

Alibaba Cloud Auto Scaling

API Reference

Issue: 20190918

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 due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall 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 the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. 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 bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, 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 content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, 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 the Alibaba Cloud website, product programs, or content 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 contact Alibaba Cloud directly if you discover any errors in this document.

Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information 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.
	This warning information 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 restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 Notice: Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 Note: You can use Ctrl + A to select all files.
>	Multi-level menu cascade.	Settings > Network > Set network type
Bold	It is used for buttons, menus, page names, and other UI elements.	Click OK.
<code>Courier</code> font	It is used for commands.	Run the <code>cd / d C :/ windows</code> command to enter the Windows system folder.
<i>Italics</i>	It is used for parameters and variables.	<code>bae log list --instanceid Instance_ID</code>
[] or [a b]	It indicates that it is a optional value, and only one item can be selected.	<code>ipconfig [-all -t]</code>

Style	Description	Example
<code>{}</code> or <code>{a b}</code>	It indicates that it is a required value, and only one item can be selected.	<code>switch {stand slave}</code>

Contents

Legal disclaimer.....	I
Generic conventions.....	I
1 API usage instructions.....	1
2 API quick start.....	2
2.1 Process introduction.....	2
2.2 Create a scaling group.....	2
2.3 Create a scaling configuration.....	3
2.4 Enable a scaling group.....	3
2.5 Create a scaling rule.....	4
2.6 Create a scheduled task.....	4
3 Calling method.....	5
4 Request structure.....	6
5 Common parameters.....	8
6 Signatures.....	11
7 Response results.....	15
8 Scaling group.....	19
8.1 CreateScalingGroup.....	19
8.2 ModifyScalingGroup.....	40
8.3 EnableScalingGroup.....	50
8.4 DisableScalingGroup.....	59
8.5 DeleteScalingGroup.....	61
8.6 DescribeScalingGroups.....	64
8.7 DescribeScalingInstances.....	75
8.8 DescribeScalingActivities.....	81
8.9 AttachLoadBalancers.....	86
8.10 DetachLoadBalancers.....	90
8.11 DetachDBInstances.....	93
8.12 AttachDBInstances.....	95
8.13 AttachVServerGroups.....	98
8.14 DetachVServerGroups.....	105
9 Scaling configuration.....	109
9.1 DeleteScalingConfiguration.....	109
9.2 CreateScalingConfiguration.....	111
9.3 DescribeScalingConfigurations.....	131
9.4 ModifyScalingConfiguration.....	143
10 Scaling rule.....	161
10.1 CreateScalingRule.....	161

10.2 ModifyScalingRule.....	177
10.3 DeleteScalingRule.....	190
10.4 DescribeScalingRules.....	192
11 Trigger task.....	205
11.1 ExecuteScalingRule.....	205
11.2 AttachInstances.....	213
11.3 RemoveInstances.....	221
11.4 DetachInstances.....	225
12 Scheduled tasks.....	230
12.1 DescribeScheduledTasks.....	230
12.2 DeleteScheduledTask.....	234
12.3 CreateScheduledTask.....	236
12.4 ModifyScheduledTask.....	243
13 Lifecycle Hook.....	251
13.1 CreateLifecycleHook.....	251
13.2 ModifyLifecycleHook.....	263
13.3 DescribeLifecycleHooks.....	270
13.4 RecordLifecycleActionHeartbeat.....	274
13.5 CompleteLifecycleAction.....	277
13.6 DeleteLifecycleHook.....	281
14 Event notification.....	284
14.1 CreateNotificationConfiguration.....	284
14.2 DescribeNotificationConfigurations.....	288
14.3 ModifyNotificationConfiguration.....	291
14.4 DeleteNotificationConfiguration.....	294
14.5 DescribeNotificationTypes.....	297
15 Instance.....	299
15.1 EnterStandby.....	299
15.2 ExitStandby.....	301
15.3 RebalanceInstances.....	304
15.4 SetInstancesProtection.....	307
16 Region.....	311
16.1 DescribeRegions.....	311
17 Error codes.....	314
17.1 Client errors.....	314
17.2 Server errors.....	315
18 How to ensure idempotence.....	316

1 API usage instructions

Before calling the Auto Scaling API, you must activate the Auto Scaling service on the Alibaba Cloud official website, and grant Auto Scaling the API permission in the console.

For the detailed operations, see [access resources](#).

Prerequisite

Before calling the Auto Scaling API, you must activate the Auto Scaling service on the Alibaba Cloud official website, and grant Auto Scaling the API permission in the console. For the detailed operations, see [access resources](#).

Errors

Before calling the Auto Scaling API, you need to activate Auto Scaling at Alibaba Cloud website, and authorize your Auto Scaling to access API on the Auto Scaling console (based on Alibaba Cloud Resource Access Management [RAM], Auto Scaling uses ECS API to replace the ECS instance resources). Errors will occur in case of any inconformity:

Error codes	Error code	Description	HTTP status code
Auto Scaling is not activated and thus the API cannot be called.	Forbidden. Unsubscribed	Do not have permission to access this API.	403
You have not fully authorized Open API to Auto Scaling.	Forbidden. Unauthorized	A required authorization for the specified action is not supplied.	403

2 API quick start

2.1 Process introduction

This topic explains how to use APIs to create and configure overall scaling solutions, including scheduled, dynamic, customized, and fixed quantity scaling.

Follow the steps shown in the following figure. The first three steps are used to create a simple scaling solution:

1. Create a scaling group. Configure the minimum (Min Size) and maximum (Max Size) number of ECS instances for scaling, and select a Server Load Balancer instance and an RDS instance.
2. Create scaling configuration. Specify the specifications of ECS instances used for Auto Scaling, such as Image ID and Instance Type.
3. Enable the scaling group according to the scaling configuration created in Step 2.
4. Create a scaling rule. For example, adding N (number) ECS instances.
5. Create a scheduled task. For example, triggering the scaling rule at 13:00.
6. Create an alarm task (CloudMonitor API PutAlarmRule). For example, adding 1 ECS instance when the CPU usage is equal to or greater than 80%.

2.2 Create a scaling group

This topic gives you an example of how to make the CreateScalingGroup API request, and specifying MinSize and MaxSize.

```
http://ess.aliyuncs.com/?Action=CreateScalingGroup
&RegionId=cn-qingdao
&MaxSize=20
&MinSize=2
&LoadBalancerId=147b46d767c-cn-qingdao-cm5-a01
&DBInstanceId.1=rdszzzyun-ybaeu
&DBInstanceId.2=rdsia3u3yi-a3u3y
&<PublicRequestParameters>
```

Response example

```
<CreateScalingGroupResponse>
  <ScalingGroupId>dP8VqSd9ENXPc0ciVmbc rBT1</ScalingGroupId>
```

```
< RequestId > 536E9CAD - DB30 - 4647 - AC87 - AA5CC38C53 82 </
RequestId >
</ CreateScalingGroupResponse >
```

2.3 Create a scaling configuration

You must specify the attributes of the ECS instances to be scaled, such as the ImageID and InstanceType. You must specify the ScalingGroupId returned in step 1.

```
http :// ess . aliyuncs . com /? Action = CreateScalingConfiguration
& ScalingGroupId = dP8VqSd9EN XPc0ciVmbc rBT1
& SecurityGroupId = sg - 280ih3w4b
& ImageId = centos6u5_64_20G_ali_aegis_2014_0703 . vhd
& InstanceType = ecs . t1 . xsmall
&< common request parameters >
```

Response example

```
< CreateScalingConfigurationResponse >
  < ScalingConfigurationId > e0s27Kb0oX vQcUYjEGel JqUy </
ScalingConfigurationId >
  < RequestId > 5CC0AD41 - 08ED - 4559 - A683 - 6F56355FE0 68 </
RequestId >
</ CreateScalingConfigurationResponse >
```

2.4 Enable a scaling group

This topic gives you an example of how to make the EnableScalingGroup API request according to the scaling configuration you specified, and add an existing ECS instance.

```
http :// ess . aliyuncs . com /? Action = EnableScalingGroup
& ScalingGroupId = dP8VqSd9EN XPc0ciVmbc rBT1
& ActiveScalingConfigurationId = e0s27Kb0oX vQcUYjEGel JqUy
& InstanceId . 1 = i - 283vvytn
&< Public Request Parameters >
```

Response example

```
< EnableScalingGroupResponse >
  < RequestId > 6469DCD0 - 13AC - 487E - 85A0 - CE4922908F DE </
RequestId >
```

```
</ EnableScalingGroupResponse >
```

2.5 Create a scaling rule

This topic is an example of API request `CreateScalingRule` (create a scaling rule). You must specify the `ScalingGroupId`.

```
http://ess.aliyuncs.com/? Action = CreateScalingRule
& ScalingGroupId = dP8VqSd9ENXPc0ciVmbcrBT1
& AdjustmentType = QuantityChangeInCapacity
& AdjustmentValue = 1
&< common request parameters >
```

Response example

```
< CreateScalingRuleResponse >
< ScalingRuleAri >
  ari:acs:ess:cn-qingdao:1344371:scalingrule/
  eMKWG8SRNb9dBLAjweNI1Ik
</ ScalingRuleAri >
  < ScalingRuleId > eMKWG8SRNb9dBLAjweNI1Ik </ ScalingRuleId >
  < RequestId > 570C84F4 - A434 - 488A - AFA1 - 1E3213682B33 </
  RequestId >
</ CreateScalingRuleResponse >
```

2.6 Create a scheduled task

This topic is an example of API request `CreateScheduledTask` (create a scheduled task). You must specify the `ScalingRuleAri`.

```
http://ess.aliyuncs.com/? Action = CreateScheduledTask
& RegionId = cn-qingdao
& LaunchTime = 2014-08-17T12:00Z
& RecurrenceType = Daily
& RecurrenceValue = 1
& RecurrenceEndTime = 2014-09-17T16:55Z
& ScheduledAction = ari:acs:ess:cn-qingdao:1344371:
  scalingrule/eMKWG8SRNb9dBLAjweNI1Ik
&< common request parameters >
```

Response example

```
< CreateScheduledTaskResponse >
  < ScheduledTaskId > edRtShc57WGXdt8TlPbrjsnV </ ScheduledTaskId >
  < RequestId > 0F02D931 - 2B12 - 44D7 - A0E9 - 39925C13D15E </
  RequestId >
</ CreateScheduledTaskResponse >
```

3 Calling method

This topic introduces the calling method.

You can call the Auto Scaling API by sending an HTTP GET request to the Auto Scaling API server and adding the relevant request parameters to the request according to the interface instructions. Results are returned based on how the request is processed

.

4 Request structure

This topic introduces the request structure.

Service address

The following table describes the access endpoints of Auto Scaling API service. To reduce network latency, we recommend that you configure endpoints based on the source from which you call the service.

Region (Location)	Endpoint
Default	ess.aliyuncs.com
China (Qingdao)	ess.aliyuncs.com
China (Beijing)	ess.aliyuncs.com
China (Zhangjiakou)	ess.cn-zhangjiakou.aliyuncs.com
China (Hohhot)	ess.cn-huhehaote.aliyuncs.com
China (Hangzhou)	ess.aliyuncs.com
China (Shanghai)	ess.aliyuncs.com
China (Shenzhen)	ess.aliyuncs.com
China (Hong Kong)	ess.aliyuncs.com
Singapore	ess.aliyuncs.com
Australia (Sydney)	ess.ap-southeast-2.aliyuncs.com
Malaysia (Kuala Lumpur)	ess.ap-southeast-3.aliyuncs.com
Indonesia (Jakarta)	ess.ap-southeast-5.aliyuncs.com
Japan (Tokyo)	ess.ap-northeast-1.aliyuncs.com
UK (London)	ess.eu-west-1.aliyuncs.com
US (Silicon Valley)	ess.aliyuncs.com
US (Virginia)	ess.aliyuncs.com
Germany (Frankfurt)	ess.eu-central-1.aliyuncs.com
UAE (Dubai)	ess.me-east-1.aliyuncs.com
India (Mumbai)	ess.ap-south-1.aliyuncs.com

Communication protocols

The system supports request communication through the HTTP or HTTPS channel.

We recommend that you send requests over HTTPS for enhanced security.

Request methods

The system allows you to send HTTP GET requests. This requires request parameters to be included in the request URL.

Request parameters

For each request, you must specify the operation you want to execute, namely the Action parameter (for example, CreateScalingGroup), and each operation must contain the public and specific request parameters of the specified operation.

Character encoding

Requests and returned results are encoded using the UTF-8 character set.

5 Common parameters

This topic describes the common parameters.

Common parameters include the [common request parameters](#) and the [common request parameters](#).

Common request parameters

The following table describes the common parameters that comprise of a URL for a GET request over HTTP or HTTPS protocol.

Name	Type	Required	Description
Action	String	Yes	The target API.
AccessKeyId	String	Yes	Equivalent to a logon password. However, an AccessKey is used to call APIs, while logon password is used to log on to the console . For more information, see Create an AccessKey .
Signature	String	Yes	Your signature. For more information, see signatures .
SignatureMethod	String	Yes	Signature method. Value: HMAC-SHA1.
SignatureVersion	String	Yes	Signature algorithm version. Value: 1.0.
SignatureNonce	String	Yes	Unique random number, which is used to prevent network replay attacks. Different random numbers must be used for different requests.

Name	Type	Required	Description
Timestamp	String	Yes	Request time stamp. Use the offset from Coordinated Universal Time (UTC), the time display is based on ISO8601. Format: YYYY-MM-DDThh:mm:ssZ. For example, 2018-01-01T12:00:00Z indicates 20:00:00, Jan 01, 2018, Beijing time (UTC+8).
Version	String	Yes	The API version to use. Value: 2014-08-28.
Format	String	No	Type of the response parameters. Optional values: json xml. Default value: json.

Request example

```
https://ess.aliyuncs.com/?Action=XXXXXX
&Format=xml
&Version=2014-08-28
&Signature=Pc5WB8gokVn0xfeu%2FZV%2BiNM1dgI%3D
&SignatureMethod=HMAC-SHA1
&SignatureNonce=15215528852396
&SignatureVersion=1.0
&AccessKeyId=key-test
&Timestamp=2018-01-01T12:00:00Z
```

...

Common response parameter

Name	Type	Description
RequestId	String	The request ID. We return a unique RequestId for every API request, whether the request is successful or not.

6 Signatures

This topic introduces the signature.

Auto Scaling performs authentication on each access request. Therefore, each request, whether being sent via HTTP or HTTPS, must contain signature information. By using AccessKeyID and AccessKeySecret, Auto Scaling performs symmetric encryption to authenticate the request sender. The AccessKeyID and AccessKeySecret are officially issued to visitors by Alibaba Cloud (visitors can apply for and manage them at Alibaba Cloud's official website). The AccessKeyID indicates the identity of the visitor. The AccessKeySecret is the secret key used to encrypt and verify the signature string on the server. It must be kept confidential and should only be available to Alibaba Cloud and the user.

When a user calls a server, the following method is used to sign the request:

1. The Canonicalized Query String is constructed using the request parameters.
 - a. The request parameters are ordered alphabetically by the parameter names (this includes the "public request parameters" and user-defined parameters for the given request interfaces described in this document, but not the Signature parameter mentioned in "public request parameters").



Note:

For a request submitted using the GET method, these parameters constitute the parameter section of the request URI (that is, the section in the URI following "?" and connected by "&").

- b. The name and value of each request parameter are encoded. The names and values must be URL encoded using the UTF-8 character set. The URL encoding rules are as follows:

- English letters A–Z and a–z, digits 0–9, and characters "-", "_", ".", and "~" are not encoded.
- Other characters are encoded in the "%XY" format, with "XY" representing the characters' ASCII code in hexadecimal notation. For example, the English double quotes are encoded as "%22".
- Extended UTF-8 characters are encoded in the "%XY%ZA..." format.
- Note that a space is encoded into "%20" instead of a plus sign "+".



Note:

Generally, libraries that support URL encoding (for example, Java's `java.net.URLEncoder`) are all encoded according to the rules for the "application/x-www-form-urlencoded" MIME-type. If this encoding method is used, replace the plus signs "+" in the encoded strings with "%20", the asterisks "*" with "%2A", and change "%7E" back to the tilde "~" to conform to the encoding rules described above.

- c. Connect the encoded parameter names and values with the equal sign "=".
- d. Then, sort the parameter name and value pairs connected by equal signs in alphabetical order, and connect them with the "&" symbol to produce the Canonicalized Query String.
2. Follow the rules below to construct the string used for signature calculation by using the Canonicalized Query String constructed in the previous step:

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

```
percentEncode ( CanonicalizedQueryString )
```

- HTTPMethod is the HTTP method used for request submission, for example, GET.
 - percentEncode("/") is the coded value for the character "/" according to the URL encoding rules described in 1.ii, that is, "%2F".
 - percentEncode(CanonicalizedQueryString) is the encoded string of the Canonicalized Query String constructed in Step 1, produced by following the URL encoding rules described in 1.ii.
3. According to RFC2104 definitions, use the above signature string to calculate the signature's HMAC value.

**Note:**

The Key used for calculating the signature is the AccessKeySecret held by the user, which ends with the "&" character (ASCII:38) and is based on the SHA1 hashing.

4. According to Base64 encoding rules, encode the above HMAC value into a string. This gives you the signature value.
5. Add the obtained signature value to the request parameters as the Signature parameter. This completes the request signing process.

**Note:**

When the obtained signature value is submitted to the ECS server as the final request parameter value, the value is URL encoded like other parameters according to RFC3986 rules.

Take DescribeScalingGroups as an example. The request URL prior to signing is as follows:

```
http://ess.aliyuncs.com/?TimeStamp=2014-08-15T11%3A10%3A07Z&Format=xml&AccessKeyId=testid&Action=DescribeScalingGroups&SignatureMethod=HMAC-SHA1&RegionId=cn-qingdao&SignatureNonce=1324fd0e-e2bb-4bb1-917c-bd6e437f1710&SignatureVersion=1.0&Version=2014-08-28
```

Thus, the StringToSign is:

```
GET &%2F & AccessKeyId %3Dtestid & Action %3DDescribeScalingGroups & Format %3Dxml & RegionId %3Dcn-qingdao & SignatureMethod %3DHMAC-SHA1 & SignatureNonce %3D1324fd0e-e2bb-4bb1-917c-bd6e437f1710 & SignatureVersion %3D1.
```

```
0 & TimeStamp % 3D2014 - 08 - 15T11 % 253A10 % 253A07Z & Version %  
3D2014 - 08 - 28
```

Assume that the AccessKeyID is "testid", the AccessKeySecret is "testsecret", and the Key used for HMAC calculation is "testsecret&". The calculated signature value is "SmhZuLUnXmqxSEZ%2FGqyiWgqmf%2BM=".

The signed request URL is (added with the Signature parameter):

```
http :// ess . aliyuncs . com /? TimeStamp = 2014 - 08 - 15T11  
% 3A10 % 3A07Z & Format = xml & AccessKeyI d = testid & Action  
= DescribeSc alingGroup s & SignatureM ethod = HMAC - SHA1 &  
RegionId = cn - qingdao & SignatureN once = 1324fd0e - e2bb - 4bb1  
- 917c - bd6e437f17 10 & SignatureV ersion = 1 . 0 & Version =  
2014 - 08 - 28 & Signature = SmhZuLUnXm qxSEZ % 2FGqyiWgqm f %  
2BM % 3D
```

For more information about signatures and request submission, see [how to call interfaces](#).

7 Response results

This topic introduces the response results.

We return the results in either XML or JSON format, but JSON is the default choice. You can switch to XML schema by specifying the request parameter `Format`. For more information, see [common parameters](#).



Note:

Response examples in our API documents have line breaks and indentions to make them easy to read. The actual response results are not formatted.

Success response example

Every successful response has a request ID in the `RequestId` element and other API-specific response parameters. The HTTP status code for a success response is 2XX.

XML format:

```
<? xml version =" 1 . 0 " encoding =" UTF - 8 "? > <!-- Result
  root   node -->
< ActionResp onse > <!-- Response   request   tag ->
  < RequestId > 4C467B38 - 3910 - 447D - 87BC - AC049166F2  23 </
  RequestId > <!-- Response   result   data -->
</ ActionResp onse >
```

JSON format:

```
{
  " RequestId ": " 4C467B38 - 3910 - 447D - 87BC - AC049166F2  23
" /* Response   result   data */
}
```

Error response example

Every error response consists of a request ID in the `RequestId` element and access endpoint in the `HostId` element, the error code, and the error message. The HTTP status code for an error response is 4xx or 5xx.

You can fix the exception according to the API-specific error codes or [common error codes](#) and try the request again. Alternatively, you can [open a ticket](#) and provide additional inputs such as the `HostId` and `RequestId` to get technical support from us.

XML format:

```
<? xml version =" 1 . 0 " encoding =" UTF - 8 "? ><! -- Result
root node -->
< Error >
    < RequestId > 540CFF28 - 407A - 40B5 - B6A5 - 74Bxxxxxxx xx </
RequestId > <! -- Request ID -->
    < HostId > ess . aliyuncs . com </ HostId > <! -- Endpoint -->
    < Code > InternalEr ror </ Code > <! -- Error code -->
    < Message > The request processing has failed due to
some unknown error , exception or failure . </ Message > <!
-- Error message -->
</ Error >
```

JSON format:

```
{
  " RequestId ": " 540CFF28 - 407A - 40B5 - B6A5 - 74Bxxxxxxx xx
", /* Request ID */
  " HostId ": " ess . aliyuncs . com ", /* Endpoint */
  " Code ": " InternalEr ror ", /* Error code */
  " Message ": " The request processing has failed due
to some unknown error , exception or failure ." /* Error
message */
}
```

Common error codes

Error code	Error message	HTTP status code	Description
InvalidAccessKeyId.NotFound	The Access Key ID provided does not exist in our records.	400	The specified AccessKey does not exist.
InvalidParameter	The specified value of parameter <parameter name> is not valid.	400	The specified parameter is invalid .
MissingParameter	The input parameter <parameter name> that is mandatory for processing this request is not supplied.	400	You must specify the required parameters.
NoSuchVersion	The specified version does not exist.	400	The specified API version does not exist.

Error code	Error message	HTTP status code	Description
ResourceNotAvailable	Resource you requested is not available in this region or zone.	400	ESS service is unavailable in the specified region.
Throttling	Request was denied due to request throttling.	400	You have made too many frequent requests in short time. Please try again later.
UnsupportedOperation	The specified action is not supported.	400	Unable to call the specified API.
Forbidden	Users are not authorized to operate on the specified resource.	403	You cannot perform the specified action .
Forbidden.RiskControl	This operation is forbidden by Aliyun Risk Control system.	403	The specified action is under risk control.
Forbidden.Unsubscribed	Do not have permission to access this API.	403	You must enable the ESS service before calling this API.
Forbidden.UserVerification	Your user account is not verified by Aliyun.	403	You must complete the real-name registration.
SignatureDoesNotMatch	The signature we calculated does not match the one you provided.	403	The signature calculated by us is different from the one you provide.
InternalServerError	The request processing has failed due to some unknown error, exception or failure .	500	Internal error.

Error code	Error message	HTTP status code	Description
ServiceUnavailable	The request has failed due to a temporary failure of the server.	503	The server cannot respond to your request. Please try again later.

8 Scaling group

8.1 CreateScalingGroup

You can call this operation to create a scaling group.

Description

A scaling group is a group of Elastic Compute Service (ECS) instances that are dynamically scaled based on the configured scenario.

- A single account can have up to 50 scaling groups in a region.
- A scaling group does not take effect immediately after being created. You must call the [EnableScalingGroup](#) operation to enable the scaling group to trigger scaling rules and execute scaling activities.
- Each scaling group must be in the same region as the Server Load Balancer (SLB) instance and ApsaraDB for Relationship Database Service (RDS) instance that are associated with it. For more information, see [Regions and zones](#).
- If an SLB instance is specified for a scaling group, the scaling group automatically adds ECS instances to the VServer group of the specified SLB instance. The default weight of an ECS instance added to the VServer 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.
 - Health check must be enabled on all listener ports configured for the SLB instance. Otherwise, the scaling group fails to be created.
- If an RDS instance is specified for a scaling group, the scaling group automatically adds the internal IP addresses of its ECS instances to the whitelist of the specified RDS instance. The specified RDS instance must meet the following requirements:
 - The RDS instance must be in the Running state. You can call the [DescribeDBInstances](#) operation to query the status of the specified RDS instance.
 - The number of IP addresses in the RDS instance whitelist cannot exceed the upper limit. For more information, see [Configure a whitelist](#).

- If the MultiAZPolicy parameter of a scaling group is set to COST_OPTIMIZED (the cost optimization policy):
 - When the OnDemandBaseCapacity, OnDemandPercentageAboveBaseCapacity, and SpotInstancePools parameters are specified,

the scaling group will use the instance allocation method based on the cost optimization policy. This allocation method is prioritized during scaling activities.
 - When at least one of the OnDemandBaseCapacity, OnDemandPercentageAboveBaseCapacity, and SpotInstancePools parameters is not specified,

the instance types available at the lowest cost are selected to create instances based on the cost optimization policy.

Debugging

Alibaba Cloud provides OpenAPI Explorer to simplify API usage. You can use OpenAPI Explorer to search for APIs, call APIs, and dynamically generate SDK example code.

