunchTemplateVersion	String	Default	The version of the launch template used by the scaling group.
LifecycleState	String	Active	The status of the scaling group. Valid values: • Active • Inactive • Deleting
LoadBalancerIds	List	lb-bp19byhscefk3x0li****	The IDs of the Server Load Balancer (SLB) instances that are associated with the scaling group.

Parameter	Type	Example	Description
MaxSize	Integer	2	The maximum number of ECS instances in the scaling group.
MinSize	Integer	1	The minimum number of ECS instances in the scaling group.
ModificationTime	String	2014-08-14T10:58Z	The time at which the scaling group was modified.

Parameter	Type	Example	Description
MultiAZPolicy	String	PRIORITY	The scaling policy for the multi-zone scaling group that contains ECS instances. Valid values: • PRIORITY: ECS instances are created based on the VSwitchIds.N parameter. If an ECS instance cannot be created in the zone where the vSwitch of the highest priority resides, Auto Scaling automatically creates the ECS instance in the zone where the vSwitch of the next highest priority resides. • COST_OPTIMIZED: ECS instances are created based on the unit prices of their vCPUs. ECS instances that use the vCPU of the lowest price are preferentially created. Preemptible instances are created first if preemptible instance types are specified in the scaling configuration. Pay-as-you-go instances are created if no preemptible instances are available due to issues such as insufficient resources. Note The COST_OPTIMIZED parameter takes effect only when multiple instance types are specified or preemptible instances are used. • BALANCE: ECS instances are evenly distributed across zones that are specified for the scaling group. If ECS instances are unevenly distributed across multiple zones due to issues such as insufficient resources, you can call the RebalanceInstance operation to evenly distribute the instances across the zones.
OnDemandBaseCapacity	Integer	30	The minimum number of pay-as-you-go instances that are required in the scaling group. Valid values: 0 to 1000. If the number of pay-as-you-go instances is less than the value of this parameter, Auto Scaling preferentially creates pay-as-you-go instances.

Parameter	Type	Example	Description
OnDemandPercentageAboveBaseCapacity	Integer	20	The percentage of pay-as-you-go instances among the extra instances that exceed the number specified by the OnDemandBaseCapacity parameter. Valid values: 0 to 100.
PendingCapacity	Integer	0	The number of ECS instances that are being added to the scaling group and are being configured.
PendingWaitCapacity	Integer	1	The number of ECS instances that are pending to be added to the scaling group.
ProtectedCapacity	Integer	1	The number of ECS instances that are in the Protected state in the scaling group.
RegionId	String	cn-qingdao-env17-d01	The ID of the region to which the scaling group belongs.
RemovalPolicies	List	OldestScalingConfiguration	The policies used to remove ECS instances from the scaling group. Valid values: • OldestInstance: Auto Scaling removes ECS instances that are created at the earliest point in time. • NewestInstance: Auto Scaling removes ECS instances that are created at the latest point in time. • OldestScalingConfiguration: Auto Scaling removes ECS instances that are created based on the earliest scaling configuration.
RemovingCapacity	Integer	0	The number of ECS instances that are being removed from the scaling group.
RemovingWaitCapacity	Integer	1	The number of ECS instances that are pending to be removed from the scaling group.
ScalingGroupId	String	asg-bp14wlu85wrpchm0****	The ID of the scaling group.

Parameter	Type	Example	Description
ScalingGroupName	String	dyrSuvBOtO1dEdIllbp****	The name of the scaling group.
ScalingPolicy	String	recycle	The instance reclaim mode of the scaling group. Valid values: • recycle • release You can call the RemoveInstances operation to remove instances from the scaling group.
SpotInstancePools	Integer	5	The number of instance types that are supported. Auto Scaling creates preemptible instances of multiple instance types that are provided at the lowest cost. Valid values: 0 to 10.
SpotInstanceRemedy	Boolean	true	Indicates whether preemptible instances can be supplemented. If this parameter is set to true, Auto Scaling creates an instance to replace a preemptible instance when Auto Scaling receives a system message which indicates that the preemptible instance is to be reclaimed.
StandbyCapacity	Integer	1	The number of instances that are in the Standby state in the scaling group.
StoppedCapacity	Integer	1	The number of instances that are stopped in Economical Mode in the scaling group.
SuspendedProcesses	List	ScaleIn	The process that is suspended. If no process is suspended, the return value is empty. Valid values: • ScaleIn • ScaleOut • HealthCheck • AlarmNotification • ScheduledAction

Parameter	Type	Example	Description
SystemSuspended	Boolean	true	Indicates whether Auto Scaling stops executing scaling activities in the scaling group. • true: Auto Scaling stops executing scaling activities in the scaling group if the scaling activities failed for more than seven consecutive days. You must modify the scaling group or scaling configuration to resume the execution of scaling activities. • false: Auto Scaling does not stop executing scaling activities in the scaling group.
TotalCapacity	Integer	1	The total weighted capacity of all ECS instances in the scaling group if the Weighted parameter is specified. if the Weighted parameter is not specified, the value of this parameter indicates the total number of ECS instances in the scaling group.
TotalInstanceCount	Integer	1	The total number of ECS instances in the scaling group.
VServerGroups	Array of VServerGroup		The backend server groups.
VServerGroup			
LoadBalancerId	String	147b46d767c-cn-qingdao-cm5****	The ID of the SLB instance to which the backend server group belongs.
VServerGroupAttributes	Array of VServerGroupAttribute		The attributes of the backend server group.
VServerGroupAttribute			
Port	Integer	22	The port number used by the SLB instance to provide external services.
VServerGroupId	String	rsp-bp12bjrny****	The ID of the backend server group.
Weight	Integer	1	The weight of the backend server group.

Parameter	Type	Example	Description
VSwitchId	String	vsw-bp1whw2u46cn8zubm****	The ID of the vSwitch that is associated with the scaling group. If you specify the VSwitchIds parameter, the VSwitchId parameter is ignored.
VSwitchIds	List	vsw-bp1whw2u46cn8zubm****	The IDs of the vSwitches that are associated with the scaling group. If you specify the VSwitchIds parameter, the VSwitchId parameter is ignored.
VpcId	String	vpc-****	The ID of the VPC in which the scaling group resides.
TotalCount	Integer	1	The total number of scaling groups.

Examples

Sample requests

```
http://ess.example.com/?Action=DescribeScalingGroups
&RegionId=cn-qingdao-env17-d01
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DescribeScalingGroupsResponse>
      <RequestId>68386699-8B9E-4D5B-BC4C-75A28F6C2A00</RequestId>
      <TotalCount>1</TotalCount>
      <PageSize>10</PageSize>
      <PageNumber>1</PageNumber>
      <ScalingGroups>
            <ScalingGroup>
                  <ScalingGroupId>asg-bp14wlu85wrpchm0****</ScalingGroupId>
                  <ScalingGroupName>b8pYCVbIV5k9cz4PWpbe****</ScalingGroupName>
                  <RegionId>cn-qingdao</RegionId>
                  <RemovingCapacity>0</RemovingCapacity>
                  <DefaultCooldown>300</DefaultCooldown>
                  <MinSize>1</MinSize>
                  <MaxSize>2</MaxSize>
                  <LifecycleState>Inactive</LifecycleState>
                  <ActiveScalingConfigurationId>asc-bp1et2qekq3ojr33****</ActiveScalingConf
igurationId>
                  <LoadBalancerIds>
                        <LoadBalancerId>lb-bp19byhscefk3x0li****</LoadBalancerId>
                  </LoadBalancerIds>
                  <PendingCapacity>0</PendingCapacity>
                  <TotalCapacity>0</TotalCapacity>
                  <ActiveCapacity>0</ActiveCapacity>
                  <CreationTime>2014-08-14T10:58Z</CreationTime>
                  <DBInstanceIds>
                        <DBInstanceId>rm-bp15556qzebg1****</DBInstanceId>
                        <DBInstanceId>rm-bp52316qzhbg6****</DBInstanceId>
                  </DBInstanceIds>
                  <VSwitchId>vsw-bp1whw2u46cn8zubm****</VSwitchId>
                  <RemovalPolicies>
                        <RemovalPolicy>OldestScalingConfiguration</RemovalPolicy>
                        <RemovalPolicy>OldestInstance</RemovalPolicy>
                  </RemovalPolicies>
            </ScalingGroup>
      </ScalingGroups>
</DescribeScalingGroupsResponse>
```

`JSON` format

```
{
    "RequestId": "68386699-8B9E-4D5B-BC4C-75A28F6C2A00",
    "TotalCount": 1,
    "PageSize": 10,
    "PageNumber": 1,
    "ScalingGroups": {
        "ScalingGroup": [
            {
                "ScalingGroupId": "asg-bp14wlu85wrpchm0****",
                "ScalingGroupName": "b8pYCVbIV5k9cz4PWpbe****",
                "RegionId": "cn-qingdao",
                "RemovingCapacity": 0,
                "DefaultCooldown": 300,
                "MinSize": 1,
                "MaxSize": 2,
                "LifecycleState": "Inactive",
                "ActiveScalingConfigurationId": "asc-bp1et2qekq3ojr33****",
                "LoadBalancerIds": {
                    "LoadBalancerId": [
                        "lb-bp19byhscefk3x0li****"
                    ]
                },
                "PendingCapacity": 0,
                "TotalCapacity": 0,
                "ActiveCapacity": 0,
                "CreationTime": "2014-08-14T10:58Z",
                "DBInstanceIds": {
                    "DBInstanceId": [
                        "rm-bp15556qzebg1****",
                        "rm-bp52316qzhbg6****"
                    ]
                },
                "VSwitchId": "vsw-bp1whw2u46cn8zubm****",
                "RemovalPolicies": {
                    "RemovalPolicy": [
                        "OldestScalingConfiguration",
                        "OldestInstance"
                    ]
                }
            }
        ]
    }
}
```


6.1.7.7. DescribeScalingInstances

Queries the Elastic Compute Service (ECS) instances in a scaling group and the details of the instances.

Description

You can query the ECS instances in a scaling group by specifying the scaling group ID, scaling configuration ID, health status, lifecycle status, and how an ECS instance is created.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DescribeScalingInstances	The operation that you want to perform. Set the value to DescribeScalingInstances.
RegionId	String	Yes	cn-qingdao-env17-d01	The region ID of the scaling group.
ScalingGroupId	String	No	asg-****	The ID of the scaling group.
ScalingConfigurationId	String	No	asc-****	The ID of the scaling configuration that is associated with the scaling group.
HealthStatus	String	No	Healthy	The health status of the ECS instance in the scaling group. If the ECS instance is not in the Running state, the instance is unhealthy. Valid values: • Healthy • Unhealthy Auto Scaling automatically removes unhealthy ECS instances from the scaling group. Auto Scaling stops and releases only the ECS instances that are automatically created.
LifecycleState	String	No	InService	The lifecycle status of the ECS instance in the scaling group. Valid values: • InService: The ECS instance is added to the scaling group and can provide services. • Pending: The ECS instance is being added to the scaling group, but is still being configured. For example, the ECS instance is being created, the ECS instance is being attached to the associated Server Load Balancer (SLB) instance, or the private IP address of the ECS instance is being added to the whitelist of the associated ApsaraDB RDS instance. • Removing: The ECS instance is being removed from the scaling group.

Parameter	Type	Required	Example	Description
CreationType	String	No	AutoCreated	Specifies how to create the ECS instance. Valid values: • AutoCreated: Auto Scaling automatically creates the ECS instance based on the scaling configuration and scaling rules. • Attached: You manually add the ECS instance to the scaling group.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	10	The number of entries to return on each page. Maximum value: 50. Default value: 10.
InstanceId.N	RepeatList	No	i-bp109k5j3dum1ce6****	The ID of ECS instance N. Valid values of N: 1 to 20. The IDs of inactive instances are not displayed in the query result, and no errors are reported.
ScalingActivityId	String	No	asa-bp1c9djwrgxjyk31****	The ID of the scaling activity.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	10	The number of entries returned per page.
RequestId	String	E481B3F5-F290-4F3E-AB40-F980FB8E098A	The ID of the request.
ScalingInstances	Array of ScalingInstance		The information about the ECS instances.
ScalingInstance			

Parameter	Type	Example	Description
CreatedTime	String	2019-03-06T08:06:39Z	The time when the ECS instance was added to the scaling group. The value is accurate to the second.
CreationTime	String	2019-03-06T08:06Z	The time when the ECS instance was created in the scaling group.
CreationType	String	AutoCreated	Indicates how the ECS instance is created. Valid values: • AutoCreated: The ECS instance is automatically created by Auto Scaling based on the scaling configuration and scaling rules. • Attached: The ECS instance is manually added to the scaling group.
Entrusted	Boolean	false	Indicates whether the scaling group can manage the instance lifecycles when you manually add ECS instances to the scaling group. If the scaling group can manage the instance lifecycles, Auto Scaling releases the ECS instances when the ECS instances are automatically removed from the scaling group. This rule does not apply to ECS instances that are manually removed from the scaling group. Valid values: • true • false
HealthStatus	String	Healthy	The health status of the ECS instance in the scaling group. Valid values: • Healthy • Unhealthy
InstanceId	String	i-283vv****	The ID of the ECS instance.
LaunchTemplateId	String	lt-m5e3ofjr1zn1aw7****	The ID of the launch template.

Parameter	Type	Example	Description
LaunchTemplateVersion	String	1	The version number of the launch template.
LifecycleState	String	InService	The lifecycle status of the ECS instance in the scaling group. Valid values: • InService: The ECS instance is added to the scaling group and can provide services. • Pending: The ECS instance is being added to the scaling group, but is still being configured. For example, the ECS instance is being created, the ECS instance is being attached to the associated SLB instance, or the private IP address of the ECS instance is being added to the whitelist of the associated ApsaraDB RDS instance. • Removing: The ECS instance is being removed from the scaling group.
LoadBalancerWeight	Integer	1	The weight of the ECS instance as an SLB backend server.
ScalingActivityId	String	asa-bp1c9djwrgxjyk31****	The ID of the scaling activity during which the ECS instance is added to the scaling group.
ScalingConfigurationId	String	asc-****	The ID of the scaling configuration that is associated with the scaling group.
ScalingGroupId	String	asg-****	The ID of the scaling group to which the ECS instance belongs.
SpotStrategy	String	SpotWithPriceLimit	The bidding policy for the preemptible instance. Valid values: • SpotWithPriceLimit: The instance is created as a preemptible instance that has a user-defined maximum hourly price. • SpotAsPriceGo: The instance is a preemptible instance for which the market price at the time of purchase is used as the bid price.

Parameter	Type	Example	Description
WarmupState	String	NoNeedWarmup	The warm-up status of the ECS instance. Valid values: • NoNeedWarmup: No warm-up is required. • WaitingForInstanceWarmup: The instance is being warmed up. • InstanceWarmupFinish: The warm-up is complete.
WeightedCapacity	Integer	4	The weight of the instance type. The weight indicates the capacity of an instance of the specified instance type in the scaling group. A higher weight indicates that a smaller number of instances of this instance type are required to meet the expected capacity.
ZoneId	String	cn-qingdao-env17-g	The ID of the zone where the ECS instance resides.
TotalCount	Integer	1	The total number of the ECS instances in the scaling group.
TotalSpotCount	Integer	4	The total number of preemptible instances that are in the Running state in the scaling group.

Examples

Sample requests

```
https://ess.example.com/?Action=DescribeScalingInstances
&RegionId=cn-qingdao-env17-d01
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DescribeScalingInstancesResponse>
      <RequestId>13305F2D-A4C2-4E6B-B7C7-0F2150842EA3</RequestId>
      <TotalCount>1</TotalCount>
      <PageNumber>1</PageNumber>
      <PageSize>50</PageSize>
      <ScalingInstances>
            <ScalingInstance>
                  <ScalingConfigurationId>asc-****</ScalingConfigurationId>
                  <CreationType>AutoCreated</CreationType>
                  <InstanceId>i-28sov****</InstanceId>
                  <CreationTime>2014-08-14T10:59Z</CreationTime>
                  <HealthStatus>Healthy</HealthStatus>
                  <LifecycleState>InService</LifecycleState>
                  <Entrusted>false</Entrusted>
                  <ScalingGroupId>asg-****</ScalingGroupId>
            </ScalingInstance>
      </ScalingInstances>
</DescribeScalingInstancesResponse>
```

`JSON` format

```
{
    "RequestId": "13305F2D-A4C2-4E6B-B7C7-0F2150842EA3",
    "TotalCount": 1,
    "PageNumber": 1,
    "PageSize": 50,
    "ScalingInstances": {
        "ScalingInstance": [
            {
                "ScalingConfigurationId": "asc-****",
                "CreationType": "AutoCreated",
                "InstanceId": "i-28sov****",
                "CreationTime": "2014-08-14T10:59Z",
                "HealthStatus": "Healthy",
                "LifecycleState": "InService",
                "Entrusted": "false",
                "ScalingGroupId": "asg-****"
            }
        ]
    }
}
```

6.1.7.8. DescribeScalingActivities

Queries one or more scaling activities.

Description

- You can specify a scaling group ID to query all scaling activities in the scaling group.
- You can filter query results based on the status of scaling activities.
- You can query scaling activities that are executed in the previous 30 days.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DescribeScalingActivities	The operation that you want to perform. Set the value to DescribeScalingActivities.
RegionId	String	Yes	cn-qingdao-env17-d01	The region ID of the scaling group to which the scaling activity that you want to query belongs.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	50	The number of entries to return on each page. Maximum value: 50. Default value: 10.
ScalingGroupId	String	No	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.
StatusCode	String	No	Successful	The status of the scaling activity. Valid values: • Successful: The scaling activity is successful. • Warning: The scaling activity is partially successful. • Failed: The scaling activity failed. • InProgress: The scaling activity is in progress. • Rejected: The scaling activity is rejected.
ScalingActivityId.N	RepeatList	No	asa-bp161xudmuxdzofe****	The ID of scaling activity N that you want to query. Valid values of N: 1 to 20. **Note** You must specify a value for the ScalingGroupId or ScalingActivityId.N parameter.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.
ScalingActivities	Array of ScalingActivity		The information about the scaling activities.
ScalingActivity			
AttachedCapacity	String	1	The total number of instances that were manually added to the scaling group after the scaling activity was complete.
AutoCreatedCapacity	String	1	The total number of instances that were automatically created by Auto Scaling after the scaling activity was complete.
Cause	String	A user requests to execute scaling rule "asr-1kx06p7kys0e6cy2****", changing the Total Capacity from "0" to "1".	The cause that triggered the scaling activity.
Description	String	Add "1" ECS instance	The description of the scaling activity.
EndTime	String	2014-08-17T12:39Z	The time when the scaling activity was complete.
Progress	Integer	100	The execution progress of the scaling activity.
ScalingActivityId	String	ebta5WbUzC8gcwUWvfch****	The ID of the scaling activity.
ScalingGroupId	String	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.
ScalingInstanceNumber	Integer	1	• The number of instances that are created or restarted from the Economical Mode during the scale-out activity. • The number of instances that are deleted or put into the Economical Mode during the scale-in activity.

Parameter	Type	Example	Description
StartTime	String	2014-08-17T12:39Z	The time when the scaling activity started.
StatusCode	String	Successful	The status of the scaling activity. Valid values: • Successful: The scaling activity is successful. • Warning: The scaling activity is partially successful. • Failed: The scaling activity failed. • InProgress: The scaling activity is in progress. • Rejected: The request to execute the scaling activity is rejected.
StatusMessage	String	"1" ECS instance is added.	The status information about the scaling activity.
TotalCapacity	String	2	The total number of instances in the scaling group after the scaling activity is complete.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
TotalCount	Integer	1	The total number of scaling activities.

Examples

Sample requests

```
http://ess.example.com/?Action=DescribeScalingActivities
&RegionId=cn-qingdao-env17-d01
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DescribeScalingActivitiesResponse>
      <RequestId>8FAAE99E-EB43-4838-85AD-93F62454904C</RequestId>
      <TotalCount>1</TotalCount>
      <PageNumber>1</PageNumber>
      <PageSize>10</PageSize>
      <ScalingActivities>
            <ScalingActivity>
                  <Cause>A scheduled task executes scaling rule "sr****", changing the Tota
l Capacity from "0" to "1".</Cause>
                  <Description>Add "1" ECS instance</Description>
                  <EndTime>2014-08-17T12:39Z</EndTime>
                  <Progress>100</Progress>
                  <ScalingActivityId>ebta5WbUzC8gcwUWvfch****</ScalingActivityId>
                  <ScalingGroupId>AG6CQdPU8OKdwLjgZcJ****</ScalingGroupId>
                  <StartTime>2014-08-17T12:39Z</StartTime>
                  <StatusCode>Successful</StatusCode>
                  <StatusMessage>"1" ECS instances is added. </StatusMessage>
            </ScalingActivity>
      </ScalingActivities>
</DescribeScalingActivitiesResponse>
```

`JSON` format

```
{
        "RequestId": "8FAAE99E-EB43-4838-85AD-93F62454904C",
        "TotalCount": 1,
        "PageNumber": 1,
        "PageSize": 10,
        "ScalingActivities": {
            "ScalingActivity": {
                "Cause": "A scheduled task executes scaling rule \"sr****\", changing the T
otal Capacity from \"0\" to \"1\".",
                "Description": "Add \"1\" ECS instance",
                "EndTime": "2014-08-17T12:39Z",
                "Progress": 100,
                "ScalingActivityId": "ebta5WbUzC8gcwUWvfch****",
                "ScalingGroupId": "AG6CQdPU8OKdwLjgZcJ****",
                "StartTime": "2014-08-17T12:39Z",
                "StatusCode": "Successful",
                "StatusMessage": "\"1\" ECS instances is added."
            }
        }
}
```

6.1.8. Scaling configurations

6.1.8.1. CreateScalingConfiguration

Creates a scaling configuration.

