

ALIBABA CLOUD

Alibaba Cloud

Auto Scaling
API Reference

Document Version: 20220629

Legal disclaimer

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

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).
5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.
6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

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

Table of Contents

1.API overview -----	08
2.API usage instructions -----	11
3.API quick start -----	12
3.1. Process introduction -----	12
3.2. Create a scaling group -----	12
3.3. Create a scaling configuration -----	13
3.4. Enable a scaling group -----	13
3.5. Create a scaling rule -----	14
3.6. Create a scheduled task -----	14
4.Calling method -----	16
5.Request structure -----	17
6.Common parameters -----	19
7.Signatures -----	21
8.Responses -----	23
9.Scaling group -----	26
9.1. CreateScalingGroup -----	26
9.2. EnableScalingGroup -----	48
9.3. DisableScalingGroup -----	58
9.4. SetGroupDeletionProtection -----	60
9.5. DeleteScalingGroup -----	62
9.6. DescribeScalingGroups -----	65
9.7. DescribeScalingInstances -----	80
9.8. DescribeScalingActivities -----	91
9.9. DescribeScalingActivityDetail -----	96
9.10. AttachLoadBalancers -----	98
9.11. DetachLoadBalancers -----	101

9.12. DetachDBInstances -----	105
9.13. SuspendProcesses -----	107
9.14. ResumeProcesses -----	109
9.15. AttachDBInstances -----	111
9.16. AttachVServerGroups -----	115
9.17. DetachVServerGroups -----	119
9.18. AttachAlbServerGroups -----	122
9.19. DetachAlbServerGroups -----	126
10. Scaling configurations -----	130
10.1. CreateScalingConfiguration -----	130
10.2. DescribeScalingConfigurations -----	160
10.3. ModifyScalingConfiguration -----	175
10.4. DeleteScalingConfiguration -----	199
10.5. CreateEciScalingConfiguration -----	201
10.6. DescribeEciScalingConfigurations -----	222
10.7. ModifyEciScalingConfiguration -----	245
10.8. DeleteEciScalingConfiguration -----	265
11. Scaling rule -----	268
11.1. CreateScalingRule -----	268
11.2. DescribeScalingRules -----	279
11.3. ModifyScalingRule -----	290
11.4. DeleteScalingRule -----	301
12. Trigger task -----	303
12.1. ExecuteScalingRule -----	303
12.2. ScaleWithAdjustment -----	309
12.3. AttachInstances -----	317
12.4. RemoveInstances -----	324
12.5. DetachInstances -----	329

13.Scheduled tasks -----	335
13.1. CreateScheduledTask -----	335
13.2. ModifyScheduledTask -----	341
13.3. DescribeScheduledTasks -----	347
13.4. DeleteScheduledTask -----	352
14.Alarm tasks -----	355
14.1. CreateAlarm -----	355
14.2. DescribeAlarms -----	365
14.3. ModifyAlarm -----	375
14.4. EnableAlarm -----	384
14.5. DisableAlarm -----	386
14.6. DeleteAlarm -----	387
15.Lifecycle Hook -----	389
15.1. CreateLifecycleHook -----	389
15.2. ModifyLifecycleHook -----	394
15.3. DescribeLifecycleHooks -----	399
15.4. RecordLifecycleActionHeartbeat -----	404
15.5. DescribeLifecycleActions -----	406
15.6. CompleteLifecycleAction -----	411
15.7. DeleteLifecycleHook -----	414
16.Event notification -----	417
16.1. CreateNotificationConfiguration -----	417
16.2. ModifyNotificationConfiguration -----	421
16.3. DescribeNotificationConfigurations -----	423
16.4. DeleteNotificationConfiguration -----	426
16.5. DescribeNotificationTypes -----	428
17.Instance -----	431
17.1. SetInstanceHealth -----	431

17.2. EnterStandby -----	432
17.3. ExitStandby -----	435
17.4. RebalanceInstances -----	438
17.5. SetInstancesProtection -----	440
18.Region -----	444
18.1. DescribeRegions -----	444
19.Tag management -----	448
19.1. TagResources -----	448
19.2. ListTagResources -----	449
19.3. UntagResources -----	453
19.4. ListTagKeys -----	455
19.5. ListTagValues -----	457
20.Error codes -----	460
20.1. Client errors -----	460
20.2. Server errors -----	461
21.How to ensure idempotence -----	462
22.Call API operations over the internal network -----	463

1. API overview

The following tables list API operations available for use in Auto Scaling.

Scaling groups

API	Description
CreateScalingGroup	Creates a scaling group.
ModifyScalingGroup	Modifies a scaling group.
EnableScalingGroup	Enables a scaling group.
DisableScalingGroup	Disables a scaling group.
DeleteScalingGroup	Deletes a scaling group.
DescribeScalingGroups	Queries scaling groups.
DescribeScalingInstances	Queries the list of ECS instances in a scaling group and lists details about the instances.
DescribeScalingActivities	Queries scaling activities.
AttachLoadBalancers	Associates one or more SLB instances with a scaling group.
DetachLoadBalancers	Disassociates one or more SLB instances from a scaling group.
AttachDBInstances	Associates one or more ApsaraDB for RDS instances with a scaling group.
DetachDBInstances	Disassociates one or more ApsaraDB for RDS instances from a scaling group.
AttachVServerGroups	Adds one or more VServer groups under an SLB instance to a scaling group.
DetachVServerGroups	Removes one or more VServer groups.
SuspendProcesses	Suspends the specified processes in a scaling group.
ResumeProcesses	Resumes the suspended processes in a scaling group.

Scaling configurations

API	Description
CreateScalingConfiguration	Creates a scaling configuration.
DescribeScalingConfigurations	Queries scaling configurations.
DeleteScalingConfiguration	Deletes a scaling configuration.

API	Description
ModifyScalingConfiguration	Modifies a scaling configuration.

Scaling rules

API	Description
CreateScalingRule	Creates a scaling rule.
ModifyScalingRule	Modifies a scaling rule.
DescribeScalingRules	Queries all scaling rules in a scaling group and lists information about the scaling rules.
DeleteScalingRule	Deletes a scaling rule.

Scaling tasks

API	Description
ExecuteScalingRule	Executes a scaling rule.
AttachInstances	Adds ECS instances to a scaling group.
RemoveInstances	Deletes ECS instances in a scaling group.
DetachInstances	Removes one or more ECS instances from a scaling group.

Scheduled tasks

API	Description
CreateScheduledTask	Creates a scheduled task.
ModifyScheduledTask	Modifies a scheduled task.
DescribeScheduledTasks	Queries scheduled tasks.
DeleteScheduledTask	Deletes a scheduled task.

Lifecycle hooks

API	Description
CreateLifecycleHook	Creates one or more lifecycle hooks in a scaling group.
ModifyLifecycleHook	Modifies a lifecycle hook.
DescribeLifecycleHooks	Queries lifecycle hooks.

API	Description
RecordLifecycleActionHeartbeat	Extends the timeout period of an ECS instance and keeps the instance in the wait state.
CompleteLifecycleAction	Takes a scaling activity out of the wait state in advance.
DeleteLifecycleHook	Deletes a lifecycle hook.

Event notifications

API	Description
CreateNotificationConfiguration	Creates an event notification.
DeleteNotificationConfiguration	Deletes an event notification.
DescribeNotificationConfigurations	Queries event notifications.
DescribeNotificationTypes	Queries the types of event notifications.
ModifyNotificationConfiguration	Modifies an event notification.

Instances

API	Description
EnterStandby	Puts ECS instances in a scaling group into the standby state.
ExitStandby	Changes the state of ECS instances in a scaling group from standby to running.
RebalanceInstances	Rebalances the distribution of ECS instances in a scaling group across multiple zones.
SetInstancesProtection	Enables or disables protection for one or more ECS instances in a scaling group.
SetInstanceHealth	Sets the health status of ECS instances in a scaling group.

Regions

API	Description
DescribeRegions	Queries the regions where Auto Scaling is available.

2. API usage instructions

Before you call Auto Scaling API, you must activate Auto Scaling on the official Alibaba Cloud website and grant Auto Scaling the API permissions in the console.

Prerequisites

Auto Scaling is activated. The following permissions are granted to Auto Scaling and RAM users:

- The permissions on related cloud resources granted to Auto Scaling. For more information, see [Manage the service-linked role for Auto Scaling](#).
- The permissions on related cloud resources granted to RAM users, including `AliyunESSFullAccess` and `AliyunECSFullAccess`. For more information, see [Grant permissions to a RAM user](#).

Errors

Before you call Auto Scaling API, you must activate Auto Scaling on the official Alibaba Cloud website and grant Auto Scaling the API permissions to scale ECS instances. The following table describes the errors that may occur if the preceding requirements are not met.

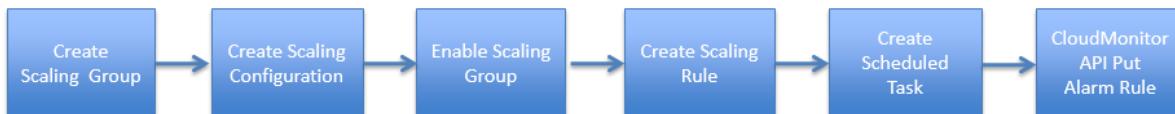
HTTP status code	Error code	Error message	Description
403	Forbidden.Unsubscribed	Do not have permission to access this API.	The error message returned because Auto Scaling is not activated and not granted the permissions to call the specified API operation.
403	Forbidden.Unauthorized	A required authorization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

3.API quick start

3.1. 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%.

3.2. Create a scaling group

This topic provides an example of how to call an API operation to create a scaling group.

Description

In this example, the MinSize, MaxSize, LoadBalancerId, and DBInstanceId parameters are specified to create a scaling group. For more information, see [CreateScalingGroup](#).

Sample requests

```
http://ess.aliyuncs.com/?Action=CreateScalingGroup
&RegionId=cn-qingdao
&MaxSize=20
&MinSize=2
&LoadBalancerId=147b46d767c-cn-qingdao-cm****
&DBInstanceId.1=rdszzzyuny****
&DBInstanceId.2=rdsia3u3yia****
&<Common request parameters>
```

Sample responses

```
<CreateScalingGroupResponse>
  <ScalingGroupId>dP8VqSd9ENXPc0ciVmbe****</ScalingGroupId>
  <RequestId>536E9CAD-DB30-4647-AC87-AA5CC38C5382</RequestId>
</CreateScalingGroupResponse>
```

3.3. Create a scaling configuration

This topic provides an example of how to call an API operation to create a scaling configuration.

Description

In this example, the `InstanceType`, `SecurityGroupId`, and `ImageId` parameters are specified to create a scaling configuration for a scaling group. For more information, see [CreateScalingConfiguration](#).

Sample requests

```
http://ess.aliyuncs.com/?Action=CreateScalingConfiguration
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbe****
&SecurityGroupId=sg-280ih****
&ImageId=centos6u5_64_20G_aliaegeis_20140703.vhd
&InstanceType=ecs.t1.xsmall
&<Common request parameters>
```

Sample responses

```
<CreateScalingConfigurationResponse>
  <ScalingConfigurationId>eOs27Kb0oXvQcUYjEGel****</ScalingConfigurationId>
  <RequestId>5CC0AD41-08ED-4559-A683-6F56355FE068</RequestId>
</CreateScalingConfigurationResponse>
```

3.4. Enable a scaling group

This topic provides an example of how to call an API operation to enable a scaling group.

Description

In this example, a scaling group is enabled based on a specified scaling configuration, and an existing ECS instance is manually added to the scaling group. For more information about the parameters, see [EnableScalingGroup](#).

Sample requests

```
http://ess.aliyuncs.com/?Action=EnableScalingGroup
&ScalingGroupId=dP8VqSd9ENXPc0ciVmbe****
&ActiveScalingConfigurationId=eOs27Kb0oXvQcUYjEGel****
&InstanceId.1=i-283vv****
&<Common request parameters>
```

Sample responses

```
<EnableScalingGroupResponse>
<RequestId>6469DCD0-13AC-487E-85A0-CE4922908FDE</RequestId>
</EnableScalingGroupResponse>
```

3.5. Create a scaling rule

This topic provides an example of how to call an API operation to create a scaling rule.

Description

In this example, the AdjustmentType and AdjustmentValue parameters are specified to create a scaling rule for a scaling group. For more information about the parameters, see [CreateScalingRule](#).

Sample requests

```
http://ess.aliyuncs.com/?Action=CreateScalingRule
&ScalingGroupId=dP8VqSd9ENXPc0ciVmhc****
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=1
&<Common request parameters>
```

Sample responses

```
<CreateScalingRuleResponse>
<ScalingRuleAri>ari:acs:ess:cn-qingdao:134437****:scalingrule/eMKWG8SRNb9dBLAjweN****</
ScalingRuleAri>
<ScalingRuleId>eMKWG8SRNb9dBLAjweN****</ScalingRuleId>
<RequestId>570C84F4-A434-488A-AFA1-1E3213682B33</RequestId>
</CreateScalingRuleResponse>
```

3.6. Create a scheduled task

This topic provides an example of how to call an API operation to create a scheduled task.

Description

In this example, the LaunchTime, RecurrenceType, RecurrenceValue, RecurrenceEndTime, and ScheduledAction parameters are specified to create a scheduled task. For more information about the parameters, see [CreateScheduledTask](#).

Sample requests

```
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:134437****:scalingrule/eMKWG8SRNb9dBLAjweN****
&<Common request parameters>
```

Sample responses

```
<CreateScheduledTaskResponse>
  <ScheduledTaskId>edRtShc57WGXd8TlPbr****</ScheduledTaskId>
  <RequestId>0F02D931-2B12-44D7-A0E9-39925C13D15E</RequestId>
</CreateScheduledTaskResponse>
```

4.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.

5. Request structure

This topic describes the request structure used when you call API operations of Auto Scaling.

Endpoints

To reduce network latency, we recommend that you configure endpoints based on your business location. The following table describes the endpoints of Auto Scaling.

Region	Public endpoint	VPC endpoint
Default	ess.aliyuncs.com	None
China (Qingdao)	ess.aliyuncs.com	ess-vpc.cn-qingdao.aliyuncs.com
China (Beijing)	ess.aliyuncs.com	ess-vpc.cn-beijing.aliyuncs.com
China (Zhangjiakou)	ess.cn-zhangjiakou.aliyuncs.com	ess-vpc.cn-zhangjiakou.aliyuncs.com
China (Hohhot)	ess.cn-huhehaote.aliyuncs.com	ess-vpc.cn-huhehaote.aliyuncs.com
China (Ulanqab)	ess.cn-wulanchabu.aliyuncs.com	ess-vpc.cn-wulanchabu.aliyuncs.com
China (Hangzhou)	ess.aliyuncs.com	ess-vpc.cn-hangzhou.aliyuncs.com
China (Shanghai)	ess.aliyuncs.com	ess-vpc.cn-shanghai.aliyuncs.com
China (Shenzhen)	ess.aliyuncs.com	ess-vpc.cn-shenzhen.aliyuncs.com
China (Heyuan)	ess.cn-heyuan.aliyuncs.com	ess-vpc.cn-heyuan.aliyuncs.com
China (Guangzhou)	ess.cn-guangzhou.aliyuncs.com	ess-vpc.cn-guangzhou.aliyuncs.com
China (Chengdu)	ess.cn-chengdu.aliyuncs.com	ess-vpc.cn-chengdu.aliyuncs.com
China (Hong Kong)	ess.aliyuncs.com	ess-vpc.cn-hongkong.aliyuncs.com
Singapore (Singapore)	ess.aliyuncs.com	ess-vpc.ap-southeast-1.aliyuncs.com
Australia (Sydney)	ess.ap-southeast-2.aliyuncs.com	ess-vpc.ap-southeast-2.aliyuncs.com
Malaysia (Kuala Lumpur)	ess.ap-southeast-3.aliyuncs.com	ess-vpc.ap-southeast-3.aliyuncs.com

Region	Public endpoint	VPC endpoint
Indonesia (Jakarta)	ess.ap-southeast-5.aliyuncs.com	ess-vpc.ap-southeast-5.aliyuncs.com
Japan (Tokyo)	ess.ap-northeast-1.aliyuncs.com	ess-vpc.ap-northeast-1.aliyuncs.com
UK (London)	ess.eu-west-1.aliyuncs.com	ess-vpc.eu-west-1.aliyuncs.com
US (Silicon Valley)	ess.aliyuncs.com	ess-vpc.us-west-1.aliyuncs.com
US (Virginia)	ess.aliyuncs.com	ess-vpc.us-east-1.aliyuncs.com
Germany (Frankfurt)	ess.eu-central-1.aliyuncs.com	ess-vpc.eu-central-1.aliyuncs.com
UAE (Dubai)	ess.me-east-1.aliyuncs.com	ess-vpc.me-east-1.aliyuncs.com
India (Mumbai)	ess.ap-south-1.aliyuncs.com	ess-vpc.ap-south-1.aliyuncs.com

Protocols

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

Methods

You can send requests by using the HTTP GET method. In an HTTP GET request, the request parameters must be included in the request URL.