Request parameters

parameter	Type	Required	Example	Description
MaxSize	Integer	Yes	20	The maximum number of ECS instances in the scaling group . Valid values: 0 to 1000. When the number of ECS instances in the scaling group exceeds the value of MaxSize, Auto Scaling removes the ECS instances from the scaling group until the number of instances is equal to the MaxSize value.

parameter	Type	Required	Example	Description
MinSize	Integer	Yes	2	The minimum number of ECS instances in the scaling group . Valid values: 0 to 1000. When the number of ECS instances in the scaling group is smaller than the value of MinSize, Auto Scaling automatically creates ECS instances until the number of instances is equal to the MinSize value.
RegionId	String	Yes	cn-qingdao	The ID of the region where the scaling group resides. For more information, see Regions and zones .
Action	String	No	CreateScalingGroup	The operation that you want to perform. Set the value to CreateScalingGroup.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The token used to ensure idempotence. The value of this parameter is generated by the client and is unique among different requests. It can contain a maximum of 64 ASCII characters. For more information, see How to ensure idempotence .
DBInstanceIds	String	No	["rm-idx", "rm-idy", "rm-idz"]	The IDs of RDS instances. This value can be a JSON array which can contain up to eight RDS instance IDs. Separate multiple IDs with commas (,).

parameter	Type	Required	Example	Description
DefaultCooldown	Integer	No	300	The cooldown period after a scaling activity (adding or removing ECS instances) is executed. Valid values: 0 to 86400. Unit: seconds. Default value: 300. During the cooldown period calculated for a scaling activity, the scaling group does not execute any other scaling activities triggered by CloudMonitor alarm tasks.
HealthCheckType	String	No	ECS	The health check mode of the scaling group. Valid values: · NONE: The system performs no health check. · ECS: The system performs the health check on ECS instances in the scaling group.
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the instance launch template from which the scaling group obtains launch configurations.
LaunchTemplateVersion	String	No	Default	The version of the instance launch template. Valid values: · A fixed template version. · Default: The default template version is always used. · Latest: The latest template version is always used.

parameter	Type	Required	Example	Description
LifecycleHook . N . DefaultResult	String	No	CONTINUE	The action that the scaling group takes when the lifecycle hook times out. Valid values: · CONTINUE: The scaling group continues the scale-in or scale-out event. · ABANDON: The scaling group releases the created ECS instances if the scaling activity type is scale-out or removes the ECS instances to be scaled-in if the scaling activity type is scale-in. Default value: CONTINUE. If the scaling group has multiple lifecycle hooks and one of them is terminated when the DefaultResult parameter is set to ABANDON during a scale-in event, the remaining lifecycle hooks in the same scaling group are also terminated. Otherwise, the scaling activity will proceed normally after the wait period expires and continue with the action specified by the DefaultResult parameter.

parameter	Type	Required	Example	Description
LifecycleHook . N . HeartbeatTimeout	Integer	No	600	The wait period before the lifecycle hook times out. If the lifecycle hook times out , the scaling group performs the action specified in the DefaultResult parameter. Valid values: 30 to 21600. Unit: seconds. Default value: 600. You can prevent the lifecycle hook from timing out by calling the RecordLifecycleActionHeartbeat operation. You can also terminate the lifecycle hook by calling the CompleteLifecycleAction operation.
LifecycleHook . N . LifecycleHookName	String	No	lifecycle_hook	The name of the lifecycle hook . The name is used to specify a lifecycle hook and cannot be modified if it is set.
LifecycleHook . N . LifecycleTransition	String	No	SCALE_OUT	The type of scaling activity to which the lifecycle hook applies. Valid values: - SCALE_OUT: scale-out events of the scaling group. - SCALE_IN: scale-in events of the scaling group.

parameter	Type	Required	Example	Description
LifecycleHook . N . NotificationArn	String	No	acs:ess:cn-hangzhou:1111111111:queue/queue2	The Alibaba Cloud Resource Name (ARN) of the notification object that Auto Scaling uses to notify you when an instance is in the transition state for the lifecycle hook. This object can be either an MNS queue or an MNS topic. The format of the parameter value is <code>acs:ess:{region}:{account-id}:{resource-relative-id}</code>. • region: the region where the scaling group resides. • account-id: the ID of the Alibaba Cloud account. Examples: • MNS queue: <code>acs:ess:{region}:{account-id}:queue/{queuename}</code> • MNS topic: <code>acs:ess:{region}:{account-id}:topic/{topicname}</code>

parameter	Type	Required	Example	Description
LifecycleHook . NotificationMetadata	String	No	Test	The fixed string that you want to include when Auto Scaling sends a message about the wait state of a scaling activity to the notification target. The parameter value cannot exceed 128 characters in length. Auto Scaling sends the specified NotificationMetadata parameter value along with the notification message so that you can easily categorize your notifications. The NotificationMetadata parameter is only valid after you set the NotificationArn parameter.
LoadBalancerIds	String	No	["lb-idx", "lb-idy", "lb-idz"]	The IDs of SLB instances. This value can be a JSON array which can contain up to five SLB instance IDs. Separate multiple IDs with commas (,).

parameter	Type	Required	Example	Description
MultiAZPolicy	String	No	PRIORITY	The ECS instance scaling policy for a multi-zone scaling group. Valid values: PRIORITY: ECS instances are scaled based on the VSwitchIds.N parameter. When an ECS instance cannot be created in the zone where the VSwitch with the highest priority resides, the system automatically uses the VSwitch with the next highest priority to create the ECS instance. COST_OPTIMIZED: ECS instances are created based on the unit price of vCPU, from low to high. Preemptible instances are created first when preemptible instance types are specified for the scaling configuration. Pay-as-you-go instances are created automatically when all types of preemptible instances are unavailable due to issues such as insufficient ECS resources.  Note: COST_OPTIMIZED takes effect only when multiple instance types are specified or at least one preemptible instance type is specified.
Issue: 20190918				BALANCE: ECS instances are distributed evenly in multiple zones specified in the scaling group. If ECS

parameter	Type	Required	Example	Description
OnDemandBaseCapacity	Integer	No	30	The minimum number of pay-as-you-go instances required in the scaling group. Valid values : 0 to 1000. When the number of pay-as-you-go instances is smaller than this value, the scaling group will attempt to create pay-as-you-go instances over other instances.
OnDemandPercentageAboveBaseCapacity	Integer	No	20	The percentage of pay-as-you-go instances to be created when instances are added to the scaling group. This parameter takes effect after the number of instances reaches the OnDemandBaseCapacity value. Valid values: 0 to 100.

parameter	Type	Required	Example	Description
RemovalPolicy . 1	String	No	OldestScalingConfiguration	Specifies policy N for removing ECS instances from the scaling group. Valid values of N: 1 to 2. For more information, see Removal policies. Valid values: • OldestInstance: removes the ECS instance that is added to the scaling group at the earliest point in time. • NewestInstance: removes the ECS instance that is added to the scaling group at the latest point in time. • OldestScalingConfiguration : removes the ECS instance that is created based on the earliest scaling configuration. Default value of RemovalPolicy .1: OldestScalingConfiguration. Default value of RemovalPolicy .2: OldestInstance.
ScalingGroupName	String	No	Testsg	The name of the scaling group • The name of a scaling group must be unique in a region • The name must be 2 to 40 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit. The default value is the ID of the scaling group.

parameter	Type	Required	Example	Description
ScalingPolicy	String	No	recycle	The reclaim policy of the scaling group. Valid values: · recycle: The scaling group is set to the Shutdown and Reclaim Mode, which stops ECS instances and reclaims resources. · release: The scaling group is set to the Release Mode, which removes or adds ECS instances. For more information, see RemoveInstances.
SpotInstancePools	Integer	No	5	The number of available instance types. The scaling group will create preemptible instances of multiple instance types available at the lowest cost. Valid values: 0 to 10.
SpotInstanceRemedy	Boolean	No	true	Specifies whether to supplement preemptible instances when the target capacity of preemptible instances is not fulfilled. When receiving a system message that a preemptible instance will be reclaimed, the scaling group will create a new instance to replace the instance to be reclaimed if this parameter is set to true.

parameter	Type	Required	Example	Description
VServerGroup . N . LoadBalancerId	String	No	lb-****	The ID of the SLB instance with which the VServer group is associated. For more information, see AttachVServerGroups .
VServerGroup . N . VServerGroupAttribute . N . Port	Integer	No	22	The port number that is used by Auto Scaling to add the ECS instances to the VServer group . Valid values: 1 to 65535. For more information, see AttachVServerGroups .
VServerGroup . N . VServerGroupAttribute . N . VServerGroupId	String	No	rsp-****	The ID of the VServer group. For more information, see AttachVServerGroups .
VServerGroup . N . VServerGroupAttribute . N . Weight	Integer	No	100	The weight set for the ECS instances that are added to the VServer group. Valid values: 0 to 100. Default value: 50. For more information, see AttachVServerGroups .
VSwitchId	String	No	vsw-****	The ID of the specified VSwitch when a VPC-type scaling group is created.

parameter	Type	Required	Example	Description
VSwitchIds . N	RepeatList	No	vsw-****	The IDs of one or more VSwitches. Valid values of N: 1 to 5. If you use the VSwitchIds.N parameter, the VSwitchId parameter is ignored. This parameter is only valid when the network type of the scaling group is VPC. The specified VSwitch and the scaling group must be in the same VPC. VSwitches can be from multiple zones. VSwitches are sorted in ascending order based on the value of N. Value 1 indicates the highest priority. When an ECS instance cannot be created in the zone where the VSwitch with the highest priority resides, the system automatically uses the VSwitch with the next highest priority to create the ECS instance.

Response parameters

Parameter	Type	Example	Description
ScalingGroupId	String	dP8VqSd9ENXpC0ciVmbc****	The ID of the scaling group.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample request

```
http :// ess . aliyuncs . com /? Action = CreateScal ingGroup
& RegionId = cn - qingdao
& MaxSize = 20
& MinSize = 2
& LoadBalanc erId = 147b46d767 c - cn - qingdao - cm5 ****
& DBInstance Id . 1 = rdszzzyyun y ****
& DBInstance Id . 2 = rdsia3u3yi a ****
&< Common request parameters >
```

Sample success response

XML format

```
< CreateScal ingGroupRe sponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
  < ScalingGro upId > dP8VqSd9EN XPc0ciVmbc ****</ ScalingGro
  upId >
</ CreateScal ingGroupRe sponse >
```

JSON format

```
{
  " ScalingGro upId ":" dP8VqSd9EN XPc0ciVmbc ****",
  " RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

For more information about error codes, visit [API Error Center](#).

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 RDS instance is not in the Running state.

HTTP status code	Error code	Error message	Description
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	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 and 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 RDS instance and the specified scaling group are not 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 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 attached to the specified SLB instance 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 are not 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 attached to the SLB instance of the scaling group are not in the same VPC.
400	InvalidParameter	The specified value of parameter "ScalingPolicy" is not valid.	The error message returned because the specified value of the ScalingPolicy parameter is invalid.

HTTP status code	Error code	Error message	Description
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 specified scaling group name already exists.
400	QuotaExceeded. DBInstance SecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because the number of IP addresses in the specified RDS instance whitelist reaches the upper limit.
400	QuotaExceeded. PrivateIpAddress	Private IP address quota exceeded in the specified virtual switch.	The error message returned because the private IP address quota of the VSwitch is exceeded.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded. ScalingGroup	Scaling group quota exceeded.	The error message returned because the number of scaling groups reaches the upper limit.
400	QuotaExceeded. VPCInstance	Instance quota exceeded in the specified VPC.	The error message returned because the number of instances in the VPC reaches the upper limit.
404	InvalidDBInstance. NotFound	DB instance "XXX" does not exist.	The error message returned because the specified 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 does not exist.	The error message returned because the specified version of the instance launch template does not exist.
400	LaunchTemplateSet.NotFound	The specified launch template set is not found.	The error message returned because the specified instance 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 required for the specified instance launch template is not specified.

HTTP status code	Error code	Error message	Description
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 required for the specified instance 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 required for the specified instance 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 specified LaunchTemplateVersion parameter of the specified instance launch template is not in digits.

8.2 ModifyScalingGroup

You can call this operation to modify a scaling group.

Description

- The following parameters cannot be modified:

- RegionId
- LoadBalancerId



Note:

If you want to attach or detach a Server Load Balancer (SLB) instance to or from the scaling group, call the [AttachLoadBalancers](#) or [DetachLoadBalancers](#) operation.

- DBInstanceId



Note:

If you want to attach or detach an ApsaraDB for Relational Database Service (RDS) instance to or from the scaling group, call the [AttachDBInstances](#) or [DetachDBInstances](#) operation.

- You can call this operation only when the scaling group is in the Active or Inactive state.
- New scaling configurations do not affect running Elastic Compute Service (ECS) instances that are created based on the previous scaling configurations.
- If you increase the value of the MaxSize parameter, and the number of ECS instances in the scaling group exceeds the new value, the scaling group automatically removes instances to ensure that the number of instances are within the modified range.
- If you decrease the value of the MinSize parameter, and the number of ECS instances in the scaling group falls below the new value, the scaling group automatically adds instances to ensure that the number of instances are within the modified range.

Debugging

Alibaba Cloud provides OpenAPI Explorer to simplify API usage. You can use OpenAPI Explorer to search for APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroup upId	String	Yes	cqS5Qbbhmv GLcJbWoDbW ****	The ID of the scaling group to be modified.
Action	String	No	ModifyScalingGroup	The operation that you want to perform. Set the value to ModifyScalingGroup.
ActiveScalingConfigurationId	String	No	bU5uZHcAgt zwcL4IeDea ****	The ID of the active scaling configuration in the scaling group.
DefaultCooldown	Integer	No	600	The cooldown period after a scaling activity (adding or removing ECS instances) is executed. Valid values: 0 to 86400. Unit: seconds. During the cooldown period calculated for a scaling activity, the scaling group does not execute any other scaling activities triggered by CloudMonitor alarm tasks.
HealthCheckType	String	No	ECS	The health check mode of the scaling group. Valid values: · NONE: The system performs no health check. · ECS: The system performs the health check on ECS instances in the scaling group.

Parameter	Type	Required	Example	Description
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the instance launch template from which the scaling group obtains launch configurations.
LaunchTemplateVersion	String	No	Default	The version of the instance launch template. Valid values: • A fixed template version. • Default: The default template version is always used. • Latest: The latest template version is always used.
MaxSize	Integer	No	99	The maximum number of ECS instances in the scaling group . Valid values: 0 to 1000. When the number of ECS instances in the scaling group exceeds the value of MaxSize, Auto Scaling removes the ECS instances from the scaling group until the number of instances is equal to the MaxSize value.
MinSize	Integer	No	1	The minimum number of ECS instances in the scaling group . Valid values: 0 to 1000. When the number of ECS instances in the scaling group is smaller than the value of MinSize, Auto Scaling automatically creates ECS instances until the number of instances is equal to the MinSize value.

Parameter	Type	Required	Example	Description
OnDemandBaseCapacity	Integer	No	30	The minimum number of pay-as-you-go instances required in the scaling group. Valid values : 0 to 1000. When the number of pay-as-you-go instances is smaller than this value, the scaling group will attempt to create pay-as-you-go instances over other instances.
OnDemandPercentageAboveBaseCapacity	Integer	No	20	The percentage of pay-as-you-go instances to be created when instances are added to the scaling group. This parameter takes effect after the number of instances reaches the OnDemandBaseCapacity value. Valid values: 0 to 100.

Parameter	Type	Required	Example	Description
RemovalPolicy . 1	String	No	OldestScalingConfiguration	Specifies policy N for removing ECS instances from the scaling group. Valid values of N: 1 to 2. For more information, see Removal policies. Valid values: • OldestInstance: removes the ECS instance that is added to the scaling group at the earliest point in time. • NewestInstance: removes the ECS instance that is added to the scaling group at the latest point in time. • OldestScalingConfiguration : removes the ECS instance that is created based on the earliest scaling configuration.
ScalingGroupName	String	No	Scaling****	The name of the scaling group • The name of a scaling group must be unique in a region • The name must be 2 to 40 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or digit.
SpotInstancePools	Integer	No	5	The number of available instance types. The scaling group will create preemptible instances of multiple instance types available at the lowest cost. Valid values: 0 to 10.

Parameter	Type	Required	Example	Description
SpotInstanceRemedy	Boolean	No	true	Specifies whether to supplement preemptible instances when the target capacity of preemptible instances is not fulfilled. When receiving a system message that a preemptible instance will be reclaimed, the scaling group will create a new instance to replace the instance to be reclaimed if this parameter is set to true.

Parameter	Type	Required	Example	Description
VSwitchIds . N	RepeatList	No	vsw-****	The IDs of one or more VSwitches. Valid values of N: 1 to 5. This parameter is only valid when the network type of the scaling group is VPC. The specified VSwitch and the scaling group must be in the same VPC. VSwitches can be from multiple zones. VSwitches are sorted in ascending order based on the value of N. Value 1 indicates the highest priority. When an ECS instance cannot be created in the zone where the VSwitch with the highest priority resides, the system automatically uses the VSwitch with the next highest priority to create the ECS instance.

Response parameters

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

Examples

Sample request

```
http://ess.aliyuncs.com/?Action=ModifyScalingGroup
&ScalingGroupId=cqS5QbbhmvGLcJbWoDbW****
&ScalingGroupName=Scaling****
```


&< Common request parameters >

Sample success response

XML format

```
< ModifyScalingGroupResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
</ ModifyScalingGroupResponse >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

For more information about error codes, visit [API Error Center](#).

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 current account.
400	InvalidScalingGroupName.Duplicate	The specified value of parameter < parameter name> is duplicated.	The error message returned because the specified scaling group name 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.

HTTP status code	Error code	Error message	Description
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 that in the active scaling configuration.
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	LaunchTemplateVersionSet.NotFound	The specific version of launch template does not exist.	The error message returned because the specified instance launch template version does not exist.
400	LaunchTemplateSet.NotFound	The specified launch template set is not found.	The error message returned because the specified instance launch template does not exist.

HTTP status code	Error code	Error message	Description
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 required for the specified instance 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 required for the specified instance 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 required for the specified instance 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 specified LaunchTemplateVersion parameter of the specified instance launch template is not in digits.

8.3 EnableScalingGroup

You can call this operation to enable a scaling group.

Description

You can call this operation to enable a specified scaling group.

- If the operation is successful, the scaling group becomes active and the specified ECS instances are added to the scaling group.
- If the number of ECS instances in the scaling group is smaller than the MinSize value after the ECS instances are added, additional ECS instances are automatically created to achieve the MinSize value. For example, a scaling group is created with the MinSize parameter set to 5. Two existing ECS instances are specified in the InstanceId.N parameter when the scaling group is enabled. Then, three additional ECS instances are automatically created after the two ECS instances are added by Auto Scaling to the scaling group.

The operation can only be called when the scaling group is inactive.

If the scaling group has no active scaling configuration, you must specify a scaling configuration when enabling the scaling group.

- A single scaling group can only have one active scaling configuration at a time.

- If an active scaling configuration has been created before the scaling group is enabled, this configuration will become inactive after you specify a new active scaling configuration by calling the operation.

The restrictions on ECS instances to be manually added are as follows:

- The ECS instances and the scaling group must be in the same region.
- The ECS instances must have the same instance type as the active scaling configuration.
- The ECS instances must be running.
- The ECS instances have not been added to any scaling group.
- The ECS instances support subscription and pay-as-you-go billing methods.
- If the VSwitchId parameter is specified for a scaling group, ECS instances that belong to a classic network or in different VPCs cannot be added to the scaling group.
- If the VSwitchId parameter is not specified for a scaling group, ECS instances that belong to a VPC cannot be added to the scaling group.

If the sum of the number of instances to be added specified by this operation and the number of ECS instances currently in the scaling group is greater than the MaxSize value, the call fails.

Debugging

OpenAPI Explorer automatically calculates the signature value. For your convenience, we recommend that you call this operation in OpenAPI Explorer. OpenAPI Explorer dynamically generates the sample code of the operation for different SDKs.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	dmIDKNcyWfzncX9MWX1****	The ID of the scaling group.
Action	String	No	EnableScalingGroup	The operation that you want to perform. Set the value to EnableScalingGroup.

Parameter	Type	Required	Example	Description
ActiveScalingConfigurationId	String	No	cGsGHrdMBa3DcDrrBVcc****	The ID of the scaling configuration to be activated in the scaling group.
InstanceId . 1	String	No	i-283vv****	The ID of enabled ECS instance N to be added to the scaling group. Valid values of N: 1 to 20.
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the instance launch template, from which the specified scaling group can obtain launch configurations.
LaunchTemplateVersion	String	No	Default	The version number of the instance launch template. Valid values: • A fixed template version number. • Default: always uses the default template version. • Latest: always uses the latest template version.
LoadBalancerWeight . 1	Integer	No	50	The weight of ECS backend server N. Valid values of N: 1 to 20. Valid values for this parameter: 0 to 100. Default value: 50.

Response parameters

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

Examples

Sample requests

```
http :// ess . aliyuncs . com /? Action = EnableScal ingGroup
& ScalingGro upId = dmIDKNcyWf zncX9MWX1 ****
& InstanceId . 1 = i - 283vv ****
&< Common request parameters >
```

Sample success responses

XML format

```
< EnableScal ingGroupRe sponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
</ EnableScal ingGroupRe sponse >
```

JSON format

```
{
  " RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

For more information about error codes, visit [API Error Center](#).

HTTP status code	Error code	Error message	Description
404	InvalidScal ingGroupId. NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist in the current 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.

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 currently 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 that 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.

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 in the current 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 are not in the same region.
400	InvalidInstanceId.InstanceTypeMismatch	Instance "XXX" and existing active scaling configuration have different instance type.	The error message returned because the type of the specified ECS instance is different from that of the active scaling configuration.
400	IncorrectInstanceState	The current status of instance "XXX" does not support this action.	The error message returned because the specified ECS instance is not running.

HTTP status code	Error code	Error message	Description
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 that 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 are not 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 is already added to another scaling group.
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 active.

HTTP status code	Error code	Error message	Description
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 the ECS instance attached to the specified SLB instance is different from that 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 attached to the specified SLB instance is not in the same VPC as the VSwitch specified by VSwitchID.
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified RDS instance is not running.

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 (Total Capacity) exceeds the specified value of MaxSize after 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 instance launch template does not exist.
400	LaunchTemplateSet. NotFound	The specified launch template set is not found.	The error message returned because the specified instance 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 required for the specified instance launch template is not specified.

HTTP status code	Error code	Error message	Description
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 required for the specified instance 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 required for the specified instance 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 specified LaunchTemplateVersion parameter of the specified launch template is not in digits.

8.4 DisableScalingGroup

You can call this operation to disable a scaling group.

Description

This operation disables a specified scaling group.

- Scaling activities that are currently being executed by the specified scaling group before this operation is called will continue until they are completed. Scaling activity requests that are made to the target scaling group after this operation has been called will be rejected.
- When you call this operation, note that the scaling group must be active.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	dmIDKNcyWfzncX9MWX1****	The ID of the scaling group.
Action	String	No	DisableScalingGroup	The operation that you want to perform. Set the value to DisableScalingGroup.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

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

Successful response examples

XML format

```
< DisableScalingGroupResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
    RequestId >
</ DisableScalingGroupResponse >
```

JSON format

```
{
  "RequestId": "473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E"
}
```

Error codes[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.

8.5 DeleteScalingGroup

You can call this operation to delete a scaling group.

Description

This operation deletes a specified scaling group.

- **ForceDelete** indicates whether to forcibly delete a scaling group and remove or release its associated ECS instances regardless of whether it is undergoing a scaling activity.

- If ForceDelete is set to false, the scaling group can be deleted only when the following conditions are met:
 - Condition 1: The scaling group is not currently undergoing a scaling activity.
 - Condition 2: The number of ECS instances currently in the scaling group (Total Capacity) is 0.
 - If both of these conditions are met, the scaling group can be disabled and deleted.
- If ForceDelete is set to true, we recommend that you perform the following procedure:
 - Disable the scaling group to reject new scaling activity requests. Remove all ECS instances from the scaling group after the ongoing scaling activities are completed and then delete the scaling group. Manually added ECS instances are only removed from the scaling group, whereas automatically created instances are deleted.
- Note that deleting a scaling group also deletes its scaling configurations, rules, activities, and requests.
- The following tasks and instances are not deleted when a scaling group is deleted: scheduled tasks, CloudMonitor alert tasks, SLB instances, and RDS instances.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	dmIDKNcyWfzncX9MWX1****	The ID of the scaling group.
Action	String	No	DeleteScalingGroup	The operation that you want to perform. Set the value to DeleteScalingGroup.

Parameter	Type	Required	Example	Description
ForceDelete	Boolean	No	false	Indicates whether to forcibly delete the scaling group and remove or release its associated ECS instances regardless of whether it is undergoing a scaling activity. Default value: false, indicating that the scaling group is not to be forcibly deleted.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DeleteScalingGroup
&ScalingGroupId=dmIDKNcyWf_zncX9MWX1 ****
&<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned when 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 when the specified scaling group still contains ECS instances and cannot be deleted.

8.6 DescribeScalingGroups

You can call this operation to query scaling groups.

Debugging

Alibaba Cloud provides OpenAPI Explorer to simplify API usage. You can use OpenAPI Explorer to search for APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The ID of the region where a scaling group resides.
Action	String	No	DescribeScalingGroups	The operation that you want to perform. Set the value to DescribeScalingGroups.
PageNumber	Integer	No	1	The page number of the scaling group list 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.
ScalingGroupId . 1	String	No	dyrSuvBOtO1dEdIllbp****	The ID of scaling group N to be queried. Valid values of N: 1 to 20. The IDs of inactive scaling groups will not be displayed in the query result and no errors will be returned.
ScalingGroupName . 1	String	No	dyrSuvBOtO1dEdIllbp****	The name of scaling group N to be queried. Valid values of N: 1 to 20. The names of inactive scaling groups will not be displayed in the query result and no errors will be returned.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The current page number.
PageSize	Integer	50	The number of entries returned on each page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingGroups			The set of scaling groups.
ActiveCapacity	Integer	1	The number of Elastic Compute Service (ECS) instances that have been added to the scaling group and are running properly.
ActiveScalingConfigurationId	String	dyo713cNYIB4ddeVlKbc****	The ID of the active scaling configuration in the scaling group.
CreationTime	String	2014-08-14T10:58Z	The time when the scaling group was created.
DBInstanceIds		rdszzzyyuny****	The IDs of ApsaraDB for Relational Database Service (RDS) instances associated with the scaling group.
DefaultCooldown	Integer	60	The default cooldown period of the scaling group. During the cooldown period calculated for a scaling activity, the scaling group does not execute any other scaling activities triggered by CloudMonitor alarm tasks.

Parameter	Type	Example	Description
HealthCheckType	String	ECS	The health check mode of the scaling group. Valid values: - NONE: The system performs no health check. - ECS: The system performs the health check on ECS instances in the scaling group.
LaunchTemplateId	String	lt-m5e3ofjr1zn1aw7****	The ID of the instance launch template used by the scaling group.
LaunchTemplateVersion	String	true	The version of the instance launch template used by the scaling group.
LifecycleState	String	Active	The status of the scaling group. Valid values: - Active: Active scaling groups can receive requests to execute scaling rules and trigger scaling activities. - Inactive: Inactive scaling groups cannot receive any requests to execute scaling rules. - Deleting: Scaling groups that are being deleted cannot receive any requests to execute scaling rules and their parameters cannot be modified.
LoadBalancerIds		147b46d767c-cn-qingdao-cm5****	The IDs of Server Load Balancer (SLB) instances associated with the scaling group.
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.

Parameter	Type	Example	Description
ModificationTime	String	2014-08-14T10:58Z	The time when the scaling group was modified.

Parameter	Type	Example	Description
MultiAZPolicy	String	PRIORITY	The ECS instance scaling policy for a multi-zone scaling group. Valid values: • PRIORITY: ECS instances are scaled based on the VSwitchIds.N parameter. When an ECS instance cannot be created in the zone where the VSwitch with the highest priority resides, the system automatically uses the VSwitch with the next highest priority to create the ECS instance. • COST_OPTIMIZED: ECS instances are created based on the unit price of vCPU, from low to high. Preemptible instances are created first when preemptible instance types are specified for the scaling configuration. Pay-as-you-go instances are created automatically when all types of preemptible instances are unavailable due to issues such as insufficient ECS resources.  Note: COST_OPTIMIZED takes effect only when multiple instance types are specified or at least one preemptible instance type is specified. • BALANCE: ECS instances are distributed evenly in multiple zones specified in the scaling group. If ECS instances are unevenly distributed among the

Parameter	Type	Example	Description
OnDemandBaseCapacity	Integer	30	The minimum number of pay-as-you-go instances required in the scaling group. Valid values: 0 to 1000. When the number of pay-as-you-go instances is smaller than this value, the scaling group will attempt to create pay-as-you-go instances over other instances.
OnDemandPercentageAboveBaseCapacity	Integer	20	The percentage of pay-as-you-go instances to be created when instances are added to the scaling group. This parameter takes effect after the number of instances reaches the OnDemandBaseCapacity value. Valid values: 0 to 100.
PendingCapacity	Integer	0	The number of ECS instances that are being added to the scaling group, but are still being configured.
PendingWaitCapacity	Integer	1	The number of ECS instances that are in the pending state in 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	The ID of the region where the scaling group resides.

Parameter	Type	Example	Description
RemovalPolicies		OldestScalingConfiguration	The set of policies for removing ECS instances from the scaling group.
RemovingCapacity	Integer	0	The number of ECS instances that are currently 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	dyrSuvBOtO1dEdIlIbp****	The ID of the scaling group.
ScalingGroupName	String	dyrSuvBOtO1dEdIlIbp****	The name of the scaling group.
ScalingPolicy	String	recycle	The reclaim policy of the scaling group. Valid values: · recycle: The scaling group is set to the Shutdown and Reclaim Mode, which stops ECS instances and reclaims resources. · release: The scaling group is set to the Release Mode, which removes or adds ECS instances. For more information, see RemoveInstances.
SpotInstancePools	Integer	5	The number of available instance types. The scaling group will create preemptible instances of multiple instance types available at the lowest cost. Valid values: 0 to 10.

Parameter	Type	Example	Description
SpotInstanceRemedy	Boolean	true	Indicates whether supplementing preemptible instances was enabled when the target capacity of preemptible instances is not fulfilled. When receiving a system message that a preemptible instance will be reclaimed, the scaling group will create a new instance to replace the instance to be reclaimed if this parameter is set to true.
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 in the stopped state in the scaling group.
TotalCapacity	Integer	1	The total number of ECS instances in the scaling group.
VServerGroups			The list of VServer groups.
LoadBalancerId	String	147b46d767c-cn-qingdao-cm5****	The ID of the SLB instance with which the VServer group is associated.
VServerGroupAttributes			The attributes of the VServer group.
Port	Integer	22	The number of the port that is used by the SLB instance to provide public-facing services.

Parameter	Type	Example	Description
VServerGroupId	String	rsp-bp12bjrny****	The ID of the VServer group.
Weight	Integer	1	The weight of the VServer group.
VSwitchId	String	vpc-25j4g****	The ID of the VSwitch associated with the scaling group.
VSwitchIds		vpc-25j4g****	The set of IDs of the VSwitches associated with the scaling group. If you use the VSwitchIds parameter, the VSwitchId parameter is ignored.
VpcId	String	vpc-****	The ID of the VPC to which the scaling group belongs.
TotalCount	Integer	1	The total number of scaling groups.

Examples

Sample request

```
http://ess.aliyuncs.com/?Action=DescribeScalingGroups
&RegionId=cn-qingdao
&PageSize=50
&<Common request parameters>
```

Sample success response

XML format