Description

Auto Scaling automatically scales out Elastic Compute Service (ECS) instances based on a scaling configuration. ECS instances can be configured in the following modes:

- InstancePatternInfo.N: intelligent configuration mode. In this mode, you need to only specify the number of vCPUs, memory size, instance family, and price limit. The system creates ECS instances of an instance type that is provided at the lowest price based on your configurations. This mode is available only for scaling groups that reside in virtual private clouds (VPCs). This mode helps reduce the failures of scale-out activities caused by insufficient inventory of instance types.
- InstanceType: In this mode, you must specify an instance type.
- InstanceTypes.N: In this mode, you must specify multiple instance types.
- InstanceTypeOverride.N: In this mode, you must specify multiple instance types and weights for the instance types.
- Cpu and Memory: In this mode, you can specify the Cpu and Memory parameters to configure the range of instance types. If you specify the parameters, Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price. This mode is available only if the Scaling Policy parameter is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.

> **Note** You cannot specify the InstanceType, InstanceTypes.N, InstanceTypeOverride.N, and Cpu and Memory parameters at the same time. If you specify the InstanceTypes.N parameter, the InstanceType parameter is ignored. You can specify only the InstanceType parameter and the InstancePatternInfo.N parameter or the InstanceTypes.N parameter and the InstancePatternInfo.N parameter at the same time. In this case, Auto Scaling preferentially creates instances of the instance type specified by the InstanceType parameter or the InstanceTypes.N parameter. If the specified instance type does not have sufficient inventory, Auto Scaling creates instances of the instance type that is specified by the InstancePatternInfo.N parameter.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	CreateScalingConfiguration	The operation that you want to perform. Set the value to CreateScalingConfiguration.
ScalingGroupId	String	Yes	asg-bp14wlu85wrpchm0****	The ID of the scaling group in which you want to create the scaling configuration.
ImageId	String	No	centos6u5_64_20G_aliaegis****.vhd	The ID of the image that is used by Auto Scaling to create instances.

Parameter	Type	Required	Example	Description
ImageName	String	No	myimage	The name of the image. The name must be unique within a region. If you specify the ImageId parameter, the ImageName parameter is ignored. You cannot use the ImageName parameter to specify Alibaba Cloud Marketplace images.
InstanceType	String	No	ecs.t1.xsmall	The instance type that you want to specify in the scaling configuration. For more information about the instance types, see the *Instance type families* topic in ECS documentation.
Cpu	Integer	No	2	The number of vCPUs. You can specify the number of vCPUs and the memory size to configure the range of instance types. For example, you can set Cpu to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price. **Note** You can specify CPU and memory specifications to create instances only if the Scaling Policy parameter is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.

Parameter	Type	Required	Example	Description
Memory	Integer	No	16	The size of the memory. You can specify the number of vCPUs and the memory size to specify the range of instance types. For example, you can set Cpu to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price. **Note** You can specify CPU and memory specifications to create instances only if the Scaling Policy parameter is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.
DeploymentSetId	String	No	ds-bp1frxuzdg87zh4pz****	The ID of the deployment set to which the ECS instance belongs.
InstanceTypes.N	RepeatList	No	ecs.t1.xsmall.1	Instance type N of ECS instances. • If you specify the InstanceTypes.N parameter, the InstanceType parameter is ignored. • Valid values of N in the InstanceTypes.N parameter: 1 to 10. You can specify up to 10 instance types in a scaling group. A smaller value of N specifies a higher priority and a greater value specifies a lower priority. The instance type whose N is set to 1 has the highest priority. • Auto Scaling creates instances based on the priorities of instance types. If Auto Scaling cannot create instances of the instance type that has the highest priority, Auto Scaling creates instances of the instance type that has the next highest priority.

Parameter	Type	Required	Example	Description
InstanceTypeOverride.N.InstanceType	String	No	ecs.c5.xlarge	If you want to specify the capacity of instance types in the scaling configuration, you must specify the InstanceTypeOverride.N.InstanceType and InstanceTypeOverride.N.WeightedCapacity parameters. The InstanceTypeOverride.N.InstanceType parameter is used to specify instance types. You can specify multiple instance types. The InstanceTypeOverride.N.WeightedCapacity parameter is used to specify weights for instance types. Valid values of N: 1 to 10. **Note** If you specify the InstanceTypeOverride.N.InstanceType parameter, you cannot specify the InstanceTypes.N or InstanceType parameter. For information about the valid values of InstanceType in the InstanceTypeOverride.N.InstanceType parameter, see the *Instance type families* topic in ECS documentation.

Parameter	Type	Required	Example	Description
InstanceTypeOverride.N.WeightedCapacity	Integer	No	4	If you want to specify the capacity of instance types in the scaling configuration, you must specify the InstanceTypeOverride.N.InstanceType and InstanceTypeOverride.N.WeightedCapacity parameters. The values of N in the two parameters must be the same. This parameter specifies the weight of the instance type. The weight specifies the capacity of a single instance of the specified instance type in the scaling group. A higher weight specifies that a smaller number of instances of the specified instance type are required to meet the expected capacity. Performance metrics such as the number of vCPUs and the memory size of each instance type may vary. You can specify different weights for different instance types based on your business requirements. Example: • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 To meet the expected capacity, Auto Scaling creates two ecs.c5.xlarge instances. **Note** The capacity of the scaling group cannot exceed the sum of the maximum capacity that is specified by the MaxSize parameter and the maximum weight of the instance type. Valid values of WeightedCapacity in the InstanceTypeOverride.N.WeightedCapacity parameter: 1 to 500.

Parameter	Type	Required	Example	Description
SecurityGroupId	String	No	sg-280ih****	The ID of the security group with which the ECS instance is associated. The ECS instances that are associated with the same security group can access each other.
IoOptimized	String	No	optimized	Specifies whether the instance is I/O optimized. For instances of retired instance types, the default value is none. For instances of other instance types, the default value is optimized. Valid values: • none: non-I/O optimized instances. • optimized: I/O optimized instances.
InternetChargeType	String	No	PayByTraffic	The billing method for network usage. Valid values: • PayByBandwidth: You are charged for the maximum available bandwidth that is specified by the InternetMaxBandwidthOut parameter. • PayByTraffic: You are charged for the actual traffic that you use. The InternetMaxBandwidthOut parameter specifies only the maximum available bandwidth. Default value for the classic network: PayByBandwidth. Default value for VPCs: PayByTraffic.
InternetMaxBandwidthIn	Integer	No	100	The maximum inbound bandwidth from the Internet. Unit: Mbit/s. Valid values: 1 to 200. Default value: 200. This parameter is not used for billing because inbound traffic to instances is free of charge.

Parameter	Type	Required	Example	Description
InternetMaxBandwidthOut	Integer	No	50	The maximum outbound bandwidth to the Internet. Unit: Mbit/s. • Valid values if you set the InternetChargeType parameter to PayByBandwidth: 0 to 100. If this parameter is not specified, a value of 0 is returned. • Valid values if you set the InternetChargeType parameter to PayByTraffic: 0 to 100. If this parameter is not specified, an error is reported.
SystemDisk.Category	String	No	cloud_ssd	The category of the system disk. Valid values: • cloud: basic disk • cloud_efficiency: ultra disk • cloud_ssd: standard SSD • ephemeral_ssd: local SSD • cloud_essd: enhanced SSD (ESSD) For non-I/O optimized instances of Generation I instance types, the default value is cloud. For instances of other types, the default value is cloud_efficiency.
SystemDisk.Size	Integer	No	100	The size of the system disk. Unit: GiB. Valid values: 20 to 500. Default value: max{40, ImageSize}. The value of the SystemDisk.Size parameter must be greater than or equal to max{20, ImageSize}.
SystemDisk.DiskName	String	No	cloud_ssdSystem	The name of the system disk. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with `http://` or `https://`. The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).

Parameter	Type	Required	Example	Description
SystemDisk.Description	String	No	Test system disk.	The description of the system disk. The description must be 2 to 256 characters in length and cannot start with `http://` or `https://`.
SystemDisk.AutoSnapshotPolicyId	String	No	sp-bp12m37ccmxvbmi5****	The ID of the automatic snapshot policy that is applied to the system disk.
SystemDisk.PerformanceLevel	String	No	PL0	The performance level (PL) of the system disk of the ESSD category. Default value: PL1. Valid values: • PL0: An ESSD can deliver up to 10,000 random read/write IOPS. • PL1: An ESSD can deliver up to 50,000 random read/write IOPS. • PL2: An ESSD can deliver up to 100,000 random read/write IOPS. • PL3: An ESSD can deliver up to 1,000,000 random read/write IOPS. Default value: PL0. **Note** For more information about how to select ESSD PLs, see the *ESSDs* topic in ECS documentation.
ScalingConfigurationName	String	No	test-sc	The name of the scaling configuration. 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. The name of the scaling configuration must be unique within a scaling group in a region. If you do not specify this parameter, the value of the ScalingConfigurationId parameter is used.

Parameter	Type	Required	Example	Description
DataDisk.N.Size	Integer	No	100	The size of data disk N. Unit: GiB. Valid values of N: 1 to 16. • Valid values if you set the Category parameter to cloud: 5 to 2000. • Valid values if you set the Category parameter to cloud_efficiency: 20 to 32768. • Valid values if you set the Category parameter to cloud_ssd: 20 to 32768. • Valid values if you set the Category parameter to cloud_essd: 20 to 32768. • Valid values if you set the DataDisk.N.Category parameter to ephemeral_ssd: 5 to 800. The value of the DataDisk.N.Size parameter must be greater than or equal to the size of the snapshot that is specified by the SnapshotId parameter.
DataDisk.N.SnapshotId	String	No	s-280s7****	The ID of the snapshot that you want to use to create data disk N. Valid values of N: 1 to 16. If you specify this parameter, the DataDisk.N.Size parameter is ignored. The size of the disk is the same as the size of the specified snapshot. If you specify a snapshot created on or before July 15, 2013, the operation fails and the system returns InvalidSnapshot.TooOld.

Parameter	Type	Required	Example	Description
DataDisk.N.Category	String	No	cloud_ssd	The category of data disk N. Valid values of N: 1 to 16. Valid values: • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • ephemeral_ssd: local SSD. • cloud_essd: ESSD. For I/O optimized instances, the default value is cloud_efficiency. For non-I/O optimized instances, the default value is cloud.
DataDisk.N.Device	String	No	/dev/xvdb	The mount target of data disk N. Valid values of N: 1 to 16. If you do not specify this parameter, the system automatically assigns a mount target when you create the ECS instance. The name of the mount target ranges from /dev/xvdb to /dev/xvdz.
DataDisk.N.DeleteWithInstance	Boolean	No	true	Specifies whether to release data disk N if the instance to which data disk N is attached is released. Valid values of N: 1 to 16. Valid values: • true: If the instance is released, the data disk is released. • false: If the instance is released, the data disk is retained. Default value: true. **Note** This parameter is available only for independent disks whose DataDisk.N.Category parameter is set to cloud, cloud_efficiency, cloud_ssd, or cloud_essd. If you specify this parameter for other disks, an error is returned.

Parameter	Type	Required	Example	Description
DataDisk.N.Encrypted	String	No	false	Specifies whether to encrypt data disk N. Valid values of N: 1 to 16. Valid values: • true: encrypts data disk N. • false: does not encrypt data disk N. Default value: false.
DataDisk.N.KMSKeyId	String	No	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the Key Management Service (KMS) key of data disk N. Valid values of N: 1 to 16.
DataDisk.N.DiskName	String	No	cloud_ssdData	The name of data disk N. Valid values of N: 1 to 16. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with `http://` or `https://`. The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).
DataDisk.N.Description	String	No	Test data disk.	The description of data disk N. Valid values of N: 1 to 16. The description must be 2 to 256 characters in length and cannot start with `http://` or `https://`.
DataDisk.N.AutoSnapshotPolicyId	String	No	sp-bp19nq9enxqkomib****	The ID of the automatic snapshot policy that you want to apply to data disk N. Valid values of N: 1 to 16.

Parameter	Type	Required	Example	Description
DataDisk.N.PerformanceLevel	String	No	PL1	The PL of data disk N that is of the ESSD category. If you set the DataDisk.N.Category parameter to cloud_essd, the N value must be the same as the N value in DataDisk.N.Category. Valid values: • PL0: An ESSD can deliver up to 10,000 random read/write IOPS. • PL1: An ESSD can deliver up to 50,000 random read/write IOPS. • PL2: An ESSD can deliver up to 100,000 random read/write IOPS. • PL3: An ESSD can deliver up to 1,000,000 random read/write IOPS. Default value: PL1. **Note** For more information about how to select ESSD PLs, see the *ESSDs* topic in ECS documentation.

Parameter	Type	Required	Example	Description
DataDisk.N.Categories.N	RepeatList	No	cloud	Category N of data disk N. The first N in DataDisk.N.Categories.N represents the serial number of the data disk. Valid values of N: 1 to 16. The second N in DataDisk.N.Categories.N represents the serial number of the data disk category. Valid values of N: 1 to 4. You can specify up to four data disk categories in a scaling configuration. A smaller value of N specifies a higher priority. If Auto Scaling cannot create instances by using the data disk that has the highest priority, Auto Scaling creates instances by using the data disk that has the next highest priority. Valid values: • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • cloud_essd: ESSD. Note You cannot specify the DataDisk.N.Categories.N and DataDisk.N.Category parameters at the same time.
LoadBalancerWeight	Integer	No	50	The weight of the instance as a backend server. Valid values: 1 to 100. Default value: 50.

Parameter	Type	Required	Example	Description
Tags	String	No	{"key1":"value1","key2":"value2", ..."key5":"value5"}	The tags of the ECS instance. Tags must be specified as key-value pairs. A maximum of five tags can be specified. Take note of the following items when you specify tag keys and tag values: • A tag key can be up to 64 characters in length. The key cannot start with acs: or aliyun and cannot contain `http://` or `https://`. You cannot specify an empty string as a tag key. • A tag value can be up to 128 characters in length. The value cannot start with acs: or aliyun and cannot contain `http://` or `https://`. You can specify an empty string as a tag value.
UserData	String	No	echo hello ecs!	The user data of the ECS instance. The data must be encoded in Base64. The maximum size of the data before encoding is 16 KB.
KeyPairName	String	No	KeyPairTest	The name of the key pair that is used to log on to the ECS instance. • Windows instances do not use this parameter. • By default, the username and password authentication method is disabled for Linux instances.
RamRoleName	String	No	RamRoleTest	The name of the RAM role that is associated with the ECS instance. The name is provided and maintained by Resource Access Management (RAM). You can call the ListRoles operation to query the available RAM roles. You can call the CreateRole operation to create RAM roles.

Parameter	Type	Required	Example	Description
SecurityEnhancementStrategy	String	No	Active	Specifies whether to enable security hardening. Valid values: • Active: enables security hardening. This value takes effect only for public images. • Deactive: disables security hardening. This value takes effect only for all images.
InstanceName	String	No	Instance****	The name of the instance that is automatically created based on the scaling configuration.
HostName	String	No	Host****	The hostname of the instance. The hostname cannot start or end with a period (.) or a hyphen (-). The hostname cannot contain consecutive periods (.) or hyphens (-). Naming conventions for different types of instances: • Windows instances: The hostname must be 2 to 15 characters in length and can contain letters, digits, and hyphens (-). The hostname cannot contain periods (.) or contain only digits. • Other instances such as Linux instances: The hostname must be 2 to 64 characters in length. You can use periods (.) to separate a hostname into multiple segments. Each segment can contain letters, digits, and hyphens (-).

Parameter	Type	Required	Example	Description
SpotStrategy	String	No	NoSpot	The preemption policy that you want to apply to pay-as-you-go instances and preemptible instances. Valid values: • NoSpot: The instance is created as a pay-as-you-go instance. • SpotWithPriceLimit: The instance is created as a preemptible instance with a user-defined maximum hourly price. • SpotAsPriceGo: The instance is created as a preemptible instance for which the market price at the time of purchase is automatically used as the bid price. Default value: NoSpot.
PasswordInherit	Boolean	No	false	Specifies whether to use the password that is preconfigured in the image. Before you use this parameter, make sure that a password is preconfigured for the image.
SpotPriceLimit.N.InstanceType	String	No	ecs.t1.xsmall	The instance type of preemptible instance N. Valid values of N: 1 to 10. This parameter takes effect only if you set the SpotStrategy parameter to SpotWithPriceLimit.
SpotPriceLimit.N.PriceLimit	Float	No	0.5	The price limit of preemptible instance N. Valid values of N: 1 to 10. This parameter takes effect only if you set the SpotStrategy parameter to SpotWithPriceLimit.

Parameter	Type	Required	Example	Description
Password	String	No	123-abcABC	The password that is used to log on to the ECS instance. The password must be 8 to 30 characters in length and contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. Special characters: `()` ~!@#$%^&*-_+=|{}[]:;'<>,.?/` The password of a Windows instance cannot start with a forward slash (/). **Note** If you specify the Password parameter, we recommend that you use HTTPS to send requests for security purposes.
ResourceGroupId	String	No	rg-resource****	The ID of the resource group to which the ECS instance belongs.
SecurityGroupIds.N	RepeatList	No	sg-280ih****	The ID of the security group with which the ECS instance is associated. The ECS instances that are associated with the same security group can access each other.
HpcClusterId	String	No	hpc-clusterid	The ID of the Elastic High Performance Computing (E-HPC) cluster to which the ECS instance belongs.
InstanceDescription	String	No	Test instance.	The description of the ECS instance. The description must be 2 to 256 characters in length and cannot start with `http://` or `https://`.

Parameter	Type	Required	Example	Description
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see *How to ensure idempotence* in ECS documentation.
Ipv6AddressCount	Integer	No	1	The number of randomly generated IPv6 addresses that are allocated to the elastic network interface (ENI).
CreditSpecification	String	No	Standard	The performance mode of the burstable instance. Valid values: • Standard: standard mode • Unlimited: unlimited mode For more information, see the *Performance modes* section in the *Overview* topic of ECS documentation.
ImageFamily	String	No	hangzhou-daily-update	The name of the image family. If you specify this parameter, the latest custom images that are available in the specified image family are returned. You can use the images to create instances. If you specify the ImageId parameter, you cannot specify the ImageFamily parameter.
DedicatedHostId	String	No	dh-bp67acfmxazb4p****	The ID of the dedicated host on which the ECS instance runs. You cannot create preemptible instances on dedicated hosts. If you specify the DedicatedHostId parameter, the SpotStrategy and SpotPriceLimit parameters are ignored. You can call the DescribeDedicatedHosts operation to query dedicated host IDs.

Parameter	Type	Required	Example	Description
Affinity	String	No	default	Specifies whether to associate the instance on a dedicated host with the dedicated host. Valid values: • default: does not associate the instance with the dedicated host. If you start an instance that was stopped in economical mode when the available resources of the original dedicated host are insufficient, the instance is automatically deployed to another dedicated host in the automatic deployment resource pool. • host: associates the instance with the dedicated host. If you start an instance that was stopped in economical mode, the instance remains on the original dedicated host. If the original dedicated host has insufficient resources, the instance fails to start. Default value: default.
Tenancy	String	No	default	Specifies whether to create the instance on a dedicated host. Valid values: • default: does not create the instance on a dedicated host. • host: creates the instance on a dedicated host. If you do not specify the DedicatedHostId parameter, the system selects a dedicated host for the instance that you want to create. Default value: default.

Parameter	Type	Required	Example	Description
PrivatePoolOptions.MatchCriteria	String	No	Open	The type of the private pool. A private pool is generated when an elasticity assurance or a capacity reservation takes effect. You can select a private pool when you create an instance. Valid values: • Open: open private pool. The system selects a matching open private pool to create instances. If no matching open private pools are found, the resources in the public pool are used. In this case, you do not need to specify the PrivatePoolOptions.Id parameter. • Target: specified private pool. The system uses the capacity of a specified private pool to create the instance. If the specified private pool is unavailable, the instance cannot be created. If you set this parameter to Target, you must specify the PrivatePoolOptions.Id parameter. • None: no private pool. The capacity of private pools is not used.
PrivatePoolOptions.Id	String	No	eap-bp67acfmxazb4****	The ID of the private pool. The ID of a private pool is the same as the ID of the elasticity assurance or capacity reservation for which the private pool is generated.
SpotDuration	Integer	No	1	The retention period of the preemptible instance. Unit: hours. Valid values: 0, 1, 2, 3, 4, 5, and 6. • The following retention periods are available in invitational preview: 2, 3, 4, 5, and 6 hours. If you want to set this parameter to one of these values, submit a ticket. • If you set this parameter to 0, no retention period is specified for the preemptible instance. Default value: 1.