Parameters

In each request, you must specify the operation such as CreateScalingGroup that you want to perform by using the Action parameter. You must also specify the common request parameters and operation-specific parameters.

Encoding

All requests and responses are encoded in UTF-8.

6. Common parameters

This topic describes common parameters.

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

Common request parameters

The following table describes the common request parameters that are included in a URL for a GET request.

Parameter	Type	Required	Description
Action	String	Yes	The operation that you want to perform.
AccessKeyId	String	Yes	The AccessKey ID provided to you by Alibaba Cloud. The AccessKey pair is used to call API operations for authentication, similar to how the user password is used to log on to the Auto Scaling console . For more information, see Obtain an AccessKey pair .
Signature	String	Yes	The signature string of the current request. For more information about how signatures are calculated, see Signatures .
SignatureMethod	String	Yes	The encryption method of the signature string. Set the value to HMAC-SHA1.
SignatureVersion	String	Yes	The version of the signature encryption algorithm. Set the value to 1.0.
SignatureNonce	String	Yes	A unique, random number used to prevent replay attacks. You must use different numbers for different requests.

Parameter	Type	Required	Description
Timestamp	String	Yes	<p>The timestamp of the request. Specify the time in the ISO 8601 standard in the YYYY-MM-DDThh:mm:ssZ format. The time must be in UTC.</p> <p>For example, 20:00:00 on January 1, 2018 (UTC+8) is written as 2018-01-01T12:00:00Z .</p>
Version	String	Yes	<p>The version number of the API. The value must be in the YYYY-MM-DD format. Set the value to 2014-08-28.</p>
Format	String	No	<p>The format in which to return the response. Valid values: JSON and XML. Default value: JSON.</p>

Examples

```
https://ess.aliyuncs.com/?Action=XXXXXX
&Format=xml
&Version=2014-08-28
&Signature=Pc5WB8gokVn0xfeu%2FZV%2BiNM1dg****
&SignatureMethod=HMAC-SHA1
&SignatureNonce=1521552885****
&SignatureVersion=1.0
&AccessKeyId=key-test
&Timestamp=2018-01-01T12:00:00Z
...
...
```

Common response parameters

Parameter	Type	Description
RequestId	String	The ID of the request. This parameter is returned regardless of whether the call is successful.

7.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.
 - i. 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 "&").
 - ii. 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.
 - iii. Connect the encoded parameter names and values with the equal sign "=".
 - iv. 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%25A10%25A07Z&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&AccessKeyId=testid&Action=DescribeScalingGroups&SignatureMethod=HMAC-SHA1&RegionId=cn-qingdao&SignatureNonce=1324fd0e-e2bb-4bb1-917c-bd6e437f1710&SignatureVersion=1.0&Version=2014-08-28&Signature=SmhZuLUnXmqxSEZ%2FGqyiwGqmf%2BM%3D
```

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

8. Responses

This topic describes the responses returned after API operations are called.

Responses can be returned in either the JSON or XML format. The default response format is JSON. You can specify the `Format` common parameter in the request to change the format of responses. For more information, see [Common parameters](#).

 **Note** Sample responses provided in API Reference are formatted with line breaks and indentations for better readability and clarity. However, the actual responses are not formatted with line breaks or indentations.

Sample success responses

Every response returns a unique Request ID regardless of whether the call is successful. API responses use the HTTP response format where a 2xx status code indicates a successful call.

XML format

```
<?xml version="1.0" encoding="UTF-8"?> <!--The root node of the response-->
<ActionResponse> <!--Returned request ID-->
    <RequestId>4C467B38-3910-447D-87BC-AC049166F223</RequestId> <!--Response data-->
</ActionResponse>
```

JSON format

```
{
    "RequestId": "4C467B38-3910-447D-87BC-AC049166F223" /* Response data */
}
```

Sample error responses

If an error occurs when you call an operation, an error response that consists of the error code, error message, and request ID is returned. API responses use the HTTP response format where a 4xx or 5xx status code indicates a failed call.

You can troubleshoot an error based on the error code returned and [common error codes](#). If the error persists, you can [submit a ticket](#) and provide the `HostId` and `RequestId` values to obtain technical support from Alibaba Cloud.

XML format