```
<DescribeScalingGroupsResponse>
  <RequestId>6393C3A8-B611-42F2-AFA6-F080FC45D5D0</RequestId>
  <TotalCount>1</TotalCount>
  <PageNumber>1</PageNumber>
  <PageSize>10</PageSize>
  <ScalingGroups>
    <ScalingGroup>
      <ActiveCapacity>1</ActiveCapacity>
      <ActiveScalingConfigurationId>d90713cNYIB4ddEVLKbc****</ActiveScalingConfigurationId>
      <DBInstanceIds>
        <DBInstanceId>rdszzzyyun****</DBInstanceId>
      </DBInstanceIds>
    </ScalingGroup>
  </ScalingGroups>
</DescribeScalingGroupsResponse>
```

```

        </ DBInstance Ids >
        < VSwitchId > vpc - 25j4g ****</ VSwitchId >
        < DefaultCoolDown > 20 </ DefaultCoolDown >
        < LifecycleState > Active </ LifecycleState >
        < LoadBalancerIds >
            < LoadBalancerId > 147b46d767 c - cn -
qingdao - cm5 ****</ LoadBalancerId >
        </ LoadBalancerIds >
        < MaxSize > 1 </ MaxSize >
        < MinSize > 0 </ MinSize >
        < PendingCapacity > 0 </ PendingCapacity >
        < RegionId > cn - qingdao </ RegionId >
        < RemovingCapacity > 0 </ RemovingCapacity >
        < ScalingGroupId > dyrSuvB0t0 1dEdIlIbp ****</
ScalingGroupId >
        < ScalingGroupName > dyrSuvB0t0 1dEdIlIbp ****</
ScalingGroupName >
        < RemovalPolicies >
            < RemovalPolicy > OldestScalingConfigu
ration </ RemovalPolicy >
            < RemovalPolicy > OldestInstance </
RemovalPolicy >
        </ RemovalPolicies >
        < TotalCapacity > 1 </ TotalCapacity >
        < CreationTime > 2014 - 08 - 14T10 : 58Z </
CreationTime >
    </ ScalingGroup >
</ ScalingGroups >
</ DescribeScalingGroupsResponse >

```

JSON format

```

{
  "PageNumber": 1,
  "TotalCount": 1,
  "PageSize": 10,
  "RequestId": "68386699 - 8B9E - 4D5B - BC4C - 75A28F6C2A 00",
  "ScalingGroups": {
    "ScalingGroup": [
      {
        "DefaultCoolDown": 300,
        "MaxSize": 2,
        "RemovalPolicies": {
          "RemovalPolicy": [
            "OldestScalingConfiguration",
            "OldestInstance"
          ]
        },
        "PendingCapacity": 0,
        "RemovingCapacity": 0,
        "VSwitchId": "vpc - 25j4g ****",
        "ScalingGroupName": "b8pYCVbIV5 k9cz4PWpbe ****",
        "CreationTime": "2014 - 08 - 14T10 : 58Z",
        "ActiveCapacity": 0,
        "ActiveScalingConfigurationId": "d9o713cNYI B4ddEVlKbc
****",
        "ScalingGroupId": "b8pYCVbIV5 k9cz4PWpbe ****",
        "RegionId": "cn - qingdao",
        "TotalCapacity": 0,
        "DBInstanceIds": {
          "DBInstanceId": [
            "rdsia3u3yi a ****",
            "rdszzzyyun y ****"
          ]
        }
      }
    ]
  }
}

```

```
]
},
"LoadBalancerIds":{
  "LoadBalancerId":[
    "147b46d767c-cn-qingdao-cm5-****"
  ]
},
"MinSize":1,
"LifecycleState":"Inactive"
}
]
```

Error codes

For more information about error codes, visit [API Error Center](#).

8.7 DescribeScalingInstances

You can call this operation to query the list of ECS instances in a scaling group and list details about the instances.

Description

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

Debugging

[OpenAPI Explorer](#) automatically calculates the signature value. For your convenience, we recommend that you call this operation in [OpenAPI Explorer](#). [OpenAPI Explorer](#) dynamically generates the sample code of the operation for different SDKs.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group.
Action	String	No	DescribeScalingInstances	The operation that you want to perform. Set the value to DescribeScalingInstances.

Parameter	Type	Required	Example	Description
CreationType	String	No	AutoCreated	Specifies how to create an ECS instance. Valid values: • AutoCreated: The ECS instance is automatically created by a scaling group based on scaling configurations and rules. • Attached: The ECS instance is manually added to a scaling group.
HealthStatus	String	No	Healthy	The health status of the ECS instance in the scaling group. The ECS instance is unhealthy if it is not in the Running state. Valid values: • Healthy: a healthy ECS instance. • Unhealthy: an unhealthy ECS instance. Auto Scaling automatically removes unhealthy ECS instances from the scaling group. Auto Scaling stops and releases only automatically created ECS instances.
InstanceId • 1	String	No	i-281vv****	The ID of ECS instance N. Valid values of N: 1 to 20. The IDs of inactive instances will not be displayed in the query result and no errors will be returned.

Parameter	Type	Required	Example	Description
LifecycleState	String	No	InService	The lifecycle status of the ECS instance in the scaling group. Valid values: • InService: The ECS instance has been added to the scaling group and is providing services normally. • Pending: The ECS instance is being added to the scaling group, but is still being configured. For example, the instance is being created, attached to an SLB instance, or added to an RDS whitelist. • Removing: The ECS instance is being removed from the scaling group.
PageNumber	Integer	No	1	The page number of the ECS instance list 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.
ScalingConfigurationId	String	No	asc-****	The ID of the associated scaling configuration.
ScalingGroupId	String	No	asg-****	The ID of the scaling group.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The current page number.
PageSize	Integer	10	The number of entries returned on each page.
RequestId	String	E481B3F5-F290-4F3E-AB40-F980FB8E098A	The ID of the request.
ScalingInstances			The set of ECS instances.
CreationTime	String	2019-03-06T08:06Z	The time when the ECS instance was added to the scaling group.
CreationType	String	AutoCreated	Indicates how the ECS instance was created. Valid values: - AutoCreated: The ECS instance is automatically created by a scaling group based on scaling configurations and rules. - Attached: The ECS instance is manually added to a scaling group.
HealthStatus	String	Healthy	The health status of the ECS instance in the scaling group. Valid values: - Healthy: a healthy ECS instance. - Unhealthy: an unhealthy ECS instance.
InstanceId	String	i-283vv****	The ID of the ECS instance.
LaunchTemplateId	String	lt-m5e3ofjr1zn1aw7****	The ID of the instance launch template.

Parameter	Type	Example	Description
LaunchTemplateVersion	String	1	The version number of the instance launch template.
LifecycleState	String	InService	The lifecycle status of the ECS instance in the scaling group. Valid values: • InService: The ECS instance has been added to the scaling group and is providing services normally. • Pending: The ECS instance is being added to the scaling group, but is still being configured. For example, the instance is being created, attached to an SLB instance, or added to an RDS whitelist. • Removing: The ECS instance is being removed from the scaling group.
LoadBalancerWeight	Integer	1	The weight of the SLB instance.
ScalingConfigurationId	String	asg-****	The ID of the associated scaling configuration.
ScalingGroupId	String	asg-****	The ID of the scaling group to which an ECS instance belongs.
WarmupState	String	NoNeedWarmup	The warmup state of the ECS instance. Valid values: • NoNeedWarmup: No warmup is required. • WaitingForInstanceWarmup: You must wait for the instance to finish warmup. • InstanceWarmupFinish: The warmup is completed.

Parameter	Type	Example	Description
TotalCount	Integer	1	The total number of ECS instances.

Examples

Sample requests

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

Sample success responses

XML format

```
<DescribeScalingInstancesResponse>
  <RequestId>DFF8797F-5B73-4BD7-A7D0-03479C458F7A</RequestId>
  <TotalCount>1</TotalCount>
  <PageNumber>1</PageNumber>
  <PageSize>10</PageSize>
  <ScalingInstances>
    <ScalingInstance>
      <CreationTime>2014-08-15T17:37Z</CreationTime>
      <CreationType>AutoCreated</CreationType>
      <HealthStatus>Healthy</HealthStatus>
      <InstanceId>i-283vv****</InstanceId>
      <LifecycleState>InService</LifecycleState>
      <ScalingConfigurationId>cGsGHrdMBa3DcDrrBVcc****</ScalingConfigurationId>
      <ScalingGroupId>dBCYxE26IH-KcGq1xPcT****</ScalingGroupId>
    </ScalingInstance>
  </ScalingInstances>
</DescribeScalingInstancesResponse>
```

JSON format

```
{
  "PageNumber": 1,
  "TotalCount": 1,
  "PageSize": 50,
  "RequestId": "13305F2D-A4C2-4E6B-B7C7-0F2150842EA3",
  "ScalingInstances": {
    "ScalingInstance": [
      {
        "CreationTime": "2014-08-14T10:59Z",
        "ScalingConfigurationId": "bU5uZHcAgtzwcL4IeDe****",
        "CreationType": "AutoCreated",
        "HealthStatus": "Healthy",
        "ScalingGroupId": "dE9YbOdCHqaFdFZXHVdD****",
        "InstanceId": "i-28sov****",
        "LifecycleState": "InService"
      }
    ]
  }
}
```

```
]
}
```

Error codes

For more information about error codes, visit [API Error Center](#).

8.8 DescribeScalingActivities

You can call this operation to query scaling activities.

Description

You can specify a scaling group ID to query all scaling activities in this scaling group.

You can filter query results based on the status of scaling activities.

Only scaling activities during the last 30 days can be returned.

Debugging

[OpenAPI Explorer](#) automatically calculates the signature value. For your convenience , we recommend that you call this operation in [OpenAPI Explorer](#). [OpenAPI Explorer](#) dynamically generates the sample code of the operation for different SDKs.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group to which a scaling activity belongs.
Action	String	No	DescribeScalingActivities	The operation that you want to perform. Set the value to DescribeScalingActivities.
PageNumber	Integer	No	1	The page number of the scaling activity list to return. Pages start from page 1. Default value: 1.

Parameter	Type	Required	Example	Description
PageSize	Integer	No	50	The number of entries to return on each page. Maximum value: 50. Default value: 10.
ScalingActivityId . 1	String	No	ebta5WbUzC8gcwUWvfch****	The ID of scaling activity N to be queried. Valid values of N: 1 to 20.
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 was successfully executed. • Warning: The scaling activity was partially executed. • Failed: The scaling activity failed. • InProgress: The scaling activity is currently being executed. • Rejected: The scaling activity request was rejected.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The current page number.
PageSize	Integer	50	The number of entries returned on each page.

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingActivities			The set of scaling activities.
AttachedCapacity	String	1	The total number of instances that you added manually to the scaling group after the scaling activity was completed.
AutoCreatedCapacity	String	1	The total number of the instances that were added automatically by the scaling group after the scaling activity was completed.
Cause	String	A scheduled task execute scaling rule \"sr****\", 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 ended.
Progress	Integer	100	The execution progress of the scaling activity.
ScalingActivityId	String	ebta5WbUzC8gcwUWvfch****	The ID of the scaling activity.
ScalingGroupId	String	AG6CQdPU80KdwLjgZcJ****	The ID of the scaling group.

Parameter	Type	Example	Description
ScalingInstanceNumber	Integer	1	The number of instances that were created or restarted from the shutdown and reclaim mode during the scale-out event. The number of instances that were deleted or switched to the shutdown and reclaim mode during the scale-in event.
StartTime	String	2014-08-17T12:39Z	The time when the scaling activity started.
StatusCode	String	2014-08-17T12:39Z	The status of the scaling activity. Valid values: • Successful: The scaling activity was successfully executed. • Warning: The scaling activity was partially executed. • Failed: The scaling activity failed. • InProgress: The scaling activity is currently being executed. • Rejected: The scaling activity request was rejected.
StatusMessage	String	"1" ECS instances is added.	The details about the scaling activity in a state.
TotalCapacity	String	2	The total number of instances in the scaling group after the scaling activity was completed.
TotalCount	Integer	1	The total number of scaling activities.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DescribeScalingActivities
&RegionId=cn-qingdao
&PageSize=50
<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 execute scaling rule "sr****", changing the Total 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>AG6CQdPU80KdwLjgZcJ****</ScalingGroupId>
      <StartTime>2014-08-17T12:39Z</StartTime>
      <StatusCode>Successful</StatusCode>
      <StatusMessage>"1" ECS instances is added.</StatusMessage>
    </ScalingActivity>
  </ScalingActivities>
</DescribeScalingActivitiesResponse>
```

JSON format

```
{
  "PageNumber": 1,
  "TotalCount": 2582,
  "PageSize": 1,
  "RequestId": "0A016AD6-A91F-4210-9DEE-D5BBD4270F45",
  "ScalingActivities": {
    "ScalingActivity": [
      {
        "StatusCode": "Successful",
        "Description": "Add \"1\" ECS instance",
        "ScalingGroupId": "c7acXJbAJm pPcGE7G3bw****",
        "StatusMessage": "\"1\" ECS instances is added.",
        "EndTime": "2014-08-18T20:49Z",
        "StartTime": "2014-08-18T20:49Z",
        "Progress": 100,

```

```

    "Cause ":" A scheduled task execute scaling rule \"
sr ****\", changing the Total Capacity from \" 0 \" to \"
1 \".",
    "ScalingActivityId ":" bBx0R5dhKF occGkbTrcy ****"
  }
}
}
}

```

Error codes

For more information about error codes, visit [API Error Center](#).

8.9 AttachLoadBalancers

You can call this operation to add one or more SLB instances.

Description

Due to [limits](#), the following conditions must be met when you add an SLB instance to a scaling group:

- The SLB instance and the scaling group must be in the same account.
- The SLB instance and the scaling group must be in the same region.
- The SLB instance must be in the Running state.
- The SLB instance must have at least one listener with health check enabled.
- The SLB instance and the scaling group must be in the same VPC if the network type of both of them is VPC.
- If the network type of the scaling group is VPC, the network type of the SLB instance is classic network, and the SLB back-end server contains a VPC instance, the VPC instance and the scaling group must be in the same VPC.
- The number of added SLB instances must be less than the quota of instances in the scaling group.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
LoadBalancerIds	RepeatList	Yes	lb-2zeur05gfs****	The IDs of SLB instances. Up to five SLB instances can be added at a time.
ScalingGroupId	String	Yes	AG6CQdPU80KdwLjgZcJ****	The ID of the scaling group.
Action	String	No	AttachLoadBalancers	The operation that you want to perform. Set the value to AttachLoadBalancers.
ForceAttach	Boolean	No	false	Indicates whether to add all instances in the current scaling group to the SLB back-end server: · true: Add all instances to the back-end server. · false: Do not add any instance to the back-end server. Default value: false.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=AttachLoadBalancers&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
```

```
& LoadBalancer.1=lb-2zeur05gfs ****
< Common request parameters >
```

Successful response examples

XML format

```
< AttachLoadBalancersResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
</ AttachLoadBalancersResponse >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

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 when Auto Scaling is not authorized to call the specified operation.
404	InvalidScalingGroupId. NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded. LoadBalancer	LoadBalancer quota exceeded in the scaling group "%s".	The error message returned when the number of SLB instances in the scaling group exceeds the quota.
404	InvalidLoadBalancerId. NotFound	The load balancer "%s" does not exist.	The error message returned when the specified SLB instance does not exist.
400	InvalidLoadBalancerId. RegionMismatch	The load balancer "%s" and the specified scaling group are not in the same Region.	The error message returned when the SLB instance and the scaling group are not in the same region.
400	IncorrectLoadBalancerStatus	The current status of the load balancer "%s" does not support this action.	The error message returned when the SLB instance is in a state that does not support the current operation.
400	IncorrectLoadBalancerHealthCheck	The current health check type of the load balancer "%s" does not support this action.	The error message returned when health check is not enabled for the specified SLB instance.

HTTP status code	Error code	Error message	Description
400	InvalidLoadBalancerId. VPCMismatch	The specified virtual switch and the instance in the load balancer "%s" are not in the same VPC.	The error message returned when the SLB instance and the scaling group are not in the same VPC.
400	QuotaExceeded. BackendServer	Backend server quota exceeded in the load balancer "%s".	The error message returned when the number of back-end servers of the SLB instance exceeds the quota.
404	InvalidScalingConfigurationId. NotFound	The specified scaling configuration does not exist.	The error message returned when the specified scaling configuration that is enabled for the current scaling group does not exist.

8.10 DetachLoadBalancers

You can call this operation to remove one or more SLB instances.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
LoadBalancerIds	RepeatList	Yes	lb-2zeur05gfs****	The IDs of SLB instances. Up to five SLB instances can be removed at a time.
ScalingGroupId	String	Yes	AG6CQdPU80KdwLjgZcJ****	The ID of the scaling group.
Action	String	No	DetachLoadBalancers	The operation that you want to perform. Set the value to DetachLoadBalancers.
ForceDetach	Boolean	No	false	Indicates whether to remove instances in the current scaling group from the back-end servers of the SLB instance: · true: Remove the instances. · false: Do not remove the instances. Default value: false.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DetachLoadBalancers
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&LoadBalancerIds.1=lb-2zeur05gfs****
```

&< Common request parameters >

Successful response examples

XML format

```
< DetachLoad BalancersR esponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E
</ RequestId >
</ DetachLoad BalancersR esponse >
```

JSON format

```
{
  " RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

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 when Auto Scaling is not authorized to call the specified operation.
404	InvalidScalingGroupId. NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.
404	InvalidLoadBalancerId. NotFound	The Load Balancer "%s" does not exist.	The error message returned when the specified SLB instance does not exist in the scaling group.

8.11 DetachDBInstances

You can call this operation to remove one or more RDS instances.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
DBInstance · N	RepeatList	Yes	rm-bp12cy3****	The IDs of RDS instances. Up to five RDS instances can be removed at a time.
ScalingGroup upId	String	Yes	AG6CQdPU8OKdwLjgZcj****	The ID of the scaling group.
Action	String	No	DetachDBInstances	The operation that you want to perform. Set the value to DetachDBInstances.
ForceDetach	Boolean	No	false	Indicates whether to remove the private IP addresses of instances in the scaling group from the IP whitelist of the RDS instance: · true: Remove the private IP addresses. · false: Do not remove the private IP addresses. Default value: false.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DetachDBInstances
&ScalingGroupId=AG6CQdPU80KdWLjgZcJ****
&DBInstanceId=rm-bp12cy3****
&<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.

HTTP status code	Error code	Error message	Description
400	InvalidDBInstanceId. NotFound	DB instance "%s" does not exist.	The error message returned when the specified RDS instance does not exist.

8.12 AttachDBInstances

You can call this operation to add one or more RDS instances.

Description

Due to limits, the following conditions must be met when you add an RDS instance to a scaling group:

- The SLB instance and the scaling group must belong to the same account.
- The RDS instance must be in the Unlocked state. For more information about the lock policy, see [RDS usage instructions](#).
- The SLB instance must be in the Running state.
- After an RDS instance is added, the default group of the RDS IP whitelist can contain up to 1,000 IP addresses. For more information about the IP whitelist, see [Configure the whitelist](#).

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
DBInstanceIds • N	RepeatList	Yes	rm-bp12cy3****	The IDs of RDS instances. Up to five RDS instances can be added at a time.
ScalingGroupID	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
Action	String	No	AttachDBInstances	The operation that you want to perform. Set the value to AttachDBInstances.
ForceAttach	Boolean	No	false	Indicates whether to add the private IP addresses of all instances in the current scaling group to the IP whitelist of the RDS instance: · true: Add the private IP addresses to the IP whitelist. · false: Do not add the private IP addresses to the IP whitelist. Default value: false.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=AttachDBInstances
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&DBInstanceId=rm-bp12cy3****
&<Common request parameters>
```

Successful response examples

XML format

```
<AttachDBInstancesResponse>
```

```
< RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
RequestId >
</ AttachDBIn stancesRes ponse >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.
400	QuotaExceeded.RDS	"RDS" quota exceeded.	The error message returned when the number of RDS instances in the scaling group exceeds the quota.
400	InvalidDBInstanceId.NotFound	The specified value of parameter "%s" is not valid.	The error message returned when the specified RDS instance does not exist.

HTTP status code	Error code	Error message	Description
400	IncorrectDBInstanceStatus	The current status of DB instance "%s" does not support this action.	The error message returned when the RDS instance is in a state that does not support the current operation.
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "%s".	The error message returned when the number of back-end IP addresses in the whitelist of the RDS instance exceeds the quota.
400	InvalidInstanceIds.PrivateIpNotFound	Can not find all private ips of instances in specific scaling group.	The error message returned when the private IP addresses of an RDS instance in the group cannot be obtained.

8.13 AttachVServerGroups

You can call this operation to add one or more VServer groups to an SLB instance.

Description

Due to the [limits](#) on SLB, the following conditions must be met when you add a VServer group to a scaling group:

- The SLB instance and the scaling group must belong to the same account.
- The SLB instance and the scaling group must be in the same region.
- The SLB instance must be in theRunningstate.
- The SLB instance must have at least one listener with health check enabled.

- The SLB instance and the scaling group must be in the same VPC if both of them are of the VPC network type.
- If the network type of the scaling group is VPC, the network type of the SLB instance is classic network, and the VServer group contains a VPC instance, the VPC instance and the scaling group must be in the same VPC.
- The VServer group that you want to add to the scaling group must be associated with the SLB instance.
- Currently, up to five VServer groups can be added to a scaling group.

**Note:**

A VServer group is uniquely identified by the combination of the SLB instance ID (LoadBalancerId),

- VServer group ID (VServerGroupId),
- and port number of the VServer group (Port). If you have added a VServer group to a scaling group by using different port numbers, Auto Scaling considers that multiple VServer groups are added to the scaling group. If multiple VServer groups that have the same group ID and port number are specified in the request parameters,
- only the first VServer group is used. The other VServer groups are ignored.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-hangzhou	The ID of the region where the scaling group was created, such as cn-hangzhou or cn-shanghai. For more information, see Regions and zones .

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	asg-****	The ID of the scaling group.
Action	String	No	AttachVServerGroups	The operation that you want to perform. Set the value to AttachVServerGroups.
ForceAttach	Boolean	No	false	Indicates whether the ECS instances that belong to the scaling group are to be added to the newly attached VServer group. · true: Add the ECS instances to the VServer group. · false: Do not add the ECS instances to the VServer group. Default value: false.
VServerGroup.N.LoadBalancerId	String	No	lb-****	The ID of the SLB instance with which the VServer group is associated. N represents the SLB instance number. Valid values of N: 1 to 5.

Parameter	Type	Required	Example	Description
VServerGroup . N . VServerGroupAttribute . N . Port	Integer	No	22	The port number used by Auto Scaling to add the ECS instances to the VServer group . Valid values: 1 to 65,535. N represents the SLB instance number. Valid values of N : 1 to 5. M represents the VServer group number that is associated with the SLB instance. Valid values of M: 1 to 5.
VServerGroup . N . VServerGroupAttribute . N . VServerGroupId	String	No	rsp-****	The ID of the VServer group. N represents the SLB instance number. Valid values of N : 1 to 5. M represents the VServer group number that is associated with the SLB instance. Valid values of M: 1 to 5.
VServerGroup . N . VServerGroupAttribute . N . Weight	Integer	No	100	The weight set for the ECS instances that are added to the VServer group. Valid values: 0 to 100. Default value: 50. N represents the SLB instance number. Valid values of N : 1 to 5. M represents the VServer group number that is associated with the SLB instance. Valid values of M: 1 to 5.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=AttachVServerGroups
&ScalingGroupUpId=asg-****
&RegionId=cn-hangzhou
&VServerGroupUpId=1-LoadBalancerId=lb-****
&VServerGroupUpId=1-VServerGroupUpAttribute.1.VServerGroupUpId=rsp-****
&VServerGroupUpId=1-VServerGroupUpAttribute.1.Port=22
&VServerGroupUpId=1-VServerGroupUpAttribute.1.Weight=100
&<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

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 when Auto Scaling is not authorized to call the specified operation.
404	InvalidScalingGroupId. NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.
404	InvalidLoadBalancerId. NotFound	The load balancer "%s" does not exist.	The error message returned when the specified SLB instance does not exist in the current account.
400	InvalidLoadBalancerId. RegionMismatch	The load balancer "%s" and the specified scaling group are not in the same Region.	The error message returned when the SLB instance and the scaling group are not in the same region.
400	IncorrectLoadBalancerStatus	The current status of the load balancer "%s" does not support this action.	The error message returned when the SLB instance is in a state that does not support the current operation.

HTTP status code	Error code	Error message	Description
400	IncorrectLoadBalancerHealthCheck	The current health check type of the load balancer "%s" does not support this action.	The error message returned when health check is not enabled for the specified SLB instance.
400	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in the load balancer "%s" are not in the same VPC.	The error message returned when the SLB instance and the scaling group are not in the same VPC.
400	InvalidVServerGroupId.ForLoadBalancer	Invalid VServerGroupId For LoadBalancer "%s".	The error message returned when the VServer group that is specified by the VServerGroupId parameter is not associated with the specified SLB instance.
400	QuotaExceeded.VServerGroup	VServerGroup quota exceeded in the specified scaling group.	The error message returned when the maximum number of VServer groups that can be attached to the scaling group has been reached.

8.14 DetachVServerGroups

You can call this operation to remove one or more VServer groups.

Description

To remove a VServer group from a scaling group, you must specify the SLB instance ID (LoadBalancerId),

- the VServer group ID (VServerGroupId),
- and the port number of the VServer group (Port). The VServer group is removed if the VServer group specified by the request parameters matches the VServer group in the scaling group. Otherwise, the request is ignored and no error message will be returned by the operation.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-hangzhou	The ID of the region where the scaling group is created, such as cn-hangzhou or cn-shanghai. For more information, see Regions and zones .
ScalingGroupId	String	Yes	asg-****	The ID of the scaling group.
Action	String	No	DetachVServerGroups	The operation that you want to perform. Set the value to DetachVServerGroups.

Parameter	Type	Required	Example	Description
ForceDetach	Boolean	No	false	Indicates whether the ECS instances that belong to the specified scaling group are to be removed from the VServer being detached. · true: Remove the ECS instances. · false: Do not remove the ECS instances. Default value: false.
VServerGroup.N.LoadBalancerId	String	No	lb-****	The ID of the SLB instance with which the VServer group is associated. N represents the SLB instance number. Valid values of N: 1 to 5.
VServerGroup.N.VServerGroupAttribute.N.Port	Integer	No	22	The port number used by Auto Scaling to add the ECS instances to the VServer group. Valid values: 1 to 65,535. N represents the SLB instance number. Valid values of N: 1 to 5. M represents the VServer group number that is associated with the SLB instance. Valid values of M: 1 to 5.

Parameter	Type	Required	Example	Description
VServerGroup.N.VServerGroupAttribute.N.VServerGroupId	String	No	rsp-****	The ID of the VServer group. N represents the SLB instance number. Valid values of N: 1 to 5. M represents the VServer group number that is associated with the SLB instance. Valid values of M: 1 to 5.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DetachVServerGroups
&ScalingGroupId=asg-****
&RegionId=cn-hangzhou
&VServerGroupId.1.LoadBalancerId=lb-****
&VServerGroupId.1.VServerGroupAttribute.1.VServerGroupId=rsp-****
&VServerGroupId.1.VServerGroupAttribute.1.Port=22
&VServerGroupId.1.VServerGroupAttribute.2.VServerGroupId=rsp-****
&VServerGroupId.1.VServerGroupAttribute.2.Port=80
&<Common request parameters>
```

Successful response examples

XML format

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

```
</ DetachVServerGroupsResponse >
```

JSON format

```
{
  "requestId": "473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

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 when Auto Scaling is not authorized to call the specified operation.
404	InvalidScalingGroupId. NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.
400	InvalidParameter	The specified value of parameter "%s" is not valid.	The error message returned when the value of the parameter is invalid.
400	MissingParameter	The input parameter "%s" that is mandatory for processing this request is not supplied.	The error message returned when the required parameter is not specified.

9 Scaling configuration

9.1 DeleteScalingConfiguration

You can call this operation to delete a scaling configuration.

Description

Active scaling configurations in a scaling group cannot be deleted.

If any ECS instances created based on a scaling configuration are still in the scaling group, the scaling configuration cannot be deleted.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingConfigurationId	String	Yes	eOs27Kb0oXvQcUYjEGel****	The ID of the scaling configuration.
Action	String	No	DeleteScalingConfiguration	The operation that you want to perform. Set the value to DeleteScalingConfiguration.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DeleteScalingConfiguration
&ScalingConfigurationId=e0s27Kb0oXvQcUYjEGel****
&<Common request parameters>
```

Successful response examples

XML format

```
<DeleteScalingConfigurationResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8-3E</
  RequestId>
</DeleteScalingConfigurationResponse>
```

JSON format

```
{
  "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8-3E"
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	The error message returned when the specified scaling configuration does not exist in the current account.
400	IncorrectScalingConfigurationLifecycleState	The current lifecycle state of specified scaling configuration does not support this action.	The error message returned when the specified scaling configuration is in a state other than inactive.