Parameter	Type	Required	Example	Description
SpotInterruptionBehavior	String	No	Terminate	The interruption mode of the preemptible instance. Default value: Terminate. Set the value to Terminate, which specifies that the instance is released.
InstancePatternInfo.N.Cores	Integer	No	2	The number of vCPUs that you want to specify in an instance type in intelligent configuration mode. This parameter is used to filter available instance types that meet the specified criteria. For more information, see the *Instance type families* topic in ECS documentation. Valid values of N: 1 to 10. Take note of the following items: • The InstancePatternInfo.N parameter is available only for scaling groups that reside in VPCs. • You must specify the InstancePatternInfo.N, InstancePatternInfo.N.Cores, and InstancePatternInfo.N.Memory parameters at the same time. • If you specify an instance type in the InstanceType or InstanceTypes.N parameter, Auto Scaling preferentially uses the instance type that is specified by the InstanceType or InstanceTypes.N parameter for scale-out activities. If the specified instance type does not have sufficient inventory, Auto Scaling creates instances of the instance type that is specified by the InstancePatternInfo.N parameter based on the price in ascending order.
InstancePatternInfo.N.Memory	Float	No	4	The memory size that you want to specify in an instance type in intelligent configuration mode. Unit: GiB. This parameter is used to filter the available instance types that meet the specified criteria. For more information, see the *Instance type families* topic in ECS documentation. Valid values of N: 1 to 10.

Parameter	Type	Required	Example	Description
InstancePatternInfo.N.MaxPrice	Float	No	2	The maximum hourly price for a pay-as-you-go instance or a preemptible instance in intelligent configuration mode. This parameter is used to filter the available instance types that meet the specified criteria. Valid values of N: 1 to 10. **Note** If you set the SpotStrategy parameter to SpotWithPriceLimit, you must specify this parameter. In other cases, this parameter is optional.
InstancePatternInfo.N.InstanceFamilyLevel	String	No	EnterpriseLevel	The instance family level in intelligent configuration mode. This parameter is used to filter available instance types that meet the specified criteria. Valid values of N: 1 to 10. Valid values: • EntryLevel: shared instance type. Instances of this level are the most cost-effective but may not provide stable computing performance in a consistent manner. Instances of this level are suitable for business scenarios in which the CPU utilization is low. For more information, see the *Shared instance families* topic in ECS documentation. • EnterpriseLevel: Instances of this level provide stable performance and dedicated resources, and are suitable for business scenarios that require high stability. For more information, see the *Instance type families* topic in ECS documentation. • CreditEntryLevel: This value is available only for burstable instances. CPU credits are used to ensure computing performance. Instances of this level are suitable for scenarios in which the CPU utilization is low but may fluctuate in specific cases. For more information, see the *Burstable instance types* topic in ECS documentation.

Parameter	Type	Required	Example	Description
SystemDiskCategories.N	RepeatList	No	cloud	The categories of the system disk. Valid values of N: 1 to 4. You can specify up to four system disk categories in a scaling configuration. A smaller value of N specifies a higher priority. If Auto Scaling cannot create instances by using the system disk that has the highest priority, Auto Scaling creates instances by using the system disk that has the next highest priority. Valid values: • cloud: basic disk. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • cloud_essd: ESSD. **Note** You cannot specify the SystemDiskCategories.N and SystemDisk.Category parameters at the same time.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingConfigurationId	String	asc-bp1ffogfdauy0nu5****	The ID of the scaling configuration.

Examples

Sample requests

```
http://ess.example.com/?Action=CreateScalingConfiguration
&ScalingGroupId=asg-bp14wlu85wrpchm0****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<CreateScalingConfigurationResponse>
      <ScalingConfigurationId>asc-bp1ffogfdauy0nu5****</ScalingConfigurationId>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</CreateScalingConfigurationResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "ScalingConfigurationId": "asc-bp1ffogfdauy0nu5****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
400	InstanceType.Mismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	The error message returned because the instance type in the specified scaling configuration is different from the instance type in the current scaling configuration.
404	InvalidDataDiskSnapshotId.NotFound	Snapshot "XXX" does not exist.	The error message returned because the specified snapshot does not exist.
400	InvalidDataDiskSnapshotId.SizeNotSupported	The capacity of snapshot "XXX" exceeds the size limit of the specified disk category.	The error message returned because the size of the specified snapshot exceeds the maximum size allowed for the specified disk.
403	InvalidDevice.InUse	Device "XXX" has been occupied.	The error message returned because the mount target of the data disk is occupied.

HTTP status code	Error code	Error message	Description
400	InvalidImageId.InstanceTypeMismatch	The specified image does not support the specified instance type.	The error message returned because the specified image does not support the specified instance type.
404	InvalidImageId.NotFound	The specified image does not exist.	The error message returned because the specified image does not exist within the Apsara Stack tenant account.
400	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	The error message returned because the specified key pair does not exist.
400	InvalidNetworkType.ForRAMRole	RAMRole can't be used For classic instance.	The error message returned because the network type of the instance is classic network. The classic network does not support the RamRoleName parameter.
400	InvalidParameter	The specified value of parameter KeyPairName is not valid.	The error message returned because the OS of the specified instance is Windows. Windows instances do not support the KeyPairName parameter.
400	InvalidParameter.Conflict	The value of parameter SystemDisk.Category and parameter DataDisk.N.Category are conflict.	The error message returned because the specified system disk category conflicts with the data disk category.

HTTP status code	Error code	Error message	Description
400	InvalidRamRole.NotFound	The specified RamRoleName does not exist.	The error message returned because the specified value of the RamRoleName parameter does not exist.
400	InvalidScalingConfigurationName.Duplicate	The specified value of parameter ScalingConfigurationName is duplicated.	The error message returned because the specified scaling configuration name already exists.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Apsara Stack tenant account.
400	InvalidSecurityGroupId.IncorrectNetworkType	The network type of specified security Group does not support this action.	The error message returned because the network type of the specified security group is different from the network type of the scaling group.
404	InvalidSecurityGroupId.NotFound	The specified security group does not exist.	The error message returned because the specified security group does not exist within the Apsara Stack tenant account.
400	InvalidSecurityGroupId.VPCMismatch	The specified security group and the specified virtual switch are not in the same VPC.	The error message returned because the specified security group and vSwitch are not in the same VPC.

HTTP status code	Error code	Error message	Description
403	InvalidSnapshot.TooOld	This operation is denied because the specified snapshot is created before 2013-07-15.	The error message returned because the snapshot was created on or before July 15, 2013. Therefore, the request is rejected.
403	InvalidSystemDiskCategory.ValueUnauthorized	The system disk category is not authorized.	The error message returned because you are not authorized to create an ephemeral system disk.
400	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded.	The error message returned because the specified user data is not encoded in Base64.
400	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	The error message returned because the user data size exceeds the limit.
403	QuotaExceeded.EphemeralDiskSize	Ephemeral disk size quota exceeded.	The error message returned because the total capacity of mounted ephemeral data disks exceeds 2 TiB (2,048 GiB).
400	QuotaExceeded.ScalingConfiguration	Scaling configuration quota exceeded in the specified scaling group.	The error message returned because the maximum number of scaling configurations has been reached.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the maximum number of ECS instances that can be associated with the specified security group is reached.

6.1.8.2. DescribeScalingConfigurations

Queries one or more scaling configurations.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DescribeScalingConfigurations	The operation that you want to perform. Set the value to DescribeScalingConfigurations.
RegionId	String	Yes	cn-qingdao-env17-d01	The region ID of the scaling group to which the scaling configuration that you want to query belongs.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	50	The number of entries to return on each page. Maximum value: 50. Default value: 10
ScalingGroupId	String	No	asg-bp17pelvl720x3v7****	The ID of the scaling group. You can query all scaling configurations in the scaling group.

Parameter	Type	Required	Example	Description
ScalingConfigurationId.N	RepeatList	No	asc-bp17pelvl720x5ub****	The ID of scaling configuration N that you want to query. Valid values of N: 1 to 10. The IDs of active and inactive scaling configurations are displayed in the query results. You can differentiate between active and inactive scaling configurations based on the value of the LifecycleState parameter.
ScalingConfigurationName.N	RepeatList	No	scalingcon****	The name of scaling configuration N that you want to query. Valid values of N: 1 to 10. The names of inactive scaling configurations are not displayed in the query results and no error is returned.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingConfigurations	Array of ScalingConfiguration		Details of the scaling configurations.
ScalingConfiguration			

Parameter	Type	Example	Description
Affinity	String	default	Indicates whether an instance on a dedicated host is associated with the dedicated host. Valid values: • default: The instance is not associated with the dedicated host. If you restart an instance that was stopped in economical mode when the original dedicated host of the instance has insufficient resources, the instance is deployed to another dedicated host in the automatic deployment resource pool. • host: The instance is associated with the dedicated host. If you restart an instance that was stopped in economical mode, the instance remains on the original dedicated host. If the original dedicated host has insufficient resources, the instance cannot be restarted.
Cpu	Integer	2	The number of vCPUs. You can specify the number of vCPUs and the memory size to determine the range of instance types. For example, you can set the Cpu parameter to 2 and the Memory parameter to 16 to specify the instance types that have 2 vCPUs and 16 GiB of memory. Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price. **Note** You can specify the number of vCPUs and the memory size to determine the range of instance types only if the Scaling Policy parameter is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.
CreationTime	String	2014-08-14T10:58Z	The time when the scaling configuration was created.

Parameter	Type	Example	Description
CreditSpecification	String	Standard	The performance mode of the burstable instance. Valid values: • Standard: the standard mode. For more information, see the "Performance modes" section in the *Overview* topic of ECS documentation. • Unlimited: the unlimited mode. For more information, see the "Performance modes" section in the *Overview* topic of ECS documentation.
DataDisks	Array of DataDisk		Details of the data disks.
DataDisk			
AutoSnapshotPolicyId	String	sp-bp19nq9enxqkomib****	The ID of the automatic snapshot policy that is applied to the data disk.
Categories	List	cloud_essd	The categories of the data disks. The values are sorted based on their priorities. The first value has the highest priority. If Auto Scaling cannot use the disk category that has the highest priority to create instances, Auto Scaling uses the disk category that has the next highest priority to create instances. Valid values: • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • cloud_essd: enhanced SSD (ESSD).
Category	String	cloud	The category of the data disk. Valid values: • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • ephemeral_ssd: local SSD. • cloud_essd: ESSD.

Parameter	Type	Example	Description
DeleteWithInstance	Boolean	true	Indicates whether the data disk is released with the instance. Valid values: • true: The data disk is released when the instance is released. • false: The data disk is retained when the instance is released.
Description	String	FinanceDept	The description of the data disk.
Device	String	/dev/xvdb	The mount target of the data disk.
DiskName	String	cloud_ssdData	The name of the data disk.
Encrypted	String	false	Indicates whether the data disk is encrypted. Valid values: • true • false
KMSKeyId	String	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the Key Management Service (KMS) key that is used to encrypt the data disk.
PerformanceLevel	String	PL1	The performance level (PL) of the data disk that is an ESSD.
Size	Integer	200	The size of the data disk. Unit: GiB. • Valid values if you set the Category parameter to cloud: 5 to 2000. • Valid values if you set the Category parameter to cloud_efficiency: 20 to 32768. • Valid values if you set the Category parameter to cloud_ssd: 20 to 32768. • Valid values if you set the Category parameter to cloud_essd: 20 to 32768. • Valid values if you set the Category parameter to ephemeral_ssd: 5 to 800.
SnapshotId	String	s-23f2i****	The ID of the snapshot that is used to create the data disk.

Parameter	Type	Example	Description
DedicatedHostId	String	dh-bp67acfmxazb4p****	The ID of the dedicated host on which the ECS instance runs. You cannot create preemptible instances on dedicated hosts. If you specify the DedicatedHostId parameter, the SpotStrategy and SpotPriceLimit parameters are ignored. You can call the DescribeDedicatedHosts operation to query the list of dedicated host IDs.
DeploymentSetId	String	ds-bp1frxuzdg87zh4p****	The ID of the deployment set to which the ECS instance belongs.
HostName	String	LocalHost	The hostname of the instance.
HpcClusterId	String	hpc-clusterid	The ID of the Elastic High Performance Computing (E-HPC) cluster to which the ECS instance belongs.
ImageFamily	String	null	**Note** This parameter is unavailable.
ImageId	String	centos6u5_64_20G_aliaegis_2014****.vhd	The ID of the image that is used by Auto Scaling to automatically create ECS instances.
ImageName	String	centos6u5_64_20G_aliaegis_20140703.vhd	The name of the image.
InstanceDescription	String	FinaceDept	The description of the ECS instance.
InstanceGeneration	String	ecs-3	The generation of the ECS instance.
InstanceName	String	fortest	The name of the ECS instance.
InstancePatternInfos	Array of InstancePatternInfo		Information about the intelligent configuration settings, which determines the available instance types.

Parameter	Type	Example	Description
InstancePatternInfo			
Cores	Integer	2	The number of vCPUs that are supported by the instance type.
InstanceFamilyLevel	String	EnterpriseLevel	The level of the instance family. • EntryLevel: shared instance type. Instances of this level are the most cost-effective but may not provide stable computing performance in a consistent manner. Instances of this level are suitable for business scenarios in which the CPU utilization is low. For more information, see the *Shared instance families* topic in ECS documentation. • EnterpriseLevel: Instances of this level provide stable performance and dedicated resources, and are suitable for business scenarios that require high stability. For more information, see the *Instance type families* topic in ECS documentation. • CreditEntryLevel: This value is available only for burstable instances. CPU credits are used to ensure computing performance. Instances of this level are suitable for scenarios in which the CPU utilization is low but may fluctuate in specific cases. For more information, see the *Burstable instance types* topic in ECS documentation.
MaxPrice	Float	2	The maximum hourly price for pay-as-you-go instances or preemptible instances.
Memory	Float	4	The memory size that is supported by the instance type. Unit: GiB.
InstanceType	String	ecs.t1.xsmall	The instance type of the ECS instance.
InstanceTypes	List	ecs.t1.xsmall	The list of ECS instance types.

Parameter	Type	Example	Description
InternetChargeType	String	PayByTraffic	The billing method for network usage. Valid values: • PayByBandwidth: You are charged for the maximum available bandwidth that is specified by the InternetMaxBandwidthOut parameter. • PayByTraffic: You are charged for the actual traffic that you use. The InternetMaxBandwidthOut parameter specifies only the maximum available bandwidth.
InternetMaxBandwidthIn	Integer	200	The maximum inbound bandwidth from the Internet. Unit: Mbit/s. Valid values: 1 to 200.
InternetMaxBandwidthOut	Integer	0	The maximum outbound bandwidth to the Internet. Unit: Mbit/s. • Valid values if you set the InternetChargeType parameter to PayByBandwidth: 0 to 100. If this parameter is not specified, the value of this parameter is 0. • Valid values if you set the InternetChargeType parameter to PayByTraffic: 0 to 100. If this parameter is not specified, an error is reported.
IoOptimized	String	none	Indicates whether the instance is I/O optimized. Valid values: • none: The instance is not I/O optimized. • optimized: The instance is I/O optimized.
Ipv6AddressCount	Integer	1	The number of randomly generated IPv6 addresses that are allocated to the elastic network interface (ENI).
KeyPairName	String	fortest	The name of the key pair used to log on to the ECS instance.

Parameter	Type	Example	Description
LifecycleState	String	Active	The status of the scaling configuration in the scaling group. Valid values: • Active: The scaling configuration is active in the scaling group. Auto Scaling uses the active scaling configuration to automatically create ECS instances. • Inactive: The scaling configuration is inactive in the scaling group. Inactive scaling configurations are still retained in the scaling group, but are not used by Auto Scaling to create ECS instances.
LoadBalancerWeight	Integer	1	The weight of the instance as a backend server. Valid values: 1 to 100.
Memory	Integer	16	The size of the memory. You can specify the number of vCPUs and the memory size to determine the range of instance types. For example, you can set the Cpu parameter to 2 and the Memory parameter to 16 to specify the instance types that have 2 vCPUs and 16 GiB of memory. If you specify the Cpu and Memory parameters, Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price. **Note** You can specify the number of vCPUs and the memory size to determine the range of instance types only if the Scaling Policy parameter is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.

Parameter	Type	Example	Description
PasswordInherit	Boolean	true	Indicates whether the password preset in the image is used.
PrivatePoolOptions.Id	String	eap-bp67acfmxazb4****	The ID of the private pool. The ID of a private pool is the same as the ID of the elasticity assurance or capacity reservation for which the private pool is generated.
PrivatePoolOptions.MatchCriteria	String	Open	The type of the private pool. A private pool is generated after an elasticity assurance or a capacity reservation takes effect. You can select the private pool when you start an instance. Valid values: • Open: open private pool. The system selects a matching open private pool to start the instance. If no matching open private pools are available, the resources in the public pool are used. • Target: specified private pool. The system uses the capacity of a specified private pool to start the instance. If the specified private pool is unavailable, the instances cannot be started. • None: no private pool. The capacity of private pools is not used.
RamRoleName	String	RamRoleTest	The name of the RAM role that is associated with the ECS instance. The name is provided and maintained by Resource Access Management (RAM). You can call the ListRoles operation to query the available RAM roles. You can call the CreateRole operation to create RAM roles.
ResourceGroupId	String	rg-resourcegroupid1	The ID of the resource group to which the ECS instance belongs.
ScalingConfigurationId	String	asc-bp1ezrfgoyn5kijl****	The ID of the scaling configuration.
ScalingConfigurationName	String	c1908dd1-690f-4c9b-ab73-350f1f06****	The name of the scaling configuration.
ScalingGroupId	String	sg-280ih****	The ID of the scaling group for which the scaling configuration is created.

Parameter	Type	Example	Description
SchedulerOptions	Struct		Note This parameter is unavailable.
ManagedPrivateSpaceId	String	testManagedPrivateSpaceId	Note This parameter is unavailable.
SecurityEnhancementStrategy	String	Active	Indicates whether security hardening is enabled. Valid values: • Active: Security hardening is enabled. This value is available only for public images. • Deactive: Security hardening is disabled. This value is available for all types of images.
SecurityGroupId	String	sg-bp18kz60mefs****	The ID of the security group with which the ECS instance is associated. The ECS instances that are associated with the same security group can access each other.
SecurityGroupIds	List	sg-bp18kz60mefs****	The IDs of the security groups with which the ECS instance is associated. The ECS instances that are associated with the same security group can access each other.
SpotDuration	Integer	1	The retention period of the preemptible instance. Unit: hours.
SpotInterruptionBehavior	String	Terminate	The interruption mode of the preemptible instance.
SpotPriceLimit	Array of SpotPriceModel		Information about the preemptible instance.
SpotPriceModel			
InstanceType	String	ecs.t1.xsmall	The instance type of the preemptible instance.
PriceLimit	Float	0.125	The price limit of the preemptible instance.

Parameter	Type	Example	Description
SpotStrategy	String	NoSpot	The preemption policy that is applied to pay-as-you-go instances and preemptible instances. Valid values: • NoSpot: The instance is created as a pay-as-you-go instance. • SpotWithPriceLimit: The instance is a preemptible instance that has a user-defined maximum hourly price. • SpotAsPriceGo: The instance is a preemptible instance for which the market price at the time of purchase is used as the bid price.
SystemDiskAutoSnapshotPolicyId	String	sp-bp12m37ccmxvbmi5****	The ID of the automatic snapshot policy that is applied to the system disk.
SystemDiskCategories	List	cloud	The categories of the system disks. The values are sorted based on their priorities. The first value has the highest priority. If Auto Scaling cannot use the disk category that has the highest priority to create instances, Auto Scaling uses the disk category that has the next highest priority to create instances. Valid values: • cloud: basic disk • cloud_efficiency: ultra disk • cloud_ssd: standard SSD • cloud_essd: ESSD
SystemDiskCategory	String	cloud	The category of the system disk. Valid values: • cloud: basic disk • cloud_efficiency: ultra disk • cloud_ssd: standard SSD • ephemeral_ssd: local SSD • cloud_essd: ESSD
SystemDiskDescription	String	TestDescription	The description of the system disk.
SystemDiskName	String	cloud_ssd_Test	The name of the system disk.
SystemDiskPerformanceLevel	String	PL1	The PL of the system disk that is an ESSD.

Parameter	Type	Example	Description
SystemDiskSize	Integer	100	The size of the system disk.
Tags	Array of Tag		Information about tags.
Tag			
Key	String	binary	The key of the tag.
Value	String	alterTable	The value of the tag.
Tenancy	String	default	Indicates whether the instance is created on a dedicated host. Valid values: • default: The instance is not created on a dedicated host. • host: The instance is created on a dedicated host. If you do not specify the DedicatedHostId parameter, the system selects a dedicated host for the instance that you created.
UserData	String	echo hello ecs!	The user data of the ECS instance.
WeightedCapacities	List	4	The weight of the instance type. This parameter indicates the capacity of an instance of this instance type in the scaling group. A higher weight indicates that a smaller number of instances of this instance type are required to meet the expected capacity.
ZoneId	String	cn-hangzhou-env17-d01-g	The ID of the zone in which the instance is created. You can call the DescribeZones operation to query the available zones.
TotalCount	Integer	1	The total number of scaling configurations.

Examples

Sample requests