```
<?xml version="1.0" encoding="UTF-8"?> <!--The root node of the response-->
<Error>
    <RequestId>540CFF28-407A-40B5-B6A5-74Bxxxxxxxxx</RequestId> <!--Request ID-->
    <HostId>ess.aliyuncs.com</HostId> <!--Service endpoint-->
    <Code>InternalError</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-b6a5-74bxxxxxxxxx", /* Request ID */
  "HostId": "ess.aliyuncs.com", /* Service endpoint */
  "Code": "InternalError", /* Error code */
  "Message": "The request processing has failed due to some unknown error, exception or failure." /* Error message */
}
```

Common error codes

HTTP status code	Error code	Error message	Description
400	InvalidAccessKeyId.NotFound	The Access Key ID provided does not exist in our records.	The error message returned because the specified AccessKeyId parameter does not exist.
400	InvalidParameter	The specified value of parameter <parameter name> is not valid.	The error message returned because the specified value of the parameter is invalid.
400	MissingParameter	The input parameter <parameter name> that is mandatory for processing this request is not supplied.	The error message returned because the required parameter is not specified.
400	NoSuchVersion	The specified version does not exist.	The error message returned because the specified version does not exist.
400	ResourceNotAvailable	Resource you requested is not available in this region or zone.	The error message returned because the ESS service is not available in the specified region.
400	Throttling	Request was denied due to request throttling.	The error message returned because the system is under throttling condition. Try again later.
400	UnsupportedOperation	The specified action is not supported.	The error message returned because the specified operation is not supported.

HTTP status code	Error code	Error message	Description
403	Forbidden	Users are not authorized to operate on the specified resource.	The error message returned because you are not authorized to perform the specified operation.
403	Forbidden.RiskControl	This operation is forbidden by Aliyun Risk Control system.	The error message returned because the specified operation is not allowed.
403	Forbidden.Unsubscribed	Do not have permission to access this API.	The error message returned because you have not activated the ESS service to call the specified API operation.
403	Forbidden.UserVerification	Your user account is not verified by Aliyun.	The error message returned because you have not completed real-name verification.
403	SignatureDoesNotMatch	The signature we calculated does not match the one you provided.	The error message returned because the calculated signature is different from the one provided by you.
500	InternalError	The request processing has failed due to some unknown error, exception or failure.	The error message returned because a system error occurred.
503	ServiceUnavailable	The request has failed due to a temporary failure of the server.	The error message returned because the server cannot respond to your request. Try again later.

9. Scaling group

9.1. CreateScalingGroup

Creates a scaling group.

Description

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

The number of scaling groups that can be created in a region depends on your usage of Auto Scaling. You can go to the [Quota Center](#) to check the quota of the **Total Scaling Groups**.

A scaling group does not immediately take effect after it is created. You must call the [EnableScalingGroup](#) operation to enable the scaling group. After the scaling group is enabled, Auto Scaling can execute scaling rules to trigger scaling activities in the scaling group.

The scaling group, Classic Load Balancer (CLB) instances to be associated, and ApsaraDB RDS instances to be associated must be in the same region. CLB instances are formerly known as Server Load Balancer (SLB) instances. For more information, see [Regions and zones](#).

If you associate a CLB instance when you create a scaling group, Auto Scaling automatically adds ECS instances in the scaling group to the backend server group of the associated CLB instance. You can specify a server group to which ECS instances can be added. You can add ECS instances to the following types of server groups:

- Default server group: a group of ECS instances that are used to receive requests. If you do not specify a vServer group or a primary/secondary server group for a listener, requests are forwarded to the ECS instances in the default server group.
- vServer group: If you want to forward requests to different backend servers or configure domain name- or URL-based routing methods, you can use vServer groups.

 **Note** If you specify the default server group and multiple vServer groups at the same time, ECS instances are added to all the specified server groups.

The default weight of an ECS instance added to the backend server group of a CLB instance is 50. The CLB instance must meet the following requirements:

- The CLB instance must be in the Active state. You can call the [DescribeLoadBalancers](#) operation to query the status of the specified CLB instance.
- Health check must be enabled on all listener ports configured for the CLB instance. Otherwise, the scaling group fails to be created.

If you associate an Application Load Balancer (ALB) server group with a scaling group, Auto Scaling automatically adds ECS instances that are added to the scaling group to the ALB server group as backend servers to process access requests distributed by the ALB instance to which the ALB server group belongs. You can specify multiple ALB server groups, but the server groups must belong to the same virtual private cloud (VPC) as the scaling group. For more information, see [AttachAlbServerGroups](#).

If you associate an ApsaraDB RDS instance when you create a scaling group, Auto Scaling automatically adds the internal IP addresses of the ECS instances that are added to the scaling group to the whitelist of the associated ApsaraDB RDS instance. The ApsaraDB RDS instance must meet the following requirements:

- The ApsaraDB RDS instance must be in the Running state. You can call the [DescribeDBInstances](#) operation to query the status of the specified ApsaraDB RDS instance.
- The number of IP addresses in the whitelist of the ApsaraDB RDS instance cannot exceed the upper limit. For more information, see [Configure a whitelist](#) in the ApsaraDB RDS documentation.

If the MultiAZPolicy parameter of a scaling group is set to COST_OPTIMIZED, the following rules apply:

- You can use the OnDemandBaseCapacity, OnDemandPercentageAboveBaseCapacity, and SpotInstancePools parameters to specify the percentages of pay-as-you-go instances and preemptible instances based on the cost optimization policy. This instance allocation method is prioritized during scaling.
- When OnDemandBaseCapacity, OnDemandPercentageAboveBaseCapacity, or SpotInstancePools is not specified, the instance types available at the lowest cost are used to create instances based on the cost optimization policy.

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
Action	String	Yes	CreateScalingGroup	The operation that you want to perform. Set the value to CreateScalingGroup .
ScalingGroupName	String	No	scalinggroup****	<p>The name of the scaling group. The name of each scaling group must be unique in a region. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.</p> <p>The default value is the value of ScalingGroupId.</p>
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the launch template that is used by Auto Scaling to create instances.
LaunchTemplateVersion	String	No	Default	<p>The version number of the launch template. Valid values:</p> <ul style="list-style-type: none"> A fixed template version number. Default: the default template version. Latest: the latest template version.

Parameter	Type	Required	Example	Description
InstanceId	String	No	i-28wt4***	The ID of the ECS instance from which Auto Scaling obtains configuration information and uses this information to create a scaling configuration.
RegionId	String	Yes	cn-qingdao	The ID of the region where the scaling group resides. For more information, see Regions and zones .
MinSize	Integer	Yes	2	<p>The minimum number of ECS instances in the scaling group. When the number of existing ECS instances in the scaling group is less than the MinSize value, Auto Scaling automatically creates ECS instances until the number of instances is equal to the MinSize value.</p> <div style="background-color: #e1f5fe; padding: 5px; border-radius: 5px;"> ? Note The value of MinSize must be less than or equal to the value of MaxSize. </div>
MaxSize	Integer	Yes	20	<p>The maximum number of ECS instances in the scaling group. When the number of existing ECS instances in the scaling group is greater than the MaxSize value, Auto Scaling automatically removes ECS instances until the number of instances is equal to the MaxSize value.</p> <p>The value range of MaxSize depends on your usage of Auto Scaling. You can go to the Quota Center to check the quota of Instances That Can Be Configured for a Scaling Group. For example, if the quota of Instances That Can Be Configured for a Scaling Group is 2000, the value range of MaxSize is 0 to 2000.</p>

Parameter	Type	Required	Example	Description
DefaultCooldown	Integer	No	300	<p>The cooldown time after a scale-in or scale-out activity is executed. Valid values: 0 to 86400. Unit: seconds.</p> <p>During the cooldown time, Auto Scaling executes only scaling activities that are triggered by CloudMonitor event-triggered tasks.</p> <p>Default value: 300.</p>
LoadBalancerIds	String	No	<pre>["lb- bp1u7etiogg38yv wz****", "lb- bp168cqrux9ai9l7 f****", "lb- bp1jv3m9zvj22uf xp****"]</pre>	<p>The IDs of CLB instances. CLB instances are formerly known as SLB instances. The value can be a JSON array that contains multiple CLB instance IDs. Separate multiple IDs with commas (,).</p> <p>The number of CLB instances that can be associated with a scaling group depends on your usage of Auto Scaling. You can go to the Quota Center to check the quota of SLB Instances That Can Be Associated with a Scaling Group.</p>
DBInstancelds	String	No	<pre>["rm- bp142f86de0t7**", **, "rm- bp18l1z42ar4o***", **, "rm- bp1lqr97h4aqk***"]</pre>	<p>The IDs of ApsaraDB RDS instances. The value can be a JSON array that contains multiple ApsaraDB RDS instance IDs. Separate multiple IDs with commas (,).</p> <p>The number of ApsaraDB RDS instances that can be associated with a scaling group depends on your usage of Auto Scaling. You can go to the Quota Center to check the quota of ApsaraDB RDS Instances That Can Be Associated with a Scaling Group.</p>

Parameter	Type	Required	Example	Description
RemovalPolicy.1	String	No	OldestScalingConfiguration	<p>Specifies the first step for the instance removing policy. Valid values:</p> <ul style="list-style-type: none">• OldestInstance: removes ECS instances that are added to the scaling group at the earliest point in time.• NewestInstance: removes ECS instances that are most recently added to the scaling group.• OldestScalingConfiguration: removes ECS instances that are created based on the earliest scaling configuration. <p>Note The scaling configuration mentioned in OldestScalingConfiguration refers to the instance configuration source of a scaling group, which can be a scaling configuration or a launch template.</p> <p>The version of a launch template does not indicate the sequence in which the template is added. For example, you use the lt-foress V2 template to create a scaling group, and then you replace the template with the lt-foress V1 template when you modify the scaling group. In this case, the scaling group considers the lt-foress V2 launch template as the earliest template.</p> <p>The default value of RemovalPolicy.1 is used only when both RemovalPolicy.1 and RemovalPolicy.2 are not specified. Default value: OldestScalingConfiguration.</p> <p>Note The removal of ECS instances from a scaling group is also affected by the scaling policy (MultiAZPolicy) specified for the scaling group. For more information, see Configure a combination policy for removing instances.</p>

Parameter	Type	Required	Example	Description
RemovalPolicy.2	String	No	OldestInstance	<p>Specifies the second step for the instance removing policy. This parameter value cannot be the same as the RemovalPolicy.1 value. Valid values:</p> <ul style="list-style-type: none"> • OldestInstance: removes ECS instances that are added to the scaling group at the earliest point in time. • NewestInstance: removes ECS instances that are most recently added to the scaling group. • OldestScalingConfiguration: removes ECS instances that are created based on the earliest scaling configuration. <p>The default value of RemovalPolicy.2 is used only when both RemovalPolicy.1 and RemovalPolicy.2 are not specified. Default value: OldestInstance.</p> <div style="background-color: #e0f2fd; padding: 10px;"> <p>? Note The removal of ECS instances from a scaling group is also affected by the scaling policy (MultiAZPolicy) specified for the scaling group. For more information, see Configure a combination policy for removing instances.</p> </div>
VSwitchId	String	No	VSW-bp14zolna43z266bq***	<p>The ID of the vSwitch. If VSwitchId is specified, the network type of the scaling group is VPC.</p> <div style="background-color: #e0f2fd; padding: 10px;"> <p>? Note If VSwitchId or VSwitchIds.N is not specified, the network type of the scaling group is classic network.</p> </div>

Parameter	Type	Required	Example	Description
MultiAZPolicy	String	No	PRIORITY	<p>The ECS instance scaling policy for a multi-zone scaling group. Valid values:</p> <ul style="list-style-type: none"> • PRIORITY: ECS instances are scaled based on the vSwitchIds.N parameter. Auto Scaling preferentially scales instances in the zone where the vSwitch that has the highest priority resides. If the scaling fails, Auto Scaling attempts to scale instances in the zone where the vSwitch that has the next highest priority resides. • COST_OPTIMIZED: During a scale-out activity, Auto Scaling attempts to create ECS instances that have vCPUs provided at the lowest price. During a scale-in activity, Auto Scaling attempts to remove ECS instances that have vCPUs provided at the highest price. Preemptible instances are preferentially created when preemptible instance types are specified in the active scaling configuration. You can configure the CompensateWithOnDemand parameter to specify whether to automatically create pay-as-you-go instances when preemptible instances cannot be created due to insufficient resources. <p>Note COST_OPTIMIZED is valid when multiple instance types are specified or at least one preemptible instance type is specified.</p> <ul style="list-style-type: none"> • BALANCE: ECS instances are evenly distributed across zones that are specified in the scaling group. If ECS instances are unevenly distributed across zones due to insufficient resources, you can call the RebalanceInstance operation to balance the instances across zones. <p>Default value: PRIORITY.</p>

Parameter	Type	Required	Example	Description
HealthCheckType	String	No	ECS	<p>Specifies whether to perform health checks on instances in the scaling group. Valid values:</p> <ul style="list-style-type: none"> • NONE: Auto Scaling does not perform health checks on instances in the scaling group. • ECS: Auto Scaling performs health checks on ECS instances in the scaling group. <p>Default value: ECS.</p>
ScalingPolicy	String	No	recycle	<p>The reclaim mode of the scaling group. Valid values:</p> <ul style="list-style-type: none"> • recycle • release <p>ScalingPolicy specifies the reclaim modes of scaling groups, but the policy that is used to remove ECS instances from scaling groups is determined by the RemovePolicy parameter of the RemoveInstances operation. For more information, see RemoveInstances.</p>
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see Ensure idempotence.</p>
OnDemandBaseCapacity	Integer	No	30	<p>The minimum number of pay-as-you-go instances required in the scaling group. Valid values: 0 to 1000. If the number of pay-as-you-go instances is less than the value of this parameter, Auto Scaling preferentially creates pay-as-you-go instances.</p>

Parameter	Type	Required	Example	Description
OnDemandPercentageAboveBaseCapacity	Integer	No	20	The percentage of pay-as-you-go instances that can be created when instances are added to the scaling group. This parameter takes effect when the number of pay-as-you-go instances reaches the value for the OnDemandBaseCapacity parameter. Valid values: 0 to 100.
SpotInstanceRecovery	Boolean	No	true	Specifies whether to supplement preemptible instances. If this parameter is set to true, Auto Scaling attempts to create an instance to replace a preemptible instance when Auto Scaling receives a system message which indicates that the preemptible instance is to be reclaimed.
CompensateWithOnDemand	Boolean	No	true	Specifies whether to automatically create pay-as-you-go instances to meet the requirements on the number of instances when the expected capacity of preemptible instances cannot be fulfilled due to reasons such as high prices or insufficient resources. This parameter takes effect only when MultiAZPolicy is set to COST_OPTIMIZED. Valid values: <ul style="list-style-type: none">• true• false Default value: true.
SpotInstancePools	Integer	No	5	The number of instance types that are available. The system creates preemptible instances of multiple instance types that are available at the lowest cost in the scaling group. Valid values: 1 to 10.
DesiredCapacity	Integer	No	5	The expected number of ECS instances in the scaling group. Auto Scaling automatically maintains the specified expected number of ECS instances. The expected number cannot be greater than the MaxSize value and cannot be less than the MinSize value.

Parameter	Type	Required	Example	Description
GroupDeletionProtection	Boolean	No	true	<p>Specifies whether to enable deletion protection for the scaling group. Valid values:</p> <ul style="list-style-type: none">• true: Deletion protection is enabled for the scaling group, and the scaling group cannot be deleted.• false: Deletion protection is disabled for the scaling group. <p>Default value: false.</p>
GroupType	String	No	ECS	<p>The type of instances that are managed by the scaling group. Valid values:</p> <ul style="list-style-type: none">• ECS• ECI <p>Default value: ECS.</p>
ContainerGroupId	String	No	eci-uf6fonnghi50u374****	The ID of the elastic container instance.

Parameter	Type	Required	Example	Description
VSwitchIds.N	String	No	vsw-bp14zolna43z266bq****	<p>The ID of vSwitch N. Valid values of N: 1 to 5. If you use the VSwitchIds.N parameter, the VSwitchId parameter is ignored. If VSwitchId is specified, the network type of the scaling group is VPC.</p> <p>When you specify multiple vSwitches, take note of the following items:</p> <ul style="list-style-type: none">• The vSwitches must belong to the same VPC.• The vSwitches can belong to different zones.• vSwitches are sorted in ascending order based on the value of N. 1 indicates the highest priority. When an ECS instance cannot be created in the zone where the vSwitch with the highest priority resides, Auto Scaling creates instances in the zone where the vSwitch that has the next highest priority resides. <div style="background-color: #e1f5fe; padding: 10px; border-radius: 5px;"><p>? Note If VSwitchId or VSwitchIds.N is not specified, the network type of the scaling group is classic network.</p></div>

Parameter	Type	Required	Example	Description
LifecycleHook.N.DefaultResult	String	No	CONTINUE	<p>The action that the scaling group takes when the lifecycle hook times out. Valid values:</p> <ul style="list-style-type: none"> CONTINUE: Auto Scaling continues to respond to a scale-in or scale-out request. ABANDON: If a scale-out activity is executed, Auto Scaling releases the created ECS instances. If a scale-in activity is executed, Auto Scaling removes the ECS instances. <p>If the scaling group has multiple lifecycle hooks and one of them times out when the DefaultResult parameter is set to ABANDON during a scale-in activity, the remaining lifecycle hooks in the scaling group also time out. Otherwise, the scaling activity proceeds as expected after the lifecycle hook times out and continues with the action specified by the DefaultResult parameter.</p> <p>Default value: CONTINUE.</p>
LifecycleHook.N.LifecycleHookName	String	No	lifecyclehook****	The name of lifecycle hook N. After you specify this parameter, you cannot modify the name of the lifecycle hook. If you do not specify this parameter, the name of the lifecycle hook is the same as the ID of the lifecycle hook.

Parameter	Type	Required	Example	Description
LifecycleHook.N.LifecycleTransition	String	No	SCALE_OUT	<p>The type of scaling activities to which the lifecycle hook applies. Valid values:</p> <ul style="list-style-type: none"> • SCALE_OUT • SCALE_IN <div style="background-color: #e0f2ff; padding: 10px;"> ? Note If lifecycle hooks are specified for the scaling group, LifecycleHook.N.LifecycleTransition is required and other related parameters are optional. </div>
LifecycleHook.N.NotificationMetadata	String	No	Test	<p>The fixed string to be included when Auto Scaling sends a notification about the wait state of a scaling activity. The parameter value cannot exceed 128 characters in length. Auto Scaling sends the specified NotificationMetadata parameter value together with the notification message. This way, you can categorize your notifications. The LifecycleHook.N.NotificationMetadata parameter is valid only after you specify the LifecycleHook.N.NotificationArn parameter.</p>
LifecycleHook.N.NotificationArn	String	No	acs:ess:cn-hangzhou:11111111:queue/queue2	<p>The Alibaba Cloud Resource Name (ARN) of the notification object that Auto Scaling uses to notify you when an instance enters the pending state for the lifecycle hook. This object can be a Message Service (MNS) queue or an MNS topic. The parameter value is in the following format: acs:ess:{region}:{account-id}:{resource-relative-id}.</p> <ul style="list-style-type: none"> • region: the region where the scaling group resides • account-id: the ID of the Alibaba Cloud account <p>Examples:</p> <ul style="list-style-type: none"> • MNS queue: acs:ess:{region}:{account-id}:queue/{queuename} • MNS topic: acs:ess:{region}:{account-id}:topic/{topicname}

Parameter	Type	Required	Example	Description
LifecycleHook.N.HeartbeatTimeout	Integer	No	600	<p>The period of time before the lifecycle hook times out. When the lifecycle hook times out, Auto Scaling performs the default action. Valid values: 30 to 21600. Unit: seconds.</p> <p>After you create a lifecycle hook, you can call the RecordLifecycleActionHeartbeat operation to extend the timeout period and keep the instances in the pending state. You can also call the CompleteLifecycleAction operation to remove a scaling activity from the pending state.</p> <p>Default value: 600.</p>
VServerGroup.N.ServerGroupAttribute.N.VServerGroupUpdateId	String	No	rsp-bp1443g77****	<p>The ID of the vServer group.</p> <p>For more information, see AttachVServerGroups.</p>
VServerGroup.N.ServerGroupAttribute.N.Weight	Integer	No	100	<p>The weight of the ECS instance as a backend server after the instance is added to vServer group N. If you increase the weight of an ECS instance in a vServer group, the number of access requests that are forwarded to the ECS instance increases. If the weight is set to 0, no access requests are forwarded to the ECS instance.</p> <p>Valid values: 0 to 100. Default value: 50.</p> <p>Valid values of N: 1 to 5.</p> <p>For more information, see AttachVServerGroups.</p>
VServerGroup.N.ServerGroupAttribute.N.Port	Integer	No	22	<p>The port number used by the ECS instance after the instance is added to vServer group N. Valid values: 1 to 65535.</p> <p>For more information, see AttachVServerGroups.</p>

Parameter	Type	Required	Example	Description
VServerGroup.N.LoadBalancerId	String	No	lb-bp1u7etiogg38ywz****	<p>The ID of the CLB instance to which vServer group N belongs. CLB instances are formerly known as SLB instances.</p> <p>For more information, see AttachVServerGroups.</p>
Tag.N.Key	String	No	Department	The key of tag N of the scaling group.
Tag.N.Value	String	No	Finance	The value of tag N of the scaling group.
LaunchTemplateOverride.N.InstanceType	String	No	ecs.c5.xlarge	<p>If you want to scale instances in the scaling group based on the capacity, you must specify both the LaunchTemplateOverride.N.InstanceType and LaunchTemplateOverride.N.WeightedCapacity parameters.</p> <p>This parameter specifies instance type N that overrides the instance type specified in the launch template. You can specify N instance types for the extended configurations. Valid values of N: 1 to 10.</p> <p>Note This parameter takes effect only when the LaunchTemplateId parameter is specified.</p> <p>For information about the valid values of InstanceType in InstanceTypeOverride.N.InstanceType, see Instance families.</p>

Parameter	Type	Required	Example	Description
LaunchTemplateOverride.N.WeightedCapacity	Integer	No	4	<p>If you want to scale instances in the scaling group based on the capacity, you must specify the LaunchTemplateOverride.N.WeightedCapacity parameter after the LaunchTemplateOverride.N.InstanceType parameter is specified. The two parameters have a one-to-one correspondence. The value of N in the two parameters must be the same.</p> <p>This parameter specifies the weight of the instance type. The weight specifies the capacity of a single instance of the specified instance type in the scaling group. A greater weight indicates that a smaller number of instances of the specified instance type are required to meet the expected capacity.</p> <p>Performance metrics such as the number of vCPUs and the memory size of each instance type may vary. You can specify different weights for different instance types based on your business requirements.</p> <p>Example:</p> <ul style="list-style-type: none"> • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 <p>To meet the expected capacity, Auto Scaling creates two ecs.c5.xlarge instances.</p> <div style="background-color: #e0f2f1; padding: 10px; margin-top: 10px;"> ? Note The capacity of the scaling group cannot exceed the sum of the maximum capacity specified by MaxSize and the maximum weight of the instance type. </div> <p>Valid values of WeightedCapacity in InstanceTypeOverride.N.WeightedCapacity: 1 to 500.</p>

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

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingGroupId	String	asg-bp14wlu85wrpchm0****	The ID of the scaling group.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action/CreateScalingGroup
&AlbServerGroup.1.AlbServerGroupId=sgp-ddwb0y0g6y9bjm****
&AlbServerGroup.1.Port=22
&AlbServerGroup.1.Weight=100
&MaxSize=20
&MinSize=2
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<CreateScalingGroupResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
  <ScalingGroupId>asg-bp14wlu85wrpchm0****</ScalingGroupId>
</CreateScalingGroupResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
  "ScalingGroupId" : "asg-bp14wlu85wrpchm0****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	IncorrectDBInstanceState	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance is not in the Running state.

HTTP status code	Error code	Error message	Description
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned because the health check feature is not enabled for the specified CLB instance.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified CLB 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 ApsaraDB RDS instance and the specified scaling group are not in the same region.
400	InvalidLoadBalancerId.InvalidAddressType	The current address type of specified load balancer does not support this action.	The error message returned because the network type of the specified CLB instance does not match the network type of 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 added to the backend server group of the specified CLB instance is different from the network type 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 CLB instance and the specified 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 added to the backend server group of the CLB instance associated with 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 ScalingPolicy parameter does not exist.

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.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because the maximum number of IP addresses in the whitelist that manages access to the associated ApsaraDB RDS instance has been reached.
400	QuotaExceeded.PrivateIpAddress	Private IP address quota exceeded in the specified virtual switch.	The error message returned because no idle private IP addresses are available in the CIDR block of the vSwitch.
400	QuotaExceeded.ScalingGroup	Scaling group quota exceeded.	The error message returned because the maximum number of scaling groups has been reached.
400	QuotaExceeded.VPCInstance	Instance quota exceeded in the specified VPC.	The error message returned because the maximum number of instances in the VPC has been reached.

HTTP status code	Error code	Error message	Description
404	InvalidDBInstanceId.NotFound	DB instance "XXX" does not exist.	The error message returned because the specified ApsaraDB RDS instance does not exist.
404	InvalidLoadBalancerId.NotFound	The specified Load Balancer does not exist.	The error message returned because the specified CLB instance does not exist.
404	InvalidRegionId.NotFound	The specified region does not exist.	The error message returned because the specified region does not exist.
404	InvalidVSwitchId.NotFound	The specified virtual switch does not exist.	The error message returned because the specified vSwitch does not exist.
400	LaunchTemplateVersionSet.NotFound	The specific version of launch template is not exist.	The error message returned because the specified version of the launch template does not exist.
400	LaunchTemplateSet.NotFound	The specified launch template set is not found.	The error message returned because the specified launch template does not exist.
400	TemplateMissingParameter.ImageId	The input parameter "ImageId" that is mandatory for processing this request is not supplied.	The error message returned because the ImageId parameter required for the specified 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 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 launch template is not specified.
400	TemplateVersion.NotNumber	The input parameter "LaunchTemplateVersion" is supposed to be a string representing the version number.	The error message returned because the fixed version number specified for the LaunchTemplateVersion parameter of the launch template contains non-digit characters.
400	AlbServerGroup.NotExist	The ServerGroup "%s" do(es) not exist.	The error message returned because the specified ALB server group does not exist within the account.

9.2. EnableScalingGroup

Enables a scaling group.

Description

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

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

If you specify a value for the `Instances.N` parameter when you call the operation, Auto Scaling checks whether the total number of ECS instances is within the range allowed for the scaling group after you call the operation.

- If the total number of ECS instances is less than the minimum number of instances allowed in the scaling group after you call the operation, Auto Scaling automatically creates the required number of pay-as-you-go ECS instances to reach the minimum number. For example, if the minimum number of instances allowed in your scaling group is five, and you specify the `Instances.N` parameter to add two created ECS instances to the scaling group, Auto Scaling automatically creates three instances in the scaling group after the two instances are added.
- If the total number of instances is greater than the maximum number of instances allowed in the scaling group, the operation fails to be called.

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
Action	String	Yes	EnableScalingGroup	The operation that you want to perform. Set the value to EnableScalingGroup .
ScalingGroupId	String	Yes	asg-bp14wlu85wrpchm0****	The ID of the scaling group.
ActiveScalingConfigurationId	String	No	asc-bp1ffogfdauy0nu5****	The ID of the scaling configuration that you need to put into the Active state.
LaunchTemplateId	String	No	lt-m5e3ofjr1zn1aw7****	The ID of the launch template that you want to use to create instances.

Parameter	Type	Required	Example	Description
LaunchTemplateVersion	String	No	Default	<p>The version number of the launch template. Valid values:</p> <ul style="list-style-type: none"> • A fixed template version number. • Default: the default template version. • Latest: the latest template version.
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.
InstanceId.N	String	No	i-283vv***	<p>The ID of instance N that you want to add to the scaling group after the scaling group is enabled. Valid values of N: 1 to 20.</p> <p>Before you add ECS instances to a scaling group, make sure that the instances meet the following requirements:</p> <ul style="list-style-type: none"> • The instances and the scaling group must reside in the same region. • The instances are in the Running state. • The instances are not in other scaling groups. • The instances are billed on a subscription or pay-as-you-go basis, or the instances are preemptible instances. • If the VswitchID parameter is specified for the scaling group, the instances must reside in the same virtual private cloud (VPC) as the specified vSwitch. Instances reside in the classic network or other VPCs cannot be added to the scaling group. • If the VswitchID parameter is not specified for the scaling group, the ECS instances must not reside in VPCs.

Parameter	Type	Required	Example	Description
LoadBalancerWeight.N	Integer	No	50	<p>The weight of instance N that serves as a backend server of the associated Server Load Balancer (SLB) instance. Valid values of N: 1 to 20. Valid values of this parameter: 1 to 100. Default value: 50.</p>
LaunchTemplateOverride.N.InstanceType	String	No	ecs.c5.xlarge	<p>If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.InstanceType and LaunchTemplateOverride.N.WeightedCapacity parameters at the same time.</p> <p>This parameter specifies instance type N that overrides the instance type specified in the launch template. You can specify N instance types in the extended configurations of the launch template. Valid values of N: 1 to 10.</p> <p>Note This parameter takes effect only when the LaunchTemplateId parameter is specified.</p> <p>For information about the valid values of InstanceType in InstanceTypeOverride.N.InstanceType, see Instance families.</p>
				<p>If you want to scale instances in the scaling group based on the weighted capacity of instance types, you must specify the LaunchTemplateOverride.N.WeightedCapacity parameter after you specify the LaunchTemplateOverride.N.InstanceType parameter. The two parameters have a one-to-one correspondence. The value of N in the two parameters must be the same.</p>

Parameter	Type	Required	Example	Description
LaunchTemplateOverride.N.WeightedCapacity	Integer	No	4	<p>This parameter specifies the weight of the instance type. The weight specifies the capacity of a single instance of the specified instance type in the scaling group. A higher weight specifies that a smaller number of instances of the specified instance type are required to meet the expected capacity.</p> <p>Performance metrics such as the number of vCPUs and the memory size of each instance type may vary. You can specify different weights for different instance types based on your business requirements.</p> <p>Examples:</p> <ul style="list-style-type: none"> • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 <p>To meet the expected capacity, Auto Scaling creates two ecs.c5.xlarge instances.</p> <p>Note The capacity of the scaling group cannot exceed the sum of the maximum capacity that is specified by MaxSize and the maximum weight of the instance type.</p> <p>Valid values of WeightedCapacity in LaunchTemplateOverride.N.WeightedCapacity: 1 to 500.</p>

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=EnableScalingGroup  
&ScalingGroupId=asg-bp14wlu85wrpchm0****  
&ActiveScalingConfigurationId=asc-bp1ffogfdauy0nu5****  
&LaunchTemplateId=lt-m5e3ofjrlzn1aw7****  
&LaunchTemplateVersion=Default  
&InstanceId=["i-283vv****"]  
&LoadBalancerWeight=[50]  
&LaunchTemplateOverride=[{"InstanceType":"ecs.c5.xlarge","WeightedCapacity":4}]  
&RegionId=cn-qingdao  
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK  
Content-Type:application/xml  
<EnableScalingGroupResponse>  
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>  
</EnableScalingGroupResponse>
```