HTTP status code	Error code	Error message	Description
400	InstanceInUse	You cannot delete a scaling configuration or scaling group while there is an instance associated with it.	The error message returned when the specified scaling configuration still has associated ECS instances.

9.2 CreateScalingConfiguration

You can call this operation to create a scaling configuration.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group to which the scaling configuration to be created belongs.
SecurityGroupId	String	Yes	sg-280ih****	The ID of the security group to which created instances belong. Mutual access is allowed between instances in a security group.
Action	String	No	CreateScalingConfiguration	The operation that you want to perform. Set the value to CreateScalingConfiguration.

Parameter	Type	Required	Example	Description
Cpu	Integer	No	2	The number of vCPUs. You can specify the number of CPU cores and the amount of memory to define the range of instance types. For example, instance types with a specified value of CPU as 2 and Memory as 16 can be defined as 2 vCPU 16 GiB. Auto Scaling uses factors such as I/O optimization and zone to determine a set of available instance types. Auto Scaling then creates instances of the type that is most cost-effective out of the available types.  Note: This instance type range takes effect only when cost optimization is enabled and the scaling configuration does not have a specified instance type.

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: indicates a basic disk. The DeleteWithInstance property of a basic disk created along with the instance is true. - cloud_efficiency: indicates an ultra disk. - cloud_ssd: indicates an SSD disk. - ephemeral_ssd: indicates a local SSD disk. Default value: cloud.
DataDisk. N. DeleteWithInstance	Boolean	No	true	Indicates whether data disk N is to be released with its attached instance. Valid values: - true: Release the data disk with its attached instance. - false: Retain the data disk when the attached instance is released. Default value: true. This parameter is valid only for independent disks, for which the DataDisk.N.Category parameter is set to cloud, cloud_efficiency, or cloud_ssd. An error will be returned if you set this parameter for other disks.

Parameter	Type	Required	Example	Description
DataDisk . N . Description	String	No	FinanceDept	The description of data disk N . It must be 2 to 256 characters in length and cannot start with "http://" or "https://."
DataDisk . N . Device	String	No	/dev/xvdb	The mount point of data disk N. Valid values of N: 1 to 4. If this parameter is not specified , ECS automatically allocates a mount point to created ECS instances. The name of the mount point ranges from /dev/xvdb to /dev/xvdz in an alphabetical order.
DataDisk . N . DiskName	String	No	cloud_ssdData	The name of data disk N. It must be 2 to 128 characters in length. It must start with a letter and cannot start with "http://" or "https://." It can contain letters, digits, colons (:), underscores (_), and hyphens (-). Default value: empty.
DataDisk . N . Encrypted	String	No	false	Indicates whether data disk N is to be encrypted. Default value: false.
DataDisk . N . KMSKeyId	String	No	0e478b7a-4262-4802-b8cb-00d3fb40826X	The KMS key ID corresponding to the data disk.

Parameter	Type	Required	Example	Description
DataDisk . N . Size	Integer	No	100	The size of data disk N. Unit: GB. Valid values: · cloud: 5 to 2000 · cloud_efficiency: 20 to 32768 · · cloud_ssd: 20 to 32768. · ephemeral_ssd: 5 to 800. If this parameter is specified , the data disk size must be greater than or equal to the size of the snapshot specified by SnapshotId.
DataDisk . N . SnapshotId	String	No	s-280s7****	The ID of the snapshot used to create data disk N. After this parameter is specified , DataDisk.N.Size is ignored and the size of the created disk is the size of the specified snapshot. You cannot specify a snapshot that was created on or before July 15, 2013.
Deployment SetId	String	No	ds-bp1frxuzdg 87zh4pz****	The ID of the deployment set.

Parameter	Type	Required	Example	Description
HostName	String	No	Host****	The hostname of the ECS instance. It cannot start or end with a period (.) or hyphen (-). It cannot have two or more consecutive periods (.) or hyphens (-). Naming rules: - Windows-based instances: It must be 2 to 15 characters in length. It can contain uppercase or lowercase letters, digits, and hyphens (-). It cannot contain periods (.) or be entirely composed of digits. - Other instances such as Linux instances: It must be 2 to 64 characters in length. It can be segments separated by periods (.). Each segment can contain uppercase or lowercase letters, digits, and hyphens (-).
ImageId	String	No	centos6u5_64_20G_ali_aegis_2014_0703.vhd	The ID of the image file that was specified when the instance was started.
ImageName	String	No	myimage	The name of the image file. Image names must be unique within a region. This parameter is ignored if ImageId is specified. Marketplace images cannot be specified by the ImageName parameter.

Parameter	Type	Required	Example	Description
InstanceName	String	No	Instance****	The name of the instance that is created based on the current scaling configuration.
InstanceType	String	No	ecs.t1.xsmall	The single instance type from which ECS instances are to be created.
InstanceTypes . N	RepeatList	No	ecs.t1.xsmall. 1	The multiple instance types from which ECS instances are to be created. InstanceType is ignored if this parameter is specified. Valid values of N: 1 to 10. This means that up to 10 instance types can be configured for a scaling configuration. N represents the priority of the instance type in the scaling configuration. A lower value of N indicates a higher priority for the instance type. If an instance of the highest priority type cannot be created, Auto Scaling will create an instance of the next highest priority type.

Parameter	Type	Required	Example	Description
InternetChargeType	String	No	PayByBandwidth	The bandwidth billing type of created instances. Valid values: • PayByBandwidth: Users pay for the maximum available bandwidth specified by the <code>InternetMaxBandwidthOut</code> parameter. • PayByTraffic Users pay for the actual traffic used. The <code>InternetMaxBandwidthOut</code> parameter only specifies the upper limit of available bandwidth when this parameter is specified. Default value: <code>PayByBandwidth</code> for classic networks or <code>PayByTraffic</code> for VPCs.
InternetMaxBandwidthIn	Integer	No	100	The maximum inbound bandwidth from the public network. Unit: Mbit/s. Valid values: 1 to 200. Default value: 200. This parameter is not used for billing because the inbound traffic to instances is free of charge.

Parameter	Type	Required	Example	Description
InternetMaxBandwidthOut	Integer	No	50	The maximum outbound bandwidth to the public network. Unit: Mbit/s. Valid values: 0 to 100. • If InternetChargeType is set to PayByBandwidth and this parameter is not specified, this parameter is automatically set to 0. • If InternetChargeType is set to PayByTraffic and this parameter is not specified, an error is returned.
IoOptimized	String	No	optimized	Indicates whether the instance is I/O optimized. This parameter is unavailable for non I/O-optimized instances. Valid value: • optimized • Empty: - Generation I instance types are not I/O optimized by default. - Instance types of later generations are I/O optimized by default.
KeyPairName	String	No	KeyPairTest	The name of the key pair. • Ignore this parameter if you are creating a Windows-based ECS instance. This parameter is null by default. • Password logon is disabled for Linux-based ECS instances by default.

Parameter	Type	Required	Example	Description
LoadBalancerWeight	Integer	No	50	The weight of the back-end server. Valid values: 0 to 100. Default value: 50.
Memory	Integer	No	16	The size of memory. You can specify the number of CPU cores and the amount of memory to define the range of instance types. For example, instance types with a specified value of CPU as 2 and Memory as 16 can be defined as 2 vCPU 16 GiB. Auto Scaling uses factors such as I/O optimization and zone to determine a set of available instance types. Auto Scaling then creates instances of the type that is most cost-effective out of the available types.  Note: This instance type range takes effect only when cost optimization is enabled and the scaling configuration does not have a specified instance type.

Parameter	Type	Required	Example	Description
Password	String	No	123-abcABC	The password of the instance . It must be 8 to 30 characters in length. It must include at least three of the following characters types: uppercase letters, lowercase letters, digits, and special characters. Special characters include: <pre>() ` ~ ! @ # \$ % ^ & * - _ + = \ { } [] : ; ' < > , . ? /</pre> Passwords for Windows-based instances cannot start with a forward slash (/).  Note: We recommend that you use HTTPS to send requests if the Password parameter is specified to avoid password leakage.
PasswordInherit	Boolean	No	false	Indicates whether to use the preconfigured password of the specified image. To use this parameter, ensure that a password is configured for the specified image.
RamRoleName	String	No	RamRoleTest	The name of the instance RAM role. This name is provided and maintained by RAM. Call ListRoles to query the list of RAM role names. For more information, see CreateRole .

Parameter	Type	Required	Example	Description
ResourceGroupId	String	No	rg-resource****	The ID of the resource group.
ScalingConfigurationName	String	No	Test_sc	The name of the scaling configuration. It must be 2 to 40 characters in length. It must start with a digit, uppercase letter, or lowercase letter. It can contain letters, digits, underscores (_), hyphens (-), and periods (.). Scaling configuration names must be unique within a scaling group. If this parameter is not specified, the value of ScalingConfigurationId is used.
SecurityEnhancementsStrategy	String	No	Active	Indicates whether to enable security hardening. Valid values: - Active: Enables security hardening for public images. - Deactive: Disables security hardening for all images.
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 when the SpotStrategy parameter is set to SpotWithPriceLimit.

Parameter	Type	Required	Example	Description
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 when the SpotStrategy parameter is set to SpotWithPriceLimit.
SpotStrategy	String	No	NoSpot	The preemption policy to be used for pay-as-you-go instances. Valid values: • NoSpot: indicates a pay-as-you-go instance. • SpotWithPriceLimit: indicates a preemptible instance with a maximum price. • SpotAsPriceGo: indicates an instance with its price based on the market price at the time of purchase. Default value: NoSpot.
SystemDisk . Category	String	No	cloud_ssd	The type of the system disk. Valid values: • cloud: indicates a basic disk. • cloud_efficiency: indicates an ultra disk. • cloud_ssd: indicates an SSD disk. • ephemeral_ssd: indicates a local SSD disk. Default value: cloud for Generation I type instances that are not I/O optimized, and cloud_efficiency for other types.

Parameter	Type	Required	Example	Description
SystemDisk · Description	String	No	FinanceDept	The description of the system disk. It must be 2 to 256 characters in length and cannot start with "http://" or "https://."
SystemDisk · DiskName	String	No	cloud_ssdSystem	The name of the system disk. It must be 2 to 128 characters in length. It must start with a letter and cannot start with "http://" or "https://." It can contain letters, digits, colons (:), underscores (_), and hyphens (-). Default value: null.
SystemDisk · Size	Integer	No	100	The size of the system disk. Unit: GB. Valid values: · cloud: 40 to 500. · cloud_efficiency: 40 to 500. · cloud_ssd: 40 to 500. · ephemeral_ssd: 40 to 500. Default value: the greater value out of ImageSize and 40. The specified value must be greater than or equal to the default value.

Parameter	Type	Required	Example	Description
Tags	String	No	<pre>{"key1": "value1", "key2": "value2", ... "key5": "value5"}</pre>	The tags of the instance. Tags must be specified as key-value pairs. Up to five tags can be used. Requirements for keys and values: · A key can be up to 64 characters in length. It cannot start with "aliyun", "http://", or "https://." It cannot be left empty. · A value can be up to 128 characters in length. It cannot start with "aliyun", "http://", or "https://." It can be left empty.
UserData	String	No	echo hello ecs!	The custom data of the instance. It must be encoded in the Base64 format. The maximum size of the raw data is 16 KB.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.
ScalingConfigurationId	String	eOs27Kb0oXvQcUYjEGel****	The ID of the scaling configuration. It is generated by the system and must be globally unique.

Examples

Sample requests

```
http ( s )://[ Endpoint ]/? Action = CreateScalingConfiguration
& ScalingGroupId = AG6CQdPU80 KdwLjgZcJ ****
& SecurityGroupId = sg - 280ih ****
&< Common request parameters >
```

Successful response examples

XML format

```
< CreateScalingConfigurationResponse >
  < ScalingConfigurationId > e0s27Kb0oX vQcUYjEGel ****</
ScalingConfigurationId >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
RequestId >
</ CreateScalingConfigurationResponse >
```

JSON format

```
{
  "ScalingConfigurationId ":" e0s27Kb0oX vQcUYjEGel ****",
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InstanceTypeMismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	The error message returned when the instance types of the specified scaling configuration and the current scaling configuration are different.

HTTP status code	Error code	Error message	Description
404	InvalidDataDiskSnapshotId.NotFound	Snapshot “XXX” does not exist.	The error message returned when 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 when the size of the specified snapshot exceeds the maximum size of the specified disk category.
403	InvalidDevice.InUse	Device “XXX” has been occupied.	The error message returned when the mount point of the data disk is already occupied.
400	InvalidImageId.InstanceTypeMismatch	The specified image does not support the specified instance type.	The error message returned when the specified image does not support the specified instance type.
404	InvalidImageId.NotFound	The specified image does not exist.	The error message returned when the specified image does not exist in the current account.

HTTP status code	Error code	Error message	Description
400	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	The error message returned when the specified key pair name does not exist .
400	InvalidNetworkType.ForRAMRole	RAMRole can't be used For classic instance.	The error message returned when the network type of the instance is classic network and does not support the RamRoleName parameter.
400	InvalidParameter	The specified value of parameter KeyPairName is not valid.	The error message returned when the specified instance OS is Windows and does 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 when the specified system disk type conflicts with the data disk type.

HTTP status code	Error code	Error message	Description
400	InvalidRamRole.NotFound	The specified RamRoleName does not exist.	The error message returned when the specified RAM role name does not exist.
400	InvalidScalingConfigurationName.Duplicate	The specified value of parameter ScalingConfigurationName is duplicated.	The error message returned when the scaling configuration name already exists.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.
400	InvalidSecurityGroupId.IncorrectNetworkType	The network type of specified security group does not support this action.	The error message returned when the network types of the specified security group and scaling group are different.
404	InvalidSecurityGroupId.NotFound	The specified security group does not exist.	The error message returned when the specified security group does not exist in the current account.

HTTP status code	Error code	Error message	Description
400	InvalidSecurityGroupId.VPCMismatch	The specified security group and the specified virtual switch are not in the same VPC.	The error message returned when the specified security group and VSwitch are not in the same VPC.
403	InvalidSnapshot.TooOld	This operation is denied because the specified snapshot is created before 2013-07-15.	The error message returned when the snapshot was created on or before July 15, 2013 and cannot be called.
403	InvalidSystemDiskCategory.ValueUnauthorized	The system disk category is not authorized.	The error message returned when you are not authorized to create an ephemeral disk.
400	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded.	The error message returned when the specified user data is not encoded in the Base64 format.
400	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	The error message returned when the specified user data exceeds 16 KB.

HTTP status code	Error code	Error message	Description
403	QuotaExceeded. EphemeralDiskSize	Ephemeral disk size quota exceeded.	The error message returned when the total capacity of mounted ephemeral disks exceeds 2,048 GB.
400	QuotaExceeded .ScalingCon figuration	Scaling configurat ion quota exceeded in the specified scaling group.	The error message returned when the maximum number of scaling configurations has been reached.
400	QuotaExcee ded.SecurityGr oupInstance	Instance quota exceeded in the specified security group.	The error message returned when the maximum number of ECS instances that can be added to the specified security group has been reached.

9.3 DescribeScalingConfigurations

You can call this operation to query scaling configurations.

Debugging

OpenAPI Explorer automatically calculates the signature value. For your convenience, we recommend that you call this operation in OpenAPI Explorer. OpenAPI Explorer dynamically generates the sample code of the operation for different SDKs.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group to which a scaling configuration belongs.
Action	String	No	DescribeScalingConfigurations	The operation that you want to perform. Set the value to DescribeScalingConfigurations.
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.
ScalingConfigurationId . 1	String	No	bU5uZHcAgtzwcL4IeDea****	The ID of scaling configuration N to be queried. Valid values of N: 1 to 10. The IDs of active and inactive scaling configurations are displayed in the query result, and can be differentiated by LifecycleState.
ScalingConfigurationName . 1	String	No	c1908dd1-690f-4c9b-ab73-350f1f06****	The name of scaling configuration N to be queried. Valid values of N: 1 to 20. The names of inactive scaling configurations will not be displayed in the query result and no errors will be returned.

Parameter	Type	Required	Example	Description
ScalingGroupId	String	No	dE9YbOdCHq aFdFZXHVdD ****	The ID of the scaling group . You can query all scaling configurations in the scaling group.

Response parameters

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

Parameter	Type	Example	Description
Cpu	Integer	2	The number of vCPUs. You can specify the number of vCPUs and the amount of memory to define the range of instance types. For example, to specify instances that have 2 vCPUs and 16 GiB memory, set the value of Cpu to 2 and the value of Memory to 16. Auto Scaling uses factors such as I/O optimization and zone to determine a set of available instance types. Auto Scaling then creates instances of the type that is most cost-effective out of the available types.  Note: This instance type range takes effect only when cost optimization is enabled and you have not specified an instance type in the scaling configuration.
CreationTime	String	2014-08-14T10:58Z	The time when a scaling configuration was created.
DataDisks			The set of data disks.

Parameter	Type	Example	Description
Category	String	cloud	The disk type of a data disk. Valid values: - cloud: basic disk The DeleteWith Instance attribute of a basic disk created along with the instance is true. - cloud_efficiency: ultra disk - cloud_ssd: standard SSD - ephemeral_ssd: local SSD - cloud_essd: enhanced SSD
DeleteWith Instance	Boolean	true	Indicates whether the data disk will be released with the instance. Valid values: - true: releases the data disk when the instance is released. - false: retains the data disk when the instance is released.
Description	String	FinanceDept	The description of the data disk.
Device	String	/dev/xvdb	The mount point 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: The disk is encrypted. - false: The disk is not encrypted.
KMSKeyId	String	0e478b7a-4262-4802-b8cb-00d3fb40826X	The ID of the KMS key corresponding to the data disk.

Parameter	Type	Example	Description
Size	Integer	200	The size of the data disk. Unit: GiB Valid values: - cloud: 5 to 2000 - cloud_efficiency: 20 to 32768 - cloud_ssd: 20 to 32768 - ephemeral_ssd: 5 to 800
SnapshotId	String	s-23f2i****	The ID of the snapshot used to create data disks.
DeploymentSetId	String	ds-bp1frxuzdg87zh4p****	The ID of the deployment set to which an ECS instance belongs.
HostName	String	LocalHost	The hostname of an ECS instance.
HpcClusterId	String	hpc-clusterid	The ID of the Elastic HPC cluster to which an ECS instance belongs.
ImageId	String	centos6u5_64_20G_aliaegis_20140703.vhd	The ID of the image file used to automatically create an instance.
ImageName	String	centos6u5_64_20G_aliaegis_20140703.vhd	The name of the image file.
InstanceDescription	String	FinaceDept	The description of an ECS instance.
InstanceGeneration	String	ecs-3	The generation of an ECS instance.
InstanceName	String	fortest	The name of an ECS instance.
InstanceType	String	ecs.t1.xsmall	The type of an ECS instance.
InstanceTypes		ecs.t1.xsmall	The set of ECS instance types.

Parameter	Type	Example	Description
InternetChargeType	String	PayByTraffic	The bandwidth billing method of created instances. Valid values: PayByTraffic: You pay for the actual traffic used. The InternetMaxBandwidthOut parameter only specifies the upper limit of available bandwidth when this parameter is specified.
InternetMaxBandwidthIn	Integer	200	The maximum inbound bandwidth from the public network. Unit: Mbit/s. Valid values: 1 to 200.
InternetMaxBandwidthOut	Integer	0	The maximum outbound bandwidth to the public network. Unit: Mbit/s. Valid values: 0 to 100. If InternetChargeType is set to PayByTraffic and this parameter is not specified, an error is returned.
IoOptimized	String	none	Indicates whether the instance is I/O optimized.
KeyPairName	String	fortest	The name of the key pair used to log on to an ECS instance.

Parameter	Type	Example	Description
LifecycleState	String	Active	The state of the scaling configuration in a scaling group. Valid values: • Active: Active scaling configurations are used by a scaling group to automatically create ECS instances. • Inactive: Inactive scaling configurations are still retained in a scaling group, but are not used by the group to create ECS instances.
LoadBalancerWeight	Integer	1	The weight of the backend server. Valid values: 0 to 100.

Parameter	Type	Example	Description
Memory	Integer	16	The size of memory. You can specify the number of vCPUs and the amount of memory to define the range of instance types. For example, to specify instances that have 2 vCPUs and 16 GiB memory, set the value of Cpu to 2 and the value of Memory to 16. Auto Scaling uses factors such as I/O optimization and zone to determine a set of available instance types. Auto Scaling then creates instances of the type that is most cost-effective out of the available types.  Note: This instance type range takes effect only when cost optimization is enabled and you have not specified an instance type in the scaling configuration.
PasswordInherit	Boolean	true	Indicates whether the password predefined in the specified image is used.
RamRoleName	String	RamRoleTest	The name of the instance RAM role. This name is provided and maintained by RAM. You can call the ListRoles operation to query the list of RAM role names. For more information about how to create a RAM role, see CreateRole .

Parameter	Type	Example	Description
ResourceGroupId	String	abcd1234abcd****	The ID of the resource group to which an ECS instance belongs.
ScalingConfigurationId	String	bU5uZHcAgtzwcL4IeDea****	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 to which the scaling configuration belongs.
SecurityEnhancementStrategy	String	Active	Indicates whether security hardening is enabled. Valid values: - Active: enables security hardening for public images. - Deactive: disables security hardening for all images.
SecurityGroupId	String	sg-280ih****	The ID of the security group to which an ECS instance belongs. ECS instances in the same security group can access each other.
SecurityGroupIds		sg-bp18kz60mefs****	The IDs of the security groups to which an ECS instance belongs. ECS instances in the same security group can access each other.
SpotPriceLimit			The set of the preemptible instances.
InstanceType	String	ecs.t1.xsmall	The type of a preemptible instance.
PriceLimit	Float	0.125	The price limit of the preemptible instance.

Parameter	Type	Example	Description
SpotStrategy	String	NoSpot	The preemption strategy for pay-as-you-go instances. Valid values: · NoSpot: a pay-as-you-go instance · SpotWithPriceLimit: a preemptible instance with a maximum hourly price. · SpotAsPriceGo: an instance priced at the market price at the time of purchase.
SystemDisk Category	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: enhanced SSD
SystemDisk Description	String	FinanceDept	The description of the system disk.
SystemDisk Name	String	cloud_ssdSystem	The name of the system disk.
SystemDisk Size	Integer	100	The size of the system disk.
Tags			The set of tags.
Key	String	binary	The key of the tag.
Value	String	alterTable	The value of the tag.
UserData	String	echo hello ecs!	The user data of an ECS instance.
TotalCount	Integer	1	The total number of scaling configurations.

Examples

Sample requests

```
http ( s )://[ Endpoint ]/? Action = DescribeScalingConfigurations
& RegionId = cn - qingdao
&< Common request parameters >
```

Sample success responses

XML format

```
< DescribeScalingConfigurationsResponse >
  < RequestId > 804F240A - 8D3E - 40A1 - BD68 - 6B333DEA2C A8 </
RequestId >
  < TotalCount > 1 </ TotalCount >
  < PageNumber > 1 </ PageNumber >
  < PageSize > 50 </ PageSize >
  < ScalingConfigurations >
    < ScalingConfiguration >
      < CreationTime > 2014 - 08 - 14T10 : 58Z </
CreationTime >
      < ImageId > centos6u5_ 64_20G_ali aegis_2014 0703
. vhd </ ImageId >
      < InstanceType > ecs . t1 . xsmall </ InstanceTy
pe >
      < InternetChargeType > PayByTraffic </
InternetChargeType >
      < InternetMaxBandwidthIn > 200 </ InternetMa
xBandwidthIn >
      < InternetMaxBandwidthOut > 0 </ InternetMa
xBandwidthOut >
      < LifecycleState > Active </ LifecycleS tate >
      < ScalingConfigurationId > bU5uZHcAgt
zwCL4IeDea ****</ ScalingCon figuration Id >
      < ScalingConfigurationName > c1908dd1 - 690f -
4c9b - ab73 - 350f1f06 ****</ ScalingCon figuration Name >
      < ScalingGroupId > dE9Yb0dCHq aFdFZXHVdD ****</
ScalingGro upId >
      < SecurityGroupId > sg - 280ih ****</ SecurityGr
oupId >
      < SystemDiskCategory > cloud </ SystemDisk
Category >
      < DataDisks >
        < DataDisk >
          < Size > 200 </ Size >
          < Category > cloud </ Category >
          < SnapshotId > s - 280s7 ****</ SnapshotId >
          < Device > / dev / xvdb </ Device >
        </ DataDisk >
      </ DataDisks >
    </ ScalingCon figuration >
  </ ScalingCon figuration s >
</ DescribeScalingConfigurationsResponse >
```

JSON format