```
http://ess.example.com/?Action=DescribeScalingConfigurations
&RegionId=cn-qingdao-env17-d01
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DescribeScalingConfigurationsResponse>
      <RequestId>804F240A-8D3E-40A1-BD68-6B333DEA2CA8</RequestId>
      <TotalCount>1</TotalCount>
      <PageNumber>1</PageNumber>
      <PageSize>50</PageSize>
      <ScalingConfigurations>
            <ScalingConfiguration>
                  <CreationTime>2014-08-14T10:58Z</CreationTime>
                  <ImageId>centos6u5_64_20G_aliaegis_20140703.vhd</ImageId>
                  <InstanceType>ecs.t1.xsmall</InstanceType>
                  <InternetChargeType>PayByTraffic</InternetChargeType>
                  <InternetMaxBandwidthIn>200</InternetMaxBandwidthIn>
                  <InternetMaxBandwidthOut>0</InternetMaxBandwidthOut>
                  <LifecycleState>Active</LifecycleState>
                  <ScalingConfigurationId>asc-bp1ezrfgoyn5kijl****</ScalingConfigurationId>
                  <ScalingConfigurationName>c1908dd1-690f-4c9b-ab73-350f1f06****</ScalingCo
nfigurationName>
                  <ScalingGroupId>sg-280ih****</ScalingGroupId>
                  <SecurityGroupId>sg-bp18kz60mefs****</SecurityGroupId>
                  <SystemDiskCategory>cloud</SystemDiskCategory>
                  <DataDisks>
                        <DataDisk>
                              <Size>200</Size>
                              <Category>cloud</Category>
                              <SnapshotId>s-280s7****</SnapshotId>
                              <Device>/dev/xvdb</Device>
                        </DataDisk>
                  </DataDisks>
            </ScalingConfiguration>
      </ScalingConfigurations>
</DescribeScalingConfigurationsResponse>
```

`JSON` format

```
{
    "RequestId": "804F240A-8D3E-40A1-BD68-6B333DEA2CA8",
    "TotalCount": "1",
    "PageNumber": "1",
    "PageSize": "50",
    "ScalingConfigurations": {
        "ScalingConfiguration": {
            "CreationTime": "2014-08-14T10:58Z",
            "ImageId": "centos6u5_64_20G_aliaegis_20140703.vhd",
            "InstanceType": "ecs.t1.xsmall",
            "InternetChargeType": "PayByTraffic",
            "InternetMaxBandwidthIn": "200",
            "InternetMaxBandwidthOut": "0",
            "LifecycleState": "Active",
            "ScalingConfigurationId": "asc-bp1ezrfgoyn5kijl****",
            "ScalingConfigurationName": "c1908dd1-690f-4c9b-ab73-350f1f06****",
            "ScalingGroupId": "sg-280ih****",
            "SecurityGroupId": "sg-bp18kz60mefs****",
            "SystemDiskCategory": "cloud",
            "DataDisks": {
                "DataDisk": {
                    "Size": "200",
                    "Category": "cloud",
                    "SnapshotId": "s-280s7****",
                    "Device": "/dev/xvdb"
                }
            }
        }
    }
}
```


6.1.8.3. DeleteScalingConfiguration

Deletes a scaling configuration.

Description

You cannot delete a scaling configuration in one of the following scenarios:

- The scaling configuration in your scaling group is in the Active state.
- Your scaling group contains Elastic Compute Service (ECS) instances that are created based on the scaling configuration.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DeleteScalingConfiguration	The operation that you want to perform. Set the value to DeleteScalingConfiguration.

Parameter	Type	Required	Example	Description
ScalingConfigurationId	String	Yes	eOs27Kb0oXvQcUYjEGel****	The ID of the scaling configuration that you want to delete. You can call the DescribeScalingConfigurations operation to query the ID of the scaling configuration.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. Every response returns a unique request ID regardless of whether the call is successful.

Examples

Sample requests

```
http://ess.example.com/?Action=DeleteScalingConfiguration
&ScalingConfigurationId=eOs27Kb0oXvQcUYjEGel****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DeleteScalingConfigurationResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DeleteScalingConfigurationResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description

HTTP status code	Error code	Error message	Description
404	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	The error message returned because the specified scaling configuration does not exist within your Apsara Stack tenant account.
400	IncorrectScalingConfigurationLifecycleState	The current lifecycle state of specified scaling configuration does not support this action.	The error message returned because the specified scaling configuration is in the Active state.
400	InstanceInUse	You cannot delete a scaling configuration or scaling group while there is an instance associated with it.	The error message returned because the scaling group contains ECS instances that are created based on the specified scaling configuration.

6.1.9. Scaling rules

6.1.9.1. CreateScalingRule

Creates a scaling rule.

Description

A scaling rule defines a specific scaling activity, such as adding or removing N instances.

If the number of Elastic Compute Service (ECS) instances in a scaling group is less than the minimum number allowed or greater than the maximum number allowed after a scaling rule is executed, Auto Scaling adjusts the number of ECS instances. This way, the number of ECS instances is still within the specified range after the scaling rule is executed. The number of ECS instances that is specified in the scaling rule remains unchanged.

- If your scaling group contains two ECS instances and allows up to three ECS instances, only one ECS instance is added to your scaling group after you execute a scale-out rule in which three ECS instances are specified. The number of ECS instances that is specified in the scaling rule remains unchanged.
- If your scaling group contains three ECS instances and requires at least two ECS instances, only one ECS instance is removed from your scaling group after you execute a scale-in rule in which five ECS instances are specified. The number of ECS instances that is specified in the scaling rule remains unchanged.

You can create scaling rules based on input parameters.

- If you set the AdjustmentType parameter to TotalCapacity, the number of ECS instances in the scaling group is adjusted to the specified value. The value of the AdjustmentValue parameter must be greater than or equal to 0.
- If you set the AdjustmentType parameter to QuantityChangeInCapacity or PercentChangeInCapacity, a positive value of AdjustmentValue specifies the number of ECS instances that are added to the scaling group, and a negative value of AdjustmentValue specifies the number of ECS instances that are removed from the scaling group.
- If you set the AdjustmentType parameter to PercentChangeInCapacity, Auto Scaling uses the following formula to calculate a value, and then rounds the value to the nearest integer to obtain the number of ECS instances that must be scaled: Value of TotalCapacity × Value of AdjustmentValue/100.
- If the cooldown time (Cooldown) is specified in a scaling rule, the specified time applies to the scaling group after the rule is executed. Otherwise, the value of the DefaultCooldown parameter of the scaling group applies to the scaling group.
- A scaling group can have up to 10 scaling rules.
- The following API operations use the unique identifier that is indicated by the ScalingRuleAri parameter of a scaling rule returned by this operation:
  - ExecuteScalingRule: You can call this operation to manually execute a specific scaling rule by setting the ScalingRuleAri parameter to the unique identifier.
  - CreateScheduledTask: You can call this operation to create a scheduled task for a specific scaling rule by setting the ScheduledAction parameter to the unique identifier.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	CreateScalingRule	The operation that you want to perform. Set the value to CreateScalingRule.
ScalingGroupId	String	Yes	asg-bp1ffogfdauy0jw0****	The ID of the scaling group for which you want to create the scaling rule.
ScalingRuleName	String	No	scalingrule****	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. The name of a scaling rule must be unique in the scaling group to which the scaling rule belongs within an Alibaba Cloud account. If you do not specify this parameter, the value of the ScalingRuleId parameter is used.

Parameter	Type	Required	Example	Description
Cooldown	Integer	No	60	The cooldown time of the scaling rule. This parameter is applicable only to simple scaling rules. Valid values: 0 to 86400. Unit: seconds. By default, this parameter is left empty.
MinAdjustmentMagnitude	Integer	No	1	The minimum number of instances that you want to scale when the adjustment type is PercentChangeInCapacity. This parameter takes effect only when the ScalingRuleType parameter is set to SimpleScalingRule or StepScalingRule and the AdjustmentType parameter is set to PercentChangeInCapacity.
AdjustmentType	String	No	QuantityChangeInCapacity	The adjustment method of the scaling rule. This parameter is required and applicable only to simple scaling rules. Valid values: • QuantityChangeInCapacity: adds the specified number of ECS instances to or removes the specified number of ECS instances from the scaling group. • PercentChangeInCapacity: adds the specified percentage of ECS instances to or removes the specified percentage of ECS instances from the scaling group. • TotalCapacity: adjusts the number of ECS instances in the scaling group to a specified number.

Parameter	Type	Required	Example	Description
AdjustmentValue	Integer	No	100	The number of ECS instances that you want to scale in the scaling rule. This parameter is required and applicable only to simple scaling rules. The number of ECS instances that you want to scale in a single scaling activity cannot exceed 500. • Valid values if you set the AdjustmentType parameter to QuantityChangeInCapacity: -500 to 500 • Valid values if you set the AdjustmentType parameter to PercentChangeInCapacity: -100 to 10000 • Valid values if you set the AdjustmentType parameter to TotalCapacity: 0 to 1000
ScalingRuleType	String	No	SimpleScalingRule	The type of the scaling rule. Valid values: • SimpleScalingRule: scales the number of ECS instances based on the values of the AdjustmentType and AdjustmentValue parameters. • TargetTrackingScalingRule: calculates the number of ECS instances that must be scaled and maintains the value of a predefined metric close to the value of the TargetValue parameter. • StepScalingRule: scales ECS instances in steps based on specified thresholds and metric values. • PredictiveScalingRule: analyzes historical monitoring data of the scaling group and uses machine learning to predict the future values of metrics. If you set the ScalingRuleType parameter to PredictiveScalingRule, Auto Scaling creates scheduled tasks to specify the boundary values for the scaling group. Default value: SimpleScalingRule.

Parameter	Type	Required	Example	Description
EstimatedInstanceWarmup	Integer	No	300	The warmup period of the ECS instance. This parameter is available only if you set the ScalingRuleType parameter to TargetTrackingScalingRule or PredictiveScalingRule. Auto Scaling adds ECS instances that are in the warmup state to a scaling group but does not report monitoring data to CloudMonitor during the warmup period. **Note** Auto Scaling calculates the number of ECS instances that must be scaled. ECS instances in the warmup state are not included in the current capacity of the scaling group. Valid values: 0 to 86400. Unit: seconds. Default value: 300.

Parameter	Type	Required	Example	Description
MetricName	String	No	CpuUtilization	The predefined metric that you want to monitor. This parameter is required and applicable only to target tracking scaling rules and predictive scaling rules. Valid values if you set the ScalingRuleType parameter to TargetTrackingScalingRule: • CpuUtilization: the average CPU utilization • ClassicInternetRx: the average inbound Internet traffic over the classic network • ClassicInternetTx: the average outbound Internet traffic over the classic network • VpcInternetRx: the average inbound Internet traffic over a virtual private cloud (VPC) • VpcInternetTx: the average outbound Internet traffic over a VPC • IntranetRx: the average inbound traffic over the internal network • IntranetTx: the average outbound traffic over the internal network Valid values if you set the ScalingRuleType parameter to PredictiveScalingRule: • CpuUtilization: the average CPU utilization • IntranetRx: the average inbound traffic over the internal network • IntranetTx: the average outbound traffic over the internal network
TargetValue	Float	No	0.125	The target value of the metric. This parameter is required and applicable only to target tracking scaling rules and predictive scaling rules. The value must be greater than 0 and can have up to three decimal places.
DisableScaleIn	Boolean	No	false	Specifies whether to disable scale-in. This parameter is applicable only to target tracking scaling rules. Default value: false.

Parameter	Type	Required	Example	Description
ScaleInEvaluationCount	Integer	No	15	The number of consecutive times that the event-triggered task created for scale-in activities meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule. Default value: 15.
ScaleOutEvaluationCount	Integer	No	3	The number of consecutive times that the event-triggered task created for scale-out activities meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule. Default value: 3.

Parameter	Type	Required	Example	Description
StepAdjustment.N.MetricIntervalLowerBound	Float	No	1.0	The lower limit that is specified in a step adjustment. This parameter takes effect only if you set the ScalingRuleType parameter to StepScalingRule. Valid values: -9.999999E18 to 9.999999E18.
StepAdjustment.N.MetricIntervalUpperBound	Float	No	5.0	The upper limit value that is specified in a step adjustment. This parameter takes effect only if you set the ScalingRuleType parameter to StepScalingRule. Valid values: -9.999999E18 to 9.999999E18.
StepAdjustment.N.ScalingAdjustment	Integer	No	1	The number of ECS instances that you want to scale in a step adjustment. This parameter takes effect only if you set the ScalingRuleType parameter to StepScalingRule.
PredictiveScalingMode	String	No	PredictAndScale	The mode of the predictive scaling rule. Valid values: • PredictAndScale: provides predictions and creates prediction tasks. • PredictOnly: provides predictions but does not create prediction tasks. Default value: PredictAndScale.

Parameter	Type	Required	Example	Description
PredictiveValueBehavior	String	No	MaxOverridePredictiveValue	The action on the predicted maximum value. Valid values: • MaxOverridePredictiveValue: uses the initial maximum capacity as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMax: uses the predicted value as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMaxWithBuffer: increases the predicted value by a percentage that is specified by the PredictiveValueBuffer parameter, and then uses the increased value as the maximum value for prediction tasks if the predicted value increased by this percentage is greater than the initial maximum capacity. Default value: MaxOverridePredictiveValue.
PredictiveValueBuffer	Integer	No	50	The percentage of the increment to the predicted value when you set the PredictiveValueBehavior parameter to PredictiveValueOverrideMaxWithBuffer. If the predicted value increased by this percentage is greater than the initial maximum capacity, the increased value is used as the maximum value for prediction tasks. Valid values: 0 to 100. Default value: 0.

Parameter	Type	Required	Example	Description
PredictiveTaskBufferTime	Integer	No	30	The amount of buffer time before the prediction task is executed. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed on the hour. You can specify an amount of buffer time for resource preparation before prediction tasks are executed. Valid values: 0 to 60. Unit: minutes. Default value: 0.
InitialMaxSize	Integer	No	100	The maximum number of ECS instances in the scaling group. This parameter is used together with the PredictiveValueBehavior parameter. The default value of this parameter is the value of the MaxSize parameter.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingRuleAri	String	ari:acs:ess:cn-hangzhou-env17-d01:140692647406****:scalingrule/asr-bp1dvirgwkoowxk7****	The unique identifier of the scaling rule.
ScalingRuleId	String	asr-bp1dvirgwkoowxk7****	The ID of the scaling rule, which is generated by the system and is globally unique.

Examples

Sample requests

```
http://ess.example.com/?Action=CreateScalingRule
&ScalingGroupId=asg-bp1ffogfdauy0jw0****
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=-10
&<Common request parameters>
```