JSON format

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"  
}
```

Error codes

For a list of error codes, visit the [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 within the Alibaba Cloud account.

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the operation.
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the specified scaling group is being deleted.
404	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	The error message returned because the specified scaling configuration does not exist in the 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 scaling configuration that is in the Active state.
400	MissingActiveScalingConfiguration	An active scaling configuration for the specified scaling group is not supplied.	The error message returned because no scaling configuration in the scaling group is in the Active state.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the specified ECS instance does not exist within the Alibaba Cloud account.

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

HTTP status code	Error code	Error message	Description
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified SLB instance is not in the Active state.
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned because the health check feature is not enabled for the specified SLB instance.
400	InvalidLoadBalancerId.InvalidInstanceNetworkType	The network type of the instance in specified Load Balancer does not support this action.	The error message returned because the network type of the ECS instance that is attached to the 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 that is attached to the specified SLB instance does not reside in the same VPC as that of the vSwitch specified by VswitchID.
400	IncorrectDBInstanceState	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance is not in the Running state.
400	IncorrectCapacity.MaxValue	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 is greater than the maximum number of instances allowed after a specified number of instances are added to the scaling group.

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

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

9.3. DisableScalingGroup

You can call this operation to disable a scaling group.

Description

When you disable a scaling group, take note of the following items:

- If scaling activities are being executed in the specified scaling group when you call this operation, these activities will continue until they are complete. However, scaling activities that are triggered after this operation is called will be rejected.
- This operation can be called only when the scaling group is in the Active state.

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
Action	String	Yes	DisableScalingGroup	The operation that you want to perform. Set the value to DisableScalingGroup.
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dloq****	The ID of the scaling group.

Response parameters

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

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=DisableScalingGroup
&ScalingGroupId=asg-bp18p2yfxow2dloq****
&<Common request parameters>
```

Sample success responses

XML format

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

JSON format

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

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
------------------	------------	---------------	-------------

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.

9.4. SetGroupDeletionProtection

Enables or disables deletion protection for a 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
Action	String	Yes	SetGroupDeletionProtection	The operation that you want to perform. Set the value to SetGroupDeletionProtection .
GroupDeletionProtection	Boolean	Yes	true	<p>Specifies whether to enable deletion protection for the scaling group. Valid values:</p> <ul style="list-style-type: none"> • true: enables deletion protection. After deletion protection is enabled, you cannot delete the scaling group by using the console or by calling an API operation. Before you delete a scaling group, you must disable deletion protection for the scaling group. • false: disables deletion protection. <p>Default value: false.</p>
ScalingGroupId	String	Yes	asg-bp1igpak5ft1flyp****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.

Response parameters

Parameter	Type	Example	Description
RequestId	String	CCC29E24-3AEC-4F2C-8A14-78B14FA738B7	The ID of the request.

Examples

Sample request

```
http(s)://ess.aliyuncs.com/?Action=SetGroupDeletionProtection
&GroupDeletionProtection=true
&ScalingGroupId=asg-bpligpak5ft1flyp****
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<SetGroupDeletionProtectionResponse>
  <RequestId>CCC29E24-3AEC-4F2C-8A14-78B14FA738B7</RequestId>
</SetGroupDeletionProtectionResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "CCC29E24-3AEC-4F2C-8A14-78B14FA738B7"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description

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.
400	InvalidOperationException.Conflict	Specific operation may conflicts with other operations, please retry later.	The error message returned because the operation conflicts with another operation that is in progress. Try again later.

9.5. DeleteScalingGroup

Deletes a scaling group.

Description

Before you call the DeleteScalingGroup operation to delete a scaling group, take note of the following items:

- If you delete a scaling group, the scaling configuration, scaling rules, scaling activities, and scaling requests related to the scaling group are also deleted.
- If you delete a scaling group, the scheduled tasks and event-triggered tasks of the scaling group are not deleted. The Server Load Balancer (SLB) instances and ApsaraDB RDS instances with which the scaling group is associated are also not deleted.

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
Action	String	Yes	DeleteScalingGroup	The operation that you want to perform. Set the value to DeleteScalingGroup .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dl0q***	The ID of the scaling group.

Parameter	Type	Required	Example	Description
ForceDelete	Boolean	No	false	<p>Specifies whether to forcibly delete the scaling group and remove and release Elastic Compute Service (ECS) instances in the scaling group when the ECS instances or ongoing scaling activities exist in the scaling group. Valid values:</p> <ul style="list-style-type: none"> • true: The scaling group is disabled and new scaling requests are rejected. After all existing scaling requests are processed, the ECS instances in the scaling group are removed. Then, the scaling group is deleted. If the ECS instances are manually added to the scaling group, the ECS instances are removed. If the ECS instances are automatically created in the scaling group, the ECS instances are removed and released. • false: You can set this parameter to false if the following conditions are met: <ul style="list-style-type: none"> ◦ No ongoing scaling activities exist in the scaling group. ◦ The Total Capacity parameter is set to 0, which specifies that no ECS instances exist in the scaling group. <p>Default value: false.</p>
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3C83E	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DeleteScalingGroup  
&ScalingGroupId=asg-bp18p2yfxow2dlog****  
&ForceDelete=false  
&RegionId=cn-qingdao  
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK  
Content-Type:application/xml  
<DeleteScalingGroupResponse>  
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>  
</DeleteScalingGroupResponse>
```