```
{
```



```

" PageNumber ":" 1 ",
" TotalCount ":" 1 ",
" ScalingConfiguration s ":{
  " ScalingConfiguration ":{
    " ImageId ":" centos6u5_64_20G_ali_aegis_2014_0703 . vhd ",
    " SecurityGroupId ":" sg - 280ih ****",
    " InternetMaxBandwidth In ":" 200 ",
    " DataDisks ":{
      " DataDisk ":[]
      " Category ":" cloud ",
      " Device ":" / dev / xvdb ",
      " SnapshotId ":" s - 280s7 ****",
      " Size ":" 200 "
    }
  },
  " InternetChargeType ":" PayByTraffic ",
  " InstanceType ":" ecs . t1 . xsmall ",
  " CreationTime ":" 2014 - 08 - 14T10 : 58Z ",
  " ScalingConfigurationId ":" bU5uZHcAgt_zwCL4IeDea ****",
  " InternetMaxBandwidth Out ":" 0 ",
  " ScalingGroupUpId ":" dE9YbOdCHq_aFdFZXHVdD ****",
  " ScalingConfigurationName ":" c1908dd1 - 690f - 4c9b - ab73 - 350f1f06 ****",
  " SystemDiskCategory ":" cloud ",
  " LifecycleState ":" Active "
}
},
" PageSize ":" 50 ",
" RequestId ":" 804F240A - 8D3E - 40A1 - BD68 - 6B333DEA2C_A8 "
}

```

Error codes

For more information about error codes, visit [API Error Center](#).

9.4 ModifyScalingConfiguration

You can call this operation to modify a scaling configuration. If you are modifying the name of a scaling configuration, ensure that the new name is unique within the scaling group.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingConfigurationId	String	Yes	asc-****	The ID of the scaling configuration to be modified.
Action	String	No	ModifyScalingConfiguration	The operation that you want to perform. Set the value to ModifyScalingConfiguration.
Cpu	Integer	No	2	The number of vCPUs. You can specify the number of CPU cores and the amount of memory to define the range of instance types. For example, instance types with a specified value of CPU as 2 and Memory as 16 can be defined as 2 vCPU 16 GiB. Auto Scaling uses factors such as I/O optimization and zone to determine a set of available instance types. Auto Scaling then creates instances of the type that is most cost-effective out of the available types.  Note: This instance type range takes effect only when cost optimization is enabled and the scaling configuration does not have a specified instance type.

Parameter	Type	Required	Example	Description
DataDisk. N. Category	String	No	cloud_ssd	The category of data disk N. Valid values: - cloud: indicates a basic disk. The DeleteWithInstance property of a basic disk created along with the instance is true. - cloud_efficiency: indicates an ultra disk. - cloud_ssd: indicates an SSD disk. - ephemeral_ssd: indicates a local SSD disk. Default value: cloud.
DataDisk. N. DeleteWith Instance	Boolean	No	true	Indicates whether data disk N is to be released with its attached instance. Valid values: - true: Release the data disk with its attached instance. - false: Retain the data disk when the attached instance is released. Default value: true. This parameter is valid only for independent disks, for which the DataDisk.N.Category parameter is set to cloud, cloud_efficiency, or cloud_ssd. An error will be returned if you set this parameter for other disks.

Parameter	Type	Required	Example	Description
DataDisk . N . Description	String	No	FinanceDept	The description of data disk N . It must be 2 to 256 characters in length and cannot start with "http://" or "https://."
DataDisk . N . Device	String	No	/dev/xvdb	The mount point of data disk N. Valid values of N: 1 to 4. If this parameter is not specified , ECS automatically allocates a mount point to created ECS instances. The name of the mount point ranges from /dev/xvdb to /dev/xvdz in an alphabetical order.
DataDisk . N . DiskName	String	No	cloud_ssdData	The name of data disk N. It must be 2 to 128 characters in length. It must start with a letter and cannot start with "http://" or "https://." It can contain letters, digits, colons (:), underscores (_), and hyphens (-). Default value: empty.
DataDisk . N . Encrypted	String	No	false	Indicates whether data disk N is to be encrypted. Default value: false.
DataDisk . N . KMSKeyId	String	No	0e478b7a-4262-4802-b8cb-00d3fb40826X	The KMS key ID corresponding to the data disk.

Parameter	Type	Required	Example	Description
DataDisk . N . Size	Integer	No	100	The size of data disk N. Unit: GB. Valid values of N: 1 to 16. Valid values: · cloud: 5 to 2000. · cloud_efficiency: 20 to 32768 · · cloud_ssd: 20 to 32768. · ephemeral_ssd: 5 to 800. If this parameter is specified, the data disk size must be greater than or equal to the size of the snapshot specified by SnapshotId.
DataDisk . N . SnapshotId	String	No	s-snapshot****	The ID of the snapshot used to create data disk N. Valid values of N: 1 to 4. If this parameter is specified, DataDisk.N.Size is ignored and the size of the created disk is the size of the specified snapshot. You cannot specify a snapshot that was created on or before July 15, 2013.
Deployment SetId	String	No	ds-bp13v7bjnj 9gis****	The ID of the deployment set.

Parameter	Type	Required	Example	Description
HostName	String	No	Joshu****	The name of the host server where created ECS instances reside. It cannot start or end with a period (.) or hyphen (-). It cannot contain consecutive periods (.) or hyphens (-). Naming rules: • Windows-based instances: It must be 2 to 15 characters in length. It can contain uppercase or lowercase letters, digits, and hyphens (-). It cannot contain periods (.) or be entirely composed of digits. • Other instances such as Linux-based instances: It must be 2 to 128 characters in length. It can be segments separated by periods (.). Each segment can contain uppercase or lowercase letters, digits, and hyphens (-).
ImageId	String	No	centos6u5_64_20G_aliaegis_20140703.vhd	The ID of the image file to be specified when created instances are started.

Parameter	Type	Required	Example	Description
ImageName	String	No	suse11sp3_64_20G_aliaegis_20150428.vhd	The name of the image file. Image names must be unique within a region. This parameter is ignored if ImageId is specified. If you specify this parameter instead of ImageId, the image with the specified name is used for the scaling configuration.
InstanceName	String	No	Joshu****	The name of the instance created based on the current scaling configuration.
InstanceTypes . N	RepeatList	No	ecs.t1.xsmall	The multiple instance types from which ECS instances are to be created. InstanceType is ignored if this parameter is specified. Valid values of N: 1 to 10. Up to 10 instance types can be configured for a scaling configuration. N represents the priority of the instance type in the scaling configuration. A lower value of N indicates a higher priority for the instance type. If an instance of the highest priority type cannot be created, Auto Scaling will create an instance of the next highest priority type.

Parameter	Type	Required	Example	Description
InternetChargeType	String	No	PayByBandwidth	The bandwidth billing type of created instances. Valid values: • PayByBandwidth: Users pay for the maximum available bandwidth. • PayByTraffic: Users pay for the actual traffic used. The InternetMaxBandwidthOut parameter only specifies the upper limit of available bandwidth when this parameter is specified. Default value: PayByBandwidth for classic networks or PayByTraffic for VPCs.
InternetMaxBandwidthOut	Integer	No	50	The maximum outbound bandwidth to the public network. Unit: Mbit/s. Valid values: 1 to 100. • If InternetChargeType is set to PayByBandwidth and this parameter is not specified, this parameter is automatically set to 0. • If InternetChargeType is set to PayByTraffic and this parameter is not specified, an error is returned.

Parameter	Type	Required	Example	Description
IoOptimized	String	No	none	Indicates whether created instances are I/O optimized. Valid values: · none · optimized The value of this parameter is none if InstanceType is set to an instance type which has been phased out. The value of this parameter is optimized if InstanceType is set to an instance types later than Generation I.
KeyPairName	String	No	null	The name of the key pair. · Ignore this parameter if you are creating a Windows-based ECS instance. This parameter is null by default. · Password logon is disabled for Linux-based ECS instances by default.
LoadBalancerWeight	Integer	No	50	The weight of the back-end server. Valid values: 0 to 100. Default value: 50.

Parameter	Type	Required	Example	Description
Memory	Integer	No	16	The size of memory. You can specify the number of CPU cores and the amount of memory to define the range of instance types. For example, instance types with a specified value of CPU as 2 and Memory as 16 can be defined as 2 vCPU 16 GiB. Auto Scaling uses factors such as I/O optimization and zone to determine a set of available instance types. Auto Scaling then creates instances of the type that is most cost-effective out of the available types.  Note: This instance type range takes effect only when cost optimization is enabled and the scaling configuration does not have a specified instance type.
Override	Boolean	No	true	Indicates whether to overwrite the existing data. Valid values: · true · false

Parameter	Type	Required	Example	Description
PasswordInherit	Boolean	No	false	Indicates whether to use the preconfigured password of the specified image. To use this parameter, ensure that a password is configured for the specified image.
RamRoleName	String	No	RamRoleTest	The name of the instance RAM role. This name is provided and maintained by RAM. Call ListRoles to query the list of RAM role names. For more information, see CreateRole .
ResourceGroupId	String	No	abcd1234abcd****	The ID of the resource group.
ScalingConfigurationName	String	No	test-modify	The new name of the scaling configuration. It must be 2 to 40 characters in length. It must start with a digit, uppercase letter, or lowercase letter. It can contain letters, digits, underscores (_), hyphens (-), and periods (.). Scaling configuration names must be unique within a scaling group. If this parameter is not specified, the value of ScalingConfigurationId is used.
SecurityGroupId	String	No	sg-F876F****	The ID of the security group.

Parameter	Type	Required	Example	Description
SpotPriceLimit.N.InstanceType	String	No	ecs.t1.small	The instance type of preemptible instance N. Valid values of N: 1 to 10. This parameter takes effect only when the SpotStrategy parameter is set to SpotWithPriceLimit.
SpotPriceLimit.N.PriceLimit	Float	No	0.125	The price limit of preemptible instance N. Valid values of N: 1 to 10. This parameter takes effect only when the SpotStrategy parameter is set to SpotWithPriceLimit.
SpotStrategy	String	No	NoSpot	The preemption policy to be used for pay-as-you-go instances. It takes effect only when the InstanceChargeType parameter is set to PostPaid. Valid values: • NoSpot: indicates a pay-as-you-go instance. • SpotWithPriceLimit: indicates a preemptible instance with a maximum price. • SpotAsPriceGo: indicates an instance with its price based on the market price at the time of purchase. Default value: NoSpot.

Parameter	Type	Required	Example	Description
SystemDisk · Category	String	No	cloud_efficiency	The type of the system disk. Valid values: · cloud: indicates a basic disk. · cloud_efficiency: indicates an ultra disk. · cloud_ssd: indicates an SSD disk. · ephemeral_ssd: indicates a local SSD disk. Default value: cloud for Generation I type instances that are not I/O optimized, and cloud_efficiency for other types.
SystemDisk · Description	String	No	FinanceDept	The description of the system disk. It must be 2 to 256 characters in length and cannot start with "http://" or "https://."
SystemDisk · DiskName	String	No	cloud_ssdSystem	The name of the system disk. It must be 2 to 128 characters in length. It must start with a letter and cannot start with "http://" or "https://." It can contain letters, digits, colons (:), underscores (_), and hyphens (-). Default value: null.

Parameter	Type	Required	Example	Description
SystemDisk . Size	Integer	No	50	The size of the system disk. Unit: GB. Valid values: - cloud: 40 to 500. - cloud_efficiency: 40 to 500. - cloud_ssd: 40 to 500. - ephemeral_ssd: 40 to 500. Default value: the greater value out of ImageSize and 40. The specified value must be greater than or equal to the default value.
Tags	String	No	"key1" :" value1"	The tags of the instance. Tags must be specified as key-value pairs. Up to five tags in the { "key1" :" value1" , "key2" :" value2" , ... "key5" :" value5" } format can be specified. Requirements for keys and values: - A key can be up to 64 characters in length. It cannot start with "aliyun", "http://", or "https://." It cannot be left empty. - A value can be up to 128 characters in length. It cannot start with "aliyun", "http://", or "https://." It can be left empty.
UserData	String	No	echo hello ecs!	The custom data of the instance. It must be encoded in the Base64 format. The maximum size of the raw data is 16 KB.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http ( s )://[ Endpoint ]/? Action = ModifyScalingConfiguration
& ScalingConfigurationId = asc -****
&< Common request parameters >
```

Successful response examples

XML format

```
< ModifyScalingConfigurationResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
</ ModifyScalingConfigurationResponse >
```

JSON format

```
{
  " requestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

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 when you are not authorized to perform the specified operation.

HTTP status code	Error code	Error message	Description
404	InvalidDataDiskSnapshotId.NotFound	Snapshot “XXX” does not exist.	The error message returned when 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 when the size of the specified snapshot exceeds the maximum size of the specified disk category.
404	InvalidImageId.NotFound	The specified image does not exist.	The error message returned when the specified image does not exist.
400	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	The error message returned when the specified key pair name does not exist.
400	InvalidNetworkType.ForRAMRole	RAMRole can't be used For classic instance.	The error message returned when the network type of the instance is classic network and does not support the RamRoleName parameter.

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned when the specified parameter value is invalid.
400	InvalidScalingConfigurationName. Duplicate	The specified value of parameter is duplicated.	The error message returned when the specified scaling configuration name already exists.
400	InvalidSecurityGroupId. IncorrectNetworkType	The network type of specified Security Group does not support this action.	The error message returned when the network types of the specified security group and scaling group are different.
400	InvalidSecurityGroupId. VPCMismatch	The specified security group and the specified virtual switch are not in the same VPC.	The error message returned when the specified security group and VSwitch are not in the same VPC.
400	InvalidTags. KeyValue	The specified tags key/value cannot be empty.	The error message returned when the Tags parameter is not specified.

HTTP status code	Error code	Error message	Description
400	InvalidTags.ListSize	The specified tags list size cannot be more than “5” .	The error message returned when the length of the Tags list exceeds the limit.
400	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded.	The error message returned when the specified user data is not encoded in the Base64 format.
400	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	The error message returned when the specified user data exceeds 16 KB.

10 Scaling rule

10.1 CreateScalingRule

You can call this operation to create a scaling rule, so as to define a scaling action.

Description

A scaling rule defines the scaling action to be executed, such as adding or removing a certain number of ECS instances. If the number of ECS instances after a scaling rule is executed is smaller than MinSize or greater than MaxSize of a scaling group, Auto Scaling automatically adds or removes ECS instances. In this case, Auto Scaling keeps the number of ECS instances in the scaling group to MinSize or MaxSize while the specified number of ECS instances to be adjusted in the scaling rule is ignored and remains unchanged. You can create scaling rules based on input parameters.

- **Example 1:** A scaling group has a MaxSize of 3, a TotalCapacity of 2, and a scaling rule of adding three ECS instances. When executing this scaling rule, Auto Scaling adds only one ECS instance to the group to prevent TotalCapacity (the current number of ECS instances in the group) from exceeding MaxSize (the maximum number of ECS instances allowed in the group).
- **Example 2:** A scaling group has a MinSize of 2, a TotalCapacity of 3, and a scaling rule of removing five ECS instances. When executing this scaling rule, Auto Scaling removes only one ECS instance from the group to prevent TotalCapacity (the current number of ECS instances in the group) from becoming smaller than MinSize (the minimum number of ECS instances allowed in the group).

You can create scaling rules based on input parameters.

- When AdjustmentType is set to TotalCapacity, the number of ECS instances currently in the scaling group is adjusted to the specified value. The value of AdjustmentValue must be greater than or equal to 0.
- When AdjustmentType is set to QuantityChangeInCapacity or PercentChangeInCapacity, a positive value of AdjustmentValue indicates the number of ECS instances to be added, and a negative value of AdjustmentValue indicates the number of ECS instances to be removed.

- When AdjustmentType is set to PercentChangeInCapacity, Auto Scaling rounds the number of ECS instances to be adjusted to a nearest integer based on the formula: Number of ECS instances to be adjusted = Current number of ECS instances (Total Capacity) x AdjustmentValue/100.
- When the cooldown time (Cooldown) is specified in a scaling rule, the specified time applies to the scaling group after this rule is executed. Otherwise, DefaultCooldown applies to the scaling group.
- A scaling group can have a maximum of 50 scaling rules.
- The following operations use the unique identifier of a scaling rule (ScalingRuleAri) returned by CreateScalingRule:
 - ExecuteScalingRule: You can call this operation to manually execute a specified scaling rule by setting ScalingRuleAri to the identifier.
 - CreateScheduledTask: You can call this operation to create a scheduled task for a specified scaling rule by setting ScheduledAction to the identifier.

Debugging

Alibaba Cloud provides [OpenAPI Explorer](#) to simplify API usage. You can use OpenAPI Explorer to search for APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcj****	The ID of the scaling group to which the scaling rule belongs.
Action	String	No	CreateScalingRule	The operation that you want to perform. Set the value to CreateScalingRule.

Parameter	Type	Required	Example	Description
ScalingRuleName	String	No	TestSr	The display name of the scaling rule. The name must be 2 to 40 characters in length. It must start with a letter or digit. It can contain digits, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). This parameter must be unique within a scaling group. The default value is the ID of the scaling rule.
Cooldown	Integer	No	60	The cooldown time in the scaling rule. This parameter is only applicable to simple scaling rules. Valid values: 0 to 86400. Unit: seconds. Default value: null.
MinAdjustmentMagnitude	Integer	No	1	The minimum number of ECS instances to be adjusted in a scaling rule. This parameter only takes effect when the scaling rule type is SimpleScalingRule or StepScalingRule and AdjustmentType is PercentChangeInCapacity.

Parameter	Type	Required	Example	Description
Adjustment Type	String	No	QuantityChangeInCapacity	The adjustment method of the scaling rule. This parameter is required and only applicable to simple scaling rules. Valid values: - QuantityChangeInCapacity: adds or removes a specified number of ECS instances. - PercentChangeInCapacity: adds or removes a specified proportion of ECS instances. - TotalCapacity: adds or removes ECS instances to ensure that the current scaling group has a specified number of ECS instances.
Adjustment Value	Integer	No	100	The number of ECS instances to be adjusted in the scaling rule. This parameter is required and only applicable to simple scaling rules. The number of ECS instances to be adjusted in a single scaling activity cannot exceed 500. Valid values: - If AdjustmentType is set to QuantityChangeInCapacity, the value range is -500 to 500. - If AdjustmentType is set to PercentChangeInCapacity, the value range is -100 to 10000. - If AdjustmentType is set to TotalCapacity, the value range is 0 to 1000.

Parameter	Type	Required	Example	Description
ScalingRuleType	String	No	SimpleScalingRule	The type of the scaling rule. Valid values: • SimpleScalingRule: scales ECS instances based on the values of AdjustmentType and AdjustmentValue. • TargetTrackingScalingRule: dynamically calculates the number of ECS instances to be adjusted and tries to keep the value of a predefined monitoring metric close to TargetValue. • StepScalingRule: scales ECS instances in steps based on specified thresholds and metric values. • PredictiveScalingRule: analyzes historical monitoring data of the scaling group and then predicts the future values of monitored metrics based on machine learning, then automatically creates scheduled tasks to set the boundary values for the scaling group. Default value: SimpleScalingRule.

Parameter	Type	Required	Example	Description
EstimatedInstanceWarmup	Integer	No	300	The warm-up period of the ECS instances. This parameter is applicable to target tracking scaling rules and step scaling rules. The system adds ECS instances that are in the warm-up state to the scaling group, but does not report monitoring data to CloudMonitor during the warm-up period.  Note: When calculating the number of ECS instances to be adjusted, the system does not count ECS instances in the warm-up state as part of 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 specify to monitor. This parameter is required and only applicable to target tracking scaling rules and predictive scaling rules. Valid values for a target tracking scaling rule: • CpuUtilization: the average CPU utilization • ClassicInternetRx: the average Internet inbound traffic over the classic network • ClassicInternetTx: the average Internet outbound traffic over the classic network • VpcInternetRx: the average Internet inbound traffic over the VPC network • VpcInternetTx: the average Internet outbound traffic over the VPC network • IntranetRx: the average intranet inbound traffic • IntranetTx: the average intranet outbound traffic Valid values for a predictive scaling rule: • CpuUtilization: the average CPU utilization • IntranetRx: the average intranet inbound traffic • IntranetTx: the average intranet outbound traffic

Parameter	Type	Required	Example	Description
TargetValue	Float	No	0.125	The target value of a metric . This parameter is required and only applicable to target tracking scaling rules and predictive scaling rules. The value of TargetValue must be greater than 0 and can have a maximum of three decimal places.
DisableScaleIn	Boolean	No	false	Specifies whether to disable scale-in. This parameter is only applicable to target tracking scaling rules. Default value: false.
InitialMaxSize	Integer	No	100	The maximum number of ECS instances in the scaling group , which is used together with PredictiveValueBehavior. Default value: the same as the value of MaxSize.
PredictiveScalingMode	String	No	PredictAndScale	The mode of the predictive scaling rule. Valid values: · PredictAndScale: generates both forecasts and forecast tasks. · PredictOnly: generates forecasts but does not generate forecast tasks. Default value: PredictAndScale.

Parameter	Type	Required	Example	Description
Predictive TaskBuffer Time	Integer	No	30	The amount of buffer time ahead of the forecast task execution time. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed at the beginning of each hour. You can set a buffer time to execute forecast tasks ahead of schedule, so that resources can be prepared in advance. Valid values: 0 to 60. Default value: 0.

Parameter	Type	Required	Example	Description
PredictiveValueBehavior	String	No	MaxOverridePredictiveValue	The action taken on the predicted maximum value. Valid values: · MaxOverridePredictiveValue : uses the initial maximum capacity as the maximum value for forecast tasks when the predicted value is greater than the initial maximum capacity. · PredictiveValueOverrideMax : uses the predicted value as the maximum value for forecast tasks when the predicted value is greater than the initial maximum capacity. · PredictiveValueOverrideMaxWithBuffer: increases the predicted value with a ratio, which is specified by PredictiveValueBuffer. If the value after increase is greater than the initial maximum capacity, the value after increase is used as the maximum value for forecast tasks. Default value: MaxOverridePredictiveValue.

Parameter	Type	Required	Example	Description
PredictiveValueBuffer	Integer	No	50	The ratio of the increment to the predicted value when PredictiveValueBehavior is set to PredictiveValueOverrideMaxWithBuffer. When the value after increase is greater than the initial maximum capacity, the value after increase is used for forecast tasks. Valid values: 0 to 100. Default value: 0.
StepAdjustment.N.MetricIntervalLowerBound	Float	No	1.0	The lower limit value specified in step adjustment N. Valid values: -9.999999E18 to 9.999999E18.
StepAdjustment.N.MetricIntervalUpperBound	Float	No	5.0	The upper limit value specified in step adjustment N. Valid values: -9.999999E18 to 9.999999E18.
StepAdjustment.N.ScalingAdjustment	Integer	No	1	The specified number of ECS instances to be adjusted in step adjustment N.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. The system returns a unique RequestId for each API request, regardless of whether the API operation is successful.

Parameter	Type	Example	Description
ScalingRuleAri	String	ari:acs:ess:cn-qingdao:1344371:scalingrule/eMKWG8SRNb9dBLAjweNI1Ik	The unique identifier of the scaling rule.
ScalingRuleId	String	eMKWG8SRNb9dBLAjweN****	The globally unique scaling rule ID generated by the system.

Examples

Sample request

```
http://ess.aliyuncs.com/?Action=CreateScalingRule
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=-10
&<Common request parameters>
```

Sample success response

XML format

```
<CreateScalingRuleResponse>
  <ScalingRuleAri>ari:acs:ess:cn-qingdao:1344371:
scalingrule/eMKWG8SRNb9dBLAjweNI1Ik</ScalingRuleAri>
  <ScalingRuleId>eMKWG8SRNb9dBLAjweN****</ScalingRuleId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</
RequestId>
</CreateScalingRuleResponse>
```

JSON format

```
{
  "ScalingRuleAri": "ari:acs:ess:cn-qingdao:1344371:
scalingrule/eMKWG8SRNb9dBLAjweNI1Ik",
  "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
  "ScalingRuleId": "eMKWG8SRNb9dBLAjweN****"
}
```

Error codes

[View 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 current 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 number of scaling rules in the specified scaling group has reached the upper limit.
400	TargetTrackingScalingRule.UnsupportedMetric	Specific metric is not supported for target tracking scaling rule.	The error message returned because target tracking scaling rules do not support the specified monitoring metric.

HTTP status code	Error code	Error message	Description
400	TargetTrackingScalingRule.DuplicateMetric	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 with an unspecified upper limit value already exists.
400	InvalidStepAdjustments.MultipleNullLowerBound	At most one StepAdjustment may have an unspecified lower bound.	The error message returned because a step adjustment with an unspecified lower limit value already exists.

HTTP status code	Error code	Error message	Description
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 value of a step adjustment is negative, but a different step adjustment with an unspecified lower limit value 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 value of a step adjustment is positive, but a different step adjustment with an unspecified upper limit value does not exist.
400	InvalidStepAdjustments.Gap	StepAdjustment intervals cannot have gaps between them.	The error message returned because the specified ranges of step adjustments have gaps between them.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.Overlap	StepAdjustment intervals cannot 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 value of a step adjustment is greater than or equal to the upper limit value.
400	InvalidStepAdjustments.BothNull	Both lower and upper bounds of a StepAdjustment cannot be left unspecified.	The error message returned because neither the upper limit value nor the lower limit value for a step adjustment is 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 a scaling group has reached the upper limit.

HTTP status code	Error code	Error message	Description
400	StepBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned because the specified upper limit value or lower limit value of a step adjustment is not within the range of valid values.

10.2 ModifyScalingRule

You can call this operation to modify an existing scaling rule, in order to redefine the scaling action.

Description

You can call this operation to modify specific attributes of an existing scaling rule for new scenarios.

Debugging

Alibaba Cloud provides [OpenAPI Explorer](#) to simplify API usage. You can use OpenAPI Explorer to search for APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingRuleId	String	Yes	eMKWG8SRNb9dBLAjweN****	The ID of the scaling rule.
Action	String	No	ModifyScalingRule	The operation that you want to perform. Set the value to ModifyScalingRule.

Parameter	Type	Required	Example	Description
Adjustment Type	String	No	QuantityChangeInCapacity	The method used by the scaling rule to adjust the number of ECS instances. This parameter is only applicable to simple scaling rules. Valid values: - QuantityChangeInCapacity: adds or removes a specified number of ECS instances. - PercentChangeInCapacity: adds or removes a specified proportion of ECS instances. - TotalCapacity: adds or removes ECS instances to ensure that the current scaling group has a specified number of ECS instances.
Adjustment Value	Integer	No	100	The specified number of ECS instances to be adjusted by the scaling rule. This parameter is only applicable to simple scaling rules. The number of ECS instances to be adjusted in a single scaling activity cannot exceed 500. Valid values: - If AdjustmentType is set to QuantityChangeInCapacity, the value range is -500 to 500. - If AdjustmentType is set to PercentChangeInCapacity, the value range is -100 to 10000. - If AdjustmentType is set to TotalCapacity, the value range is 0 to 1000.

Parameter	Type	Required	Example	Description
Cooldown	Integer	No	60	The cooldown time in the scaling rule. This parameter is only applicable to simple scaling rules. Valid values: 0 to 86400. Unit: seconds.
DisableScaleIn	Boolean	No	true	Specifies whether to disable scale-in. This parameter is only applicable to target tracking scaling rules.
EstimatedInstanceWarmup	Integer	No	60	The warm-up period of the ECS instances. This parameter is applicable to target tracking scaling rules and step scaling rules. The system adds ECS instances that are in the warm-up state to the scaling group, but does not report monitoring data to CloudMonitor during the warm-up period.  Note: When calculating the number of ECS instances to be adjusted, the system does not count ECS instances in the warm-up state as part of the current capacity of the scaling group. Valid values: 0 to 86400. Unit: seconds.

Parameter	Type	Required	Example	Description
InitialMax Size	Integer	No	100	The maximum number of ECS instances in the scaling group , which is used together with PredictiveValueBehavior.

Parameter	Type	Required	Example	Description
MetricName	String	No	CpuUtilization	The predefined metric that you specify to monitor. This parameter is required and only applicable to target tracking scaling rules and predictive scaling rules. Valid values for a target tracking scaling rule: • CpuUtilization: the average CPU utilization • ClassicInternetRx: the average Internet inbound traffic over the classic network • ClassicInternetTx: the average Internet outbound traffic over the classic network • VpcInternetRx: the average Internet inbound traffic over the VPC network • VpcInternetTx: the average Internet outbound traffic over the VPC network • IntranetRx: the average intranet inbound traffic • IntranetTx: the average intranet outbound traffic Valid values for a predictive scaling rule: • CpuUtilization: the average CPU utilization • IntranetRx: the average intranet inbound traffic • IntranetTx: the average intranet outbound traffic

Parameter	Type	Required	Example	Description
MinAdjustmentMagnitude	Integer	No	1	The minimum number of ECS instances to be adjusted in a scaling rule. This parameter only takes effect when the scaling rule type is SimpleScalingRule or StepScalingRule and AdjustmentType is PercentChangeInCapacity.
PredictiveScalingMode	String	No	PredictAndScale	The mode of the predictive scaling rule. Valid values: • PredictAndScale: generates both forecasts and forecast tasks. • PredictOnly: generates forecasts but does not generate forecast tasks.
PredictiveTaskBufferTime	Integer	No	30	The amount of buffer time ahead of the forecast task execution time. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed at the beginning of each hour. You can set a buffer time to execute forecast tasks ahead of schedule, so that resources can be prepared in advance. Valid values: 0 to 60.

Parameter	Type	Required	Example	Description
PredictiveValueBehavior	String	No	MaxOverridePredictiveValue	The action taken on the predicted maximum value. Valid values: • MaxOverridePredictiveValue : uses the initial maximum capacity as the maximum value for forecast tasks when the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMax : uses the predicted value as the maximum value for forecast tasks when the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMaxWithBuffer: increases the predicted value with a ratio, which is specified by PredictiveValueBuffer. If the value after increase is greater than the initial maximum capacity, the value after increase is used as the maximum value for forecast tasks.

Parameter	Type	Required	Example	Description
PredictiveValueBuffer	Integer	No	50	The ratio of the increment to the predicted value when PredictiveValueBehavior is set to PredictiveValueOverrideMaxWithBuffer. When the value after increase is greater than the initial maximum capacity, the value after increase is used for forecast tasks. Valid values: 0 to 100.
ScalingRuleName	String	No	TestSr	The display name of the scaling rule. The name must be 2 to 40 characters in length. It must start with a letter or digit. It can contain digits, uppercase letters, lowercase letters, underscores (_), hyphens (-), and periods (.). This parameter must be unique within a scaling group.
StepAdjustment.N.MetricIntervalLowerBound	Float	No	1.0	The lower limit value specified in step adjustment N. Valid values: -9.999999E18 to 9.999999E18.
StepAdjustment.N.MetricIntervalUpperBound	Float	No	5.0	The upper limit value specified in step adjustment N. Valid values: -9.999999E18 to 9.999999E18.

Parameter	Type	Required	Example	Description
StepAdjustment.N.ScalingAdjustment	Integer	No	1	The specified number of ECS instances to be adjusted in step adjustment N.
TargetValue	Float	No	0.125	The target value of the metric. This parameter is only applicable to target tracking scaling rules and predictive scaling. The value of TargetValue must be greater than 0 and can have a maximum of three decimal places.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. The system returns a unique RequestId for each API request, regardless of whether the API operation is successful.

Examples

Sample request

```
http://ess.aliyuncs.com/?Action=ModifyScalingRule
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=-10
&<Common request parameters>
```

Sample success response

XML format

```
<ModifyScalingRuleResponse>
  <ScalingRuleAri>ari:acs:ess:cn-qingdao:1344371:
scalingrule/eMKWG8SRNb9dBLAjweNI1Ik</ScalingRuleAri>
  <ScalingRuleId>eMKWG8SRNb9dBLAjweN****</ScalingRuleId>
```

```
< RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
RequestId >
</ ModifyScalingRuleResponse >
```

JSON format