Sample success responses

`XML` format

```
<CreateScalingRuleResponse>
      <ScalingRuleAri>ari:acs:ess:cn-hangzhou-env17-d01:140692647406****:scalingrule/asr-bp
1dvirgwkoowxk7****</ScalingRuleAri>
      <ScalingRuleId>asr-bp1dvirgwkoowxk7****</ScalingRuleId>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</CreateScalingRuleResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "ScalingRuleId": "asr-bp1dvirgwkoowxk7****",
    "ScalingRuleAri": "ari:acs:ess:cn-hangzhou-env17-d01:140692647406****:scalingrule/asr-b
p1dvirgwkoowxk7****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Alibaba Cloud account.
400	InvalidScalingRuleName.Duplicate	The specified value of parameter <parameter name> is duplicated.	The error message returned because the specified scaling rule name already exists.
400	QuotaExceeded.ScalingRule	Scaling rule quota exceeded in the specified scaling group.	The error message returned because the maximum number of scaling rules in the scaling group has been reached.

HTTP status code	Error code	Error message	Description
400	TargetTrackingScalingRule.UnsupportedMetric	Specific metric is not supported for target tracking scaling rule.	The error message returned because the target tracking scaling rule does not support the specified monitoring metric.
400	TargetTrackingScalingRule.DumplicateMetric	Only one TargetTrackingScaling rule for a given metric specification is allowed.	The error message returned because the monitoring metric is already specified for a target tracking scaling rule in the scaling group.
400	InvalidMinAdjustmentMagnitudeMismatchAdjustmentType	MinAdjustmentMagnitude is not supported by the specified adjustment type.	The error message returned because the MinAdjustmentMagnitude parameter cannot be applied to the specified adjustment method of the scaling rule.
400	InvalidStepAdjustments.MultipleNullUpperBound	At most one StepAdjustment may have an unspecified upper bound.	The error message returned because a step adjustment without an upper limit already exists.
400	InvalidStepAdjustments.MultipleNullLowerBound	At most one StepAdjustment may have an unspecified lower bound.	The error message returned because a step adjustment without a lower limit already exists.
400	InvalidStepAdjustments.NoNullLowerBound	There must be a StepAdjustment with an unspecified lower bound when one StepAdjustment has a negative lower bound.	The error message returned because the lower limit of a step adjustment is a negative value, but a different step adjustment without a lower limit does not exist.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.NoNullUpperBound	There must be a StepAdjustment with an unspecified upper bound when one StepAdjustment has a positive upper bound.	The error message returned because the upper limit of a step adjustment is a positive value, but a different step adjustment without an upper limit does not exist.
400	InvalidStepAdjustments.Gap	StepAdjustment intervals can not have gaps between them.	The error message returned because the specified ranges of step adjustments have gaps.
400	InvalidStepAdjustments.Overlap	StepAdjustment intervals can not overlap.	The error message returned because the specified ranges of step adjustments overlap.
400	InvalidStepAdjustments.LowerGtUpper	LowerBound must be less than the UpperBound for StepAdjustment :%s.	The error message returned because the lower limit of a step adjustment is greater than or equal to the upper limit.
400	InvalidStepAdjustments.BothNull	Both lower and upper bounds of a StepAdjustment can not be left unspecified.	The error message returned because the upper limit and the lower limit for the step adjustment are not specified.
400	InvalidStepAdjustments.MaxNum	Your scaling rule can have at most %s StepAdjustments.	The error message returned because the maximum number of step adjustments in the scaling group has been reached.
400	StepBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned because the specified upper limit or lower limit of the step adjustment is invalid.

6.1.9.2. ModifyScalingRule

Modifies a scaling rule.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	ModifyScalingRule	The operation that you want to perform. Set the value to ModifyScalingRule.
ScalingRuleId	String	Yes	eMKWG8SRNb9dBLAjweN****	The ID of the scaling rule that you want to modify.
ScalingRuleName	String	No	scalingrule****	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. The name of a scaling rule must be unique in the scaling group to which the scaling rule belongs in an Apsara Stack tenant account.
Cooldown	Integer	No	60	The cooldown time of the scaling rule. This parameter takes effect only for simple scaling rules. Valid values: 0 to 86400. Unit: seconds.
MinAdjustmentMagnitude	Integer	No	1	The minimum number of instances that must be scaled when the adjustment type is PercentChangeInCapacity. This parameter takes effect only if you set the ScalingRuleType parameter to SimpleScalingRule or StepScalingRule.

Parameter	Type	Required	Example	Description
AdjustmentType	String	No	QuantityChangeIn Capacity	The method used by the scaling rule to adjust the number of instances. This parameter takes effect only for simple scaling rules. Valid values: • QuantityChangeInCapacity: adds the specified number of ECS instances to or removes the specified number of ECS instances from the scaling group. • PercentChangeInCapacity: adds the specified percentage of ECS instances to or removes the specified percentage of ECS instances from the scaling group. • TotalCapacity: adjusts the number of ECS instances in the scaling group to a specified number.
AdjustmentValue	Integer	No	100	The number of ECS instances to be adjusted by the scaling rule. This parameter takes effect only for simple scaling rules. The number of ECS instances to be adjusted in a single scaling activity cannot exceed 500. • Valid values if you set the AdjustmentType parameter to QuantityChangeInCapacity: -500 to 500 • Valid values if you set the AdjustmentType parameter to PercentChangeInCapacity: -100 to 10000 • Valid values if you set the AdjustmentType parameter to TotalCapacity: 0 to 1000

Parameter	Type	Required	Example	Description
EstimatedInstanceWarmup	Integer	No	60	The warmup period of the ECS instance. This parameter takes effect only if you set the ScalingRuleType parameter to TargetTrackingScalingRule or PredictiveScalingRule. Auto Scaling adds ECS instances that are in the warmup state to a scaling group but does not report monitoring data to CloudMonitor during the warmup period. **Note** Auto Scaling calculates the number of ECS instances that must be scaled. ECS instances in the warmup state are not included in the current capacity of the scaling group. Valid values: 0 to 86400. Unit: seconds.

Parameter	Type	Required	Example	Description
MetricName	String	No	CpuUtilization	The predefined metric that you want to monitor. This parameter is required and takes effect only for target tracking scaling rules and predictive scaling rules. Valid values if you set the ScalingRuleType parameter to TargetTrackingScalingRule: • CpuUtilization: the average CPU utilization • ClassicInternetRx: the average inbound Internet traffic over the classic network • ClassicInternetTx: the average outbound Internet traffic over the classic network • VpcInternetRx: the average inbound Internet traffic over the virtual private cloud (VPC) • VpcInternetTx: the average outbound Internet traffic over the VPC • IntranetRx: the average inbound traffic over the internal network • IntranetTx: the average outbound traffic over the internal network Valid values if you set the ScalingRuleType parameter to PredictiveScalingRule: • CpuUtilization: the average CPU utilization • IntranetRx: the average inbound traffic over the internal network • IntranetTx: the average outbound traffic over the internal network
TargetValue	Float	No	0.125	The target value of the metric. This parameter takes effect only for target tracking scaling rules and predictive scaling. The value must be greater than 0 and can have a maximum of three decimal places.
DisableScaleIn	Boolean	No	true	Specifies whether to disable scale-in. This parameter takes effect only for target tracking scaling rules.

Parameter	Type	Required	Example	Description
ScaleInEvaluationCount	Integer	No	15	The number of consecutive times that the event-triggered task created for scale-in activities must meet the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule.
ScaleOutEvaluationCount	Integer	No	3	The number of consecutive times that the event-triggered task created for scale-out activities must meet the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule.
StepAdjustment.N.MetricIntervalLowerBound	Float	No	1.0	The lower limit specified in a step adjustment. This parameter takes effect only if you set the ScalingRuleType parameter to StepScalingRule. Valid values: -9.999999E18 to 9.999999E18.
StepAdjustment.N.MetricIntervalUpperBound	Float	No	5.0	The upper limit specified in a step adjustment. This parameter takes effect only if you set the ScalingRuleType parameter to StepScalingRule. Valid values: -9.999999E18 to 9.999999E18.
StepAdjustment.N.ScalingAdjustment	Integer	No	1	The specified number of ECS instances to be scaled in a step adjustment. This parameter takes effect only if you set the ScalingRuleType parameter to StepScalingRule.

Parameter	Type	Required	Example	Description
PredictiveScalingMode	String	No	PredictAndScale	The mode of the predictive scaling rule. Valid values: • PredictAndScale: produces predictions and creates prediction tasks. • PredictOnly: produces predictions but does not create prediction tasks.
PredictiveValueBehavior	String	No	MaxOverridePredictiveValue	The action that you want to take on the predicted maximum value. Valid values: • MaxOverridePredictiveValue: uses the initial maximum capacity as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMax: uses the predicted value as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMaxWithBuffer: increases the predicted value by a percentage that is specified by the PredictiveValueBuffer parameter. If the predicted value that is increased by the percentage is greater than the initial maximum capacity, the increased value is used as the maximum value for prediction tasks.
PredictiveValueBuffer	Integer	No	50	The percentage of the increment to the predicted value when the PredictiveValueBehavior parameter is set to PredictiveValueOverrideMaxWithBuffer. If the predicted value increased by this percentage is greater than the initial maximum capacity, the increased value is used as the maximum value for prediction tasks. Valid values: 0 to 100.

Parameter	Type	Required	Example	Description
PredictiveTaskBufferTime	Integer	No	30	The amount of buffer time before the prediction task is executed. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed on the hour. You can specify an amount of buffer time for resource preparation before prediction tasks are executed. Valid values: 0 to 60.
InitialMaxSize	Integer	No	100	The maximum number of ECS instances in the scaling group. This parameter is used together with the PredictiveValueBehavior parameter.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=ModifyScalingRule
&ScalingRuleId=eMKWG8SRNb9dBLAjweN****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<ModifyScalingRuleResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</ModifyScalingRuleResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist in the Apsara Stack tenant account.
400	InvalidScalingRuleName.Duplicate	The specified value of parameter <parameter name> is duplicated.	The error message returned because the scaling rule name already exists.
400	QuotaExceeded.ScalingRule	Scaling rule quota exceeded in the specified scaling group.	The error message returned because the maximum number of scaling rules in the scaling group has been reached.
400	TargetTrackingScalingRule.UnsupportedMetric	Specific metric is not supported for target tracking scaling rule.	The error message returned because the target tracking scaling rule does not support the specified monitoring metric.
400	TargetTrackingScalingRule.DumplicateMetric	Only one TargetTrackingScaling rule for a given metric specification is allowed.	The error message returned because the monitoring metric is already specified for a target tracking scaling rule in the scaling group.
400	InvalidMinAdjustmentMagnitudeMismatchAdjustmentType	MinAdjustmentMagnitude is not supported by the specified adjustment type.	The error message returned because the MinAdjustmentMagnitude parameter cannot take effect for the specified adjustment method of the scaling rule.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.MultipleNullUpperBound	At most one StepAdjustment may have an unspecified upper bound.	The error message returned because a step adjustment without an upper limit already exists.
400	InvalidStepAdjustments.MultipleNullLowerBound	At most one StepAdjustment may have an unspecified lower bound.	The error message returned because a step adjustment without a lower limit already exists.
400	InvalidStepAdjustments.NoNullLowerBound	There must be a StepAdjustment with an unspecified lower bound when one StepAdjustment has a negative lower bound.	The error message returned because the lower limit of a step adjustment is a negative value, but a different step adjustment without a lower limit does not exist.
400	InvalidStepAdjustments.NoNullUpperBound	There must be a StepAdjustment with an unspecified upper bound when one StepAdjustment has a positive upper bound.	The error message returned because the upper limit of a step adjustment is a positive value, but a different step adjustment without an upper limit does not exist.
400	InvalidStepAdjustments.Gap	StepAdjustment intervals can not have gaps between them.	The error message returned because the specified ranges of step adjustments have gaps.
400	InvalidStepAdjustments.Overlap	StepAdjustment intervals can not overlap.	The error message returned because the specified ranges of step adjustments overlap.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.LowerGtUpper	LowerBound must be less than the UpperBound for StepAdjustment :%s.	The error message returned because the lower limit of a step adjustment is greater than or equal to the upper limit.
400	InvalidStepAdjustments.BothNull	Both lower and upper bounds of a StepAdjustment can not be left unspecified.	The error message returned because the upper limit and the lower limit for the step adjustment are not specified.
400	InvalidStepAdjustments.MaxNum	Your scaling rule can have at most %s StepAdjustments.	The error message returned because the maximum number of step adjustments in the scaling group has been reached.
400	StepBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned because the specified upper limit or lower limit of the step adjustment is invalid.

6.1.9.3. DescribeScalingRules

Queries all scaling rules in a scaling group and information about the scaling rules.

Description

You can query all scaling rules in a scaling group by specifying the ID of the scaling group and filter scaling rules that are returned by scaling rule ID, name, unique identifier, and type.

Request parameters

Parameter	Type	Required	Example	Description

Parameter	Type	Required	Example	Description
Action	String	Yes	DescribeScalingRules	The operation that you want to perform. Set the value to DescribeScalingRules.
RegionId	String	Yes	cn-qingdao-env17-d01	The region ID of the scaling group to which the scaling rules that you want to query belong.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	50	The number of entries to return on each page. Maximum value: 50. Default value: 10.
ScalingGroupId	String	No	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
ScalingRuleType	String	No	SimpleScalingRule	The type of the scaling rule. Valid values: • SimpleScalingRule: scales the number of ECS instances based on the values of the AdjustmentType and AdjustmentValue parameters. • TargetTrackingScalingRule: calculates the number of ECS instances that must be scaled and maintains the value of the predefined metric close to the value of the TargetValue parameter. • StepScalingRule: scales ECS instances in steps based on specified thresholds and metric values. • PredictiveScalingRule: analyzes historical monitoring data of the scaling group and uses machine learning to predict the future values of metrics. If you set the ScalingRuleType parameter to PredictiveScalingRule, Auto Scaling creates scheduled tasks to specify the boundary values for the scaling group. Default value: SimpleScalingRule.
ShowAlarmRules	Boolean	No	false	Specifies whether to return the event-triggered task that is associated with the scaling rule. Valid values: • true • false Default value: false.
ScalingRuleId.N	RepeatList	No	asr-bp1dvirgwkoowxk7****	The ID of scaling rule N that you want to query. Valid values of N: 1 to 10.
ScalingRuleName.N	RepeatList	No	scalingrule****	The name of scaling rule N that you want to query. Valid values of N: 1 to 10.

Parameter	Type	Required	Example	Description
ScalingRuleAri.N	RepeatList	No	ari:acs:ess:cn-hangzhou-env17-d01:140692647406****:scalingrule/asr-bp1dvirgwkoowxk7****	The unique identifier of scaling rule N that you want to query. Valid values of N: 1 to 10.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingRules	Array of ScalingRule		The information about the scaling rules.
ScalingRule			
AdjustmentType	String	QuantityChangeInCapacity	The method used by the scaling rule to adjust the number of instances. Valid values: • QuantityChangeInCapacity: adds the specified number of ECS instances to or removes the specified number of ECS instances from the scaling group. • PercentChangeInCapacity: adds the specified percentage of ECS instances to or removes the specified percentage of ECS instances from the scaling group. • TotalCapacity: adjusts the number of ECS instances in the scaling group to a specified number.
AdjustmentValue	Integer	1	The adjustment value that is specified in the scaling rule.

Parameter	Type	Example	Description
Alarms	Array of Alarm		The event-triggered tasks that are associated with the scaling rule. Event-triggered tasks that are associated with the scaling rule are returned only if you set the ShowAlarmRules parameter to true. Otherwise, an empty list is returned.
Alarm			
AlarmTaskId	String	asg-bp18p2yfxow2dloq****_1f9458d1-70e1-4bee-8c7f-7a47695b****	The ID of the event-triggered task that is associated with the scaling rule.
AlarmTaskName	String	event-notified****	The name of the event-triggered task that is associated with the scaling rule.
ComparisonOperator	String	>=	The comparison operator between the metric value and the threshold for the event-triggered task that is associated with the scaling rule. The comparison operator indicates the relationship in which the metric value and the threshold can meet the condition. • Valid value if the metric value is greater than or equal to the threshold: >= • Valid value if the metric value is less than or equal to the threshold: <= • Valid value if the metric value is greater than the threshold: > • Valid value if the metric value is less than the threshold: <
Dimensions	Array of Dimension		The dimensions of the event-triggered task that is associated with the scaling rule.
Dimension			
DimensionKey	String	scaling_group	The key of the dimension that is associated with the metric. Valid values: • scaling_group: the ID of the scaling group • userId: the ID of the Alibaba Cloud account

Parameter	Type	Example	Description
DimensionValue	String	asg-bp18p2yfxow2dloq****	The value of the dimension that is associated with the metric.
EvaluationCount	Integer	3	The number of consecutive times when the event-triggered task that is associated with the scaling rule meets the threshold expression to trigger an alert.
MetricName	String	CpuUtilization	The name of the metric of the event-triggered task that is associated with the scaling rule.
MetricType	String	system	The type of the event-triggered task that is associated with the scaling rule. Valid values: • system: the system monitoring event-triggered task • custom: the custom monitoring event-triggered task
Statistics	String	Average	The statistical method of the event-triggered task that is associated with the scaling rule. Valid values: • Average • Maximum • Minimum
Threshold	Float	50	The alert threshold of the event-triggered task that is associated with the scaling rule.
Cooldown	Integer	20	The cooldown time of the scaling rule. This parameter is applicable only to simple scaling rules. Valid values: 0 to 86400. Unit: seconds.
DisableScaleIn	Boolean	true	Indicates whether scale-in is disabled. This parameter is applicable only to target tracking scaling rules. Valid values: • true: Scale-in is disabled. • false: Scale-in is enabled.

Parameter	Type	Example	Description
EstimatedInstanceWarmup	Integer	300	The warmup period of the ECS instance.
InitialMaxSize	Integer	100	The maximum number of ECS instances in the scaling group. This parameter is used together with the PredictiveValueBehavior parameter.
MaxSize	Integer	2	The maximum number of ECS instances in the scaling group.
MetricName	String	CpuUtilization	The predefined metric that you specified to monitor. This parameter is applicable only to target tracking scaling rules and predictive scaling rules. Valid values if you set the ScalingRuleType parameter to TargetTrackingScalingRule: • CpuUtilization: the average CPU utilization • ClassicInternetRx: the average inbound Internet traffic over the classic network • ClassicInternetTx: the average outbound Internet traffic over the classic network • VpcInternetRx: the average inbound Internet traffic over a virtual private cloud (VPC) • VpcInternetTx: the average outbound Internet traffic over a VPC • IntranetRx: the average inbound traffic over the internal network • IntranetTx: the average outbound traffic over the internal network Valid values if you set the ScalingRuleType parameter to PredictiveScalingRule: • CpuUtilization: the average CPU utilization • IntranetRx: the average inbound traffic over the internal network • IntranetTx: the average outbound traffic over the internal network

Parameter	Type	Example	Description
MinAdjustmentMagnitude	Integer	1	The minimum number of instances that are scaled when the adjustment type is PercentChangeInCapacity. This parameter takes effect only when the ScalingRuleType parameter is set to SimpleScalingRule or StepScalingRule and the AdjustmentType parameter is set to PercentChangeInCapacity.
MinSize	Integer	1	The minimum number of ECS instances in the scaling group.
PredictiveScalingMode	String	PredictAndScale	The mode of the predictive scaling rule. Valid values: • PredictAndScale: provides predictions and creates prediction tasks. • PredictOnly: provides predictions but does not create prediction tasks.
PredictiveTaskBufferTime	Integer	30	The amount of buffer time before the prediction task is executed. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed on the hour. You can specify an amount of buffer time for resource preparation before prediction tasks are executed. Valid values: 0 to 60. Unit: minutes.

Parameter	Type	Example	Description
PredictiveValueBehavior	String	MaxOverridePredictiveValue	The action on the predicted maximum value. Valid values: • MaxOverridePredictiveValue: uses the initial maximum capacity as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMax: uses the predicted value as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMaxWithBuffer: increases the predicted value by a percentage that is specified by the PredictiveValueBuffer parameter, and then uses the increased value as the maximum value for prediction tasks if the predicted value increased by this percentage is greater than the initial maximum capacity.
PredictiveValueBuffer	Integer	50	The percentage of the increment to the predicted value when the PredictiveValueBehavior parameter is set to PredictiveValueOverrideMaxWithBuffer. If the predicted value increased by this percentage is greater than the initial maximum capacity, the increased value is used as the maximum value for prediction tasks. Valid values: 0 to 100.
ScaleInEvaluationCount	Integer	15	The number of consecutive times that the event-triggered task created for scale-in activities meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule.

Parameter	Type	Example	Description
ScaleOutEvaluationCount	Integer	3	The number of consecutive times that the event-triggered task created for scale-out activities meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule.
ScalingGroupId	String	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.
ScalingRuleAri	String	ari:acs:ess:cn-qingdao-env17-d01:134****:scalingRule/eMKWG8SRNb9dBLAjweN****	The unique identifier of the scaling rule.
ScalingRuleId	String	eMKWG8SRNb9dBLAjweN****	The ID of the scaling rule.
ScalingRuleName	String	eMKWG8SRNb9dBLAjweN****	The name of the scaling rule.
ScalingRuleType	String	SimpleScalingRule	The type of the scaling rule. Valid values: • SimpleScalingRule: scales the number of ECS instances based on the values of the AdjustmentType and AdjustmentValue parameters. • TargetTrackingScalingRule: calculates the number of ECS instances that must be scaled and maintains the value of the predefined metric close to the value of the TargetValue parameter. • StepScalingRule: scales ECS instances in steps based on specified thresholds and metric values. • PredictiveScalingRule: analyzes historical monitoring data of the scaling group and uses machine learning to predict the future values of metrics. If you set the ScalingRuleType parameter to PredictiveScalingRule, Auto Scaling creates scheduled tasks to specify the boundary values for the scaling group.

Parameter	Type	Example	Description
StepAdjustments	Array of StepAdjustment		The step adjustments of the step scaling rule.
StepAdjustment			
MetricIntervalLowerBound	Float	1.0	The lower limit that is specified in a step adjustment. Valid values: -9.999999E18 to 9.999999E18.
MetricIntervalUpperBound	Float	5.0	The upper limit that is specified in a step adjustment. Valid values: -9.999999E18 to 9.999999E18.
ScalingAdjustment	Integer	1	The number of ECS instances that are scaled in a step adjustment.
TargetValue	Float	0.125	The target value of the metric.
TotalCount	Integer	1	The total number of scaling rules.

Examples

Sample requests

```
http://ess.example.com/?Action=DescribeScalingRules
&RegionId=cn-qingdao-env17-d01
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DescribeScalingRulesResponse>
      <PageNumber>1</PageNumber>
      <PageSize>50</PageSize>
      <ScalingRules>
            <ScalingRule>
                  <AdjustmentType>QuantityChangeInCapacity</AdjustmentType>
                  <AdjustmentValue>1</AdjustmentValue>
                  <Cooldown>20</Cooldown>
                  <ScalingGroupId>AG6CQdPU8OKdwLjgZcJ****</ScalingGroupId>
                  <ScalingRuleAri>ari:acs:ess:cn-qingdao-env17-d01:134****:scalingRule/eMKW
G8SRNb9dBLAjweN****</ScalingRuleAri>
                  <ScalingRuleId>eMKWG8SRNb9dBLAjweN****</ScalingRuleId>

                  <ScalingRuleName>eMKWG8SRNb9dBLAjweN****</ScalingRuleName>
            </ScalingRule>
      </ScalingRules>
      <TotalCount>1</TotalCount>
      <RequestId>3306A40D-3412-4101-9F19-5F81E3055DAD</RequestId>
</DescribeScalingRulesResponse>
```

`JSON` format

```
{
    "RequestId": "B583BFEF-A779-427A-9B74-262DDD249702",
    "TotalCount": 1,
    "PageNumber": 1,
    "PageSize": 10,
    "ScalingRules": {
        "ScalingRule": [
            {
                "ScalingRuleId": "efcqrZdjlookc0UkE3dA****",
                "ScalingRuleAri": "ari:acs:ess:cn-qingdao-env17-d01:134****:scalingRule/efc
qrZdjlookc0UkE3dA****",
                "Cooldown": 500,
                "ScalingGroupId": "ccMvs9dcZlE5c9CtrwbX****",
                "AdjustmentType": "TotalCapacity",
                "ScalingRuleName": "KFJoxG****",
                "AdjustmentValue": 5
            }
        ]
    }
}
```

6.1.9.4. DeleteScalingRule

Deletes a scaling rule.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DeleteScalingRule	The operation that you want to perform. Set the value to DeleteScalingRule.
ScalingRuleId	String	Yes	eMKWG8SRNb9dBLAjweN****	The ID of the scaling rule that you want to delete. You can call the DescribeScalingRules operation to query the ID of the scaling rule.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=DeleteScalingRule
&ScalingRuleId=eMKWG8SRNb9dBLAjweN****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DeleteScalingRuleResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DeleteScalingRuleResponse>
```