JSON format

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"  
}
```

Error codes

For a list of error codes, visit the [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 scaling group does not exist within the Alibaba Cloud 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 operation.

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 because the scaling group contains ECS instances.

9.6. DescribeScalingGroups

Queries scaling groups.

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
Action	String	Yes	DescribeScalingGroups	The operation that you want to perform. Set the value to DescribeScalingGroups .
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	10	The number of entries to return on each page. Maximum value: 50. Default value: 10.
ScalingGroupName	String	No	scalinggroup****	The name of the scaling group.

Parameter	Type	Required	Example	Description
ScalingGroupName.e.1	String	No	scalinggroup****	ScalingGroupName.1 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.e.2	String	No	scalinggroup****	ScalingGroupName.2 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.e.3	String	No	scalinggroup****	ScalingGroupName.3 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.e.4	String	No	scalinggroup****	ScalingGroupName.4 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.e.5	String	No	scalinggroup****	ScalingGroupName.5 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.e.6	String	No	scalinggroup****	ScalingGroupName.6 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.e.7	String	No	scalinggroup****	ScalingGroupName.7 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.

Parameter	Type	Required	Example	Description
ScalingGroupName.8	String	No	scalinggroup****	ScalingGroupName.8 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.9	String	No	scalinggroup****	ScalingGroupName.9 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.10	String	No	scalinggroup****	ScalingGroupName.10 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.11	String	No	scalinggroup****	ScalingGroupName.11 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.12	String	No	scalinggroup****	ScalingGroupName.12 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.13	String	No	scalinggroup****	ScalingGroupName.13 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.14	String	No	scalinggroup****	ScalingGroupName.14 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.

Parameter	Type	Required	Example	Description
ScalingGroupName.15	String	No	scalinggroup****	ScalingGroupName.15 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.16	String	No	scalinggroup****	ScalingGroupName.16 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.17	String	No	scalinggroup****	ScalingGroupName.17 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.18	String	No	scalinggroup****	ScalingGroupName.18 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.19	String	No	scalinggroup****	ScalingGroupName.19 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.
ScalingGroupName.20	String	No	scalinggroup****	ScalingGroupName.20 is the name of the scaling group that you want to query. The names of inactive scaling groups are not displayed in the query results, and no error is reported.

Parameter	Type	Required	Example	Description
GroupType	String	No	ECS	<p>The type of instances that are managed by the scaling group. Valid values:</p> <ul style="list-style-type: none"> • ECS: ECS instances • ECI: elastic container instances <p>Default value: ECS.</p>
ScalingGroupId.N	String	No	asg-bp14wlu85wrpchm0****	The ID of scaling group N that you want to query. Valid values of N: 1 to 20. The IDs of inactive scaling groups are not displayed in the query results, and no error is reported.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	10	The number of entries returned per page.
TotalCount	Integer	1	The total number of scaling groups.
ScalingGroups	Array of ScalingGroup		The details of the scaling groups.
ScalingGroup			
VpcId	String	vpc-bp1vwnn14rqpyicj***	The ID of the virtual private cloud (VPC) to which the scaling group belongs.
CreationTime	String	2014-08-14T10:58Z	The time when the scaling group was created.
TotalInstanceCount	Integer	1	The total number of ECS instances in the scaling group.

Parameter	Type	Example	Description
ScalingGroupName	String	dyrSuvBOtO1dEdlIibp****	The name of the scaling group.
SpotInstancePools	Integer	5	The number of instance types that are available. Auto Scaling creates preemptible instances of multiple instance types that are provided at the lowest cost. Valid values: 0 to 10.
StoppedCapacity	Integer	1	The number of instances on which the Economical Mode feature is enabled in the scaling group.
OnDemandPercentageAboveBaseCapacity	Integer	20	The percentage of pay-as-you-go instances among the extra instances that exceed the number specified by OnDemandBaseCapacity. Valid values: 0 to 100.
ModificationTime	String	2014-08-14T10:58Z	The time when the scaling group was modified.
MinSize	Integer	1	The minimum number of ECS instances in the scaling group.
ScalingGroupId	String	asg-bp14wlu85wrpchm0****	The ID of the scaling group.
CompensateWithOnDemand	Boolean	true	Indicates whether pay-as-you-go instances can be automatically created to meet the required number of ECS instances when the expected number of preemptible instances cannot be met. The expected number of preemptible instances may not be met due to reasons such as high costs or insufficient resources. This parameter is valid only when MultiAZPolicy is set to COST_OPTIMIZED. Valid values: <ul style="list-style-type: none">• true: Pay-as-you-go instances can be automatically created.• false: Pay-as-you-go instances cannot be automatically created.

Parameter	Type	Example	Description
ScalingPolicy	String	recycle	<p>The instance reclaim mode of the scaling group. Valid values:</p> <ul style="list-style-type: none"> • recycle: The instance reclaim mode of the scaling group is set to Economical Mode. • release: The instance reclaim mode of the scaling group is set to Release. <p>For more information about how to remove instances, see RemoveInstances.</p>
RemovingWaitCapacity	Integer	1	The number of ECS instances in the Pending Remove state in the scaling group.
ActiveCapacity	Integer	1	The number of ECS instances that are added to the scaling group and are running as expected.
OnDemandBaseCapacity	Integer	30	The minimum number of pay-as-you-go instances required in the scaling group. Valid values: 0 to 1000. If the number of pay-as-you-go instances is less than the value of this parameter, Auto Scaling preferentially creates pay-as-you-go instances.
ProtectedCapacity	Integer	1	The number of ECS instances that are in the Protected state in the scaling group.
HealthCheckType	String	ECS	<p>Indicates whether to perform health checks on ECS instances in the scaling group. Valid values:</p> <ul style="list-style-type: none"> • NONE: Auto Scaling does not perform health checks on ECS instances in the scaling group. • ECS: Auto Scaling performs health checks on ECS instances in the scaling group.

Parameter	Type	Example	Description
LifecycleState	String	Active	<p>The lifecycle status of the scaling group. Valid values:</p> <ul style="list-style-type: none"> • Active: The scaling group is active. Active scaling groups can receive requests to execute scaling rules and trigger scaling activities. • Inactive: The scaling group is inactive. Inactive scaling groups cannot receive requests to execute scaling rules. • Deleting: The scaling group is being deleted. Scaling groups that are being deleted cannot receive requests to execute scaling rules, and the parameter settings of the scaling groups cannot be modified.
GroupDeletionProtection	Boolean	true	<p>Indicates whether deletion protection is enabled for the scaling group. Valid values:</p> <ul style="list-style-type: none"> • true: Deletion protection is enabled for the scaling group. The scaling group cannot be deleted. • false: Deletion protection is disabled for the scaling group.
ActiveScalingConfigurationId	String	asc-bp1et2qekq3ojr33**	The ID of the active scaling configuration in the scaling group.
GroupType	String	ECS	The type of instances that are managed by the scaling group.

Parameter	Type	Example	Description
MultiAZPolicy	String	PRIORITY	<p>The scaling policy for a multi-zone scaling group that contains ECS instances. Valid values:</p> <ul style="list-style-type: none"> • PRIORITY: ECS instances are scaled based on the VSwitchIds.N parameter. If an ECS instance cannot be created in the zone where the vSwitch that has the highest priority resides, Auto Scaling creates the ECS instance in the zone where the vSwitch that has the next highest priority resides. • COST_OPTIMIZED: ECS instances are created based on the unit prices of their vCPUs. ECS instances that have vCPUs provided at the lowest price are preferentially created. Preemptible instances are preferentially created when preemptible instance types are specified in the scaling configuration. You can configure the CompensateWithOnDemand parameter to specify whether to automatically create pay-as-you-go instances when preemptible instances cannot be created due to insufficient resources. <div style="background-color: #e0f2f1; padding: 10px; margin-top: 10px;"> ? Note COST_OPTIMIZED is valid when multiple instance types are specified or at least one preemptible instance type is specified. </div> <ul style="list-style-type: none"> • BALANCE: ECS instances are evenly distributed across zones that are specified in the scaling group. If ECS instances are unevenly distributed across zones due to insufficient resources, you can call the RebalanceInstance operation to distribute the instances across zones.
RemovingCapacity	Integer	0	The number of ECS instances that are being removed from the scaling group.
PendingWaitCapacity	Integer	1	The number of ECS instances that are in the Pending Add state in the scaling group.
StandbyCapacity	Integer	1	The number of instances that are in the Standby state in the scaling group.

Parameter	Type	Example	Description
PendingCapacity	Integer	0	The number of ECS instances that are being added to the scaling group and are also being configured.
LaunchTemplateId	String	lt-m5e3ofjr1zn1aw7**	The ID of the launch template used by the scaling group.
TotalCapacity	Integer	1	If an instance type weight is specified for the scaling group, the value of this parameter indicates the total weighted capacity of all ECS instances in the scaling group. If no instance type weight is specified for the scaling group, the value of this parameter indicates the total number of ECS instances in the scaling group.
DesiredCapacity	Integer	5	The expected number of ECS instances in the scaling group. Auto Scaling automatically maintains the specified number of ECS instances.
SpotInstanceRecovery	Boolean	true	Indicates whether to supplement preemptible instances. If this parameter is set to true, Auto Scaling attempts to create an instance to replace a preemptible instance when Auto Scaling receives a system message which indicates that the preemptible instance is to be reclaimed.
LaunchTemplateVersion	String	Default	The version of the launch template used by the scaling group.
RegionId	String	cn-qingdao	The region ID of the scaling group.
vSwitchId	String	vsw-bp1whw2u46cn8zubm***	The ID of the vSwitch that is associated with the scaling group.
MaxSize	Integer	2	The maximum number of ECS instances in the scaling group.

Parameter	Type	Example	Description
DefaultCooldown	Integer	60	The default cooldown time of the scaling group. During the cooldown time, Auto Scaling executes only the scaling activities that are triggered by event-triggered tasks associated with CloudMonitor .
VServerGroups	Array of VServerGroup		The list of backend server groups.
VServerGroup			
LoadBalancerId	String	147b46d767c-cn-qingdao-cm5***	The ID of the Server Load Balancer (SLB) instance to which the backend server group belongs.
VServerGroupAttributes	Array of VServerGroupAttribute		
VServerGroupAttribute			
VServerGroupId	String	rsp-bp12bjrny***	The attributes of the backend server group.
Weight	Integer	1	The weight of the backend server group.
Port	Integer	22	The port number used by the SLB instance to provide external services.
LaunchTemplateOverrides	Array of LaunchTemplateOverride		The details of the instance type of the extended configurations.
LaunchTemplateOverride			
WeightedCapacity	Integer	4	The weight of the instance type. The value of this parameter indicates the capacity of a single instance of this instance type in the scaling group. A greater weight indicates that a smaller number of instances of the specified instance type are required to meet the expected capacity.

Parameter	Type	Example	Description
InstanceType	String	ecs.c5.xlarge	The instance type. The specified instance type overrides the instance type in the launch template.
AlbServerGroups	Array of AlbServerGroup		The details of the Application Load Balancer (ALB) server groups.
AlbServerGroup			
AlbServerGroupId	String	sgp-ddwb0y0g6y9bjm*** *	The ID of the ALB server group.
Weight	Integer	100	The weight of the ECS instance as a backend server after the instance is added to the ALB server group.
Port	Integer	80	The port number used by the ECS instance after the instance is added to the ALB server group.
RemovalPolicies	Array of String	OldestScalingConfiguration	<p>The policies that are used to remove ECS instances from the scaling group. Valid values:</p> <ul style="list-style-type: none"> • OldestInstance: Auto Scaling removes ECS instances that are added to the scaling group at the earliest point in time. • NewestInstance: Auto Scaling removes ECS instances that are most recently added to the scaling group. • OldestScalingConfiguration: Auto Scaling removes ECS instances that are created based on the earliest scaling configuration.
DBInstancelds	Array of String	rm-bp15556qzebg1****	The IDs of the ApsaraDB RDS instances that are associated with the scaling group.
LoadBalancerIds	Array of String	lb-bp19byhscefk3x0li** **	The IDs of the SLB instances that are associated with the scaling group.

Parameter	Type	Example	Description
VSwitchIds	Array of String	vsw-bp1whw2u46cn8zu****	The IDs of the vSwitches that are associated with the scaling group. If you configure the VSwitchIds parameter, the VSwitchId parameter is ignored.
SuspendedProcesses	Array of String	ScaleIn	The process that is suspended. If no process is suspended, null is returned. Valid values: <ul style="list-style-type: none">• ScaleIn• ScaleOut• HealthCheck• AlarmNotification• ScheduledAction
SystemSuspended	Boolean	true	Indicates whether Auto Scaling stops executing scaling activities in the scaling group. <ul style="list-style-type: none">• true: Auto Scaling stops executing scaling activities in the scaling group. This indicates that scaling activities failed for more than seven consecutive days in the scaling group. You must modify the scaling group or scaling configuration to resume the execution of scaling activities.• false: Auto Scaling does not stop executing scaling activities in the scaling group.
MonitorGroupId	String	1497****	The ID of the CloudMonitor application group that is associated with the scaling group.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DescribeScalingGroups
&RegionId=cn-qingdao
&PageNumber=1
&PageSize=10
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
```

```
Content-Type:application/xml
<DescribeScalingGroupsResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
  <PageNumber>1</PageNumber>
  <PageSize>10</PageSize>
  <TotalCount>1</TotalCount>
  <ScalingGroups>
    <VpcId>vpc-bp1vwnn14rqpyicj****</VpcId>
    <CreationTime>2014-08-14T10:58Z</CreationTime>
    <TotalInstanceCount>1</TotalInstanceCount>
    <ScalingGroupName>dyrSuvB0tOldEdIlIbp****</ScalingGroupName>
    <SpotInstancePools>5</SpotInstancePools>
    <StoppedCapacity>1</StoppedCapacity>
    <OnDemandPercentageAboveBaseCapacity>20</OnDemandPercentageAboveBaseCapacity>
    <ModificationTime>2014-08-14T10:58Z</ModificationTime>
    <MinSize>1</MinSize>
    <ScalingGroupId>asg-bp14wlu85wrpchm0****</ScalingGroupId>
    <CompensateWithOnDemand>true</CompensateWithOnDemand>
    <ScalingPolicy>recycle</ScalingPolicy>
    <RemovingWaitCapacity>1</RemovingWaitCapacity>
    <ActiveCapacity>1</ActiveCapacity>
    <OnDemandBaseCapacity>30</OnDemandBaseCapacity>
    <ProtectedCapacity>1</ProtectedCapacity>
    <HealthCheckType>ECS</HealthCheckType>
    <LifecycleState>Active</LifecycleState>
    <GroupDeletionProtection>true</GroupDeletionProtection>
    <ActiveScalingConfigurationId>asc-bp1et2qekq3ojr33****</ActiveScalingConfigurationId>
  <d>
    <GroupType>ECS</GroupType>
    <MultiAZPolicy>PRIORITY</MultiAZPolicy>
    <RemovingCapacity>0</RemovingCapacity>
    <PendingWaitCapacity>1</PendingWaitCapacity>
    <StandbyCapacity>1</StandbyCapacity>
    <PendingCapacity>0</PendingCapacity>
    <LaunchTemplateId>lt-m5e3ofjr1zn1aw7****</LaunchTemplateId>
    <TotalCapacity>1</TotalCapacity>
    <DesiredCapacity>5</DesiredCapacity>
    <SpotInstanceRemedy>true</SpotInstanceRemedy>
    <LaunchTemplateVersion>Default</LaunchTemplateVersion>
    <RegionId>cn-qingdao</RegionId>
    <VSwitchId>vsw-bp1whw2u46cn8zubm****</VSwitchId>
    <MaxSize>2</MaxSize>
    <DefaultCooldown>60</DefaultCooldown>
    <VServerGroups>
      <LoadBalancerId>147b46d767c-cn-qingdao-cm5****</LoadBalancerId>
      <VServerGroupAttributes>
        <VServerGroupId>rsp-bp12bjrnny****</VServerGroupId>
        <Weight>1</Weight>
        <Port>22</Port>
      </VServerGroupAttributes>
    </VServerGroups>
    <LaunchTemplateOverrides>
      <WeightedCapacity>4</WeightedCapacity>
      <InstanceType>ecs.c5.xlarge</InstanceType>
    </LaunchTemplateOverrides>
  </d>
</DescribeScalingGroupsResponse>
```

```
</LaunchTemplateOverrides>
<AlbServerGroups>
    <AlbServerGroupId>sgp-ddwb0y0g6y9bjm****</AlbServerGroupId>
    <Weight>100</Weight>
    <Port>80</Port>
</AlbServerGroups>
<RemovalPolicies>OldestScalingConfiguration</RemovalPolicies>
<DBInstanceIds>rm-bp15556qzebg1****</DBInstanceIds>
<LoadBalancerIds>lb-bp19byhscefk3x0li****</LoadBalancerIds>
<VSwitchIds>vsw-bp1whw2u46cn8zubm****</VSwitchIds>
<SuspendedProcesses>ScaleIn</SuspendedProcesses>
<SystemSuspended>true</SystemSuspended>
<MonitorGroupId>1497****</MonitorGroupId>
</ScalingGroups>
</DescribeScalingGroupsResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "PageNumber" : 1,
    "PageSize" : 10,
    "TotalCount" : 1,
    "ScalingGroups" : [ {
        "VpcId" : "vpc-bp1vwnn14rgpyicj****",
        "CreationTime" : "2014-08-14T10:58Z",
        "TotalInstanceCount" : 1,
        "ScalingGroupName" : "dyrSuvB0t01dEdI1Ibp****",
        "SpotInstancePools" : 5,
        "StoppedCapacity" : 1,
        "OnDemandPercentageAboveBaseCapacity" : 20,
        "ModificationTime" : "2014-08-14T10:58Z",
        "MinSize" : 1,
        "ScalingGroupId" : "asg-bp14wlu85wrpchm0****",
        "CompensateWithOnDemand" : true,
        "ScalingPolicy" : "recycle",
        "RemovingWaitCapacity" : 1,
        "ActiveCapacity" : 1,
        "OnDemandBaseCapacity" : 30,
        "ProtectedCapacity" : 1,
        "HealthCheckType" : "ECS",
        "LifecycleState" : "Active",
        "GroupDeletionProtection" : true,
        "ActiveScalingConfigurationId" : "asc-bplet2qekq3ojr33****",
        "GroupType" : "ECS",
        "MultiAZPolicy" : "PRIORITY",
        "RemovingCapacity" : 0,
        "PendingWaitCapacity" : 1,
        "StandbyCapacity" : 1,
        "PendingCapacity" : 0,
        "LaunchTemplateId" : "lt-m5e3ofjr1zn1aw7****",
        "TotalCapacity" : 1,
```

```
"DesiredCapacity" : 5,
"SpotInstanceRemedy" : true,
"LaunchTemplateVersion" : "Default",
"RegionId" : "cn-qingdao",
"VSwitchId" : "vsw-bp1whw2u46cn8zubm****",
"MaxSize" : 2,
"DefaultCooldown" : 60,
"VServerGroups" : [ {
    "LoadBalancerId" : "147b46d767c-cn-qingdao-cm5****",
    "VServerGroupAttributes" : [ {
        "VServerGroupId" : "rsp-bp12bjrny****",
        "Weight" : 1,
        "Port" : 22
    } ]
} ],
"LaunchTemplateOverrides" : [ {
    "WeightedCapacity" : 4,
    "InstanceType" : "ecs.c5.xlarge"
} ],
"AlbServerGroups" : [ {
    "AlbServerGroupId" : "sgp-ddwb0y0g6y9bjm****",
    "Weight" : 100,
    "Port" : 80
} ],
"RemovalPolicies" : [ "OldestScalingConfiguration" ],
"DBInstanceIds" : [ "rm-bp15556qzebg1****" ],
"LoadBalancerIds" : [ "lb-bp19byhscefk3x0li****" ],
"VSwitchIds" : [ "vsw-bp1whw2u46cn8zubm****" ],
"SuspendedProcesses" : [ "ScaleIn" ],
"SystemSuspended" : true,
"MonitorGroupId" : "1497****"
} ]
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

9.7. DescribeScalingInstances

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

Description

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

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
Action	String	Yes	DescribeScalingInstances	The operation that you want to perform. Set the value to DescribeScalingInstances .
RegionId	String	Yes	cn-hangzhou	The ID of the region to which the scaling group belongs.
ScalingGroupId	String	No	asg-bp1igpak5ft1flyp****	The ID of the scaling group.
ScalingConfigurationId	String	No	asc-bp1i65jd06v04vdh****	The ID of the associated scaling configuration.

Parameter	Type	Required	Example	Description
HealthStatus	String	No	Healthy	<p>The health status of the ECS instance in the scaling group. ECS instances that are not in the Running state are considered unhealthy. Valid values:</p> <ul style="list-style-type: none"> • Healthy • Unhealthy <p>Auto Scaling automatically removes unhealthy ECS instances from the scaling group and then releases the automatically created instances among the unhealthy instances.</p> <p>Unhealthy ECS instances that are manually added to the scaling group are released based on the management mode of the instance lifecycle. If the lifecycle of the ECS instances is not managed by the scaling group, Auto Scaling removes the instances from the scaling group but does not release them. If the lifecycle of the ECS instances is managed by the scaling group, Auto Scaling removes the instances from the scaling group and releases them.</p> <div style="background-color: #e0f2ff; padding: 10px; margin-top: 10px;"> ? Note Make sure that you have sufficient balance within your account. If you have overdue payments in your account, pay-as-you-go and preemptible instances are stopped or released. For information about how the status of ECS instances with overdue payments changes, see Overdue payments. </div>
				The lifecycle status of the ECS instance in the scaling group. Valid values:

Parameter	Type	Required	Example	Description
LifecycleState	String	No	InService	<ul style="list-style-type: none"> ● InService: The ECS instance is added to the scaling group and can provide services. ● Pending: The ECS instance is being added to the scaling group. During this process, Auto Scaling adds the ECS instance to the backend server groups of the associated Server Load Balancer (SLB) instance and adds the IP address of the ECS instance to the whitelist that manages access to the associated ApsaraDB RDS instance. ● Pending:Wait: The ECS instance is being added to the scaling group and enters the pending state. If a lifecycle hook that applies to scale-out activities is created for the scaling group, the ECS instance enters the pending state and waits for the lifecycle hook to time out before the instance is added to the scaling group. ● Protected: The ECS instance is being protected. The ECS instance can provide services as expected. However, Auto Scaling does not manage the lifecycle of the instance. You must manually manage the lifecycle of the instance. ● Standby: The ECS instance is on standby. The ECS instance is out of service and the weight of the ECS instance as an SLB backend server is set to zero. Auto Scaling does not manage the lifecycle of the instance. You must manually manage the lifecycle of the instance. ● Stopped: The ECS instance is stopped. The ECS instance is stopped and cannot provide services. ● Removing: The ECS instance is being removed from the scaling group. During this process, Auto Scaling removes the ECS instance from the backend server groups of the associated SLB instance and removes the IP address of the ECS instance from the whitelist that manages access to the associated ApsaraDB RDS instance. ● Removing:Wait: The ECS instance is being removed from the scaling

Parameter	Type	Required	Example	Description
				group and enters the pending state. If a lifecycle hook that applies to scale-in activities is created for the scaling group, the ECS instance enters the pending state and waits for the lifecycle hook to time out before the instance is removed from the scaling group.
CreationType	String	No	AutoCreated	Specifies how the ECS instance is created. Valid values: <ul style="list-style-type: none">• AutoCreated: The ECS instance is automatically created by Auto Scaling based on the instance configuration source of the scaling group.• Attached: The ECS instance is manually added to the scaling group.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	10	The number of entries to return on each page. Maximum value: 100. Default value: 10.
ScalingActivityId	String	No	asa- bp1c9djwrgxjyk3 1****	The ID of the scaling activity.
InstanceId.N	String	No	i- bp109k5j3dum1c e6****	The ID of ECS instance N. Valid values of N: 1 to 20. The IDs of inactive instances are not displayed in the query result, and no errors are returned.

Response parameters

Parameter	Type	Example	Description
RequestId	String	B13527BF-1FBD-4334-A512-20F5E9D3FB4D	The ID of the request.
PageSize	Integer	10	The number of entries returned per page.

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
TotalSpotCount	Integer	4	The total number of running preemptible instances in the scaling group.
TotalCount	Integer	1	The total number of ECS instances.
ScalingInstances	Array of ScalingInstance		The collection of information about ECS instances.
ScalingInstance			
CreationTime	String	2020-05-18T03:11Z	The time when the ECS instance was added to the scaling group. The value is accurate to the minute.
LoadBalancerWeight	Integer	50	The weight of the ECS instance as an SLB backend server.
LaunchTemplateId	String	lt-m5e3ofjr1zn1aw7**	The ID of the launch template.
InstanceId	String	i-bp109k5j3dum1ce6***	The ID of the ECS instance
SpotStrategy	String	SpotWithPriceLimit	<p>The bidding policy for the preemptible instance. Valid values:</p> <ul style="list-style-type: none"> • SpotWithPriceLimit: This policy applies to the preemptible instance with a user-defined maximum hourly price. • SpotAsPriceGo: This policy applies to the preemptible instance for which the market price at the time of purchase is automatically used as the bid price.
LaunchTemplateVersion	String	1	The version number of the launch template.

Parameter	Type	Example	Description
HealthStatus	String	Healthy	<p>The health status of the ECS instance in the scaling group. ECS instances that are not in the Running state are considered unhealthy. Valid values:</p> <ul style="list-style-type: none"> • Healthy • Unhealthy <p>Auto Scaling automatically removes unhealthy ECS instances from the scaling group and then releases the automatically created instances among the unhealthy instances.</p> <p>Unhealthy ECS instances that are manually added to the scaling group are released based on the management mode of the instance lifecycle. If the lifecycle of the ECS instances is not managed by the scaling group, Auto Scaling removes the instances from the scaling group but does not release them. If the lifecycle of the ECS instances is managed by the scaling group, Auto Scaling removes the instances from the scaling group and releases them.</p> <div style="background-color: #e1f5fe; padding: 10px;"> <p>? Note Make sure that you have sufficient balance within your account. If you have overdue payments in your account, pay-as-you-go and preemptible instances are stopped or released. For information about how the status of ECS instances with overdue payments changes, see Overdue payments.</p> </div>
ScalingGroupId	String	asg-bp1igpak5ft1flyp*** *	The ID of the scaling group.

Parameter	Type	Example	Description
WarmupState	String	NoNeedWarmup	<p>The warmup status of the ECS instance. Valid values:</p> <ul style="list-style-type: none"> • NoNeedWarmup: No warmup is required. • WaitingForInstanceWarmup: You must wait for the instance to complete warmup. • InstanceWarmupFinish: The warmup is complete.
LifecycleState	String	InService	<p>The lifecycle status of the ECS instance in the scaling group. Valid values:</p> <ul style="list-style-type: none"> • InService: The ECS instance is added to the scaling group and can provide services. • Pending: The ECS instance is being added to the scaling group. During this process, Auto Scaling adds the ECS instance to the backend server groups of the associated SLB instance and adds the IP address of the ECS instance to the whitelist that manages access to the associated ApsaraDB RDS instance. • Pending:Wait: The ECS instance is being added to the scaling group and enters the pending state. If a lifecycle hook that applies to scale-out activities is created for the scaling group, the ECS instance enters the pending state and waits for the lifecycle hook to time out before the instance is added to the scaling group. • Protected: The ECS instance is being protected. The ECS instance can provide services as expected. However, Auto Scaling does not manage the lifecycle of the instance. You must manually manage the lifecycle of the instance. • Standby: The ECS instance is on standby. The ECS instance is out of service and the weight of the ECS instance as an SLB backend server is set to zero. Auto Scaling does not manage the lifecycle of the instance. You must manually manage the lifecycle of the instance. • Stopped: The ECS instance is stopped. The ECS instance is stopped and cannot provide services. • Removing: The ECS instance is being removed from the scaling group. During this process, Auto Scaling removes the ECS instance from the backend server

Parameter	Type	Example	Description
			<p>groups of the associated SLB instance and removes the IP address of the ECS instance from the whitelist that manages access to the associated ApsaraDB RDS instance.</p> <ul style="list-style-type: none"> • Removing:Wait: The ECS instance is being removed from the scaling group and enters the pending state. If a lifecycle hook that applies to scale-in activities is created for the scaling group, the ECS instance enters the pending state and waits for the lifecycle hook to time out before the instance is removed from the scaling group.
CreationType	String	AutoCreated	<p>Indicates how the ECS instance is created. Valid values:</p> <ul style="list-style-type: none"> • AutoCreated: The ECS instance is automatically created by Auto Scaling based on the instance configuration source of the scaling group. • Attached: The ECS instance is manually added to the scaling group.
Zoneld	String	cn-hangzhou-g	The zone ID of the ECS instance.
ScalingConfigurationId	String	asc-bp1i65jd06v04vdh**	The ID of the associated scaling configuration.
Entrusted	Boolean	true	<p>Indicates whether the scaling group is allowed to manage the instance lifecycle when you manually add the instance. If the scaling group is allowed to manage the instance lifecycle, Auto Scaling can release the instance when the instance is automatically removed from the scaling group. This rule does not apply to instances that are manually removed from the scaling group. Valid values:</p> <ul style="list-style-type: none"> • true: The instance lifecycle is managed by the scaling group. • false: The instance lifecycle is not managed by the scaling group.

Parameter	Type	Example	Description
WeightedCapacity	Integer	4	The weight of the instance type, which indicates the capacity of a single instance of the specified instance type in the scaling group. A greater weight indicates that a smaller number of instances of the specified instance type are required to meet the expected capacity.
CreatedTime	String	2020-05-18T03:11:39Z	The time when the ECS instance was added to the scaling group. The value is accurate to the second.
ScalingActivityId	String	asa-bp1c9djwrgxjyk31****	The ID of the scaling activity during which the ECS instance is added to the scaling group.

Examples

Sample requests

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

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DescribeScalingInstancesResponse>
  <TotalCount>1</TotalCount>
  <PageSize>10</PageSize>
  <RequestId>D66AC79E-8299-4E0B-B681-3063C88E215B</RequestId>
  <PageNumber>1</PageNumber>
  <ScalingInstances>
    <ScalingInstance>
      <LoadBalancerWeight>50</LoadBalancerWeight>
      <CreatedTime>2020-12-21T03:11:00Z</CreatedTime>
      <WarmupState>NoNeedWarmup</WarmupState>
      <ZoneId>cn-hangzhou-g</ZoneId>
      <InstanceId>i-m5e3z5l951fibnd9****</InstanceId>
      <ScalingActivityId>asa-bp1c9djwrgxjyk31****</ScalingActivityId>
      <ScalingGroupId>asg-m5e8n5qw4atki7f6****</ScalingGroupId>
      <HealthStatus>Healthy</HealthStatus>
      <CreationTime>2020-12-21T03:11Z</CreationTime>
      <LifecycleState>InService</LifecycleState>
      <Entrusted>true</Entrusted>
      <ScalingConfigurationId>asc-m5e9vcoen45jspz7****</ScalingConfigurationId>
      <CreationType>AutoCreated</CreationType>
    </ScalingInstance>
  </ScalingInstances>
  <TotalSpotCount>0</TotalSpotCount>
</DescribeScalingInstancesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "TotalCount" : 1,
    "PageSize" : 10,
    "RequestId" : "D66AC79E-8299-4E0B-B681-3063C88E215B",
    "PageNumber" : 1,
    "ScalingInstances" : {
        "ScalingInstance" : [ {
            "LoadBalancerWeight" : 50,
            "CreatedTime" : "2020-12-21T03:11:00Z",
            "WarmupState" : "NoNeedWarmup",
            "ZoneId" : "cn-hangzhou-g",
            "InstanceId" : "i-m5e3z51951fibnd9****",
            "ScalingActivityId" : "asa-bp1c9djwrgxjyk31****",
            "ScalingGroupId" : "asg-m5e8n5qw4atki7f6****",
            "HealthStatus" : "Healthy",
            "CreationTime" : "2020-12-21T03:11Z",
            "LifecycleState" : "InService",
            "Entrusted" : true,
            "ScalingConfigurationId" : "asc-m5e9vcoen45jspz7****",
            "CreationType" : "AutoCreated"
        } ]
    },
    "TotalSpotCount" : 0
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

9.8. DescribeScalingActivities

Queries scaling activities.

Description

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

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

You can query scaling activities that are executed in the previous 30 days.

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
-----------	------	----------	---------	-------------

Parameter	Type	Required	Example	Description
Action	String	Yes	DescribeScalingActivities	The operation that you want to perform. Set the value to DescribeScalingActivities.
RegionId	String	Yes	cn-hangzhou	The region ID of the scaling group to which the scaling activity belongs.
PageNumber	Integer	No	1	<p>The page number of the scaling activity list to return. Pages start from page 1.</p> <p>Default value: 1.</p>
PageSize	Integer	No	10	<p>The number of entries returned per page. Maximum value: 50.</p> <p>Default value: 10.</p>
ScalingGroupId	String	No	asg-bp18p2yfxow2dq****	<p>The ID of the scaling group.</p> <div style="background-color: #e1f5fe; padding: 10px;"> ? Note To call the DescribeScalingActivities operation, you must specify a value for the ScalingGroupId or ScalingActivityId.N parameter. </div>
StatusCode	String	No	Successful	<p>The status of the scaling activity. Valid values:</p> <ul style="list-style-type: none"> • Successful: The scaling activity is successful. • Warning: The scaling activity is partially successful. • Failed: The scaling activity failed. • InProgress: The scaling activity is in progress. • Rejected: The request to execute the scaling activity is rejected.

Parameter	Type	Required	Example	Description
ScalingActivityId.N	RepeatList	No	asa-bp161xudmxdzofe***	<p>The ID of scaling activity N that you want to query. Valid values of N: 1 to 20.</p> <p>Note To call the <code>DescribeScalingActivities</code> operation, you must specify a value for the <code>ScalingGroupId</code> or <code>ScalingActivityId.N</code> parameter.</p>

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	10	The number of entries returned per page.
RequestId	String	CC107349-57B7-4405-B1BF-9BF5AF7F2A46	The ID of the request.
ScalingActivities	Array of ScalingActivity		Details about the scaling activities.
ScalingActivity			
AttachedCapacity	String	0	The total number of instances that were manually added to the scaling group after the scaling activity was complete.
AutoCreatedCapacity	String	2	The total number of instances that were automatically created by Auto Scaling after the scaling activity was complete.
Cause	String	A user requests to execute scaling rule \"asr-bp12tcnol686y1ik***\", changing the Total Capacity from \"1\" to \"2\".	The cause that triggered the scaling activity.
Description	String	Add \"1\" ECS instance	The description of the scaling activity.

Parameter	Type	Example	Description
EndTime	String	2020-09-10T09:54Z	The time when the scaling activity ended.
Progress	Integer	100	The execution progress of the scaling activity.
ScalingActivityId	String	asa-bp161xudmuxdzofe****	The ID of the scaling activity.
ScalingGroupId	String	asg-bp18p2yfxow2dloq****	The ID of the scaling group.
ScalingInstanceNumber	Integer	1	<p>The number of instances that were created or restarted from the Stopped state during the scale-out activity.</p> <p>The number of instances that were deleted or stopped during the scale-in activity.</p>
StartTime	String	2020-09-10T09:54Z	The time when the scaling activity started.
StatusCode	String	Successful	<p>The status of the scaling activity. Valid values:</p> <ul style="list-style-type: none"> Successful: The scaling activity was successful. Warning: The scaling activity was partially successful. Failed: The scaling activity failed. InProgress: The scaling activity was in progress. Rejected: The request to execute the scaling activity was rejected.
StatusMessage	String	\"1\" ECS instances are added	The status information about the scaling activity.
TotalCapacity	String	2	The total number of instances in the scaling group after the scaling activity was complete.
TotalCount	Integer	1	The total number of scaling activities.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=DescribeScalingActivities  
&RegionId=cn-qingdao  
&ScalingGroupId=asg-bp18p2yfxow2dloq****  
&StatusCode=Successful  
&ScalingActivityId.1=asa-bp161xudmxdzofe****  
&<Common request parameters>
```