```
{
  "ScalingRuleAri ":" ari : acs : ess : cn - qingdao : 1344371 :
scalingrule / eMKWG8SRNb 9dBLAjweNI 1Ik ",
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E ",
  "ScalingRuleId ":" eMKWG8SRNb 9dBLAjweN ****"
}
```

Error codes

[View 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 current 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 number of scaling rules in the specified scaling group has reached the upper limit.

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 target tracking scaling rules do not support the specified monitoring metric.
400	TargetTrackingScalingRule.DuplicateMetric	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 with an unspecified upper limit value already exists.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.MultipleNullLowerBound	At most one StepAdjustment may have an unspecified lower bound.	The error message returned because a step adjustment with an unspecified lower limit value 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 value of a step adjustment is negative, but a different step adjustment with an unspecified lower limit value 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 value of a step adjustment is positive, but a different step adjustment with an unspecified upper limit value does not exist.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.Gap	StepAdjustment intervals cannot have gaps between them.	The error message returned because the specified ranges of step adjustments have gaps between them.
400	InvalidStepAdjustments.Overlap	StepAdjustment intervals cannot 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 value of a step adjustment is greater than or equal to the upper limit value.
400	InvalidStepAdjustments.BothNull	Both lower and upper bounds of a StepAdjustment cannot be left unspecified.	The error message returned because neither the upper limit value nor the lower limit value for a step adjustment is specified.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments. MaxNum	Your scaling rule can have at most %s StepAdjustments.	The error message returned because the maximum number of step adjustments in a scaling group has reached the upper limit.
400	StepBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned because the specified upper limit value or lower limit value of a step adjustment is not within the range of valid values.

10.3 DeleteScalingRule

You can call this operation to delete a scaling rule.

Description

This operation deletes a specified scaling rule.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform operations such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingRuleId	String	Yes	eMKWG8SRNb9dBLAjweN****	The ID of the scaling rule.

Parameter	Type	Required	Example	Description
Action	String	No	DeleteScalingRule	The operation that you want to perform. Set the value to DeleteScalingRule.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

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

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingRuleId.NotFound	The specified scaling rule does not exist.	The error message returned when no scaling rule by the specified ID exists in the current account.

10.4 DescribeScalingRules

You can call this operation to query all scaling rules in a scaling group and list information about the scaling rules.

Description

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

Debugging

OpenAPI Explorer automatically calculates the signature value. For your convenience, we recommend that you call this operation in OpenAPI Explorer. OpenAPI Explorer dynamically generates the sample code of the operation for different SDKs.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group to which a scaling rule belongs.
Action	String	No	DescribeScalingRules	The operation that you want to perform. Set the value to DescribeScalingRules.

Parameter	Type	Required	Example	Description
PageNumber	Integer	No	1	The page number of the scaling rule list 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.
ScalingRuleArn . 1	String	No	ari:acs:ess:cn-qingdao:1344371:scalingRule/eMKWG8SRNb9dBLAjweNI1Ik	The unique identifier of scaling rule N to be queried. Valid values of N: 1 to 10.
ScalingRuleId . 1	String	No	eMKWG8SRNb9dBLAjweN****	The ID of scaling rule N to be queried. Valid values of N: 1 to 10.
ScalingRuleName . 1	String	No	eMKWG8SRNb9dBLAjweN****	The name of scaling rule N to be queried. Valid values of N: 1 to 10.

Parameter	Type	Required	Example	Description
ScalingRuleType	String	No	SimpleScalingRule	The type of the scaling rule. Valid values: • SimpleScalingRule: scales ECS instances based on the values of AdjustmentType and AdjustmentValue. • TargetTrackingScalingRule: dynamically calculates the number of ECS instances to be scaled and tries to keep the value of a predefined metric close to TargetValue. • 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. This rule automatically creates scheduled tasks to set the boundary values for the scaling group. Default value: SimpleScalingRule.

Parameter	Type	Required	Example	Description
ShowAlarmRules	Boolean	No	false	Specifies whether to return the CloudMonitor monitoring task associated with a scaling rule. Valid values: · true: The CloudMonitor monitoring task associated with a scaling rule is returned. · false: The CloudMonitor monitoring task associated with a scaling rule is not returned. Default value: false.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The current page number.
PageSize	Integer	50	The number of entries returned on each page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingRules			The set of scaling rules.

Parameter	Type	Example	Description
Adjustment Type	String	QuantityChangeInCapacity	The method used by the scaling rule to adjust the number of instances. Valid values: - QuantityChangeInCapacity: adds or removes a specified number of ECS instances. - PercentChangeInCapacity: adds or removes a specified proportion of ECS instances. - TotalCapacity: adds or removes ECS instances to ensure that the current scaling group has a specified number of ECS instances.
Adjustment Value	Integer	1	The adjustment amount specified for the scaling rule.
Alarms			The CloudMonitor monitoring tasks associated with scaling rules . The CloudMonitor monitoring tasks associated with scaling rules are returned only when the ShowAlarmRules parameter is set to true. Otherwise, an empty list is returned.
AlarmTaskId	String	asg-bp18p2yfxow2dloq****_1f9458d1-70e1-4bee-8c7f-7a47695b****	The ID of the monitoring task associated with a scaling rule.
AlarmTaskName	String	event-notified****	The name of the monitoring task associated with a scaling rule.

Parameter	Type	Example	Description
Comparison Operator	String	>=	The comparison operator of the metric value and threshold for the monitoring task associated with a scaling rule. Valid values: · <=: The metric value is smaller than or equal to the threshold. · <: The metric value is smaller than the threshold. · >: The metric value is greater than the threshold. · >=: The metric value is greater than or equal to the threshold.
Dimensions			The dimensions of the monitoring task associated with a scaling rule.
DimensionKey	String	scaling_group	The key values of dimensions associated with a metric. Valid values: · scaling_group: the ID of the scaling group. · userId: the ID of the user account.
DimensionValue	String	asg-bp18p2yfxow2dloq****	The attribute values of dimensions associated with a metric.
Evaluation Count	Integer	3	The consecutive number of times for which the monitoring task associated with a scaling rule meets the threshold expression to trigger an alert.
MetricName	String	CpuUtilization	The name of the metric of the monitoring task associated with a scaling rule.

Parameter	Type	Example	Description
Statistics	String	Average	The statistical method of the monitoring task associated with a scaling rule. Valid values: · Average · Maximum · Minimum
Threshold	Float	50	The alert threshold for the monitoring task associated with a scaling rule.
Cooldown	Integer	20	The cooldown period of the scaling rule. This parameter is only applicable to simple scaling rules . Valid values: 0 to 86400. Unit: seconds.
DisableScaleIn	Boolean	true	Indicates whether scale-in is disabled. This parameter is only applicable to target tracking scaling rules. Valid values: · true: Scale-in is disabled. · false: Scale-in is enabled.
EstimatedInstanceWarmup	Integer	300	The warmup period of an ECS instance.
InitialMaxSize	Integer	100	The maximum number of ECS instances in the scaling group. This parameter is used in combination with the PredictiveValueBehavior parameter.
MaxSize	Integer	2	The maximum number of ECS instances in the scaling group.

Parameter	Type	Example	Description
MetricName	String	CpuUtilization	The predefined metric that you specified to monitor. This parameter is applicable to target tracking scaling rules and predictive scaling rules. Valid values for target tracking scaling rules: • CpuUtilization: the average CPU utilization • ClassicInternetRx: the average public network inbound traffic over the classic network • ClassicInternetTx: the average public network outbound traffic over the classic network • VpcInternetRx: the average public network inbound traffic over the VPC • VpcInternetTx: the average public network outbound traffic over the VPC • IntranetRx: the average inbound internal traffic • IntranetTx: the average outbound internal traffic Valid values for predictive scaling rules: • CpuUtilization: the average CPU utilization • IntranetRx: the average inbound internal traffic • IntranetTx: the average outbound internal traffic

Parameter	Type	Example	Description
MinAdjustmentMagnitude	Integer	1	The minimum number of ECS instances scaled in a scaling rule . This parameter only takes effect when the scaling rule type is SimpleScalingRule or StepScalingRule and AdjustmentType is 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: produces forecasts and creates prediction tasks. • PredictOnly: produces forecasts but does not create prediction tasks.
PredictiveTaskBufferTime	Integer	30	The amount of buffer time ahead of the prediction task execution time . By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed at the beginning of each hour. You can set a buffer time to execute prediction tasks ahead of schedule , so that resources can be prepared in advance. Valid values: 0 to 60. Unit: minutes.

Parameter	Type	Example	Description
Predictive ValueBehavior	String	MaxOverridePredictiveValue	The action taken on the predicted maximum value. Valid values: - MaxOverridePredictiveValue : uses the initial maximum capacity as the maximum value for prediction tasks when the predicted value is greater than the initial maximum capacity. - PredictiveValueOverrideMax: uses the predicted value as the maximum value for prediction tasks when the predicted value is greater than the initial maximum capacity. - PredictiveValueOverrideMaxWithBuffer: increases the predicted value by a ratio which is specified by PredictiveValueBuffer. If the predicted value increased by this ratio is greater than the initial maximum capacity, the predicted value after increase is used as the maximum value for prediction tasks.
Predictive ValueBuffer	Integer	50	The ratio of the increment to the predicted value when Predictive ValueBehavior is set to Predictive ValueOverrideMaxWithBuffer. If the predicted value increased by this ratio is greater than the initial maximum capacity, the value after the increase is used for prediction tasks. Valid values: 0 to 100.
ScalingGroupUpId	String	AG6CQdPU80KdwLjgZcJ****	The ID of the scaling group.

Parameter	Type	Example	Description
ScalingRuleAri	String	ari:acs:ess:cn-qingdao:1344371:scalingRule/eMKWG8SRNb9dBLAjweNI1Ik	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 ECS instances based on the values of AdjustmentType and AdjustmentValue. • TargetTrackingScalingRule: dynamically calculates the number of ECS instances to be scaled and tries to keep the value of a predefined metric close to TargetValue. • 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. This rule automatically creates scheduled tasks to set the boundary values for the scaling group.
StepAdjustments			The step adjustments of the step scaling rule.

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

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DescribeScalingRules
&RegionId=cn-qingdao
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&PageSize=50
&<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>AG6CQdPU80KdwLjgZcJ****</ScalingGroupId>
      <ScalingRuleAri>ari:acs:ess:cn-qingdao:1344371:scalingRule/eMKWG8SRNb9dBLAjweN****</ScalingRuleAri>
      <ScalingRuleId>eMKWG8SRNb9dBLAjweN****</ScalingRuleId>
```

```

        < ScalingRule eName > eMKWG8SRNb 9dBLAjweN ****</
ScalingRule eName >
        </ ScalingRule e >
    </ ScalingRule es >
    < TotalCount > 1 </ TotalCount >
    < RequestId > 3306A40D - 3412 - 4101 - 9F19 - 5F81E3055D AD </
RequestId >
</ DescribeScalingRules Response >

```

JSON format

```

{
  " PageNumber ": 1 ,
  " TotalCount ": 1 ,
  " PageSize ": 10 ,
  " RequestId ":" B583BFEF - A779 - 427A - 9B74 - 262DDD2497 02 ",
  " ScalingRule es ":{
    " ScalingRule e ":[
      {
        " ScalingRule eAri ":" ari : acs : ess : cn - qingdao : 1344371 :
scalingRule e / efcqrZdjlo okc0UkE3dA ****",
        " Adjustment Type ":" TotalCapacity ",
        " ScalingGroupId ":" ccMvs9dcZl E5c9CtrwbX ****",
        " Adjustment Value ": 5 ,
        " ScalingRule eName ":" KFJoxG ****",
        " Cooldown ": 500 ,
        " ScalingRule eId ":" efcqrZdjlo okc0UkE3dA ****"
      }
    ]
  }
}

```

Error codes

For more information about error codes, visit [API Error Center](#).

11 Trigger task

11.1 ExecuteScalingRule

You can call this operation to execute a specified scaling rule.

Description

- The operation can be called only when the scaling group is in the Active state.
- The operation can be called only when the scaling group does not have any scaling activities in progress.
- If the scaling group does not have any scaling activities in progress, the operation can be immediately executed without the need to wait for the cooldown period to expire.
- When this operation is called, Auto Scaling has received the request. However, it is not guaranteed that the scaling activity can be executed as intended. You can determine the status of the scaling activity based on the returned `ScalingActivityId` parameter value.
- If the number of ECS instances to be attached would leave the number of instances in the scaling group greater than `MaxSize`, the rule `Total Capacity = MaxSize` is applied.
- If the number of ECS instances to be detached would leave the number of instances in the scaling group less than `MinSize`, the rule `Total Capacity = MinSize` is applied.
- A single Alibaba Cloud account can have a maximum of 1,000 automatically created ECS instances across all scaling groups for all regions. ECS instances that are manually added to scaling groups are not affected by and do not count towards this limit.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingRuleAri	String	Yes	ari:acs:ess:cn-qingdao:1344371:scalingRule/cCBpdYdQuBe2cUxOdu6****	The unique identifier of the scaling rule.
Action	String	No	ExecuteScalingRule	The operation that you want to perform. Set the value to ExecuteScalingRule.
BreachThreshold	Float	No	1.0	The threshold specified when the step scaling rule is executed. Valid values: -9.999999E18 to 9.999999E18.
ClientToken	String	No	123e4567-e89b-12d3-a456-426655440000	The client token used to guarantee the idempotency of requests. The parameter value is generated by the client and must be unique among different requests. The maximum length is 64 ASCII characters. For more information, see Appendix How to ensure idempotency .
MetricValue	Float	No	1.0	The metric value specified when the step scaling rule is executed. Valid values: -9.999999E18 to 9.999999E18.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.
ScalingActivityId	String	ebta5WbUzC8gcwUWvfch****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=ExecuteScalingRule
&ScalingRuleAri=ari:acs:ess:cn-qingdao:1344371:
scalingRule/cCBpdYdQuB_e2cUx0du6****
&<Common request parameters>
```

Successful response examples

XML format

```
<ExecuteScalingRuleResponse>
  <ScalingActivityId>ebta5WbUzC8gcwUWvfch_yT4U</ScalingActivityId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</ExecuteScalingRuleResponse>
```

JSON format

```
{
  "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
  "ScalingActivityId": "ebta5WbUzC8gcwUWvfch****"
}
```

Error codes

[View 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 when the specified scaling rule does not exist in the current account.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned when 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 when the scaling group to which the specified scaling rule belongs is not active.
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 when the scaling group to which the specified scaling rule belongs currently has a scaling activity in progress.

HTTP status code	Error code	Error message	Description
400	InsufficientBalance	Your account does not have enough balance.	The error message returned when your account balance is insufficient.
400	QuotaExceed. Instance	Living instance quota exceeded.	The error message returned when the number of ECS instances in the current account reaches the upper limit.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned when an SLB instance for the scaling group to which the specified scaling rule belongs is not active.
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned when health check is not enabled for an SLB instance of the scaling group to which the specified scaling rule belongs.

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 when the network type of the ECS instance attached to the specified SLB instance is different than that 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 when the ECS instance attached to an SLB instance for the specified scaling group is not in the same VPC as the VSwitch.
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned when an ApsaraDB for RDS instance for the scaling group to which the specified scaling rule belongs is not running.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned when the number of IP addresses in the whitelist that can access an ApsaraDB for RDS instance for the scaling group to which the specified scaling rule belongs reaches the upper limit.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned when the number of ECS instances in the specified security group reaches the upper limit.
400	IncorrectCapacity.NoChange	To execute the specified scaling rule, the total capacity will not change.	The error message returned when no changes to the number of instances in the scaling group are made after the scaling rule is executed.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded. ScalingInstance	Scaling instance quota exceeded.	The error message returned when the number of ECS instances to be scaled reaches the upper limit.
400	QuotaExceeded. AfterpayInstance	Living afterpay instance quota exceeded.	The error message returned when the number of Pay-As-You-Go ECS instances to be used reaches the upper limit.
400	ResourceNotAvailable.ECS	The specified region or zone does not offer the specified disk or instance category.	The error message returned when the specified type of ECS instance 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 when the specified type of scaling rule cannot be executed.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments. NoStepFound	No adjustment step found for a metric value of: %s.	The error message returned when no matching adjustment steps are found for the specified metric.
400	MissingParameter. MetricValue	Metric value must be specified for StepScalingRule.	The error message returned when no metric value is specified for the step scaling rule.
400	MissingParameter. BreachThreshold	Breach threshold must be specified for StepScalingRule.	The error message returned when no threshold is specified for the step scaling rule.
400	MetricValueBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned when the specified metric value is invalid.
400	BreachThresholdBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned when the specified threshold is invalid.

11.2 AttachInstances

You can call this operation to add an ECS instance to a scaling group.

Description

The restrictions on ECS instances to be added are as follows:

- The ECS instances and the scaling group must be in the same region.
- The ECS instances must be running.
- The ECS instances have not been added to any scaling group.
- The ECS instances support subscription and pay-as-you-go billing methods.
- If the VSwitchId parameter is specified for a scaling group, ECS instances that belong to a classic network or in different VPCs cannot be added to the scaling group.
- If the VSwitchId parameter is not specified for a scaling group, ECS instances that belong to a VPC cannot be added to the scaling group.

This operation can only be called when the scaling group is active.

This operation can only be called when the scaling group does not have any scaling activities in progress.

If the scaling group does not have any scaling activities in progress, the operation can be immediately executed without waiting for the cooldown period to expire.

If the call is successful, it indicates that Auto Scaling has accepted the request.

However, this does not mean that the scaling activity will succeed. You can determine the status of the scaling activity based on the returned value of the ScalingActivityId parameter.

If the sum of the number of instances to be added specified by this operation and the number of ECS instances currently in the scaling group is greater than the MaxSize value, the call fails.

Manually added ECS instances are not associated with the active scaling configuration of the scaling group.

Debugging

OpenAPI Explorer automatically calculates the signature value. For your convenience, we recommend that you call this operation in OpenAPI Explorer. OpenAPI Explorer dynamically generates the sample code of the operation for different SDKs.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
InstanceId . 1	String	Yes	i-28wt4****	The ID of ECS instance N to be added. Valid values of N: 1 to 20.
Action	String	No	AttachInstances	The operation that you want to perform. Set the value to AttachInstances.
LoadBalancerWeight . 1	Integer	No	50	The weight of ECS backend server N. Valid values of N: 1 to 20. Valid values for this parameter: 0 to 100. Default value: 50.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingActivityId	String	bybj90caOT4ucPMbFhcq****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=AttachInstances
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&InstanceId.1=i-28wt4****
&<Common request parameters>
```

Sample success responses

XML format

```
<AttachInstancesResponse>
  <ScalingActivityId>bybj90caOT4ucPMbFhcqHfA3</
ScalingActivityId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</
RequestId>
```

```
</ AttachInstancesResponse >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E ",
  "ScalingActivityId ":" bybj90caOT 4ucPMbFhcq *****"
}
```

Error codes

For more information about error codes, visit [API Error Center](#).

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 current 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 active.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the specified ECS instance does not exist in the current 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 are not in the same region.
400	InvalidInstanceId. InstanceTypeMismatch	Instance "XXX" and existing active scaling configurations have different instance types.	The error message returned because the type of the specified ECS instance is different from that 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 running.
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 that 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 are not 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 is already 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 the specified scaling group has scaling activities in progress.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the SLB instance associated with the specified scaling group is not active.

HTTP status code	Error code	Error message	Description
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 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 the ECS instance attached to the specified SLB instance is different from that 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 attached to the specified SLB instance is not in the same VPC as the VSwitch specified by VSwitchID.
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned because the RDS instance in the specified scaling group is not running.

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 number of IP addresses in the specified RDS instance whitelist reaches the upper limit.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the number of ECS instances in the specified security group reaches the upper limit.
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 (Total Capacity) exceeds the specified value of MaxSize after ECS instances are added to the specified scaling group.

11.3 RemoveInstances

You can call this operation to remove an ECS instance from a specified scaling group.

Description

- When an automatically created ECS instance is removed from a scaling group, the instance is stopped and released.
- When a manually attached ECS instance is removed from a scaling group, the instance is not stopped or released.
- The operation can be called only when the scaling group is in the Active state.
- The operation can be called only when the scaling group does not have any scaling activities in progress.
- If the scaling group does not have any scaling activities in progress, the operation can be immediately executed without the need to wait for the cooldown period to expire.
- When this operation is called, Auto Scaling has received the request. However, it is not guaranteed that the scaling activity can be executed as intended. You can determine the status of the scaling activity based on the returned `ScalingActivityId` parameter value.
- When the number of ECS instances to be removed would leave the number of ECS instances in the specified scaling group smaller than `MinSize`, the operation fails.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
<code>InstanceId</code> • N	<code>RepeatList</code>	Yes	<code>i-28wt4****</code>	The IDs of ECS instances. You can enter a maximum of 20 IDs.
<code>ScalingGroupId</code>	<code>String</code>	Yes	<code>AG6CQdPU8OKdwLjgZcJ****</code>	The ID of the scaling group.

Parameter	Type	Required	Example	Description
Action	String	No	RemoveInstances	The operation that you want to perform. Set the value to RemoveInstances.
RemovePolicy	String	No	release	Indicates the policy used to remove an instance. Valid values: - recycle: indicates that the instance is stopped and then removed.  Note: This value can be used only when the ScalingPolicy parameter is set to recycle. - release: indicates that the instance is released and then removed. Default value: release.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.
ScalingActivityId	String	bybj9OcaOT4ucPMbFhcq****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=RemoveInstances
&ScalingGroupID=AG6CQdPU80KdwLjgZcJ****
&InstanceId.1=i-28wt4****
```


&< Common request parameters >

Successful response examples

XML format

```
< RemoveInstancesResponse >
  < ScalingActivityId > bybj90ca0T4ucPMbFhcqHfA3 </ ScalingActivityId >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </ RequestId >
</ RemoveInstancesResponse >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned when the specified ECS instance does not exist in the scaling group.

HTTP status code	Error code	Error message	Description
400	InvalidParameter	The specified group does not support the specified RemovePolicy.	The error message returned when the specified scaling group does not support the specified value of the RemovePolicy parameter.
403	Forbidden. Unauthorized	A required authorization for the specified action is not supplied.	The error message returned when 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 when the specified scaling group is not active.
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 when the specified scaling group currently has a scaling activity in progress.

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 when an SLB instance for the scaling group to which the specified scaling rule belongs is not active.
400	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	The error message returned when an ApsaraDB for RDS instance for the scaling group to which the specified scaling rule belongs is not running.
400	IncorrectCapacity.MinSize	To remove the instances, the total capacity will be lesser than the MinSize.	The error message when the number of ECS instances to be removed would leave the number of ECS instances in the specified scaling group smaller than MinSize.

11.4 DetachInstances

You can call this operation to detach one or more ECS instances from a scaling group.

Description

After an ECS instance is detached, it can either exist independent of any scaling group or be attached to another scaling group. For more information, see [AttachInstances](#)

- Detaching an ECS instance does not stop or release it. You can manually [stop](#) or [release](#) an ECS instance as needed.
- The status of the specified scaling group must be Enable (`Enable`).
- Ensure that the specified scaling group does not have any scaling activities in progress.
- If the specified scaling group does not have any scaling activities in progress, the operation can be immediately executed without the need to wait for the [cooldown period](#) to expire.
- If this operation is called, Auto Scaling has received the request. However, it is not guaranteed that the scaling activity can be executed as intended. You can determine the status of the scaling activity based on the returned `ScalingActivityId` parameter value.
- The difference between the number of ECS instances currently in the specified scaling group and the number of ECS instances to be detached must not be smaller than `MinSize` .

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
InstanceId • N	RepeatList	Yes	i-28wt4****	The IDs of ECS instances. Valid values of N: 1 to 20.
ScalingGroup upId	String	Yes	AG6CQdPU8O KdwLjgZcj****	The ID of the scaling group.
Action	String	No	DetachInst ances	The operation that you want to perform. Set the value to DetachInstances.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.
ScalingActivityId	String	asa-****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DetachInstances
&ScalingGroupId=AG6CQdPU80-KdwLjgZcJ-****
&InstanceId.1=i-28wt4-****
&<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

Error codes

[View 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 when the specified scaling group is not enabled.
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 when the specified scaling group currently has a scaling activity in progress.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned when an SLB instance for the specified scaling group is not active.
400	IncorrectDBInstanceStatus	The current status of DB instance “XXX” does not support this action.	The error message returned when an ApsaraDB for RDS instance for the specified scaling group is not running.

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 returned when the number of ECS instances to be detached would leave the number of ECS instances in the specified scaling group smaller than MinSize.
403	Forbidden. Unauthorized	A required authorization for the specified action is not supplied.	The error message returned when you are not authorized to use the DetachInstances operation.
404	InvalidScalingGroupId. NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist.
404	InvalidInstanceId. NotFound	Instance “XXX” does not exist.	The error message returned when the specified ECS instance does not exist.

12 Scheduled tasks

12.1 DescribeScheduledTasks

You can call this operation to query scheduled tasks.

Description

You can query relevant scheduled tasks by specifying the ScheduledAction parameter

.

Debugging

OpenAPI Explorer automatically calculates the signature value. For your convenience, we recommend that you call this operation in OpenAPI Explorer. OpenAPI Explorer dynamically generates the sample code of the operation for different SDKs.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group to which a scheduled task belongs.
Action	String	No	DescribeScheduledTasks	The operation that you want to perform. Set the value to DescribeScheduledTasks.
PageNumber	Integer	No	1	The page number of the scheduled task list 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.

Parameter	Type	Required	Example	Description
ScheduledAction . 1	String	No	ari:acs:ess:cn-qingdao:1344371:scalingrule/cCBpdYdQuBe2cUxOdu6piOk	The operation to be performed when scheduled task N is triggered. Valid values of N: 1 to 20.
ScheduledTaskId . 1	String	No	edRtShc57WGXdt8TlPbr****	The ID of scheduled task N to be queried. Valid values of N: 1 to 20.
ScheduledTaskName . 1	String	No	edRtShc57WGXdt8TlPbr****	The name of scheduled task N to be queried. Valid values of N : 1 to 20.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The current page number.
PageSize	Integer	50	The number of entries returned on each page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScheduledTasks			The set of scheduled tasks.
Description	String	fortest	The description of the scheduled task.
LaunchExpirationTime	Integer	600	The time period during which the failed scheduled task is retried. Unit: seconds. Valid values: 0 to 21600.

Parameter	Type	Example	Description
LaunchTime	String	2014-08-18T10:52Z	The time when the scheduled task was triggered.
MaxValue	Integer	5	The MaxValue specified in the scheduled task that was created by the predictive scaling rule. The MaxValue of the scaling group is modified when the scheduled task is executed.
MinValue	Integer	1	The MinValue specified in the scheduled task that was created by the predictive scaling rule. The MinValue of the scaling group is modified when the scheduled task is executed.
RecurrenceEndTime	String	2014-08-20T16:55Z	The end time after which the scheduled task is no longer repeated.
RecurrenceType	String	Daily	Indicates the interval that a scheduled task is repeated at.
RecurrenceValue	String	1	Indicates how often the scheduled task recurs.
ScheduledAction	String	ari:acs:ess:cn-qingdao:1344371:scalingrule/cCBpdYdQuBe2cUxOdu6****	The operation performed when a scheduled task is triggered.
ScheduledTaskId	String	edRtShc57WGXdt8TlPbr****	The ID of the scheduled task.
ScheduledTaskName	String	edRtShc57WGXdt8TlPbr****	The name of the scheduled task.