`JSON` format

```
{
  "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description

HTTP status code	Error code	Error message	Description
404	InvalidScalingRuleId.NotFound	The specified scaling rule does not exist.	The error message returned because the specified scaling rule does not exist within the Apsara Stack tenant account.

6.1.10. Scaling tasks

6.1.10.1. ExecuteScalingRule

Executes a scaling rule.

Description

- Before you call this operation, make sure that the following conditions are met:
  - The scaling group is in the Active state.
  - No scaling activities are in progress in the scaling group.
- If no scaling activities in the scaling group are in progress, the operation can trigger scaling activities without the need to wait for the cooldown time to expire.
- A successful call indicates that Auto Scaling accepts the request. However, the scaling activity may still fail. You can check the returned ScalingActivityId parameter to obtain the status of the scaling activity.
- If the number of Elastic Compute Service (ECS) instances that you want to add causes the number of instances in the scaling group to exceed the MaxSize value, the rule Total Capacity = MaxSize takes effect.
- If the number of ECS instances that you want to remove causes the number of instances in the scaling group to be less than the MinSize value, the rule Total Capacity = MinSize takes effect.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	ExecuteScalingRule	The operation that you want to perform. Set the value to ExecuteScalingRule.
ScalingRuleAri	String	Yes	ari:acs:ess:cn-hangzhou-env17-d01:140692647406****:scalingrule/asr-bp1dvirgwkoowxk7****	The unique identifier of the scaling rule.

Parameter	Type	Required	Example	Description
ClientToken	String	No	123e4567-e89b-12d3-a456-426655440000	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see *How to ensure idempotence*.
BreachThreshold	Float	No	1.0	The threshold specified when the step scaling rule is executed. Valid values: -9.999999E18~9.999999E18.
MetricValue	Float	No	1.0	The metric value specified when the step scaling rule is executed. Valid values: -9.999999E18~9.999999E18.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingActivityId	String	asa-bp13o672yeautiil****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.example.com/?Action=ExecuteScalingRule
&ScalingRuleAri=ari:acs:ess:cn-hangzhou-env17-d01:140692647406****:scalingrule/asr-bp1dvirg
wkoowxk7****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<ExecuteScalingRuleResponse>
      <ScalingActivityId>asa-bp13o672yeautiil****</ScalingActivityId>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>
</ExecuteScalingRuleResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "ScalingActivityId": "asa-bp13o672yeautiil****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingRuleAri.NotFound	The specified scaling rule Ari does not exist.	The error message returned because the specified scaling rule does not exist in the Apsara Stack tenant account.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the scaling group to which the specified scaling rule belongs is not in the Active state.
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because a scaling activity is in progress in the scaling group to which the specified scaling rule belongs.
400	InsufficientBalance	Your account does not have enough balance.	This error message appears because your account balance is insufficient.

HTTP status code	Error code	Error message	Description
400	QuotaExceed.Instance	Living instance quota exceeded.	The error message returned because the maximum number of ECS instances in your account has been reached.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the Server Load Balancer (SLB) instance associated with the scaling group to which the specified scaling rule belongs is not in the Active state.
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned because health check is not enabled for the SLB instance that is associated with the scaling group to which the specified scaling rule belongs.
400	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified load balancer does not support this action.	The error message returned because the network type of the ECS instance that is attached to the associated SLB instance is different from the network type of the scaling group.
400	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in specified load balancer are not in the same VPC.	The error message returned because the ECS instance that is attached to the associated SLB instance does not reside in the same VPC as the vSwitch specified by VSwitchID.

HTTP status code	Error code	Error message	Description
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned because the ApsaraDB RDS instance that is associated with the scaling group to which the specified scaling rule belongs is not in the Running state.
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because the maximum number of IP addresses in the whitelist that manages access to the associated ApsaraDB RDS instance has been reached for the scaling group to which the specified scaling rule belongs.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the maximum number of ECS instances associated with the specified security group has been reached.

HTTP status code	Error code	Error message	Description
400	IncorrectCapacity.NoChange	To execute the specified scaling rule, the total capacity will not change.	The error message returned because the number of instances in the scaling group remains unchanged after the scaling rule is executed.
400	QuotaExceeded.ScalingInstance	Scaling instance quota exceeded.	The error message returned because the maximum number of ECS instances that can be scaled has been reached.
400	QuotaExceeded.AfterpayInstance	Living afterpay instance quota exceeded.	The error message returned because the maximum number of pay-as-you-go ECS instances has been reached.
400	ResourceNotAvailable.ECS	The specified region or zone does not offer the specified disk or instance category.	The error message returned because the specified ECS instance type or disk cannot be created in the specified region.
400	ScalingRule.InvalidScalingRuleType	Specific scaling rule type: %s can not be executed.	The error message returned because the specified type of scaling rule cannot be executed.
400	InvalidStepAdjustments.NoStepFound	No adjustment step found for a metric value of: %s.	The error message returned because no matching adjustment steps are found for the specified metric.

HTTP status code	Error code	Error message	Description
400	MissingParameter.MetricValue	Metric value must be specified for StepScalingRule.	The error message returned because no metric value is specified for the step scaling rule.
400	MissingParameter.BreachThreshold	Breach threshold must be specified for StepScalingRule.	The error message returned because no threshold is specified for the step scaling rule.
400	MetricValueBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned because the specified metric value is invalid.
400	BreachThresholdBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned because the specified threshold is invalid.

6.1.10.2. AttachInstances

Adds Elastic Compute Service (ECS) instances to a scaling group.

Description

- Before you add ECS instances to a scaling group, take note of the following items:
  - The ECS instances and the scaling group must reside in the same region.
  - The ECS instances must be in the Running state.
  - The ECS instances are not added to other scaling groups.
  - The ECS instances use the subscription or pay-as-you-go billing method.
  - If the VswitchID parameter is specified for the scaling group, the ECS instances must reside in the same virtual private cloud (VPC) as the specified vSwitch. You cannot add instances that reside in the classic network or other VPCs to the scaling group.
  - If the VswitchID parameter is not specified for the scaling group, the ECS instances must not reside in VPCs.
- You can call this operation to add ECS instances to a scaling group only if the scaling group is in the Active state.
- You can call this operation only if no scaling activities are in progress in the scaling group.
- If no scaling activities are in progress in the scaling group, the operation can trigger scaling activities even before the cooldown period ends.

- A successful call indicates that Auto Scaling accepts the request. However, the scaling activity may still fail. You can obtain the status of a scaling activity by using the value of the ScalingActivityId parameter in the response.
- If the total number of ECS instances in the scaling group is greater than the value that you specify for the MaxSize parameter after you add a specific number of ECS instances to the scaling group, the call fails.
- ECS instances that you manually add to the scaling group are not associated with the active scaling configuration in the scaling group.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	AttachInstances	The operation that you want to perform. Set the value to AttachInstances.
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dloq****	The ID of the scaling group.
InstanceId.N	RepeatList	No	i-28wt4****	The ID of ECS instance N that you want to add to the scaling group. Valid values of N: 1 to 20.
LoadBalancerWeight.N	RepeatList	No	50	The weight of ECS instance N as a Server Load Balancer (SLB) backend server. Valid values of N: 1 to 20. Valid values of this parameter: 1 to 100. Default value: 50.

Parameter	Type	Required	Example	Description
Entrusted	Boolean	No	false	Specifies whether to allow the scaling group to manage the lifecycles of ECS instances that you manually added to the scaling group. Valid values: • true: allows the scaling group to manage the lifecycles of ECS instances. The scaling group uses the same method to manage the lifecycles of ECS instances that are manually added and ECS instances that are automatically created. If you set this parameter to true, Auto Scaling automatically releases the ECS instances when the ECS instances are removed from the scaling group. Auto Scaling does not release the ECS instances that are removed from the scaling group by calling the DetachInstances operation. • false: does not allow the scaling group to manage the lifecycles of ECS instances. If you set this parameter to false, Auto Scaling does not release the ECS instances when the ECS instances are removed from the scaling group. **Note** You cannot use the scaling group to manage subscription ECS instances. Default value: false.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingActivityId	String	asa-bp1crxor24s28xf1****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.example.com/?Action=AttachInstances
&InstanceId.1=i-28wt4****
&ScalingGroupId=asg-bp18p2yfxow2dloq****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<AttachInstancesResponse>
      <ScalingActivityId>asa-bp1crxor24s28xf1****</ScalingActivityId>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</AttachInstancesResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "ScalingActivityId": "asa-bp1crxor24s28xf1****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Apsara Stack tenant account.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the specified scaling group is not in the Active state.

HTTP status code	Error code	Error message	Description
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the specified ECS instance does not exist within the Apsara Stack tenant account.
400	InvalidInstanceId.RegionMismatch	Instance "XXX" and the specified scaling group are not in the same Region.	The error message returned because the specified ECS instance and the scaling group do not reside in the same region.
400	InvalidInstanceId.InstanceTypeMismatch	Instance "XXX" and existing active scaling configurations have different instance types.	The error message returned because the instance type of the specified ECS instance is different from the instance type that is specified in the active scaling configuration.
400	IncorrectInstanceStatus	The current status of instance "XXX" does not support this action.	The error message returned because the specified ECS instance is not in the Running state.
400	InvalidInstanceId.NetworkTypeMismatch	The network type of instance "XXX" does not support this action.	The error message returned because the network type of the specified ECS instance is different from the network type of the scaling group.

HTTP status code	Error code	Error message	Description
400	InvalidInstanceId.VPCMismatch	Instance "XXX" and the specified scaling group are not in the same VPC.	The error message returned because the added ECS instance and the specified scaling group do not reside in the same VPC.
400	InvalidInstanceId.InUse	Instance "XXX" is already attached to another scaling group.	The error message returned because the specified ECS instance has been added to another scaling group.
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because scaling activities are in progress in the specified scaling group.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the SLB instance that is associated with the specified scaling group is not in the Active state.
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned because health check is not enabled for the SLB instance that is associated with the specified scaling group.
400	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified load balancer does not support this action.	The error message returned because the network type of an ECS instance that is attached to the specified SLB instance is different from the network type of the scaling group.

HTTP status code	Error code	Error message	Description
400	InvalidLoadBalancerId.VPCMismatch	he specified virtual switch and the instance in specified load balancer are not in the same VPC.	The error message returned because an ECS instance that is attached to the specified SLB instance does not reside in the same VPC as the vSwitch that is specified by the VswitchID parameter.
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned because the ApsaraDB RDS instance that is associated with the specified scaling group is not in the Running state.
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because the maximum number of IP addresses in the whitelist of the ApsaraDB RDS instance that is associated with the specified scaling group has been reached.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the maximum number of ECS instances that are associated with the specified security group has been reached.

HTTP status code	Error code	Error message	Description
400	IncorrectCapacity.MaxSize	To attach the instances, the total capacity will be greater than the MaxSize.	The error message returned because the total number of ECS instances that is indicated by the Total Capacity parameter exceeds the specified value of the MaxSize parameter after ECS instances are added to the specified scaling group.

6.1.10.3. RemoveInstances

Removes Elastic Compute Service (ECS) instances from a scaling group.

Description

- Before you call this operation, make sure that the following requirements are met:
  - The scaling group is in the Active state.
  - No scaling activity is in progress in the scaling group.
- If no scaling activities in the specified scaling group are in progress, the operation can trigger scaling activities even before the cooldown time expires. The cooldown time is specified by the DefaultCooldown parameter.
- When an automatically created ECS instance is removed from a scaling group, the instance is stopped and released.
- When a manually added ECS instance is removed from a scaling group, the ECS instance is not stopped or released.
- A successful call indicates that Auto Scaling accepts the request. However, the scaling activity may still fail. You can use the value of the ScalingActivityId parameter in the response to check the status of the scaling activity.
- If the difference between the number of existing ECS instances that is specified by the TotalCapacity parameter and the number of ECS instances that you want to remove is less than the value of the MinSize parameter, the call fails.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	RemoveInstances	The operation that you want to perform. Set the value to RemoveInstances.

Parameter	Type	Required	Example	Description
InstanceId.N	RepeatList	Yes	i-1kx0159ruc4x22ek****	The ID of ECS instance N that you want to remove. Valid values of N: 1 to 20.
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.
RemovePolicy	String	No	release	The policy used to remove instances. Valid values: • recycle: The instances are stopped. Note This setting takes effect only if you set the ScalingPolicy parameter to recycle. • release: The instances are released. Default value: release.
DecreaseDesiredCapacity	Boolean	No	true	Specifies whether to adjust the expected number of instances in the scaling group. Valid values: • true: After ECS instances are removed from the scaling group, the expected number of ECS instances in the scaling group decreases. • false: After ECS instances are removed from the scaling group, the expected number of instances in the scaling group remains unchanged. Default value: true.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingActivityId	String	bybj9OcaOT4ucPMbFhcq****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.example.com/?Action=RemoveInstances
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ****
&InstanceId.1=i-28wt4****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<RemoveInstancesResponse>
      <ScalingActivityId>bybj9OcaOT4ucPMbFhcq****</ScalingActivityId>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</RemoveInstancesResponse>
```

`JSON` format

```
{
    "ScalingActivityId":"bybj9OcaOT4ucPMbFhcq****",
    "RequestId":"473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within your Apsara Stack tenant account.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the specified ECS instance does not exist in the scaling group.
400	InvalidParameter	The specified group does not support the specified RemovePolicy.	The error message returned because the specified scaling group does not support the specified value of the RemovePolicy parameter.

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the specified scaling group is not in the Active state.
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because a scaling activity is in progress in the specified scaling group.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the Server Load Balancer (SLB) instance that is associated with the scaling group to which the specified scaling rule belongs is not in the Active state.
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance that is associated with the scaling group to which the specified scaling rule belongs is not in the Running state.

HTTP status code	Error code	Error message	Description
400	IncorrectCapacity.MinSize	To remove the instances, the total capacity will be lesser than the MinSize.	The error message the difference between the number of existing ECS instances that is specified by the TotalCapacity parameter and the number of ECS instances that you want to remove is less than the value of the MinSize parameter.

6.1.10.4. DetachInstances

Removes one or more Elastic Compute Service (ECS) instances from a scaling group.

Description

- After ECS instances are removed from a scaling group, you can call the AttachInstances operation to add the ECS instances that are removed from the scaling group to other scaling groups.
- After you remove an ECS instance by calling the DetachInstances operation, the ECS instance is not stopped or released.
- Before you call the DetachInstances operation, make sure that the following conditions are met:
  - The specified scaling group is enabled.
  - The specified scaling group cannot have a scaling activity in progress.
- If no scaling activities in the scaling group are in progress, the DetachInstances operation can trigger scaling activities without the need to wait for the cooldown period to end.
- A successful call to the DetachInstances operation indicates only that the request for the call is accepted. The scaling activity can be triggered, but may fail to be executed. You can use the ScalingActivityId parameter in the response to query the status of the scaling activity.
- The difference between the number of existing ECS instances in the specified scaling group and the number of ECS instances that you want to remove must be greater than or equal to the value of the MinSize parameter. Otherwise, an error message is returned.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DetachInstances	The operation that you want to perform. Set the value to DetachInstances.

Parameter	Type	Required	Example	Description
InstanceId.N	RepeatList	Yes	i-28wt4****	The ID of ECS instance N that you want to remove from the scaling group. Valid values of N: 1 to 20.
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.
DecreaseDesiredCapacity	Boolean	No	true	Specifies whether to adjust the expected number of instances in the scaling group. Valid values: • true: After ECS instances are removed from a scaling group, the expected number of ECS instances in the scaling group decreases. • false: After ECS instances are removed from a scaling group, the expected number of instances in the scaling group remains unchanged. Default value: true.
DetachOption	String	No	both	Specifies whether to remove the ECS instance from the default server groups and vServer groups of the Server Load Balancer (SLB) instances that are associated with the scaling group, and whether to remove the private IP address of the ECS instance from the whitelist of the ApsaraDB RDS instances that are associated with the scaling group. If you set this parameter to both, the ECS instance is removed from the default sever groups and vServer groups of the associated SLB instances, and the private IP address of the ECS instance is removed from the whitelist of the associated ApsaraDB RDS instances.

Response parameters

Parameter	Type	Example	Description

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingActivityId	String	asa-****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.example.com/?Action=DetachInstances
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ****
&InstanceId.1=i-28wt4****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DetachInstancesResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
      <ScalingActivityId>asa-****</ScalingActivityId>
</DetachInstancesResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "ScalingActivityId": "asa-****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the specified scaling group is in the Disabled state.

HTTP status code	Error code	Error message	Description
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because a scaling activity in the specified scaling group is in progress.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified SLB instance is not in the Active state.
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance is not in the Running state.
400	IncorrectCapacity.MinSize	To remove the instances, the total capacity will be lesser than the MinSize.	The error message returned because the difference between the number of ECS instances in the specified scaling group and the number of ECS instances that you want to remove is less than the value of the MinSize parameter.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because you are not authorized to call the DetachInstances operation.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist.

HTTP status code	Error code	Error message	Description
404	InvalidInstanceId.NotFound	Instance “XXX” does not exist.	The error message returned because the specified ECS instance does not exist.

6.1.11. Scheduled tasks

6.1.11.1. CreateScheduledTask

Creates a scheduled task.

Description

- You can create up to 20 scheduled tasks within an Apsara Stack tenant account.
- If a scheduled task fails to trigger a scaling rule due to an ongoing scaling activity in a scaling group or because the scaling group is disabled, the scheduled task is automatically retried during the period that is specified by the LaunchExpirationTime parameter. If the scheduled task still fails to trigger a scaling rule after the period ends, the task is automatically skipped.
- If multiple tasks are scheduled at similar points in time to execute scaling rules in the same scaling group, the earliest task executes the scaling activity first. Other tasks may attempt to execute their scaling rules during their launch expiration time, but only one scaling activity can be triggered in a scaling group at a time. If the previous scaling activity is complete and another scheduled task executes the scaling rule during the period that is specified by the LaunchExpirationTime parameter, the scaling rule in the scheduled task is executed, and a new scaling activity is triggered.
- A scheduled task supports two scaling methods.
  - If you specify the `ScheduledAction` parameter, the scaling method of the scheduled task is to specify an existing scaling rule that you want to execute when the scheduled task is triggered.
  - If you specify the `ScalingGroupId` parameter, the scaling method of the scheduled task is to specify the minimum number, maximum number, or expected number of instances in the scaling group for which you create the scheduled task.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	CreateScheduledTask	The operation that you want to perform. Set the value to CreateScheduledTask.
RegionId	String	Yes	cn-qingdao-env17-d01	The region ID of the scheduled task.

Parameter	Type	Required	Example	Description
ScheduledTaskName	String	No	scheduled****	The display 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. The name of the scheduled task must be unique in the region and within the Apsara Stack tenant account. If you do not specify this parameter, the value of the ScheduledScalingTaskId parameter is used.
Description	String	No	Test scheduled task.	The description of the scheduled task. The description must be 2 to 200 characters in length.
ScheduledAction	String	No	ari:acs:ess:cn-qingdao-env17-d01:134****:scalingRule/cCBpdYdQuBe2cUxOdu6****	The scaling rule that you want to execute when the scheduled task is triggered. Specify the unique identifier of the scaling rule. If you specify the `ScheduledAction` parameter, you must select an existing scaling rule for the scheduled task. **Note** You cannot specify the ScheduledAction and ScalingGroupId parameters at the same time.
RecurrenceEndTime	String	No	2014-08-17T16:55Z	The end time of the scheduled task. Specify the time in the ISO 8601 standard in the YYYY-MM-DDThh:mmZ format. The time must be in UTC. You cannot enter a time point later than 365 days from the scheduled task creation. You must specify the RecurrenceType, RecurrenceValue, and RecurrenceEndTime parameters.

Parameter	Type	Required	Example	Description
LaunchTime	String	No	2014-08-17T16:52Z	The point in time at which the scheduled task is triggered. Specify the time in the ISO 8601 standard in the YYYY-MM-DDThh:mmZ format. The time must be in UTC. You cannot enter a time point later than 90 days from the scheduled task creation. • If you specify the RecurrenceType parameter, the task is repeatedly executed at the point in time that is specified by the LaunchTime parameter. • If you do not specify the RecurrenceType parameter, the task is executed only once at the point in time that is specified by the LaunchTime parameter.
RecurrenceType	String	No	Daily	The interval at which the scheduled task is repeated. Valid values: • Daily: The scheduled task is executed once every specified number of days. • Weekly: The scheduled task is executed on each specified day of the week. • Monthly: The scheduled task is executed on each specified day of the month. • Cron: The scheduled task is executed based on the specified cron expression. You must specify the RecurrenceType, RecurrenceValue, and RecurrenceEndTime parameters.

Parameter	Type	Required	Example	Description
RecurrenceValue	String	No	1	The number of recurrences of the scheduled task. • Daily: If you set the RecurrenceType parameter to Daily, you can specify only one value. Valid values: 1 to 31. • Weekly: If you set the RecurrenceType parameter to Weekly, you can specify multiple values. Separate the values with commas (,). The values that correspond to Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday are 0, 1, 2, 3, 4, 5, and 6. • Monthly: If you set the RecurrenceType parameter to Monthly, you can specify two values in the A-B format. Valid values of A and B: 1 to 31. B must be greater than or equal to A. • Cron: If you set the RecurrenceType parameter to Cron, you can specify a cron expression. A cron expression is written in UTC time and consists of five fields: minute, hour, day, month, and week. The expression can contain the letters L and W and the following wildcard characters: commas (,), question marks (?), hyphens (-), asterisks (*), number signs (#), and forward slashes (/). You must specify the RecurrenceType, RecurrenceValue, and RecurrenceEndTime parameters.
TaskEnabled	Boolean	No	true	Specifies whether to enable the scheduled task. • true • false Default value: true.
LaunchExpirationTime	Integer	No	600	The period of time during which a failed scheduled task is retried. Unit: seconds. Valid values: 0 to 21600. Default value: 600.

Parameter	Type	Required	Example	Description
MinValue	Integer	No	0	The minimum number of instances in the scaling group if the scaling method of the scheduled task is to specify the number of instances.
MaxValue	Integer	No	10	The maximum number of instances in the scaling group if the scaling method of the scheduled task is to specify the number of instances.
DesiredCapacity	Integer	No	10	The expected number of instances in the scaling group if the scaling method of the scheduled task is to specify the number of instances. **Note** You must specify the DesiredCapacity parameter when you create the scaling group.
ScalingGroupIdInvalid	String	No	asg-bp18p2yfxow2dloq****	The ID of the scaling group whose number of instances is changed when the scheduled task is triggered. If you specify the ScalingGroupId parameter, the scaling method of the scheduled task is to specify the number of instance in a scaling group. You must specify at least one of the MinValue, MaxValue, and DesiredCapacity parameters. **Note** You cannot specify the ScheduledAction and ScalingGroupId parameters at the same time.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3****	The ID of the request.

Parameter	Type	Example	Description
ScheduledTaskId	String	edRtShc57WGXdt8TlPbr****	The globally unique ID of the scheduled task. The globally unique ID is generated by the system.

Examples

Sample requests

```
http://ess.example.com/?Action=CreateScheduledTask
&RegionId=cn-qingdao-env17-d01
&ScheduledAction=ari:acs:ess:cn-qingdao-env17-d01:134****:scalingRule/cCBpdYdQuBe2cUxOdu6****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<CreateScheduledTaskResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>
      <ScheduledTaskId>edRtShc57WGXdt8TlPbr****</ScheduledTaskId>
</CreateScheduledTaskResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3****",
    "ScheduledTaskId": "edRtShc57WGXdt8TlPbr****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidRegionId.NotFound	The specified region does not exist.	The error message returned because the specified region does not exist.
400	InvalidScheduledTaskName.Duplicate	The specified value of parameter ScheduledTaskName is duplicated.	The error message returned because the specified name of the scheduled task already exists.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded.ScheduledTask	Scheduled task quota exceeded.	The error message returned because the maximum number of scheduled tasks has been reached.
400	ScheduledAction.RegionMismatch	The specified scheduled task and the specified scheduled action are not in the same Region.	The error message returned because the specified scaling rule and the specified scheduled task do not reside in the same region.