Sample success responses

XML format

```
<DescribeScalingActivitiesResponse>  
  <TotalCount>1</TotalCount>  
  <PageSize>10</PageSize>  
  <RequestId>CC107349-57B7-4405-B1BF-9BF5AF7F2A46</RequestId>  
  <PageNumber>1</PageNumber>  
  <ScalingActivities>  
    <ScalingActivity>  
      <ScalingInstanceNumber>1</ScalingInstanceNumber>  
      <Progress>100</Progress>  
      <Description>Add "1" ECS instance</Description>  
      <EndTime>2020-09-10T09:54Z</EndTime>  
      <AttachedCapacity>0</AttachedCapacity>  
      <ScalingActivityId>asa-bp161xudmxdzofe****</ScalingActivityId>  
      <ScalingGroupId>asg-bp18p2yfxow2dloq****</ScalingGroupId>  
      <StartTime>2020-09-10T09:54Z</StartTime>  
      <StatusCode>Successful</StatusCode>  
      <AutoCreatedCapacity>2</AutoCreatedCapacity>  
      <StatusMessage>"1" ECS instances are added</StatusMessage>  
      <Cause>A user requests to execute scaling rule "asr-bp12tcnol686ylik****", changing the Total Capacity from "1" to "2".</Cause>  
      <TotalCapacity>2</TotalCapacity>  
    </ScalingActivity>  
  </ScalingActivities>  
</DescribeScalingActivitiesResponse>
```