Parameter	Type	Example	Description
TaskEnabled	Boolean	true	Indicates whether the scheduled task is enabled.
TotalCount	Integer	1	The total number of scheduled tasks.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DescribeScheduledTasks
&RegionId=cn-qingdao
&ScheduledTaskId.1=edRtShc57WGXdt8TlPbr****
&PageSize=50
&<Common request parameters>
```

Sample success responses

XML format

```
<DescribeScheduledTasksResponse>
  <RequestId>43434132-91C4-4264-8343-681130760A5C</RequestId>
  <TotalCount>1</TotalCount>
  <PageSize>50</PageSize>
  <PageNumber>1</PageNumber>
  <ScheduledTasks>
    <ScheduledTask>
      <TaskEnabled>true</TaskEnabled>
      <ScheduledTaskId>b27CLSc8T478c2iqHr6****</ScheduledTaskId>
      <Description>Scheduled task test</Description>
      <ScheduledTaskName>edRtShc57WGXdt8TlPbr****</ScheduledTaskName>
      <LaunchExpirationTime>120</LaunchExpirationTime>
      <RecurrenceType>Daily</RecurrenceType>
      <RecurrenceEndTime>2014-08-13T19:19Z</RecurrenceEndTime>
      <LaunchTime>2014-08-12T17:55Z</LaunchTime>
      <RecurrenceValue>1</RecurrenceValue>
      <ScheduledAction>ari:acs:ess:cn-qingdao:1344371:scalingrule/qGx9feKlgiadmp3XKer****</ScheduledAction>
    </ScheduledTask>
  </ScheduledTasks>
</DescribeScheduledTasksResponse>
```

JSON format

```
{
  "ScheduledTasks":{
```

```

    " ScheduledTask ":[
      {
        " LaunchExpirationTime ": 120 ,
        " Description ":" Scheduled task test ",
        " Recurrence EndTime ":" 2014 - 08 - 13T19 : 19Z ",
        " Recurrence Type ":" Daily ",
        " TaskEnabled ": true ,
        " ScheduledAction ":" ari : acs : ess : cn - qingdao : 1344371 :
scalingrule / qGx9feK1gi admp3XKer ****",
        " ScheduledTaskName ":" edRtShc57W GXdt8TlPbr ****",
        " LaunchTime ":" 2014 - 08 - 12T17 : 55Z ",
        " ScheduledTaskId ":" b27CLSc8T4 78c2iqHr6 ****",
        " Recurrence Value ":" 1 "
      }
    ]
  },
  " PageNumber ": 1 ,
  " TotalCount ": 1 ,
  " PageSize ": 50 ,
  " RequestId ":" 43434132 - 91C4 - 4264 - 8343 - 681130760A 5C "
}

```

Error codes

For more information about error codes, visit [API Error Center](#).

12.2 DeleteScheduledTask

You can call this operation to delete a specified scheduled task.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScheduledTaskId	String	Yes	edRtShc57W GXdt8TlPbr ****	The ID of the scheduled task.
Action	String	No	DeleteScheduledTask	The operation that you want to perform. Set the value to DeleteScheduledTask.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

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

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScheduledTaskId.NotFound	The specified scheduled task does not exist.	The error message returned when the specified scheduled task does not exist in the current account.

12.3 CreateScheduledTask

You can call this operation to create scheduled tasks based on specified input parameters.

- You can create up to 20 scheduled tasks.
- If a scaling group is disabled or executing a scaling activity, a scheduled task fails to execute a scaling rule. The scheduled task is automatically retried within the period of time specified by `LaunchExpirationTime`. After the period of time specified by `LaunchExpirationTime` expires, the task is abandoned.
- If multiple tasks in the same group are scheduled at similar points in time, the earliest task executes its scaling activity first. A scaling group can execute only one scaling activity at a time. Other tasks attempt to execute the rule within the period of time specified by `LaunchExpirationTime`. If a scaling activity is completed within the period of time specified by `LaunchExpirationTime`, the completed activity will trigger the next scheduled scaling rule and execute the scaling activity.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The ID of the region where the scheduled task is located.
ScheduledAction	String	Yes	ari:acs:ess:cn-qingdao:1344371:scalingRule/cCBpdYdQuBe2cUxOdu6piOk	The operations performed when the scheduled task is triggered. When you set this parameter, you must also enter the unique identifier of the scaling rule.

Parameter	Type	Required	Example	Description
Action	String	No	CreateScheduledTask	The operation that you want to perform. Set the value to CreateScheduledTask.
Description	String	No	fortest	The description of the scheduled task. The description must be 2 to 200 characters in length.
LaunchExpirationTime	Integer	No	600	The period of time for which the scheduled task is retried if it fails. Default value: 600. Unit: seconds. Valid values: 0 to 21600.
LaunchTime	String	No	2014-08-17T16:52Z	The time at which the scheduled task is triggered. The time follows the ISO 8601 standard and uses UTC time. The format is YYYY-MM-DDThh:mmZ. If the RecurrenceType parameter is specified, the task is executed each day at the time specified by LaunchTime. If the RecurrenceType parameter is not specified, the task is only executed once at the time specified by LaunchTime. You cannot enter a point in time later than 90 days from the date of scheduled task creation or modification.

Parameter	Type	Required	Example	Description
RecurrenceEndTime	String	No	2014-08-17T16:55Z	The end time after which the scheduled task is no longer repeated. The time follows the ISO 8601 standard and uses UTC time. The format is YYYY-MM-DDThh:mmZ. You cannot enter a point in time later than 90 days from the date of scheduled task creation or modification. If you set RecurrenceEndTime, you must also set both RecurrenceType and RecurrenceValue.

Parameter	Type	Required	Example	Description
Recurrence Type	String	No	Daily	Indicates the interval for which the scheduled task is repeated . Valid values: · Daily: The scheduled task is executed recurrently after the specified number of days · Weekly: The scheduled task is executed on each specified day of a week. · Monthly: The scheduled task is executed on each specified day of a month. · Cron: The scheduled task is executed recurrently based on the specified Cron expression. If you set RecurrenceType, you must also set both RecurrenceValue and RecurrenceEndTime.

Parameter	Type	Required	Example	Description
Recurrence Value	String	No	1	Indicates how often the scheduled task recurs. • Daily: indicates the interval of days at which the scheduled task is repeated. You can enter a single value ranging from 1 to 31. • Weekly: indicates which days of the week that the scheduled task is repeated on. The values 0 to 6 correspond to the days of the week in sequence from Sunday to Saturday. Multiple values must be separated by commas (,). • Monthly: indicates which days of the month that the scheduled task is repeated on. You can enter two values ranging from 1 to 31. The format is A-B. B must be greater than or equal to A. • Cron: indicates a user-defined Cron expression that determines when the scheduled task is repeated. A Cron expression is written in UTC time and consists of five fields: minute, hour, day of month (date), month, and day of week. The expression can contain wildcard characters including commas (,), question marks (?), hyphens (-), asterisks (*), number signs (#), forward slashes (/), and the L and W characters. If you set RecurrenceValue, you must also set both RecurrenceType and RecurrenceEndTime.

Parameter	Type	Required	Example	Description
ScheduledTaskName	String	No	test	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 (.). It must start with a letter or digit. The name of the scheduled task must be unique to an Alibaba Cloud account in a region. The default name is the ID of the scheduled scaling task.
TaskEnabled	Boolean	No	true	Indicates whether to enable the scheduled task. - When the parameter is set to true, the task is enabled. - When the parameter is set to false, the task is disabled. Default value: true.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.
ScheduledTaskId	String	edRtShc57WGXdt8TlPbr****	The ID of the scheduled task. It is a globally unique identifier (GUID) generated by the system.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=CreateScheduledTask
&RegionId=cn-qingdao
&LaunchTime=2014-08-17T16:52Z
&RecurrenceType=Daily
&RecurrenceValue=1
&RecurrenceEndTime=2014-08-17T16:55Z
&ScheduledAction=ari:acs:ess:cn-qingdao:1344371:
scalingRule/cCBpdYdQuB_e2cUx0du6p_i0k
<Common request parameters>
```

Successful response examples

XML format

```
<CreateScheduledTaskResponse>
  <ScheduledTaskId>edRtShc57W_GXdt8TlPbr ****</ScheduledTaskId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8_3E</RequestId>
</CreateScheduledTaskResponse>
```

JSON format

```
{
  "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8_3E",
  "ScheduledTaskId": "edRtShc57W_GXdt8TlPbr ****"
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidRegionId.NotFound	The specified region does not exist.	The error message returned when the specified region does not exist.
400	InvalidScheduledTaskName.Duplicate	The specified value of parameter ScheduledTaskName is duplicated.	The error message returned when 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 when the number of scheduled tasks that can be created by a user reaches the upper limit.
400	ScheduledAction. RegionMismatch	The specified scheduled task and the specified scheduled action are not in the same Region.	The error message returned when the specified value of the ScheduledAction parameter and the specified scheduled task are not in the same region.

12.4 ModifyScheduledTask

You can call this operation to modify the attributes of a scheduled task.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScheduledTaskId	String	Yes	edRtShc57WGXdt8TlPbr****	The ID of the scheduled task for which you want to modify attributes.

Parameter	Type	Required	Example	Description
Action	String	No	ModifyScheduledTask	The operation that you want to perform. Set the value to ModifyScheduledTask.
Description	String	No	fortest	The description of the scheduled task. The description must be 2 to 200 characters in length.
LaunchExpirationTime	Integer	No	600	The period of time for which the scheduled task is retried if it fails. Default value: 600. Unit : seconds. Valid values: 0 to 21600.

Parameter	Type	Required	Example	Description
LaunchTime	String	No	2014-08-18T10:52Z	The time at which the scheduled task is triggered. The time follows the ISO 8601 standard and uses UTC time . The format is YYYY-MM-DDThh:mmZ. If the RecurrenceType parameter is specified, the task is executed each day at the time specified by LaunchTime. If the RecurrenceType parameter is not specified , the task is only executed once at the time specified by LaunchTime. You cannot enter a point in time later than 90 days from the date of scheduled task creation or modification.

Parameter	Type	Required	Example	Description
Recurrence EndTime	String	No	2014-08-20T16:55Z	The end time after which the scheduled task is no longer repeated. The time follows the ISO 8601 standard and uses UTC time. The format is YYYY-MM-DDThh:mmZ. You cannot enter a point in time later than 90 days from the date of scheduled task creation or modification. If you set RecurrenceEndTime, you must also set both RecurrenceType and RecurrenceValue.
Recurrence Type	String	No	Daily	Indicates the interval for which the scheduled task is repeated . Valid values: • Daily: The scheduled task is executed recurrently after the specified number of days . • Weekly: The scheduled task is executed on each specified day of a week. • Monthly: The scheduled task is executed on each specified day of a month. • Cron: The scheduled task is executed recurrently based on the specified Cron expression. If you set RecurrenceType, you must also set both RecurrenceValue and RecurrenceEndTime.

Parameter	Type	Required	Example	Description
Recurrence Value	String	No	2	Indicates how often the scheduled task recurs. • Daily: indicates the interval of days at which the scheduled task is repeated. You can enter a single value ranging from 1 to 31. • Weekly: indicates which days of the week that the scheduled task is repeated on. The values 0 to 6 correspond to the days of the week in sequence from Sunday to Saturday. Multiple values must be separated by commas (,). • Monthly: indicates which days of the month that the scheduled task is repeated on. You can enter two values ranging from 1 to 31. The format is A-B. B must be greater than or equal to A. • Cron: indicates a user-defined Cron expression that determines when the scheduled task is repeated. A Cron expression is written in UTC time and consists of five fields: minute, hour, day of month (date), month, and day of week. The expression can contain wildcard characters including commas (,), question marks (?), hyphens (-), asterisks (*), number signs (#), and forward slashes (/). If you set RecurrenceValue, you must also set both RecurrenceType and RecurrenceEndTime.

Parameter	Type	Required	Example	Description
ScheduledAction	String	No	ari:acs:ess:cn-qingdao:1344371:scalingRule/cCBpdYdQuBe2cUxOdu6piOk	The operations performed when the scheduled task is triggered. When you set this parameter, you must also enter the unique identifier of the scaling rule.
ScheduledTaskName	String	No	test	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 (.). It must start with a letter or digit. The name of the scheduled task must be unique to an Alibaba Cloud account in a region. The default name is the ID of the scheduled scaling task.
TaskEnabled	Boolean	No	true	Indicates whether to enable the scheduled task. - When the parameter is set to true, the task is enabled. - When the parameter is set to false, the task is disabled. Default value: true.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

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

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScheduledTaskId.NotFound	The specified scheduled task does not exist.	The error message returned when the specified scheduled task does not exist in the current account.

HTTP status code	Error code	Error message	Description
400	InvalidScheduledTaskName. Duplicate	The specified value of parameter <code>ScheduledTaskName</code> is duplicated.	The error message returned when 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 when the specified value of the ScheduledAction parameter and the specified scheduled task are not in the same region.

13 Lifecycle Hook

13.1 CreateLifecycleHook

You can call this operation to create lifecycle hooks for a scaling group (CreateLifecycleHook).

Description

Each scaling group can have a maximum of six lifecycle hooks. Lifecycle hooks enable instances in a scaling group to remain in a wait state for a finite period of time during a scale-in or scale-out event. You can adjust this time by configuring the `HeartbeatTimeout` parameter. The instances remain in a wait state for a specified period of time, and then the scaling group continues the scale-in or scale-out process. While the instances are in the wait state, scaling activities are not performed until they are manually continued or until the wait period times out. During a scale-out process, ECS instances enter the wait state after their IP addresses are added to the whitelist of a specified ApsaraDB for RDS instance. When the wait state ends, the IP addresses will be added to a specified SLB instance. During a scale-in process, ECS instances enter the wait state after their IP addresses are removed from the whitelist of an SLB instance. When the wait state ends, the IP addresses will be removed from the whitelist of an ApsaraDB for RDS instance. The procedure is as follows. We recommend that you use an Alibaba Cloud Message Service (MNS) queue or topic to receive notifications about when ECS instances are started or released.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
LifecycleTransition	String	Yes	SCALE_OUT	The type of scaling activity to which lifecycle hooks apply. Valid values: - SCALE_OUT: scale-out event - SCALE_IN: scale-in event
ScalingGroupID	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.
Action	String	No	CreateLifecycleHook	The operation that you want to perform. Set the value to CreateLifecycleHook.

Parameter	Type	Required	Example	Description
DefaultResult	String	No	CONTINUE	The action that the scaling group takes when the lifecycle hook times out. Valid values: · CONTINUE: The scaling group continues the scale-in or scale-out process. · ABANDON: The scaling group stops any remaining actions of the scale-in or scale-out event. Default value: CONTINUE. During a scale-in event (<code>SCALE_IN</code>), if the scaling group has multiple lifecycle hooks and one of them is terminated by the <code>DefaultResult = ABANDON</code> parameter, the remaining lifecycle hooks in the same scaling group are also terminated. Otherwise, the scaling activity will proceed normally after the wait period expires and continue with the action specified in the DefaultResult parameter.

Parameter	Type	Required	Example	Description
HeartbeatTimeout	Integer	No	600	The time that can elapse before the lifecycle hook times out. If the lifecycle hook times out, the scaling group performs the default action (DefaultResult). Valid values: 30 to 21600. Unit: seconds. Default value: 600. You can prevent the lifecycle hook from timing out by calling the RecordLifecycleActionHeartbeat operation. You can also terminate the lifecycle action by calling the CompleteLifecycleAction operation.

Parameter	Type	Required	Example	Description
LifecycleHook.N.DefaultResult	String	No	CONTINUE	The action that the scaling group takes when the lifecycle hook times out. Valid values: · CONTINUE: The scaling group continues the scale-in or scale-out process. · ABANDON: The scaling group stops any remaining actions of the scale-in or scale-out event. Default value: CONTINUE. During a scale-in event (<code>SCALE_IN</code>), if the scaling group has multiple lifecycle hooks and one of them is terminated by the <code>DefaultResult = ABANDON</code> parameter, the remaining lifecycle hooks in the same scaling group are also terminated. Otherwise, the scaling activity will proceed normally after the wait period expires and continue with the action specified in the <code>DefaultResult</code> parameter.

Parameter	Type	Required	Example	Description
LifecycleHook.N.HeartbeatTimeout	Integer	No	600	The time that can elapse before the lifecycle hook times out. If the lifecycle hook times out, the scaling group performs the default action (DefaultResult). Valid values: 30 to 21600. Unit: seconds. Default value: 600. You can prevent the lifecycle hook from timing out by calling the RecordLifecycleActionHeartbeat operation. You can also terminate the lifecycle action by calling the CompleteLifecycleAction operation.
LifecycleHook.N.LifecycleHookName	String	No	TESTSCALE_OUT	The name of the lifecycle hook. Each lifecycle hook name must be unique within a scaling group. The name must be 2 to 40 characters in length and can contain letters, digits, and special characters including underscores (_), hyphens (-), and periods (.). The default name is the ID of the lifecycle hook.
LifecycleHook.N.LifecycleTransition	String	No	SCALE_OUT	The type of scaling activity to which lifecycle hooks apply. Valid values: - SCALE_OUT: scale-out event - SCALE_IN: scale-in event

Parameter	Type	Required	Example	Description
LifecycleHook.N.NotificationArn	String	No	acs:ess:cn-hangzhou:1111111111:queue/queue1	The Alibaba Cloud Resource Name (ARN) of the notification target that Auto Scaling uses to notify you when an instance is in the transition state for a lifecycle hook. This target can be either an MNS queue or an MNS topic. The format of the parameter value is <code>acs:ess:{region}:{account-id}:{resource-relative-id}</code>. <code>region</code> : the region where the scaling group is located. <code>account-id</code> : the ID of the Alibaba Cloud account. Example: - MNS queue: <code>acs:ess:{region}:{account-id}:queue/{queuename}</code> - MNS topic: <code>acs:ess:{region}:{account-id}:topic/{topicname}</code>

Parameter	Type	Required	Example	Description
LifecycleHookName	String	No	Test	The fixed string that you want to include when Auto Scaling sends a message about the wait state of a scaling activity to the notification target. The parameter value can be up to 128 characters in length. Auto Scaling sends the specified <code>NotificationMetadata</code> parameter value along with the notification message so that you can easily categorize notifications. After you set the <code>NotificationArn</code> parameter, the <code>NotificationMetadata</code> parameter takes effect.
LifecycleHookName	String	No	TESTSCALE_OUT	The name of the lifecycle hook. Each lifecycle hook name must be unique within a scaling group. The name must be 2 to 40 characters in length and can contain letters, digits, and special characters including underscores (_), hyphens (-), and periods (.). The default name is the ID of the lifecycle hook.

Parameter	Type	Required	Example	Description
NotificationArn	String	No	acs:ess:cn-hangzhou:1111111111:queue/queue1	The Alibaba Cloud Resource Name (ARN) of the notification target that Auto Scaling uses to notify you when an instance is in the transition state for the lifecycle hook. This target can be either an MNS queue or an MNS topic. The format of the parameter value is <code>acs:ess:{region}:{account-id}:{resource-relative-id}</code>. <code>region</code> : the region where the scaling group is located. <code>account-id</code> : the ID of the Alibaba Cloud account. Example: - MNS queue: <code>acs:ess:{region}:{account-id}:queue/{queuename}</code> - MNS topic: <code>acs:ess:{region}:{account-id}:topic/{topicname}</code>

Parameter	Type	Required	Example	Description
NotificationMetadata	String	No	Test	The fixed string that you want to include when Auto Scaling sends a message about the wait state of the scaling activity to the notification target. The parameter value can be up to 128 characters in length. Auto Scaling sends the specified <code>NotificationMetadata</code> parameter value along with the notification message so that you can easily categorize notifications. After you set the <code>NotificationArn</code> parameter, the <code>NotificationMetadata</code> parameter takes effect.

Response parameters

Parameter	Type	Example	Description
LifecycleHookId	String	ash-****	The ID of the lifecycle hook.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=CreateLifecycleHook
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&LifecycleHookName=TESTSCALE_OUT
&LifecycleTransition=SCALE_OUT
```

```
& NotificationArn = acs : ess : cn - hangzhou : 1111111111 : queue /
queue1
& NotificationMetadata = Test
< Common request parameters >
```

Successful response examples

XML format

```
< CreateLifecycleHookResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
  < LifecycleHookId > ash -****</ LifecycleHookId >
</ CreateLifecycleHookResponse >
```

JSON format

```
{
  "LifecycleHookId ":" ash -****",
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidParameter	The specified value of parameter is not valid.	The error message returned when the specified parameter value is invalid.
400	InvalidNotificationArn	The specified parameter notificationArn is invalid.	The error message returned when the specified value of the NotificationArn parameter is invalid.

HTTP status code	Error code	Error message	Description
400	UnsupportedNotificationType. CurrentRegion	The notificationType is not supported in the special region which scalingGroup belongs to.	The error message returned when the specified notification type is not supported in the region.
400	QueueNotExist	The specified queue does not exist.	The error message returned when the specified MNS queue does not exist.
400	TopicNotExist	The specified topic does not exist.	The error message returned when the specified MNS topic does not exist.
400	InvalidLifecycleHookName. Duplicate	The specified value of parameter lifecycleHookName is duplicated.	The error message returned when the specified value of the LifecycleHookName parameter already exists.
400	QuotaExceeded. LifecycleHook	Lifecycle hook quota exceeded in the specified scaling group.	The error message returned when the number of lifecycle hooks in the specified scaling group reaches the upper limit.

13.2 ModifyLifecycleHook

You can call this operation to modify the properties of a lifecycle hook.

Description

You can use either of the following methods to modify the properties of a lifecycle hook:

- Specify the lifecycle hook ID (`LifecycleHookId`). This way, you do not need to specify the `ScalingGroupId` or `LifecycleHookName` parameters.
- Specify the `ScalingGroupId` and `LifecycleHookName` parameters at the same time.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	No	ModifyLifecycleHook	The operation that you want to perform. Set the value to <code>ModifyLifecycleHook</code> .

Parameter	Type	Required	Example	Description
DefaultResult	String	No	CONTINUE	The action that the scaling group takes when the wait state terminates. Valid values: • CONTINUE: The scaling group continues with the scale-in or scale-out process. • ABANDON: The scaling group stops the ongoing scale-in or scale-out event. Default value: CONTINUE If the scaling group has multiple lifecycle hooks and one of them is terminated by the <code>SCALE_IN</code> parameter during a scale-in event (<code>DefaultResult = ABANDON</code>), the wait states triggered by remaining lifecycle hooks in the same scaling group are also terminated. Otherwise, the action following the wait state is the next action, as specified in the <code>DefaultResult</code> parameter, after the last lifecycle event in the same scaling group.

Parameter	Type	Required	Example	Description
HeartbeatTimeout	Integer	No	600	The time that can elapse before the lifecycle hook times out. If the lifecycle hook times out, the scaling group performs the default action (DefaultResult). Valid values: 30-21600. Default value: 600. Unit: seconds. You can call RecordLifecycleActionHeartbeat to prevent the lifecycle hook from timing out. You can also call CompleteLifecycleAction to terminate the lifecycle action.
LifecycleHookId	String	No	ash-***	The ID of the lifecycle hook. Each ID must be unique within a scaling group. The ID cannot be modified if it is set.
LifecycleHookName	String	No	hook_name_test	The name of the lifecycle hook. Each name must be unique within a scaling group. The name cannot be modified if it is set.
LifecycleTransition	String	No	SCALE_IN	The type of the scaling activity to which the lifecycle hook corresponds. Valid values: • SCALE_OUT: indicates a scale-out event. • SCALE_IN: indicates a scale-in event.

Parameter	Type	Required	Example	Description
NotificationArn	String	No	acs:ess:cn-hangzhou:1111111111:queue/queue2	The Alibaba Cloud Resource Name (ARN) of the notification object that Auto Scaling uses to notify you when an instance is in the transition state for the lifecycle hook. This object can be either an MNS queue or an MNS topic. The format of the parameter value is <code>acs:ess:{region}:{account-id}:{resource-relative-id}</code>. <code>region</code> : the region where the scaling group resides. <code>account-id</code> : the ID of the Alibaba Cloud account. Example: - MNS queue: <code>acs:ess:{region}:{account-id}:queue/{queuename}</code> - MNS topic: <code>acs:ess:{region}:{account-id}:topic/{topicname}</code>

Parameter	Type	Required	Example	Description
NotificationMetadata	String	No	Test	The fixed string that is included in the message that is sent by Auto Scaling about the wait state of a scaling activity to the notification object. The parameter value must be 1 to 128 characters in length. Auto Scaling sends the specified NotificationMetadata parameter value along with the notification message so that you can easily categorize your notifications. The NotificationArn parameter is applicable only after you set the NotificationMetadata parameter.
ScalingGroupID	String	No	AG6CQdPU8OKdwLjgZcJ2***	The ID of the scaling group. Each ID must be unique within a scaling group. The ID cannot be modified if it is set.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=ModifyLifecycleHook
&ScalingGroupId=AG6CQdPU80&LifecycleHookName=hook_name_test
&LifecycleTransition=SCALE_IN
&DefaultResult=ABANDON
&NotificationArn=acs:ess:cn-hangzhou:111111111:queue/queue2
&NotificationMetadata=Test
<Common request parameters>
```

Successful response examples

XML format

```
<ModifyLifecycleHookResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8-3E</RequestId>
</ModifyLifecycleHookResponse>
```

JSON format

```
{
  "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8-3E"
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidParameter	The specified value of parameter is not valid.	The error message returned when the value of the parameter is invalid.
400	InvalidLifecycleHookId.NotFound	The specified lifecycleHookId not exist.	The error message returned when the specified lifecycle hook ID does not exist.

HTTP status code	Error code	Error message	Description
400	InvalidLifecycleHookName.NotFound	The specified lifecycleHookName you provided not exist.	The error message returned when the specified lifecycle hook name does not exist.
400	InvalidNotificationArn	The specified parameter **notificationArn** is invalid.	The error message returned when the specified Alibaba Cloud Resource Name (ARN) of the notification object does not exist.
400	UnsupportedNotificationType.CurrentRegion	The **notificationType** is not supported in the special region which scalingGroup belongs to.	The error message returned when the notification type is not supported in the region where the scaling group resides.
400	LifecycleHook	The specified queue does not exist.	The error message returned when the specified MNS queue does not exist.
400	TopicNotExist	The specified topic does not exist.	The error message returned when the specified MNS topic does not exist.

13.3 DescribeLifecycleHooks

You can call this operation to query the list of lifecycle hooks that meet one or more specified conditions.

Description

You can use one of the following three methods to query lifecycle hooks:

- Specify the `LifecycleHookId . N` parameter. This way, you do not need to specify the `ScalingGroupId` or `LifecycleHookName` parameters.
- Specify the `ScalingGroupId` parameter.
- Specify the `ScalingGroupId` and `LifecycleHookName` parameters at the same time.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	No	DescribeLifecycleHooks	The operation that you want to perform. Set the value to DescribeLifecycleHooks.
LifecycleHookId . N	RepeatList	No	ash-****	The ID of the lifecycle hook.
LifecycleHookName	String	No	Test	The name of the lifecycle hook.
PageNumber	Integer	No	1	The page number that you query in the instance status list. Starting value: 1. Default value: 1.

Parameter	Type	Required	Example	Description
PageSize	Integer	No	50	The number of rows per page . Maximum value: 50. Default value: 50.
ScalingGroupId	String	No	asg-****	The ID of the scaling group.

Response parameters

Parameter	Type	Example	Description
LifecycleHooks			The list of lifecycle hook information.
↳DefaultResult	String	CONTINUE	The action that the scaling group takes when the wait state terminates.
↳HeartbeatTimeout	Integer	60	The time that can elapse before the lifecycle hook times out. If the lifecycle hook times out, the scaling group performs the default action (DefaultResult).
↳LifecycleHookId	String	ash-****	The ID of the lifecycle hook.
↳LifecycleHookName	String	Test	The name of the lifecycle hook.
↳LifecycleTransition	String	SCALE_OUT	The type of the scaling activity to which the lifecycle hook corresponds.

Parameter	Type	Example	Description
<code>NotificationArn</code>	String	acs:ess:cn-hangzhou:111111111:queue/queue1	The Alibaba Cloud Resource Name (ARN) of the notification object that Auto Scaling uses to notify you when an instance is in the transition state for the lifecycle hook.
<code>NotificationMetadata</code>	String	Test	The fixed string that is included in the message that is sent by Auto Scaling about the wait state of a scaling activity to the notification object.
<code>ScalingGroupId</code>	String	asg-****	The ID of the scaling group.
<code>PageNumber</code>	Integer	1	The initial page number for the query.
<code>PageSize</code>	Integer	50	The number of rows returned per page when you perform the query.
<code>RequestId</code>	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.
<code>TotalCount</code>	Integer	1	The total count of lifecycle hooks.

Examples

Sample requests

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

Successful response examples

XML format

```
< DescribeLifecycleHooksResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
  < PageNumber > 1 </ PageNumber >
  < PageSize > 50 </ PageSize >
  < TotalCount > 1 </ TotalCount >
  < LifecycleHooks >
    < LifecycleHook >
      < ScalingGroupUpId > asg - xxx </ ScalingGroupUpId >
      < LifecycleHookId > ash - xxx </ LifecycleHookId >
      < LifecycleHookName > Test </ LifecycleHookName >
      < DefaultResult > CONTINUE </ DefaultResult >
      < HeartbeatTimeout > 60 </ HeartbeatTimeout >
      < LifecycleTransition > SCALE_OUT </ LifecycleTransition >
      < NotificationMetadata > Test </ NotificationMetadata >
      < NotificationArn > acs : ess : cn - hangzhou : 1111111111 :
queue / queue1 </ NotificationArn >
    </ LifecycleHook >
  </ LifecycleHooks >
</ DescribeLifecycleHooksResponse >
```

JSON format

```
{
  "lifecycleHooks": [
    {
      "notificationArn": "acs : ess : cn - hangzhou : 1111111111 :
queue / queue1",
      "lifecycleHookId": "ash -****",
      "notificationMetadata": "Test",
      "lifecycleTransition": "SCALE_OUT",
      "defaultResult": "CONTINUE",
      "lifecycleHookName": "Test",
      "scalingGroupUpId": "asg -****",
      "heartbeatTimeout": 600
    }
  ],
  "totalCount": 1,
  "requestId": "473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E",
  "pageSize": 50,
  "pageNumber": 1
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned when the value of the parameter is invalid.