6.1.11.2. ModifyScheduledTask

Modifies a scheduled task.

Description

A scheduled task supports two scaling methods.

- If you specify the `ScheduledAction` parameter, the scaling method of the scheduled task is to specify an existing scaling rule that you want to execute when the scheduled task is triggered.
- If you specify the `ScalingGroupId` parameter, the scaling method of the scheduled task is to specify the minimum number, maximum number, or expected number of instances in the scaling group for which you create the scheduled task.

> **Note** You cannot specify the ScheduledAction and ScalingGroupId parameters at the same time.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	ModifyScheduledTask	The operation that you want to perform. Set the value to ModifyScheduledTask.
ScheduledTaskId	String	Yes	edRtShc57WGXdt8TlPbr****	The ID of the scheduled task.

Parameter	Type	Required	Example	Description
ScheduledTaskName	String	No	scheduled****	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. The name of the scheduled task must be unique in the region and within the Apsara Stack tenant account.
Description	String	No	fortest	The description of the scheduled task. The description must be 2 to 200 characters in length.
ScheduledAction	String	No	ari:acs:ess:cn-qingdao-env17-d01:134****:scalingRule/cCBpdYdQuBe2cUxOdu6****	The scaling rule that you want to execute when the scheduled task is triggered. Specify the unique identifier of the scaling rule. If you specify the `ScheduledAction` parameter, you must select an existing scaling rule for the scheduled task. **Note** You cannot specify the ScheduledAction and ScalingGroupId parameters at the same time.
RecurrenceEndTime	String	No	2014-08-20T16:55Z	The end time of the scheduled task. Specify the time in the ISO 8601 standard in the `yyyy-MM-ddTHH:mm:ssZ` format. The time must be in UTC. You cannot enter a time point later than 365 days from the scheduled task modification. After the scheduled task is modified, you must specify valid values for the RecurrenceType, RecurrenceValue, and RecurrenceEndTime parameters.

Parameter	Type	Required	Example	Description
LaunchTime	String	No	2014-08-18T10:52Z	The point in time at which the scheduled task is triggered. Specify the time in the ISO 8601 standard in the `yyyy-MM-ddTHH:mm:ssZ` format. The time must be in UTC. You cannot enter a time point later than 90 days from the scheduled task modification. • If you specify the RecurrenceType parameter, the task is repeatedly executed at the point in time that is specified by the LaunchTime parameter. • If you do not specify the RecurrenceType parameter, the task is executed only once at the point in time that is specified by the LaunchTime parameter.
RecurrenceType	String	No	Daily	The interval at which the scheduled task is repeated. Valid values: • Daily: The scheduled task is executed once every specified number of days. • Weekly: The scheduled task is executed on each specified day of the week. • Monthly: The scheduled task is executed on each specified day of the month. • Cron: The scheduled task is executed based on the specified cron expression. After the scheduled task is modified, the RecurrenceType, RecurrenceValue, and RecurrenceEndTime parameters must be valid.

Parameter	Type	Required	Example	Description
RecurrenceValue	String	No	2	The number of recurrences of the scheduled task. • Daily: If you set the RecurrenceType parameter to Daily, you can specify only one value. Valid values: 1 to 31. • Weekly: If you set the RecurrenceType parameter to Weekly, you can specify multiple values. Separate the values with commas (,). For example, the values that correspond to Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday are `0, 1, 2, 3, 4, 5, and 6`. • Monthly: If you set the RecurrenceType parameter to Monthly, you can specify two values in the A-B format. Valid values of A and B: 1 to 31. B must be greater than or equal to A. • Cron: If you set the RecurrenceType parameter to Cron, you can specify a cron expression. A cron expression is written in UTC time and consists of five fields: minute, hour, day, month, and week. The expression can contain the letters L and W and the following wildcard characters: commas (,), question marks (?), hyphens (-), asterisks (*), number signs (#), and forward slashes (/). After the scheduled task is modified, the RecurrenceType, RecurrenceValue, and RecurrenceEndTime parameters must be valid.
TaskEnabled	Boolean	No	true	Specifies whether to enable the scheduled task. • true • false
LaunchExpiration Time	Integer	No	600	The period of time during which the failed scheduled task is retried. Unit: seconds. Valid values: 0 to 21600.

Parameter	Type	Required	Example	Description
MinValue	Integer	No	0	The minimum number of instances in the scaling group if the scaling method of the scheduled task is to specify the number of instances.
MaxValue	Integer	No	10	The maximum number of instances in the scaling group if the scaling method of the scheduled task is to specify the number of instances.
DesiredCapacity	Integer	No	10	The expected number of instances in the scaling group if the scaling method of the scheduled task is to specify the number of instances. **Note** You must specify the DesiredCapacity parameter when you create the scaling group.
ScalingGroupIdInvalid	String	No	asg-bp18p2yfxow2dloq****	The ID of the scaling group whose number of instances is changed when the scheduled task is triggered. If you specify the ScalingGroupId parameter, the scaling method of the scheduled task is to specify a number of instances in the scaling group. In this case, you must specify at least one of the MinValue, MaxValue, and DesiredCapacity parameters. **Note** You cannot specify the ScheduledAction and ScalingGroupId parameters at the same time.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3****	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=ModifyScheduledTask
&ScheduledTaskId=edRtShc57WGXdt8TlPbr****
&LaunchTime=2014-08-18T10:52Z
&RecurrenceEndTime=2014-08-20T16:55Z
&<Common request parameters>
```

Sample success responses

`XML` format

```
<ModifyScheduledTaskResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>
</ModifyScheduledTaskResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3****"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScheduledTaskId.NotFound	The specified scheduled task does not exist.	The error message returned because the specified scheduled task does not exist within the Apsara Stack tenant account.
400	InvalidScheduledTaskName.Duplicate	The specified value of parameter `ScheduledTaskName` is duplicated.	The error message returned because the specified name of the scheduled task already exists.
400	ScheduledAction.RegionMismatch	The specified scheduled task and the specified scheduled action are not in the same Region.	The error message returned because the specified scaling rule and the specified scheduled task do not reside in the same region.

6.1.11.3. DescribeScheduledTasks

Queries scheduled tasks.

Description

You can query scheduled tasks based on scaling rule, task ID, and task name.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DescribeScheduledTasks	The operation that you want to perform. Set the value to DescribeScheduledTasks.
RegionId	String	Yes	cn-qingdao-env17-d01	The region ID of the scaling group to which the scheduled task belongs.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	50	The number of entries to return on each page. Maximum value: 50. Default value: 10.
ScheduledAction.N	RepeatList	No	ari:acs:ess:cn-qingdao-env17-d01:101582792734****:scalingrule/asr-1kx06vd223809dlb****	The operation that you want to perform when scheduled task N is triggered. Valid values of N: 1 to 20.
ScheduledTaskId.N	RepeatList	No	edRtShc57WGXdt8TlPbr****	The ID of scheduled task N that you want to query. Valid values of N: 1 to 20.
ScheduledTaskName.N	RepeatList	No	scheduled****	The name of scheduled task N that you want to query. Valid values of N: 1 to 20.
ScalingGroupId	String	No	asg-bp1bo5tca4m56nap****	The ID of the scaling group to which the scheduled task belongs.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScheduledTasks	Array of ScheduledTask		Details of the scheduled tasks.
ScheduledTask			
Description	String	fortest	The description of the scheduled task.
DesiredCapacity	Integer	10	The expected number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.
LaunchExpirationTime	Integer	600	The time period during which the failed scheduled task is retried. Unit: seconds. Valid values: 0 to 21600.
LaunchTime	String	2014-08-18T10:52Z	The point in time at which the scheduled task was triggered.
MaxValue	Integer	10	The maximum number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.
MinValue	Integer	0	The minimum number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.

Parameter	Type	Example	Description
RecurrenceEndTime	String	2014-08-20T16:55Z	The end time of the scheduled task.
RecurrenceType	String	Daily	The interval during which the scheduled task was repeatedly executed.
RecurrenceValue	String	1	The number of recurrences of the scheduled task.
ScalingGroupId	String	asg-bp1bo5tca4m56nap****	The ID of the scaling group to which the scheduled task belongs.
ScheduledAction	String	ari:acs:ess:cn-qingdao-env17-d01:134****:scalingrule/cCBpdYdQuBe2cUxOdu6****	The scaling rule that was executed when the scheduled task was triggered. **Note** You cannot specify the `ScheduledAction` and `ScalingGroupId` parameters at the same time.
ScheduledTaskId	String	edRtShc57WGXdt8TlPbr****	The ID of the scheduled task.
ScheduledTaskName	String	edRtShc57WGXdt8TlPbr****	The name of the scheduled task.
TaskEnabled	Boolean	true	Indicates whether the scheduled task feature is enabled.
TotalCount	Integer	1	The total number of scheduled tasks.

Examples

Sample requests

```
http://ess.example.com/?Action=DescribeScheduledTasks
&RegionId=cn-qingdao-env17-d01
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DescribeScheduledTasksResponse>
      <ScheduledTasks>
            <ScheduledTask>
                  <LaunchExpirationTime>600</LaunchExpirationTime>
                  <RecurrenceType>Daily</RecurrenceType>
                  <RecurrenceEndTime>2020-03-02T00:00Z</RecurrenceEndTime>
                  <ScalingGroupId>asg-****</ScalingGroupId>
                  <TaskEnabled>false</TaskEnabled>
                  <ScheduledTaskName>demo1</ScheduledTaskName>
                  <MaxValue>1</MaxValue>
                  <LaunchTime>2020-03-01T00:00Z</LaunchTime>
                  <ScheduledTaskId>cV0bJbAXpipdbFI1NbM****</ScheduledTaskId>
                  <RecurrenceValue>31</RecurrenceValue>
            </ScheduledTask>
            <ScheduledTask>
                  <LaunchExpirationTime>600</LaunchExpirationTime>
                  <MinValue>1</MinValue>
                  <RecurrenceType>Daily</RecurrenceType>
                  <RecurrenceEndTime>2020-02-27T00:00Z</RecurrenceEndTime>
                  <ScalingGroupId>asg-****</ScalingGroupId>
                  <TaskEnabled>true</TaskEnabled>
                  <ScheduledTaskName>demo2</ScheduledTaskName>
                  <LaunchTime>2020-02-26T00:00Z</LaunchTime>
                  <ScheduledTaskId>cD3G3pesE65NcSTmkld1****</ScheduledTaskId>
                  <RecurrenceValue>31</RecurrenceValue>
            </ScheduledTask>
      </ScheduledTasks>
      <PageNumber>1</PageNumber>
      <TotalCount>34</TotalCount>
      <PageSize>10</PageSize>
      <RequestId>4CF33369-7318-4BD5-BE1C-A776BA2C6627</RequestId>
</DescribeScheduledTasksResponse>
```

`JSON` format

```
{
 "ScheduledTasks": {
  "ScheduledTask": [
   {
    "LaunchExpirationTime": 600,
    "RecurrenceType": "Daily",
    "RecurrenceEndTime": "2020-03-02T00:00Z",
    "ScalingGroupId": "asg-****",
    "TaskEnabled": false,
    "ScheduledTaskName": "demo1",
    "MaxValue": 1,
    "LaunchTime": "2020-03-01T00:00Z",
    "ScheduledTaskId": "cV0bJbAXpipdbFI1NbM****",
    "RecurrenceValue": "31"
   },
   {
    "LaunchExpirationTime": 600,
    "MinValue": 1,
    "RecurrenceType": "Daily",
    "RecurrenceEndTime": "2020-02-27T00:00Z",
    "ScalingGroupId": "asg-****",
    "TaskEnabled": true,
    "ScheduledTaskName": "demo2",
    "LaunchTime": "2020-02-26T00:00Z",
    "ScheduledTaskId": "cD3G3pesE65NcSTmkld1****",
    "RecurrenceValue": "31"
   }
  ]
 },
 "PageNumber": 1,
 "TotalCount": 34,
 "PageSize": 10,
 "RequestId": "4CF33369-7318-4BD5-BE1C-A776BA2C6627"
}
```


6.1.11.4. DeleteScheduledTask

Deletes scheduled tasks.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DeleteScheduledTask	The operation that you want to perform. Set the value to DeleteScheduledTask.

Parameter	Type	Required	Example	Description
ScheduledTaskId	String	Yes	edRtShc57WGXdt8TlPbr****	The ID of the scheduled task that you want to delete. You can call the DescribeScheduledTasks operation to query the ID of the scheduled task.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http://ess.example.com/?Action=DeleteScheduledTask
&ScheduledTaskId=edRtShc57WGXdt8TlPbr****
&<Common request parameters>
```

Sample success responses

`XML` format

```
<DeleteScheduledTaskResponse>
      <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DeleteScheduledTaskResponse>
```

`JSON` format

```
{
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

HTTP status code	Error code	Error message	Description
404	InvalidScheduledTaskId.NotFound	The specified scheduled task does not exist.	The error message returned because the specified scheduled task does not exist within the Apsara Stack tenant account.

6.1.12. Error codes

6.1.12.1. Client errors

This topic describes the details of errors caused by clients.

Description

HTTP	Error code	Error message	Description
400	MissingParameter	The input parameter `<parameter name>` that is mandatory for processing this request is not supplied.	The error message returned because a required parameter is not specified.
400	InvalidParameter	The specified value of parameter `<parameter name>` is not valid.	The error message returned because the specified parameter is invalid.
400	UnsupportedOperation	The specified action is not supported.	The error message returned because the specified API operation is invalid.
400	NoSuchVersion	The specified version does not exist.	The error message returned because the specified version is invalid.
400	Throttling	Request was denied due to request throttling.	The error message returned because the specified operation is denied by the throttling system.
400	InvalidAccessKeyId.NotFound	The Access Key ID provided does not exist in our records.	The error message returned because the specified AccessKey pair is invalid.
403	Forbidden	Users are not authorized to operate on the specified resource.	The error message returned because the specified operation is forbidden.
403	Forbidden.RiskControl	This operation is forbidden by Aliyun Risk Control system.	The error message returned because the specified operation is forbidden by the risk control system.

HTTP	Error code	Error message	Description
403	SignatureDoesNotMatch	The signature we calculated does not match the one you provided.	The error message returned because the specified signature is invalid.
403	Forbidden.Unsubscribed	Do not have permission to access this API.	The error message returned because you are not authorized to call Auto Scaling API operations.
403	Forbidden.UserVerification	Your user account is not verified by Aliyun.	The error message returned because the real-name verification is not complete.
400	ResourceNotAvailable	Resource you requested is not available in this region or zone.	The error message returned because Auto Scaling is unavailable in the specified region.

6.1.12.2. Server errors

This topic describes the details of errors caused by servers.

Description

HTTP status code	Error code	Error message	Description
500	InternalError	The request processing has failed due to some unknown error, exception or failure.	The error message returned because the server fails to process the request.
503	ServiceUnavailable	The request has failed due to a temporary failure of the server.	The error message returned because the server fails to process this request.

6.1.13. Ensure idempotence

This topic describes how to ensure that requests are idempotent by using ClientToken.

If a request times out or an internal server error occurs when the ExecuteScalingRule operation is called to create or release ECS instances, the client may attempt to resend the request. In this case, the client provides the ClientToken parameter to prevent the server from creating or releasing more instances than intended. The ClientToken parameter also ensures that requests are idempotent. The ClientToken value is a unique and case-sensitive string that is generated by the client. The token can only contain ASCII characters and cannot exceed 64 characters in length.

If the same ClientToken value is used to call the CreateInstance operation, the server returns identical request results with identical ScalingActivityId values. Therefore, when the request is resent due to an error, the same ClientToken value can ensure that only one scaling activity is created and the corresponding ScalingActivityId value is obtained.

If you provide a used ClientToken value and other request parameters are different, Auto Scaling returns the `IdempotentParameterMismatch` error code. Note: You must specify the SignatureNonce, Timestamp, and Signature parameters again before the request is resent because Auto Scaling uses SignatureNonce to prevent replay attacks and uses Timestamp to mark the time of each request. Therefore, you must provide different SignatureNonce and Timestamp values. This also produces a different Signature value.

Typically, you only need to configure the client to resend a request in the case of Error 503 (ServiceUnAvailable) or no response. If Error 200 is returned, a retry can obtain the same result as that of the previous request, but the server status is not affected. If Error 4xx or 500 is returned, the request cannot be retried unless the error message states `Try it later`.

7.Appendixes

7.1. Call the ASAPI gateway

You can use the Apsara Stack API (ASAPI) gateway to call API operations of most Apsara Stack services. This topic describes how to call API operations by using the ASAPI gateway and provides links to references.

Flowchart

Procedure

1. Obtain ASAPI SDK.

   Before you use ASAPI to call an API operation, you must obtain ASAPI SDK. For more information, see Obtain ASAPI SDK.

2. Obtain the endpoint of ASAPI. For more information, see Obtain the endpoint of ASAPI .

3. Obtain an AccessKey pair.

   Obtain the AccessKey pair that is required when you call an API operation.

   > **Note**
   >
   > We recommend that you use the AccessKey pair of a personal account to call the API operations of the Apsara Uni-manager Management Console and cloud services.

4. Obtain the region ID.

   Obtain the region ID that is required when you call an API operation.

5. Obtain the organization ID and resource set ID.

   You must call an API operation to obtain an organization ID. For more information about how to obtain organization IDs, see the *GetOrganizationList* section in *Apsara Uni-manager Management Console Developer Guide*.

   You must call an API operation to obtain a resource set ID. For more information about how to obtain resource set IDs, see the *ListResourceGroup* section in *Apsara Uni-manager Management Console Developer Guide*.

   > **Note**
   >
   > When you call an API operation, you can specify the organization ID and resource set ID by using the <x-acs-organizationid> and <x-acs-resourcegroupid> parameters. This allows you to query specified resources.
   >
   > - If you do not set the <x-acs-organizationid> parameter, the organization to which the current user belongs is used by default.
   > - If you do not set the <x-acs-resourcegroupid> parameter, it is empty by default. If you specify a resource set ID, you must also specify an organization ID.

6. Refer to API references.

   Obtain the developer guide of the cloud service that you want to use. You can learn the features of API operations and how to call the API operations. For more information, see Use developer guides.

7. Use ASAPI SDK to call API operations.

   After you complete the preceding preparations, you can use ASAPI SDK to call API operations. For more information about how to use SDKs for different languages, see the relevant topics.

   The following example shows how to call the DescribeInstances operation of Elastic Compute Service (ECS) by using SDK for Java:

```
public static void main(String[] args) {
    ASClient client = new ASClient();
    // Specify the caller. The caller is recorded in logs.
    client.setSdkSource("autoTest");

    // Configure request parameters.
    Map<String, Object> parameters = new HashMap<String, Object>();;

    parameters.put("Product", "Ecs");// Specify the POP code, which is the same as the
POP code in the corresponding API directory.

    parameters.put("Action", "DescribeInstances"); // Specify the name of the API opera
tion. Use the same case as the case in the corresponding API directory.

    parameters.put("Version", "2014-05-26"); // Specify the API version.

    // Specify the region ID that you obtained.
    parameters.put("RegionId", "xxx");

    // Specify the AccessKey ID and AccessKey secret that you obtained.
    parameters.put("AccessKeyId", "xxx");
    parameters.put("AccessKeySecret", "xxx");

    // Add the parameters based on the API documentation. Take note that the parameters
are case-sensitive.
    parameters.put("PageNumber",1);
    parameters.put("PageSize",10);

    Map<String,String> headers=new HashMap<>();
    // Specify the organization ID in the Apsara Uni-manager Management Console.
    headers.put("x-acs-organizationid","3");
    // Specify the resource set ID in the Apsara Uni-manager Management Console.
    //headers.put("x-acs-resourcegroupid", "<your-resourcegroupid>");


    // Specify the endpoint of ASAPI.
    String endpoint="https://public.asapi.xxx.xxx/asapi/v3";

    // Initiate a request.
    String result = client.doPost(endpoint,headers, parameters);

    // Display the results.
    System.out.println(JSONObject.toJSONString(JSONObject.parseObject(result), Serializ
erFeature.PrettyFormat));
}
```

7.2. Use the ASAPI gateway to make API calls

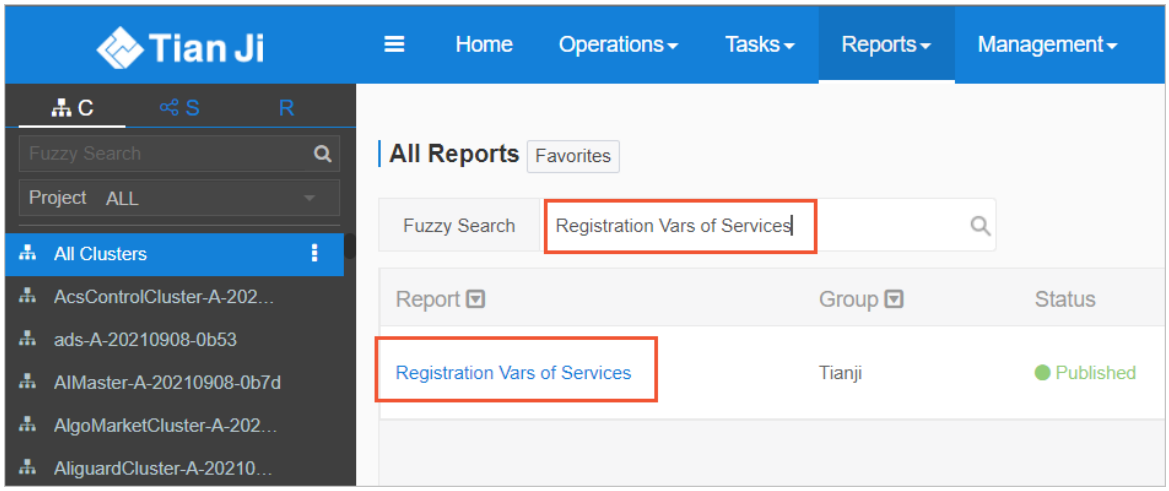
7.2.1. Obtain an endpoint of ASAPI

This topic describes three methods that you can use to obtain an endpoint of Apsara Stack API (ASAPI) and how to verify the endpoint.