JSON format

```
{  
    "TotalCount": 1,  
    "PageSize": 10,  
    "RequestId": "CC107349-57B7-4405-B1BF-9BF5AF7F2A46",  
    "PageNumber": 1,  
    "ScalingActivities": {  
        "ScalingActivity": [  
            {  
                "ScalingInstanceNumber": 1,  
                "Progress": 100,  
                "Description": "Add \"1\" ECS instance",  
                "EndTime": "2020-09-10T09:54Z",  
                "AttachedCapacity": 0,  
                "ScalingActivityId": "asa-bp161xudmuxdzofe****",  
                "ScalingGroupId": "asg-bp18p2yfxow2dloq****",  
                "StartTime": "2020-09-10T09:54Z",  
                "StatusCode": "Successful",  
                "AutoCreatedCapacity": 2,  
                "StatusMessage": "\"1\" ECS instances are added",  
                "Cause": "A user requests to execute scaling rule \"asr-bp12tcnol686ylik***  
*\", changing the Total Capacity from \"1\" to \"2\".",  
                "TotalCapacity": 2  
            }  
        ]  
    }  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

9.9. DescribeScalingActivityDetail

Queries the details about a scaling activity.

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
Action	String	Yes	DescribeScalingActivityDetail	The operation that you want to perform. Set the value to DescribeScalingActivityDetail.

Parameter	Type	Required	Example	Description
ScalingActivityId	String	Yes	asa-bp1c9djwrgxjyk31****	The ID of the scaling activity that you want to query. You can call the DescribeScalingActivities operation to query the IDs of scaling activities.

Response parameters

Parameter	Type	Example	Description
RequestId	String	B13527BF-1FBD-4334-A512-20F5E9D3FB4D	The ID of the request.
ScalingActivityId	String	asa-bp1c9djwrgxjyk31****	The ID of the scaling activity.
Detail	String	new ECS instances "i-bp16t2cgmiyimeqv****" are created.	The details about the scaling activity.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=DescribeScalingActivityDetail
&ScalingActivityId=asa-bp1c9djwrgxjyk31****
&<Common request parameters>
```

Sample success responses

XML format

```
<DescribeScalingActivityDetailResponse>
    <RequestId>B13527BF-1FBD-4334-A512-20F5E9D3FB4D</RequestId>
    <ScalingActivityId>asa-bp1c9djwrgxjyk31****</ScalingActivityId>
    <Detail>new ECS instances "i-bp16t2cgmiyimeqv****" are created.</Detail>
</DescribeScalingActivityDetailResponse>
```

JSON format

```
{
    "RequestId": "B13527BF-1FBD-4334-A512-20F5E9D3FB4D",
    "ScalingActivityId": "asa-bp1c9djwrgxjyk31****",
    "Detail": "new ECS instances \"i-bp16t2cgmiyimeqv****\" are created.\n"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HttpCode	Error code	Error message	Required
400	InvalidParameter	The input parameter \"ScalingActivityId\" that is mandatory for processing this request is not supplied	The error message returned because the ID of the scaling activity is not provided.

9.10. AttachLoadBalancers

You can call this operation to associate one or more SLB instances with a scaling group.

Description

Before you associate an SLB instance with a scaling group, make sure that the following requirements are met:

- 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 be configured with at least one listener. Health check is enabled for the SLB instance.
- 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 backend server contains VPC-type instances, the instance and the scaling group must be in the same VPC.
- The number of SLB instances cannot exceed the quota for SLB instances that can be associated with the scaling group. For more information about the quota, see [Limits](#).

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
Action	String	Yes	AttachLoadBalancers	The operation that you want to perform. Set the value to AttachLoadBalancers.

Parameter	Type	Required	Example	Description
LoadBalancer.N	RepeatList	Yes	lb-2zeur05gfs****	The ID of SLB instance N. Valid values of N: 1 to 5.
ScalingGroupId	String	Yes	asg-bp1avr6ensitts3w****	The ID of the scaling group.
ForceAttach	Boolean	No	false	<p>Specifies whether to add all instances in the specified scaling group as backend servers to the SLB instance. Valid values:</p> <ul style="list-style-type: none"> • true: adds all instances • false: does not add all instances <p>Default value: false.</p>
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .

Response parameters

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

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=AttachLoadBalancers
&ScalingGroupId=asg-bp1avr6ensitts3w****
&LoadBalancer.1=lb-2zeur05gfs****
&<Common request parameters>
```

Sample success responses

XML format

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

JSON format

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

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
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	QuotaExceeded.LoadBalancer	LoadBalancer quota exceeded in the scaling group "%s".	The error message returned because the maximum number of SLB instances associated with the scaling group has been reached.
404	InvalidLoadBalancerId.NotFound	The load balancer "%s" does not exist.	The error message returned because 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 because the SLB instance and the scaling group are not in the same region.

HTTP status code	Error code	Error message	Description
400	IncorrectLoadBalancerStatus	The current status of the load balancer "%s" does not support this action.	The error message returned because the operation is not supported while the SLB instance is in the current state.
400	IncorrectLoadBalancerHealthCheck	The current health check type of the load balancer "%s" does not support this action.	The error message returned because 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 because 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 because the maximum number of backend servers of the SLB instance has been reached.
404	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	The error message returned because the specified scaling configuration that is enabled for the specified scaling group does not exist.

9.11. DetachLoadBalancers

Disassociates one or more Server Load Balancer (SLB) instances from a 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
Action	String	Yes	DetachLoadBalancers	The operation that you want to perform. Set the value to DetachLoadBalancers .
ScalingGroupId	String	Yes	asg-bp1ffogfdauy0jw0****	The ID of the scaling group.
ForceDetach	Boolean	No	false	<p>Specifies whether to remove Elastic Compute Service (ECS) instances in the scaling group from the backend server groups of the SLB instance. Valid values:</p> <ul style="list-style-type: none">• true• false <p>Default value: false.</p>
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must make sure that it is unique among different requests.</p> <p>The token can contain only ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence.</p>

Parameter	Type	Required	Example	Description
Async	Boolean	No	false	<p>Specifies whether to perform asynchronous calls when you associate an SLB instance with the scaling group. Asynchronous calls ensure the transactional nature of operations. This means that the execution results take effect only if all operations are successfully performed. We recommend that you perform asynchronous calls.</p> <p>Valid values:</p> <ul style="list-style-type: none"> • true: performs asynchronous calls. The request returns the ID of the scaling activity. • false: does not perform asynchronous calls. <p>Default value: false.</p>
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.
LoadBalancer.N	String	Yes	lb-2zeur05gfs****	The ID of SLB instance N. You can disassociate up to five SLB instances from a scaling group at a time.

Response parameters

Parameter	Type	Example	Description
ScalingActivityId	String	asa-bp140qd7mak8k63f****	<p>The ID of the scaling activity.</p> <p>A value of this parameter is returned only when Async is set to true. You can call the DescribeScalingActivities operation to query the returned IDs of the scaling activities and view the execution status of the scaling activities.</p>
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3C83E	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DetachLoadBalancers
&ScalingGroupId=asg-bp1ffogfdauy0jw0****
&ForceDetach=false
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&Async=false
&LoadBalancer=["lb-2zeur05gfs****"]
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DetachLoadBalancersResponse>
  <ScalingActivityId>asa-bp140qd7mak8k63f****</ScalingActivityId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DetachLoadBalancersResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "ScalingActivityId" : "asa-bp140qd7mak8k63f****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the operation.

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Alibaba Cloud account.
404	InvalidLoadBalancerId.NotFound	The Load Balancer "%s" does not exist.	The error message returned because the specified SLB instance is not associated with the scaling group.

9.12. DetachDBInstances

Disassociates one or more ApsaraDB RDS instances from a 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
Action	String	Yes	DetachDBInstances	The operation that you want to perform. Set the value to DetachDBInstances .
ScalingGroupId	String	Yes	asg-bp1igpak5ft1flyp****	The ID of the scaling group.
ForceDetach	Boolean	No	false	Specifies whether to remove the private IP addresses of instances in the scaling group from the whitelist that manages access to the ApsaraDB RDS instance with which the scaling group is associated. Valid values: <ul style="list-style-type: none">• true• false Default value: false.

Parameter	Type	Required	Example	Description
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.
DBInstance.N	String	Yes	rm-bp12cy3****	The ID of ApsaraDB RDS instance N. You can disassociate up to five ApsaraDB RDS instances from a scaling group at a time.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DetachDBInstances
&ScalingGroupId=asg-bpligpak5ft1flyp****
&ForceDetach=false
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&DBInstance=["rm-bp12cy3****"]
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DetachDBInstancesResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DetachDBInstancesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [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 scaling group does not exist within the Alibaba Cloud account.
400	InvalidDBInstanceId.NotFound	DB instance "%s" does not exist.	The error message returned because the ApsaraDB RDS instance does not exist.

9.13. SuspendProcesses

Suspends a process in a 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
Action	String	Yes	SuspendProcesses	The operation that you want to perform. Set the value to SuspendProcesses .
ScalingGroupId	String	Yes	asg-bp15oubotmrq11xe***	The ID of the scaling group.

Parameter	Type	Required	Example	Description
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests.</p> <p>The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence.</p>
RegionId	String	No	cn-hangzhou	The ID of the region where the scaling group resides.
Process.N	String	Yes	ScaleIn	<p>Process N that you want to suspend. Valid values of N: 1 to 100. Valid values:</p> <ul style="list-style-type: none"> • ScaleIn • ScaleOut • HealthCheck • AlarmNotification • ScheduledAction <p>You can suspend five types of processes. If you specify more than five types, the duplicate types are automatically removed.</p>

Response parameters

Parameter	Type	Example	Description
RequestId	String	3E2033F0-03B4-419D-BCE2-C2339DB51B30	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=SuspendProcesses
&ScalingGroupId=asg-bp15oubotmrql1xe****
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&Process=["ScaleIn"]
&RegionId=cn-hangzhou
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<SuspendProcessesResponse>
  <RequestId>3E2033F0-03B4-419D-BCE2-C2339DB51B30</RequestId>
</SuspendProcessesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "3E2033F0-03B4-419D-BCE2-C2339DB51B30"
}
```

Error codes

For a list of error codes, visit the [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 within the current Alibaba Cloud account.
400	InvalidParameter	The specified value of parameter "%s" is not valid.	The error message returned because the value that you specified for a parameter is invalid.

9.14. ResumeProcesses

Resumes the suspended processes in a 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
Action	String	Yes	ResumeProcesses	The operation that you want to perform. Set the value to ResumeProcesses .
ScalingGroupId	String	Yes	asg-bp15oubotmrq11xe***	The ID of the scaling group.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.
Process.N	String	Yes	ScaleIn	<p>The type of process N that you want to resume. Valid values of N: 1 to 100.</p> <p>Valid values:</p> <ul style="list-style-type: none"> • ScaleIn • ScaleOut • HealthCheck • AlarmNotification • ScheduledAction <p>You can resume the preceding types of scaling processes. If you specify more than five types, the system removes the duplicate types.</p>

Response parameters

Parameter	Type	Example	Description
RequestId	String	E38EB733-D714-4658-8A5F-0688AB680414	The ID of the request.

Examples

Sample request

```
http(s)://ess.aliyuncs.com/?Action=ResumeProcesses
&ScalingGroupId=asg-bp15oubotmrq11xe****
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&Process=["ScaleIn"]
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<ResumeProcessesResponse>
  <RequestId>E38EB733-D714-4658-8A5F-0688AB680414</RequestId>
</ResumeProcessesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "E38EB733-D714-4658-8A5F-0688AB680414"
}
```

Error codes

For a list of error codes, visit the [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 within the current account.
400	InvalidParameter	The specified value of parameter "%s" is not valid.	The error message returned because the specified parameter is invalid.

9.15. AttachDBInstances

Associates one or more ApsaraDB RDS instances with a scaling group.

Description

Before you associate an ApsaraDB RDS instance with a scaling group, make sure that the ApsaraDB RDS instance meets the following requirements:

- The ApsaraDB RDS instance and the scaling group must be within the same account.
- The ApsaraDB RDS instance must be unlocked. For more information about the lock policy, see [Limits](#).
- The ApsaraDB RDS instance must be in the Running state.

After an ApsaraDB RDS instance is associated with the scaling group, the default whitelist that manages access to the ApsaraDB RDS instance can contain no more than 1,000 IP addresses. For more information, see the "Configure IP address whitelists for the RDS instance" section in the [Use a database client or the CLI to connect to an ApsaraDB RDS for MySQL instance](#) topic.

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
Action	String	Yes	AttachDBInstances	The operation that you want to perform. Set the value to AttachDBInstances .
ScalingGroupId	String	Yes	asg-bp1avr6ensitts3w****	The ID of the scaling group.
ForceAttach	Boolean	No	false	Specifies whether to add the internal IP addresses of all instances in the specified scaling group to the whitelist that manages access to the ApsaraDB RDS instance. Valid values: <ul style="list-style-type: none">• true• false Default value: false.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .

Parameter	Type	Required	Example	Description
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.
DBInstance.N	String	Yes	rm-bp12cy3****	The ID of ApsaraDB RDS instance N. Valid values of N: 1 to 5.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=AttachDBInstances
&ScalingGroupId=asg-bplavr6ensitts3w****
&ForceAttach=false
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&DBInstance=["rm-bp12cy3****"]
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<AttachDBInstancesResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</AttachDBInstancesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [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 within the current account.
400	QuotaExceeded.RDS	"RDS" quota exceeded.	The error message returned because the maximum number of ApsaraDB RDS instances that can be associated with the scaling group has been reached.
400	InvalidDBInstanceId.NotFound	The specified value of parameter "%s" is not valid.	The error message returned because the specified ApsaraDB RDS instance does not exist.
400	IncorrectDBInstanceStatus	The current status of DB instance "%s" does not support this action.	The error message returned because the operation is not supported while the ApsaraDB RDS instance is in the current state.
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "%s".	The error message returned because the maximum number of IP addresses in the whitelist that manages access to the ApsaraDB RDS instance has been reached.

HTTP status code	Error code	Error message	Description
400	InvalidInstanceIds.PrivateIpNotFound	Can not find all private ips of instances in specific scaling group.	The error message returned because the internal IP address of the ApsaraDB RDS instance that is associated with the scaling group cannot be obtained.

9.16. AttachVServerGroups

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

Description

Before you add a VServer group to a scaling group, make sure that the following requirements are met:

- 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 be configured with at least one listener. Health check is enabled for the SLB instance.
- 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 VServer group contains VPC-type instances, the instance and the scaling group must be in the same VPC.
- The VServer group to be added to the scaling group must be associated with the SLB instance.
- The number of VServer groups cannot exceed the quota for VServer groups in the scaling group. For more information about the quota, see [Limits](#).

You must set the following parameters to specify the VServer groups to be added:

- VServerGroup.N.LoadBalancerId: the ID of the SLB instance
- VServerGroup.N.VServerGroupAttribute.N.VServerGroupId: the ID of the VServer group
- VServerGroup.N.VServerGroupAttribute.N.Port: the port number of the VServer group

If you use different ports to add a VServer group to a scaling group, Auto Scaling considers that multiple VServer groups are added to the scaling group. If multiple VServer groups with 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 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
Action	String	Yes	AttachVServerGroups	The operation that you want to perform. Set the value to AttachVServerGroups.
RegionId	String	Yes	cn-hangzhou	The region ID of the scaling group. Example: cn-hangzhou or cn-shanghai. For more information, see Regions and zones .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dl0q****	The ID of the scaling group.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .
VServerGroup.N.LoadBalancerId	String	No	lb-bp1u7etiogg38ywz****	The ID of SLB instance N with which the VServer group is associated. N specifies the number of the SLB instance. Valid values of N: 1 to 5.
VServerGroup.N.VServerGroupAttribute.N.VServerGroupupId	String	No	rsp-bp1jp1rge****	The ID of VServer group M. N specifies the number of the SLB instance. Valid values of N: 1 to 5. M specifies the number of the VServer group that is associated with SLB instance N. Valid values of M: 1 to 5.
VServerGroup.N.VServerGroupAttribute.N.Port	Integer	No	22	The port number that is used by Auto Scaling to add the ECS instances to VServer group M. Valid values: 1 to 65535. N specifies the number of the SLB instance. Valid values of N: 1 to 5. M specifies the number of the VServer group that is associated with SLB instance N. Valid values of M: 1 to 5.

Parameter	Type	Required	Example	Description
VServerGroup.N.VServerGroupAttribute.N.Weight	Integer	No	100	<p>The weight set for the ECS instances that are added to VServer group M. Valid values: 0 to 100.</p> <p>N specifies the number of the SLB instance. Valid values of N: 1 to 5.</p> <p>M specifies the number of the VServer group that is associated with SLB instance N. Valid values of M: 1 to 5.</p> <p>Default value: 50.</p>
ForceAttach	Boolean	No	false	<p>Specifies whether to add the ECS instances in the specified scaling group to the new VServer group.</p> <ul style="list-style-type: none"> • true: adds the ECS instances to the VServer group. • false: does not add the ECS instances to the VServer group. <p>Default value: false.</p>

Response parameters

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

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=AttachVServerGroups
&ScalingGroupId=asg-****
&RegionId=cn-hangzhou
&VServerGroup.1.LoadBalancerId=lb-bp1u7etiogg38yvwz****
&VServerGroup.1.VServerGroupAttribute.1.VServerGroupId=rsp-bp1jp1rge****
&VServerGroup.1.VServerGroupAttribute.1.Port=22
&VServerGroup.1.VServerGroupAttribute.1.Weight=100
&<Common request parameters>
```