13.4 RecordLifecycleActionHeartbeat

You can call this operation to extend the timeout period during which an instance keeps in the wait state.

Description

An ECS instance can keep in the wait state for up to 6 hours. The duration of each wait state can be extended up to 20 times.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
lifecycleActionToken	String	Yes	aaaa-bbbbbb-cccc-ddddd	The token that indicates a specific scaling activity associated with an instance. You can obtain this token from an MNS queue or MNS topic specified for the lifecycle hook.
lifecycleHookId	String	Yes	ash-****	The ID of the lifecycle hook.

Parameter	Type	Required	Example	Description
Action	String	No	RecordLifecycleActionHeartbeat	The operation that you want to perform. Set the value to RecordLifecycleActionHeartbeat.
heartbeatTimeout	Integer	No	600	The time that can elapse before the lifecycle hook times out. If the lifecycle hook times out, the scaling group performs the default action (DefaultResult). Valid values: 30-21600. Default value: 600. Unit: seconds. You can call RecordLifecycleActionHeartbeat to prevent the lifecycle hook from timing out. You can also call CompleteLifecycleAction to terminate the lifecycle action.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=RecordLifecycleActionHeartbeat
&LifecycleHookId=ash-****
&LifecycleActionToken=aaaa-bbbb-cccc-dddd
&LifecycleActionResult=CONTINUE
```

&< Common request parameters >

Successful response examples

XML format

```
< RecordLife cycleActio nHeartbeat Response >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
    RequestId >
</ RecordLife cycleActio nHeartbeat Response >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned when the value of the parameter is invalid .
400	LifecycleHookIdAndLifecycleActionToken.Invalid	The specified lifecycleActionToken and lifecycleHookId you provided does not match any in process lifecycle action.	The error message returned when the specified lifecycle action token does not match the lifecycle hook ID .

HTTP status code	Error code	Error message	Description
400	LifecycleAction. TimeExceeded	The specified parameter heartbeatTime exceed lifecycleAction max suspend time.	The error message returned when the period during which the ECS instance keeps in the wait state exceeds the maximum value of 6 hours.
400	LifecycleAction. RecordTimesExceeded	The specified lifecycleAction exceed lifecycleAction max record times.	The error message returned when the duration of each wait state is extended for more than 20 times.

13.5 CompleteLifecycleAction

You can call this operation to terminate the wait state of a specified scaling activity before the corresponding lifecycle hook times out.

After the wait state is terminated, you can have the current scaling activity continue or (`CONTINUE`) roll back this scaling activity (`ABANDON`).

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
LifecycleActionToken	String	Yes	aaaa-bbbbb-cccc-ddddd	The token that indicates a specific scaling activity associated with an instance. You can obtain this token from an MNS queue or MNS topic specified for the lifecycle hook.
LifecycleHookId	String	Yes	ash-****	The ID of the lifecycle hook.
Action	String	No	CompleteLifecycleAction	The operation that you want to perform. Set the value to CompleteLifecycleAction.

Parameter	Type	Required	Example	Description
LifecycleActionResult	String	No	CONTINUE	The action that the scaling group takes when the wait state terminates. Valid values: · CONTINUE: The scaling group continues with the scale-in or scale-out process. · ABANDON: The scaling group stops the ongoing scale-in or scale-out event. Default value: CONTINUE If the scaling group has multiple lifecycle hooks and one of them is terminated by the <code>SCALE_IN</code> parameter during a scale-in event (<code>DefaultResult = ABANDON</code>), the wait states triggered by remaining lifecycle hooks in the same scaling group are also terminated. Otherwise, the action following the wait state is the next action, as specified in the <code>DefaultResult</code> parameter, after the last lifecycle event in the same scaling group.
OwnerAccount	String	No	ECSforCloud@Alibaba.com	The logon name of a RAM user.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http :// ess . aliyuncs . com /? Action = CompleteLi fecycleAct
ion
& LifecycleH ookId = ash -****
& LifecycleA ctionToken = aaaa - bbbbb - cccc - dddd
& LifecycleA ctionResul t = CONTINUE
&< Common request parameters >
```

Successful response examples

XML format

```
< CompleteLi fecycleAct ionRespons e >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
</ CompleteLi fecycleAct ionRespons e >
```

JSON format

```
{
  " RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is invalid.	The error message returned when the value of the parameter is invalid.

HTTP status code	Error code	Error message	Description
400	LifecycleHookIdAndLifecycleActionToken.Invalid	The specified lifecycleActionToken and lifecycleHookId you provided does not match any in process lifecycle action.	The error message returned when the specified lifecycle action token does not match the lifecycle hook ID.

13.6 DeleteLifecycleHook

You can call this operation to delete a lifecycle hook.

Description

The wait state is terminated if the corresponding lifecycle hook is deleted. You can use either of the following methods to delete a lifecycle hook:

- Specify the lifecycle hook ID (`LifecycleHookId`). This way, you do not need to specify the `ScalingGroupID` or `LifecycleHookName` parameter.
- Specify the `ScalingGroupID` and `LifecycleHookName` parameters at the same time.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	No	DeleteLifecycleHook	The operation that you want to perform. Set the value to DeleteLifecycleHook.
LifecycleHookId	String	No	ash-****	The ID of the lifecycle hook.

Parameter	Type	Required	Example	Description
LifecycleHookName	String	No	hook_name_test	The name of the lifecycle hook.
ScalingGroupID	String	No	dP8VqSd9EN XPc0ciVmbc *****	The ID of the scaling group.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DeleteLifecycleHook
&LifecycleHookId=ash-****
&<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is invalid.	The error message returned when the value of the parameter is invalid .
400	InvalidLifecycleHookId. NotExist	The specified lifecycleHookId does not exist.	The error message returned when the specified lifecycle hook ID does not exist.
400	InvalidLifecycleHookName. NotExist	The specified lifecycleHookName you provided does not exist.	The error message returned when the specified lifecycle hook name does not exist.

14 Event notification

14.1 CreateNotificationConfiguration

You can call this operation to configure notifications of Auto Scaling events and resource changes.

You can configure the event notification function to send notifications to CloudMonitor or system events, MNS queues, or MNS topics. When a scaling event or resource change of a specified type occurs in a scaling group, Auto Scaling sends notifications to CloudMonitor or MNS. You cannot use MNS queues or MNS topics in several Alibaba Cloud regions. For more information, see MNS FAQ or visit the MNS console.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
NotificationArn	String	Yes	acs:ess:cn-hangzhou:123456:cloudmonitor	The Alibaba Cloud Resource Name (ARN) for the notification object. The format is <code>acs : ess :{ region } :{ account - id } :{ resource - relative - id }</code>. <code>region</code> : the ID of the region where the scaling group resides. For more information, see Regions and zones. <code>account - id</code> : the ID of your account. <code>resource - relative - id</code> : the notification method. Valid values: <code>cloudmonit</code> or <code>queue/{ queuename }</code>, and <code>topic/{ topicname }</code>.
NotificationType . N	RepeatList	Yes	AUTOSCALING:SCALE_OUT_SUCCESS	The number of types of Auto Scaling events and resource changes. Valid values of N: 1 to 8. Multiple event types are listed in a column. You can also call DescribeNotificationTypes to query the parameter values.
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
Action	String	No	CreateNotificationConfiguration	The operation that you want to perform. Set the value to CreateNotificationConfiguration.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=CreateNotificationConfiguration
&ScalingGroupId=AG6CQdPU80&KdwljgZcJ****
&NotificationArn=acs:ess:cn-hangzhou:123456:cloudmonitor
&NotificationType.1=AUTOSCALING:SCALE_OUT_SUCCESS
&NotificationType.2=AUTOSCALING:SCALE_IN_SUCCESS
&NotificationType.3=AUTOSCALING:SCALE_OUT_ERROR
&NotificationType.4=AUTOSCALING:SCALE_IN_ERROR
&NotificationType.5=AUTOSCALING:SCALE_REJECT
<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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


Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidNotificationArn	The specified parameter notificationArn is invalid.	The error message returned when the specified notification ARN is invalid.
400	InvalidNotificationTypes	The specified notificationType is invalid.	The error message returned when a specified notification type is invalid.
400	NotificationConfigurationExist	The specified notificationConfiguration already exist for the scalingGroup.	The error message returned when the specified event notification function is already configured for the current scaling group.
400	NotificationConfigurationQuotaExceed.ForScalingGroup	NotificationConfiguration num exceed for the specified scalingGroup.	The error message returned when the number of event types set for the scaling group reaches the upper limit.
400	QueueNotExist	The specified queue queueName does not exist.	The error message returned when the specified MNS queue does not exist.

HTTP status code	Error code	Error message	Description
400	TopicNotExist	The specified topic topicname does not exist.	The error message returned when the specified MNS topic does not exist.
400	UnsupportedNotificationType. CurrentRegion	The notificationType is not supported in the special region which scalingGroup belongs to.	The error message returned when the notification type is not supported in the region where the scaling group resides.
404	InvalidScalingGroupId. NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist.

14.2 DescribeNotificationConfigurations

You can call this operation to query notification configurations for scaling events and resource changes.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
Action	String	No	DescribeNotificationConfigurations	The operation that you want to perform. Set the value to DescribeNotificationConfigurations.

Response parameters

Parameter	Type	Example	Description
NotificationConfigurationModels			The list of notifications for scaling events and resource changes.
↳NotificationArn	String	acs:ess:cn-hangzhou:111111111:queue/queue1	The list of notifications for scaling events and resource changes.
↳NotificationTypes		AUTOSCALING:SCALE_IN_SUCCESS	The list of different types of notifications for scaling events and resource changes.
↳ScalingGroupId	String	asg-****	The ID of the scaling group.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DescribeNotificationConfigurations
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&<Common request parameters>
```

Successful response examples

XML format

```
< DescribeNotificationConfigurationsResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
RequestId >
  < NotificationConfigurationModels >
    < NotificationConfigurationModel >
      < NotificationArn > xxxxxxxxxx </ NotificationArn >
      < ScalingGroupIds > xxxxxxxxxx </ ScalingGroupIds >
      < NotificationTypes >
        < NotificationType > AUTOSCALING : SCALE_IN_SUCCESS </
NotificationType >
        < NotificationType > AUTOSCALING : SCALE_IN_ERROR </
NotificationType >
      </ NotificationTypes >
    </ NotificationConfigurationModel >
    < NotificationConfigurationModel >
      < NotificationArn > xxxxxxxxxx </ NotificationArn >
      < ScalingGroupIds > xxxxxxxxxx </ ScalingGroupIds >
      < NotificationTypes >
        < NotificationType > AUTOSCALING : SCALE_IN_SUCCESS </
NotificationType >
        < NotificationType > AUTOSCALING : SCALE_IN_ERROR </
NotificationType >
      </ NotificationTypes >
    </ NotificationConfigurationModel >
  </ NotificationConfigurationModels >
</ DescribeNotificationConfigurationsResponse >
```

JSON format

```
{
  "requestId": "473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E",
  "notificationConfigurationModels": [
    {
      "notificationArn": "xxxxxxxxxx",
      "notificationTypes": [
        "AUTOSCALING : SCALE_OUT_SUCCESS",
        "AUTOSCALING : SCALE_OUT_ERROR",
        "AUTOSCALING : SCALE_IN_SUCCESS",
        "AUTOSCALING : SCALE_IN_ERROR",
        "AUTOSCALING : SCALE_REJECT"
      ],
      "scalingGroupIds": "xxxxxxxxxx"
    },
    {
      "notificationArn": "xxxxxxxxxx",
      "notificationTypes": [
        "AUTOSCALING : SCALE_OUT_SUCCESS",
        "AUTOSCALING : SCALE_OUT_ERROR",
        "AUTOSCALING : SCALE_IN_SUCCESS",
        "AUTOSCALING : SCALE_IN_ERROR",
        "AUTOSCALING : SCALE_REJECT"
      ],
      "scalingGroupIds": "xxxxxxxxxx"
    }
  ]
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist.

14.3 ModifyNotificationConfiguration

You can call this operation to modify notification configurations for scaling events and resource changes.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
NotificationArn	String	Yes	acs:ess:cn-hangzhou:123456:cloudmonitor	The identifier of the notification object. The value must be in the <code>acs:ess:{region}:{account-id}:{resource-relative-id}</code> format. <code>region</code> is the ID of the region where the scaling group is located. For more information, see Regions and zones. <code>account-id</code> is your account ID. <code>resource-relative-id</code> is the notification method. Valid values: <code>cloudmonitor</code> or <code>queue/{queue-name}</code>, and <code>topic/{topic-name}</code>.
NotificationType.N	RepeatList	Yes	AUTOSCALING:SCALE_OUT_SUCCESS	One or more types of scaling event notifications. Valid values of N: 1 to 5. Multiple values must be listed in ascending order. You can call the DescribeNotificationTypes operation to query parameter values.
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcj****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
Action	String	No	ModifyNotificationConfiguration	The operation that you want to perform. Set the value to ModifyNotificationConfiguration.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=ModifyNotificationConfiguration
&ScalingGroupId=AG6CQdPU80&KdwljgZcJ****
&NotificationArn=acs:ess:cn-hangzhou:123456:cloudmonitor
&NotificationType.1=AUTOSCALING:SCALE_OUT_SUCCESS
&NotificationType.2=AUTOSCALING:SCALE_IN_SUCCESS
&NotificationType.3=AUTOSCALING:SCALE_OUT_ERROR
&NotificationType.4=AUTOSCALING:SCALE_IN_ERROR
&NotificationType.5=AUTOSCALING:SCALE_REJECT
<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	InvalidNotificationTypes	The specified notificationType is invalid.	The error message returned when the specified value of the NotificationType . N parameter is invalid.
400	NotificationConfigurationNotExist	The specified notificationConfiguration not exist for the scalingGroup.	The error message returned when the specified notification configuration does not exist in the scaling group.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist .

14.4 DeleteNotificationConfiguration

You can call this operation to delete notification configurations for scaling events and resource changes.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
NotificationArn	String	Yes	acs:ess:cn-hangzhou:123456:cloudmonitor	The identifier of the notification object. The value must be in the <code>acs:ess:{region}:{account-id}:{resource-relative-id}</code> format. <code>region</code> is the ID of the region where the scaling group is located. For more information, see Regions and zones. <code>account-id</code> is your account ID. <code>resource-relative-id</code> is the notification method. Valid values: <code>cloudmonit</code> or <code>, queue/{queue-name}</code>, and <code>topic/{topic-name}</code>.
ScalingGroupId	String	Yes	AG6CQdPU8OKdwLjgZcJ****	The ID of the scaling group.
Action	String	No	DeleteNotificationConfiguration	The operation that you want to perform. Set the value to <code>DeleteNotificationConfiguration</code> .

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DeleteNotificationConfiguration
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&NotificationArn=acs:ess:cn-hangzhou:123456:cloudmonitor
or
<Common request parameters>
```

Successful response examples

XML format

```
<DeleteNotificationConfigurationResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8-3E</
  RequestId>
</DeleteNotificationConfigurationResponse>
```

JSON format

```
{
  "requestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C8-3E"
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
400	NotificationConfigurationNotExist	The specified notification configuration not exist for the scalingGroup.	The error message returned when the specified notification configuration does not exist in the scaling group.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist.

14.5 DescribeNotificationTypes

You can call this operation to query the types of notifications for scaling events and resource changes.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
Action	String	No	DescribeNotificationTypes	The operation that you want to perform. Set the value to DescribeNotificationTypes.

Response parameters

Parameter	Type	Example	Description
NotificationTypes		AUTOSCALING:SCALE_OUT_SUCCESS	The list of different types of notifications for scaling events and resource changes.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=DescribeNotificationTypes
&<Common request parameters>
```

Successful response examples

XML format

```
< DescribeNotificationTypesResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
  RequestId >
    < NotificationTypes >
      < NotificationType > AUTOSCALING : SCALE_OUT_ SUCCESS </
      NotificationType >
      < NotificationType > AUTOSCALING : SCALE_IN_S UCESS </
      NotificationType >
      < NotificationType > AUTOSCALING : SCALE_OUT_ ERROR </
      NotificationType >
      < NotificationType > AUTOSCALING : SCALE_IN_E RROR </
      NotificationType >
      < NotificationType > AUTOSCALING : SCALE_REJECT CT </
      NotificationType >
    </ NotificationTypes >
  </ DescribeNotificationTypesResponse >
```

JSON format

```
{
  "requestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E ",
  "notificationTypes ":[
    " AUTOSCALING : SCALE_OUT_ SUCCESS ",
    " AUTOSCALING : SCALE_IN_S UCESS ",
    " AUTOSCALING : SCALE_OUT_ ERROR ",
    " AUTOSCALING : SCALE_IN_E RROR ",
    " AUTOSCALING : SCALE_REJECT CT "
  ]
}
```

Error codes

[View error codes](#)

15 Instance

15.1 EnterStandby

You can call this operation to put specified instances in a scaling group into the standby state.

Description

If a Server Load Balancer (SLB) instance is specified in the scaling group, the weight of an ECS instance attached to the SLB instance is set to zero.

- When an instance is in the standby state, you can manually remove it from the scaling group and delete it.
- The changes in the number of instances in a scaling group and scale-in activities triggered by monitoring tasks will not remove ECS instances in standby state.
- When an instance in standby state is unhealthy (stopped or restarted), its health check status is not updated and its scaling activities are not removed. The health check status is updated only after the instance is [taken out of the standby state](#).

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
InstanceId . N	RepeatList	Yes	i-28wt4****	The IDs of ECS instances.
ScalingGroup upId	String	Yes	AG6CQdPU8O KdwLjgZcj****	The ID of the scaling group.
Action	String	No	EnterStandby	The operation that you want to perform. Set the value to EnterStandby.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=EnterStandby&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****&InstanceId=i-28wt4****&<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

Error codes

[View error codes](#)

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 when the RAM user does not have permission to call this operation . Contact the owner of the Alibaba Cloud account to grant permission to the RAM user before trying again.
404	InvalidInstanceId. NotFound	Instance "XXX" does not exist.	The error message returned when the specified ECS instance does not exist.
404	InvalidScalingGroupId. NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist .

15.2 ExitStandby

You can call this operation to take specified instances in a scaling group out of the standby state.

Description

If a Server Load Balancer (SLB) instance is specified in the scaling group, the weight of an ECS instance attached to the SLB instance is set to the value defined in the scaling configuration.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
InstanceId . N	RepeatList	Yes	i-28wt4****	The IDs of ECS instances. The value can be a JSON array consisting of multiple instance IDs. You can specify up to 20 instance IDs. Separate multiple IDs by using commas (,).
ScalingGroup upId	String	Yes	AG6CQdPU80 KdwLjgZcJ****	The ID of the scaling group.
Action	String	No	ExitStandby	The operation that you want to perform. Set the value to ExitStandby.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F- -4DC5-B3DB- A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=ExitStandby
&ScalingGroupUpId=AG6CQdPU80KdwLjgZcJ****
&InstanceId.1=i-28wt4****
&<Common request parameters>
```

Successful response examples

XML format

```
< ExitStandbyResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
    RequestId >
</ ExitStandbyResponse >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
}
```

Error codes[View error codes](#)

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 when the RAM user does not have permission to call this operation . Contact the owner of the Alibaba Cloud account to grant permission to the RAM user before trying again.
404	InvalidInstanceId. NotFound	A required authorization for the specified action is not supplied.	The error message returned when the specified ECS instance does not exist.

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.

15.3 RebalanceInstances

You can call this operation to rebalance the distribution of ECS instances in a multi-zone scaling group.

Description

When rebalancing, Auto Scaling starts new instances before terminating the old ones. This ensures that the performance and availability of your application is not compromised by rebalancing instances.

- You can perform the rebalance operation for ECS instances in a multi-zone scaling group only when you set `MultiAZPolicy = Balance`.
- You can perform the rebalance operation only when the instances in the scaling group are significantly imbalanced.
- A single rebalance operation can redistribute up to 20 ECS instances.
- The system can temporarily exceed the maximum capacity of a scaling group specified by `MaxSize` by a 10% margin. The minimum number of ECS instances in a scaling group is 1. You can exceed the maximum capacity for as long as it takes for the rebalance operation to complete, typically one to six minutes.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	AG6CQdPU80KdwLjgZcJ****	The ID of the scaling group.
Action	String	No	RebalanceInstances	The operation that you want to perform. Set the value to RebalanceInstances.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful.
ScalingActivityId	String	asa-kjgffgdfadah****	The ID of the scaling activity.

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=RebalanceInstances
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&<Common request parameters>
```

Successful response examples

XML format

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

JSON format

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

}

Error codes

[View 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 when you have not enabled the specified scaling group.
400	OperationDenied	This operation is denied because the specified scaling group does not support this action.	The error message returned when the MultiAZPolicy parameter of the specified scaling group is not set to Balance, or when the distribution of ECS instances is not sufficiently imbalanced to perform this operation.
403	Forbidden. Unauthorized	A required authorization for the specified action is not supplied.	The error message returned when you are not authorized to use the RebalanceInstances operation.

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist in the current account.

15.4 SetInstancesProtection

You can call this operation to enable or disable protection for one or more ECS instances in a scaling group.

Description

After protection is enabled for an ECS instance:

- The instance keeps running in the protected state until you disable the protection.
- The changes in the number of instances in a scaling group and scale-in activities triggered by monitoring tasks will not remove protected ECS instances. You must manually [remove an ECS instance](#) to release the instance.
- When an ECS instance is stopped or restarted, its health check status is not updated.

Debugging

[OpenAPI Explorer](#) simplifies API usage. You can use OpenAPI Explorer to perform debugging operations, such as retrieve APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
InstanceId • N	RepeatList	Yes	i-28wt4****	The IDs of ECS instances. Valid values of N: 1 to 20.

Parameter	Type	Required	Example	Description
ProtectedFromScaleIn	Boolean	Yes	true	Indicates whether protection is enabled for an ECS instance . Active protection prevents the ECS instance from being stopped or removed from a scaling group during scale-in activities. Valid values: · true · false
ScalingGroupId	String	Yes	AG6CQdPU80KdwLjgZcJ****	The ID of the scaling group.
Action	String	No	SetInstancesProtection	The operation that you want to perform. Set the value to SetInstancesProtection.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request. This parameter is returned regardless of whether the operation is successful .

Examples

Sample requests

```
http://ess.aliyuncs.com/?Action=SetInstancesProtection
&ScalingGroupId=AG6CQdPU80KdwLjgZcJ****
&InstanceId.1=i-28wt4****
&InstanceId.2=i-28wt4****
&ProtectedFromScaleIn=true
&<Common request parameters>
```

Successful response examples

XML format

```
< SetInstanceProtectionResponse >
  < RequestId > 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E </
    RequestId >
</ SetInstanceProtectionResponse >
```

JSON format

```
{
  "RequestId ":" 473469C7 - AA6F - 4DC5 - B3DB - A3DC0DE3C8 3E "
```

Error codes[View 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 when you have not enabled the specified scaling group.
403	Forbidden. Unauthorized	A required authorization for the specified action is not supplied.	The error message returned when you are not authorized to use the SetInstanceProtection operation.
404	InvalidInstanceId. NotFound	Instance "XXX" does not exist.	The error message returned when the specified ECS instance does not exist.

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned when the specified scaling group does not exist.

16 Region

16.1 DescribeRegions

You can call this operation to query the regions that can use Auto Scaling.

Debugging

Alibaba Cloud provides [OpenAPI Explorer](#) to simplify API usage. You can use OpenAPI Explorer to search for APIs, call APIs, and dynamically generate SDK example code.

Request parameters

Parameter	Type	Required	Example	Description
AcceptLanguage	String	No	zh-CN	The language of the results returned. For more information, see RFC7231 . Valid values: - zh-CN: Chinese - en-US: English - ja: Japanese Default value: zh-CN.
Action	String	No	DescribeRegions	The operation that you want to perform. Set the value to DescribeRegions.

Response parameters

Parameter	Type	Example	Description
Regions			A collection of regions.
└LocalName	String	China (Beijing)	The name of a region.

Parameter	Type	Example	Description
<code>RegionEndpoint</code>	String	ess.aliyuncs.com	The endpoint of a region.
<code>RegionId</code>	String	cn-beijing	The ID of a region.
<code>RequestId</code>	String	73469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample request

```
https://ess.aliyuncs.com/?Action=DescribeRegions
&AcceptLanguage=zh-CN&<Common request parameters>
```

Sample success response

XML format

```
<DescribeRegionsResponse>
  <RequestId> 38EC7366 - F5A9 - 46B1 - BDB1 - 0FDC2E2963 97 </
RequestId>
  <Regions>
    <Region>
      <RegionId> cn - beijing </RegionId>
      <RegionEndpoint> ess . aliyuncs . com </RegionEndpoint>
      <LocalName> China ( Beijing )</LocalName>
    </Region>
    <Region>
      <RegionId> cn - shanghai </RegionId>
      <RegionEndpoint> ess . aliyuncs . com </RegionEndpoint>
      <LocalName> China ( Shanghai )</LocalName>
    </Region>
  </Regions>
</DescribeRegionsResponse>
```

JSON format

```
{
  "RequestId": " 38EC7366 - F5A9 - 46B1 - BDB1 - 0FDC2E2963 97 ",
  "Regions": {
    "Region": [
      {
        "RegionId": " cn - beijing ",
        "RegionEndpoint": " ess . aliyuncs . com ",
        "LocalName": " China ( Beijing )"
      },
      {
        "RegionId": " cn - shanghai ",
        "RegionEndpoint": " ess . aliyuncs . com ",
        "LocalName": " China ( Shanghai )"
      }
    ]
  }
}
```

```
}  
]  
}  
}
```

Error codes

[View error codes](#)

HTTP status code	Error code	Error message	Description
404	InvalidAcceptLanguage. NotFound	Only Chinese (zh-CN), English (en-US), and Japanese (ja) are allowed.	The error message returned because the specified language is not supported.

17 Error codes

17.1 Client errors

This topic describes the client errors.

Error message	Error code	Description	HTTP status code
The parameter is missing.	MissingParameter	The input parameter <code><parameter name></code> that is mandatory for processing this request is not supplied	400
The parameter value is invalid.	InvalidParameter	The specified value of parameter <code><parameter name></code> is not valid.	400
The interface is invalid.	UnsupportedOperation	The specified action is not supported.	400
The version is invalid.	NoSuchVersion	The specified version does not exist.	400
The operation is rejected by the traffic control system.	Throttling	Request was denied due to request throttling.	400
The AccessKey is invalid.	InvalidAccessKeyId.NotFound	The AccessKey ID provided does not exist in our records.	400
The operation is forbidden.	Forbidden	Users are not authorized to operate on the specified resource.	403

Error message	Error code	Description	HTTP status code
The operation is forbidden by the risk control system.	Forbidden. RiskControl	This operation is forbidden by Aliyun Risk Control system.	403
The signature is invalid.	SignatureDoesNotMatch	The signature we calculated does not match the one you provided.	403
Auto Scaling is not activated and thus the API cannot be called.	Forbidden. Unsubscribed	Do not have permission to access this API.	403
Real-name verification is missing.	Forbidden. UserVerification	Your user account is not verified by Aliyun.	403
Auto Scaling is not supported by the specified region.	ResourceNotAvailable	Resource you requested is not available in this region or zone.	400

17.2 Server errors

This topic introduces the server errors.

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

18 How to ensure idempotence

This topic introduces how to ensure idempotence.

If a request timeout or internal server error is encountered when the “Execute a Scaling Rule” interface is called to create or release an ECS instance, the client might try to resend the request. In this case, the client can prevent the server from creating more instances than expected by providing the optional `ClientToken` parameter. This parameter also ensures the idempotence of the request. `ClientToken` is a unique and case sensitive string which is generated by the client and cannot contain more than 64 ASCII characters.

If you use the same `ClientToken` value to call the `CreateInstance` interface, the server returns identical request results that contain the same `ScalingActivityId`. Therefore, when you encounter an error and try again, by providing the same `ClientToken` value, you can ensure that only one scaling activity is created and the corresponding `ScalingActivityId` is obtained.

If you provide a `ClientToken` that has already been used, but with different request parameters, the Auto Scaling service returns the `IdempotentParameterMismatch` error code. However, note that you must change the `SignatureNonce`, `TimeStamp`, and `Signature` parameters. Because the Auto Scaling service uses `SignatureNonce` to prevent replay attacks and `TimeStamp` to mark the time of each request, the second request must provide different `SignatureNonce` and `TimeStamp` parameter values. This also produces a different `Signature` value.

Generally, you only need to retry the client in the case of Error 503 (ServiceUnAvailable) or no response. When Error 200 is returned, a retry generates the same results as the last time, without affecting the server status. When Error 4xx or 500 is returned, a retry usually may fail unless the message clearly indicates try it later.