Methods

Method 1: Use Registration Vars of Services to obtain an endpoint

1. Log on to the Apsara Infrastructure Management console.
   i. Log on to the Apsara Uni-manager Operations Console.
   ii. In the top navigation bar, click **Products**.
   iii. Choose **Platforms** > **Apsara Infrastructure Management Framework**.
2. In the left-side navigation pane, click **Reports**.
3. On the **All Reports** page, search for `Registration Vars of Services` and click the report name.



4. On the **Registration Vars of Services** page, click the icon next to **Service** and search for `asapi`.
5. Find the asapi service, right-click in the **Service Registration** column, and then select **Show More** from the shortcut menu.

   In the **Details** dialog box, obtain the value of **asapi.public.endpoint**.
6. Generate an endpoint of ASAPI based on the following format:

```
https://<asapi.public.endpoint>/asapi/v3
```

Method 2: Use cluster resources to generate an endpoint

1. Log on to the Apsara Infrastructure Management console.
   i. Log on to the Apsara Uni-manager Operations Console.
   ii. In the top navigation bar, click **Products**.
   iii. Choose **Platforms** > **Apsara Infrastructure Management Framework**.
2. Go to the Clusters page.
   i. In the left-side navigation pane, choose **Operations** > **Cluster Operations**.

ii. In the **Cluster** search box, enter `ascm`.

iii. Click **Operations** in the **Actions** column.

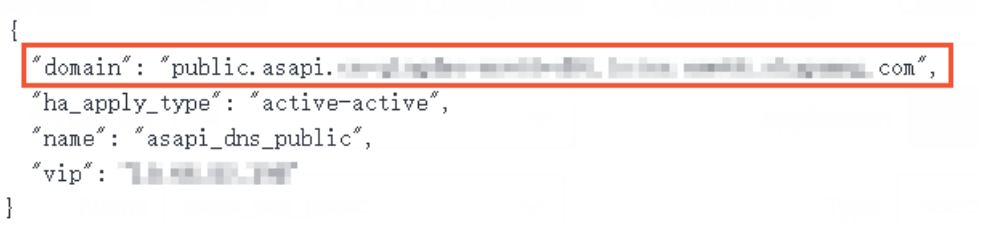
3. Obtain the value of the domain parameter.

i. Click the **Cluster Resources** tab.

ii. In the **Name** field, search for `asapi_dns_public`.

iii. Click **Details** in the **Application Parameter** column.

In the **Application Parameter** message, obtain the value of the **domain** parameter.



4. Generate an endpoint of ASAPI based on the following format:

```
https://<domain>/asapi/v3
```

Method 3: Concatenate strings to generate an endpoint

You can also concatenate strings to generate an endpoint based on the following format:

```
https://public.asapi.<region>.<internet-domain>/asapi/v3
```

> **Note**
>
> 1. Replace <region> with the actual region ID.
> 2. Replace <internet-domain> with the root domain of your Apsara Stack service. To obtain the root domain, contact Apsara Stack O&M personnel. Take note of this point: Use the internet-domain parameter instead of the intranet-domain parameter.

If the **region** value is `cn-****-****-d01` and the **internet-domain** value is `inter.****.****.com`, the following endpoint is used:

```
https://public.asapi.cn-****-****-d01.inter.****.****.com/asapi/v3
```

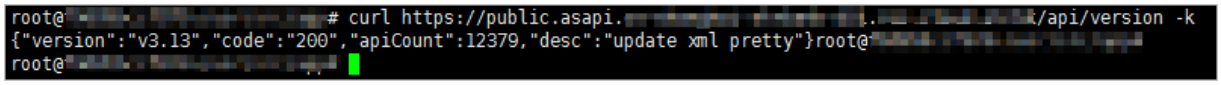
Verify the endpoint

Enter the following code. If you cannot access the endpoint, the endpoint is unavailable. You can use the historical endpoint.

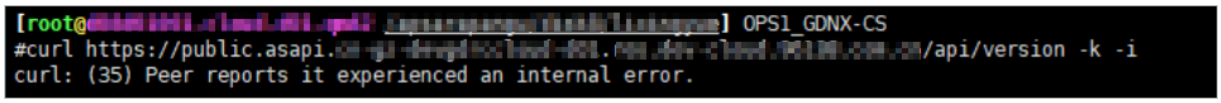
```
curl https://public.asapi.${global:region}.${global:internet-domain}/api/version -k
```

> **Note**
>
> `${global:region}` and `${global:internet-domain}` are variables. You must replace them with the actual values that you obtained.

Available endpoint example:



Unavailable endpoint example:



Historical endpoint

Only the HTTP protocol is supported. The historical endpoint is suffixed with `intranet-domain` and in the following format:

```
http://server.asapi.${global:region}.${global:intranet-domain}/asapi/v3
```

> **Note**
>
> `${global:region}` and `${global:intranet-domain}` are variables. You must replace them with the actual values that you obtained.

Procedure

To obtain the actual values of the variables ${global:region}, ${global:internet-domain}, and ${global:intranet-domain}, perform the following operations:

1. In the address bar, enter the *region-id*.aso.*intranet-domain-id*.com URL and press the Enter key.

   > **Note**
   >
   > You can select a language from the drop-down list in the upper-right corner of the page.

2. Enter your username and password.

   > **Note**
   >
   > Obtain the username and password used to log on to the Apsara Uni-manager Operations Console from the deployment personnel or an administrator.

   When you log on to the Apsara Uni-manager Operations Console for the first time, you must change the password of your username.

   To enhance security, your password must meet the following requirements:

   - The password contains uppercase and lowercase letters.

- The password contains digits.
- The password contains special characters such as exclamation points (!), at signs (@), number signs (#), dollar signs ($), and percent signs (%).
- The password is 10 to 20 characters in length.

3. Click **Log On** to log on to the **Apsara Uni-manager Operations Console**.
4. In the top navigation bar, choose **Products** > **Platforms** > **Apsara Infrastructure Management Framework** to go to Apsara Infrastructure Management.
5. In the left-side navigation pane, choose **Operations** > **Cluster Operations**.
6. On the **Clusters** page, enter **ascm-A** in the Cluster search box to find the cluster. Click the cluster name.
7. On the **Cluster Dashboard** page, click **Operations Menu** > **Configuration Files**.
8. In the **File List** window on the left side of the **Cluster Configuration** page, click **kv.conf**, click the file content in the right window, and then press Ctrl + F. Search for region, internet-domain, and intranet-domain.

   Replace `${global:region}` with the obtained region value, `${global:internet-domain}` with the obtained internet-domain value, and `${global:intranet-domain}` with the obtained intranet-domain value.

   > **Note**
   >
   > `${global:}` is not included in the endpoint of ASAPI that you can use.

7.2.2. Generate an endpoint for ASAPI

This topic describes how to generate an endpoint for ASAPI.

1. Log on to the OPS1 server.
2. Run the following command to obtain three different domain names:

   ```
   kubectl get ingress -n ascm |grep asapiserver
   ```

3. Record the domain name that starts with `public` and use the domain name as the value of `domain` in the next step.

4. Generate an endpoint for ASAPI in the following format:

   ```
   https://<domain>/asapi/v3
   ```

7.2.3. Obtain Apsara Stack SDKs

Apsara Stack API (ASAPI) is a unified portal provided by Apsara Stack. You can use ASAPI to connect to all Apsara Stack services. Apsara Stack provides SDKs for API calls.

Public environment

If you are using the public environment, you can go to the Apsara Stack documentation page and download the Apsara Stack SDKs in the **Other Resources** section at the lower part of the page.

Private environment

If you are using the private environment, you must switch to the public environment and then go to the Apsara Stack documentation page. At the lower part of the page, download the Apsara Stack SDKs in the **Other Resources** section.

Apsara Stack SDK for Java

greopId	artifactId	Version
com.alibaba	alicrypto-java-aliyun	1.0.4
com.alibaba	fastjson	1.2.68
org.bouncycastle	bcprov-jdk15to18	1.66
org.bouncycastle	bcpkix-jdk15on	1.65
org.bouncycastle	bcprov-jdk15on	1.65
com.aliyun	gmsse	1.1.0
org.apache.commons	commons-lang3	3.0
commons-codec	commons-codec	1.11
commons-logging	commons-logging	1.2
org.apache.httpcomponents	httpcore	4.4.11
org.apache.httpcomponents	httpclient	4.5.7

Apsara Stack SDK for Python

Package	Version
aliyun-python-sdk-core	2.13.35
certifi	2021.5.30
cffi	1.14.6
charset-normalizer	2.0.4
cryptography	3.4.8
idna	3.2
jmespath	0.10.0

Package	Version
pycparser	2.20
requests	2.26.0
urllib3	1.26.6

Note You can use the version for each package listed in the preceding table or a later version.

7.3. STS overview

7.3.1. What is STS?

Security Token Service (STS) allows you to manage temporary credentials to your Apsara Stack resources. Resource Access Management (RAM) provides RAM users and RAM roles. A RAM role does not have permanent identity credentials. A RAM role can be assumed to access Apsara Stack resources by using an issued STS token. When the STS token is issued, you can specify a validity period and access permissions for the STS token.

Use a RAM user to obtain an STS token

An authorized RAM user can use an AccessKey pair to call the AssumeRole operation. This way, the RAM user can obtain an STS token of a RAM role and can use the STS token to access Apsara Stack resources.

Benefits

- STS tokens help reduce the risks of AccessKey pair leaks. An AccessKey pair is a long-term credential for a RAM user.
- STS tokens are temporary credentials. You can specify the validity period for STS tokens. After STS tokens expire, they become invalid. Therefore, you do not need to rotate the STS tokens on a regular basis.
- You can attach custom policies to STS tokens for flexible and fine-grained authorization.

7.3.2. AssumeRole

Obtains a Security Token Service (STS) token that is used by a RAM user to assume a RAM role. In this example, an Apsara Stack tenant account is used as the trusted entity of the RAM role.

Request parameters

Parameter	Type	Required	Example	Description

Parameter	Type	Required	Example	Description
Action	String	Yes	AssumeRole	The operation that you want to perform. Set the value to AssumeRole.
RoleArn	String	Yes	acs:ram::123456789012****:role/adminrole	The Apsara Stack Resource Name (ARN) of the RAM role. The trusted entity of the RAM role is an Apsara Stack tenant account.
RoleSessionName	String	Yes	alice	The name of the role session. Set this parameter based on your business requirements. In most cases, you can set this parameter to the identity of the user who calls the operation. For example, specify a username. In ActionTrail logs, you can distinguish the users who assume the same RAM role to perform operations based on the value of the **RoleSessionName** parameter. This way, you can perform user-specific auditing. The name must be 2 to 64 characters in length and can contain letters, digits, periods (.), at signs (@), hyphens (-), and underscores (_).
Policy	String	No	{"Statement": [{"Action": ["*"],"Effect": "Allow","Resource": ["*"]}],"Version":"1"}	The policy that specifies the permissions of the STS token. You can use this parameter to grant the STS token fewer permissions than the permissions granted to the RAM role. The policy must be 1 to 1,024 characters in length.

Responses

Parameter	Type	Example	Description
RequestId	String	6894B13B-6D71-4EF5-88FA-F32781734A7F	The ID of the request.

Parameter	Type	Example	Description
Credentials	None	None	The access credentials.
AccessKeyId	String	STS.L4aBSCSJVMuKg5U1****	The AccessKey ID.
AccessKeySecret	String	wyLTSmsyPGP1ohvvw8xYgB29dlGI8KMiH2pK****	The AccessKey secret.
SecurityToken	String	********	The security token.
Expiration	String	2015-04-09T11:52:19Z	The time when the STS token expires.
AssumedRoleUser	None	None	The temporary identity that you use to assume the RAM role.
Arn	String	acs:ram::123456789012****:role/adminrole/alice	The ARN of the temporary identity that you use to assume the RAM role.
AssumedRoleId	String	34458433936495****:alice	The ID of the temporary identity that you use to assume the RAM role.

Examples

Add Project Object Model (POM) dependencies

The Alibaba Cloud core library is required. Perform the following steps to add the Alibaba Cloud core library to POM:

```
<dependency>
  <groupId>com.aliyun</groupId>
  <artifactId>aliyun-java-sdk-core</artifactId>
  <version>[4.4.9,5.0.0)</version>
</dependency>
```

Call an API operation to obtain the STS AccessKey 3-tuple.

In the following sample code, the `Sts::AssumeRole::2015-04-01` operation is used to obtain the 3-tuple.

```
package com.aliyuns.test;
```

```
import com.alibaba.fastjson.JSONObject;
import com.aliyuncs.CommonRequest;
import com.aliyuncs.CommonResponse;
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.exceptions.ServerException;
import com.aliyuncs.http.FormatType;
import com.aliyuncs.http.HttpClientConfig;
import com.aliyuncs.http.MethodType;
import com.aliyuncs.http.ProtocolType;
import com.aliyuncs.profile.DefaultProfile;
import org.apache.commons.lang3.StringUtils;

public class STSAccessKeyDemo {
    public static void main(String[] args) {
        // Create and initialize a DefaultAcsClient instance.
        DefaultProfile profile = DefaultProfile.getProfile("<yourRegionID>", "<yourAccessKe
yID>", "<yourAccessKeySecret>");
        HttpClientConfig clientConfig = HttpClientConfig.getDefault();
        clientConfig.setIgnoreSSLCerts(true);
        clientConfig.setProtocolType(ProtocolType.HTTPS);
        profile.setHttpClientConfig(clientConfig);
        IAcsClient client = new DefaultAcsClient(profile);
        CommonRequest request = new CommonRequest();
        // Replace "sts.aliyuns.com" with the actual endpoint of STS. In most cases, the en
dpoint is in the following format: sts-vpc.${global:region}.${global:internet-domain}.
        request.setSysDomain("sts.aliyuns.com");
        request.setSysProduct("Sts");
        request.setSysAction("AssumeRole");
        request.setSysVersion("2015-04-01");
        request.setSysMethod(MethodType.POST);
        request.setHttpContentType(FormatType.FORM);
        // Replace "<yourRoleSessionName>" with the alias that you specify for the current
role session.
        request.putQueryParameter("RoleSessionName", "<yourRoleSessionName>");
        // Replace "<yourRoleArn>" with the actual ARN that you obtained for the RAM role i
n the preceding step.
        request.putQueryParameter("RoleArn", "<yourRoleArn>");
        try {
            CommonResponse response = client.getCommonResponse(request);
            String assumeRole = response.getData();
            if (StringUtils.isEmpty(assumeRole)) return;
            JSONObject assumeRoleJson = JSONObject.parseObject(assumeRole);
            JSONObject credentials = assumeRoleJson.getJSONObject("Credentials");
            // Obtain the STS AccessKey 3-tuple.
            String stsAK = credentials.getString("AccessKeyId");
            String stsSK = credentials.getString("AccessKeySecret");
            String stsToken = credentials.getString("SecurityToken");
            System.out.println("stsAk: " + stsAK +"\nstsSk: " + stsSK + "\nstsToken: " + st
sToken);
        } catch (ServerException e) {
            System.out.println(e.getErrCode());
```

```
            System.out.println(e.getErrMsg());
        } catch (ClientException e) {
            System.out.println(e.getErrCode());
            System.out.println(e.getErrMsg());
        }
    }
}
```

Call examples

In the following sample code, the `DescribeZones` operation in Elastic Compute Service (ECS) is used.

```
package com.aliyuns.test;

import com.aliyuncs.CommonRequest;
import com.aliyuncs.CommonResponse;
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.exceptions.ServerException;
import com.aliyuncs.http.FormatType;
import com.aliyuncs.http.HttpClientConfig;
import com.aliyuncs.http.MethodType;
import com.aliyuncs.http.ProtocolType;
import com.aliyuncs.profile.DefaultProfile;

public class DescribeZonesDemo {
    public static void main(String[] args) {
        // Replace "<yourRegionID>" with the actual region ID.
        // Replace "<yourStsAccesskey>" and "<yourStsAccessKeySecret>" with the AccessKey I
D and secret in the 3-tuple that you obtained in the preceding step.
        DefaultProfile profile = DefaultProfile.getProfile("<yourRegionID>", "<yourStsAcces
sKey>", "<yourStsAccessKeySecret>");
        HttpClientConfig clientConfig = HttpClientConfig.getDefault();
        clientConfig.setIgnoreSSLCerts(true);
        clientConfig.setProtocolType(ProtocolType.HTTPS);
        profile.setHttpClientConfig(clientConfig);
        IAcsClient client = new DefaultAcsClient(profile);
        CommonRequest request = new CommonRequest();
        request.setSysDomain("ecs.aliyuns.com");
        request.setSysProduct("Ecs");
        request.setSysAction("DescribeZones");
        request.setSysVersion("2014-05-26");
        request.setSysMethod(MethodType.POST);
        request.setHttpContentType(FormatType.FORM);
        //Replace "<yourRegionID>" with the actual region ID.
        request.putHeadParameter("x-acs-regionid", "<yourRegionID>");
        // Replace "<yourStsToken>" with the STS token in the 3-tuple that you obtained in
the preceding step.
        request.putQueryParameter("SecurityToken", "<yourStsToken>");
        // Optional. Specify the resource set ID.
        // request.putHeadParameter("x-acs-resourcegroupid", "<yourResourceGroupID>");
        // Required. Specify the organization ID.
```

```
        request.putHeadParameter("x-acs-organizationid", "<yourOrganizationID>");
        try {
            CommonResponse response = client.getCommonResponse(request);
            System.out.println(response.getData());
        } catch (ServerException e) {
            System.out.println(e.getErrCode());
            System.out.println(e.getErrMsg());
        } catch (ClientException e) {
            System.out.println(e.getErrCode());
            System.out.println(e.getErrMsg());
        }
    }
}
```

Error codes

HttpCode	Error code	Error message	Description
400	InvalidParameter	The parameter RoleArn is wrongly formed.	The error message returned because the ARN format of the RAM role is invalid.
400	InvalidParameter.RoleArn	The parameter RoleArn is wrongly formed.	The error message returned because the ARN format of the RAM role is invalid.
400	InvalidParameter.RoleSessionName	The parameter RoleSessionName is wrongly formed.	The error message returned because the value of the **RoleSessionName** parameter is in an invalid format. The value must be 2 to 64 characters in length and can contain letters, digits, periods (.), at signs (@), hyphens (-), and underscores (_).

HttpCode	Error code	Error message	Description
400	InvalidParameter.DurationSeconds	The Min/Max value of DurationSeconds is 15min/1hr.	The error message returned because the value of the **DurationSeconds** parameter is invalid. The minimum value is 900, and the maximum value is equal to the value of the **MaxSessionDuration** parameter.
400	InvalidParameter.PolicyGrammar	The parameter Policy has not passed grammar check.	The error message returned because the syntax of the policy is invalid.
400	InvalidParameter.PolicySize	The size of Policy must be smaller than 1024 bytes.	The error message returned because the length of the specified policy string exceeds the upper limit. The policy string can be up to 1,024 characters in length.
403	NoPermission	You are not authorized to do this action. You should be authorized by RAM.	The error message returned because the STS token does not have the required permissions.
404	EntityNotExist.Role	The specified Role not exists.	The error message returned because the specified RAM role does not exist.
500	InternalError	STS Server Internal Error happened.	The error message returned because an internal server error occurred.