Sample success responses

XML format

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

JSON format

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

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
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.
404	InvalidLoadBalancerId.NotFound	The load balancer "%s" does not exist.	The error message returned because 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 because 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 because the operation is not supported while the SLB instance is in the current state.

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 because 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 because 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 because the VServer group that is specified by the VServerGroup.N.VServerGroupAttribute.N.VServerGroupId parameter does not belong to the specified SLB instance.
400	QuotaExceeded.VServerGroup	VServerGroup quota exceeded in the specified scaling group.	The error message returned because the maximum number of VServer groups that can be added to the scaling group has been reached.

9.17. DetachVServerGroups

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

Description

You must set the following parameters to specify the VServer groups to be removed:

- VServerGroup.N.LoadBalancerId: the ID of the SLB instance
- VServerGroup.N.VServerGroupAttribute.N.VServerGroupId: the ID of the VServer group
- VServerGroup.N.VServerGroupAttribute.N.Port: the port number of the VServer group

If the VServer groups specified in the request parameters match VServer groups in the scaling group, the matched VServer groups are removed. If no VServer groups specified in the request parameters match VServer groups in the scaling group, the request is ignored and no error is reported.

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
Action	String	Yes	DetachVServerGroups	The operation that you want to perform. Set the value to DetachVServerGroups.
RegionId	String	Yes	cn-hangzhou	The region ID of the scaling group. Example: cn-hangzhou or cn-shanghai. For more information, see Regions and zones .
ScalingGroupId	String	Yes	asg-bp1fo0dbtsbmqa9h***	The ID of the scaling group.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544***	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .
VServerGroup.N.LoadBalancerId	String	No	lb-bp1p90y3ya9h8s62d***	The ID of SLB instance N with which the VServer group is associated. N specifies the number of the SLB instance. Valid values of N: 1 to 5.
VServerGroup.N.VServerGroupAttribute.N.VServerGroupupId	String	No	rsp-bp1jp1rge***	The ID of VServer group M. N specifies the number of the SLB instance. Valid values of N: 1 to 5. M specifies the number of the VServer group that is associated with SLB instance N. Valid values of M: 1 to 5.

Parameter	Type	Required	Example	Description
VServerGroup.N.VServerGroupAttribute.N.Port	Integer	No	22	<p>The port number that is used by Auto Scaling to add the ECS instances to VServer group M. Valid values: 1 to 65535.</p> <p>N specifies the number of the SLB instance. Valid values of N: 1 to 5.</p> <p>M specifies the number of the VServer group that is associated with the SLB instance. Valid values of M: 1 to 5.</p>
ForceDetach	Boolean	No	false	<p>Specifies whether to remove the ECS instances in the specified scaling group from the VServer group to be removed.</p> <ul style="list-style-type: none"> • true: removes the ECS instances. • false: does not remove the ECS instances. <p>Default value: false.</p>

Response parameters

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

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=DetachVServerGroups
&ScalingGroupId=asg-bp1fo0dbtsbmqa9h****
&RegionId=cn-hangzhou
&VServerGroup.1.LoadBalancerId=lb-bp1p90y3ya9h8s62d****
&VServerGroup.1.VServerGroupAttribute.1.VServerGroupId=rsp-bp1jplrg****
&VServerGroup.1.VServerGroupAttribute.1.Port=22
&VServerGroup.1.VServerGroupAttribute.2.VServerGroupId=rsp-bp1lczye5****
&VServerGroup.1.VServerGroupAttribute.2.Port=80
&<Common request parameters>
```

Sample success responses

XML format

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

JSON format

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

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
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	InvalidParameter	The specified value of parameter "%s" is not valid.	The error message returned because a specified parameter is invalid.
400	MissingParameter	The input parameter "%s" that is mandatory for processing this request is not supplied.	The error message returned because a required parameter is not specified.

9.18. AttachAlbServerGroups

Associates one or more Application Load Balancer (ALB) server groups with a scaling group.

Description

Before you associate an ALB server group with a scaling group, make sure that the following requirements are met:

- The scaling group resides in a virtual private cloud (VPC). The scaling group and the ALB server group

that you want to associate must reside in the same VPC.

- The ALB server group is in the Available state.
- You can associate only a limited number of ALB server groups with a scaling group. To view the available quota or manually request a quota increase, go to the [Quota Center](#).

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
Action	String	Yes	AttachAlbServerGroups	The operation that you want to perform. Set the value to AttachAlbServerGroups .
RegionId	String	Yes	cn-hangzhou	The ID of the region to which the scaling group belongs, such as cn-hangzhou and cn-shanghai. For more information, see Regions and zones .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dloq****	The ID of the scaling group.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .

Parameter	Type	Required	Example	Description
ForceAttach	Boolean	No	false	<p>Specifies whether to add the Elastic Compute Service (ECS) instances in the scaling group to the newly associated ALB server group.</p> <ul style="list-style-type: none"> • true: adds the ECS instances and returns the <code>ScalingActivityId</code> value. You can view the ID of the scaling activity to determine whether the existing instances are added. • false: does not add the ECS instances. <p>Default value: false.</p>
AlbServerGroup.N. <code>.AlbServerGroupId</code>	String	Yes	sgp-ddwb0y0g6y9bjm****	<p>The ID of the ALB server group.</p> <p>N indicates the serial number of the ALB server group. You can associate only a limited number of ALB server groups with a scaling group. To view the available quota or manually request a quota increase, go to the Quota Center.</p>
AlbServerGroup.N. <code>.Weight</code>	Integer	Yes	100	<p>The weight of the ECS instance as a backend server after the instance is added to the ALB server group. If you increase the weight of an ECS instance, the number of access requests that are forwarded to the ECS instance increases. If you set the weight to 0, no access requests are forwarded to the ECS instance. Valid values: 0 to 100.</p> <p>N indicates the serial number of the ALB server group.</p>

Parameter	Type	Required	Example	Description
AlbServerGroup.N.Port	Integer	Yes	22	<p>The port number used by the ECS instance after the instance is added to ALB server group N. Valid values: 1 to 65535.</p> <p>N indicates the serial number of the ALB server group.</p> <p>Note If the N values are the same but the port numbers are different, the system associates multiple ports in the scaling group with ALB server group N.</p>

Response parameters

Parameter	Type	Example	Description
ScalingActivityId	String	asa-2ze6wxj8vsohn6j9**	<p>The ID of the scaling activity in which Auto Scaling associates the ALB server group with the scaling group and adds ECS instances in the scaling group to the ALB server group.</p> <p>This parameter is valid only when you set <code>ForceAttach</code> to <code>true</code>.</p>
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=AttachAlbServerGroups
&RegionId=cn-hangzhou
&ScalingGroupId=asg-bp18p2yfxow2dloq****
&AlbServerGroup.1.AlbServerGroupId=sgp-ddwb0y0g6y9bjm****
&AlbServerGroup.1.Port=22
&AlbServerGroup.1.Weight=100
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<AttachAlbServerGroupsResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
  <ScalingActivityId>asa-2ze6wxj8vsohn6j9****</ScalingActivityId>
</AttachAlbServerGroupsResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
  "ScalingActivityId" : "asa-2ze6wxj8vsohn6j9****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	AlbServerGroup.NotExist	The ServerGroup "%s" do(es) not exist.	The error message returned because the specified ALB server group does not exist within the account.
400	AlbServerGroup.AlreadyAttached	The ALB ServerGroups are already attached.	The error message returned because the ALB server group is already associated with the scaling group.

9.19. DetachAlbServerGroups

Disassociates one or more Application Load Balancer (ALB) server groups from a 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
Action	String	Yes	DetachAlbServerGroups	The operation that you want to perform. Set the value to DetachAlbServerGroups .
RegionId	String	Yes	cn-hangzhou	The ID of the region where the scaling group resides. Examples: cn-hangzhou and cn-shanghai. For more information, see Regions and zones .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dl0q***	The ID of the scaling group.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544***	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests.</p> <p>The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence.</p>
ForceDetach	Boolean	No	false	<p>Specifies whether to remove existing Elastic Compute Service (ECS) instances from the ALB server group that you want to disassociate from the scaling group.</p> <ul style="list-style-type: none"> • true: removes existing ECS instances and returns the <code>ScalingActivity.Id</code> value. You can view the ID of the scaling activity to determine whether the instances are removed. • false: does not remove existing ECS instances. <p>Default value: false.</p>
AlbServerGroup.N.AlbServerGroupId	String	Yes	sgp-ddwb0y0g6y9bjm***	The ID of ALB server group N. N indicates the serial number of the ALB server group.

Parameter	Type	Required	Example	Description
AlbServerGroup.N.Port	Integer	Yes	22	The port number used by the ECS instances in ALB server group N. N indicates the serial number of the ALB server group.

Response parameters

Parameter	Type	Example	Description
ScalingActivityId	String	asa-2ze6wxj8vsohn6j9**	The ID of the scaling activity in which the ALB server group is disassociated from the scaling group and the ECS instances in the ALB server group are removed from the ALB server group. This parameter is valid only when you set <code>ForceDetach</code> to <code>true</code> .
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DetachAlbServerGroups
&RegionId=cn-hangzhou
&ScalingGroupId=asg-bp18p2yfxow2dlog****
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&ForceDetach=false
&AlbServerGroup=[{"AlbServerGroupId":"sgp-ddwb0y0g6y9bjm****","Port":22}]
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DetachAlbServerGroupsResponse>
  <ScalingActivityId>asa-2ze6wxj8vsohn6j9****</ScalingActivityId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DetachAlbServerGroupsResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "ScalingActivityId" : "asa-2ze6wxj8vsohn6j9****",
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	AlbServerGroup.NotAttached	The ALB server groups are not attached to specific ScalingGroup.	The error message returned because the ALB server group is not associated with the scaling group.

10. Scaling configurations

10.1. CreateScalingConfiguration

Creates a scaling configuration.

Description

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

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

 **Note** You cannot use `InstanceType`, `InstanceTypes.N`, `InstanceTypeOverride.N`, and `Cpu and Memory` at the same time. You can use `InstanceType` and `InstancePatternInfo.N` or use `InstanceTypes.N` and `InstancePatternInfo.N` at the same time. If you use `InstanceType` and `InstancePatternInfo.N` or use `InstanceTypes.N` and `InstancePatternInfo.N` at the same time, Auto Scaling preferentially uses the instance types that are specified by `InstanceType` or `InstanceTypes.N` for scale-out activities. If the instance types that are specified by `InstanceType` or `InstanceTypes.N` do not have sufficient inventory, Auto Scaling uses the instance types that are specified by `InstancePatternInfo.N` for scale-out activities.

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
Action	String	Yes	<code>CreateScalingConfiguration</code>	The operation that you want to perform. Set the value to <code>CreateScalingConfiguration</code> .

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	asg-bp14wlu85wrpchm0****	The ID of the scaling group for which you want to create the scaling configuration.
ImageId	String	No	centos6u5_64_20G_aliaeis****.vhd	The ID of the image that is used by Auto Scaling to automatically create ECS instances.
ImageName	String	No	image****	<p>The name of the image. The name must be unique within a region. If you specify the ImageId parameter, the ImageName parameter is ignored.</p> <p>You cannot use the ImageName parameter to specify Alibaba Cloud Marketplace images.</p>
InstanceType	String	No	ecs.g6.large	<p>The instance type of the ECS instance. For more information, see Instance families.</p>
Cpu	Integer	No	2	<p>The number of vCPUs.</p> <p>You can specify the number of vCPUs and the memory size to determine the range of instance types. For example, you can set CPU to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. Auto Scaling determines a set of available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates ECS instances of the instance type that has the lowest unit price.</p> <div style="background-color: #e1f5fe; padding: 10px; margin-top: 10px;"> ? Note This instance type range is available only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration. </div>

Parameter	Type	Required	Example	Description
Memory	Integer	No	16	<p>The memory size. You can specify the number of vCPUs and the memory size to determine the range of instance types. For example, you can set CPU to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. Auto Scaling determines a set of available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates ECS instances of the instance type that has the lowest unit price.</p> <div style="background-color: #e1f5fe; padding: 10px; border-radius: 5px;"><p>? Note This instance type range is available only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.</p></div>
DeploymentSetId	String	No	ds-bp1frxuzdg87zh4pz****	The ID of the deployment set to which the ECS instances belong.
SecurityGroupId	String	No	sg-280ih****	The ID of the security group to which the ECS instances belong. The ECS instances that belong to the same security group can access each other.

Parameter	Type	Required	Example	Description
IoOptimized	String	No	optimized	<p>Specifies whether the instance that you want to create is an I/O optimized instance. Valid values:</p> <ul style="list-style-type: none"> • none: non-I/O optimized instance. • optimized: I/O optimized instance. <p>For instances of retired instance types, the default value is none. For instances of other instance types, the default value is optimized.</p>
InternetChargeType	String	No	PayByTraffic	<p>The billing method for network usage. Valid values:</p> <ul style="list-style-type: none"> • PayByBandwidth. You are charged for the maximum available bandwidth that is specified by InternetMaxBandwidthOut. • PayByTraffic. You are charged for the actual traffic that you use. InternetMaxBandwidthOut specifies only the maximum available bandwidth. <p>Default value for the classic network: PayByBandwidth. Default value for VPCs: PayByTraffic.</p>
InternetMaxBandwidthIn	Integer	No	100	<p>The maximum inbound public bandwidth. Unit: Mbit/s. Valid values: 1 to 200.</p> <p>Default value: 200. This parameter is not used for billing because inbound traffic of instances is free of charge.</p>
InternetMaxBandwidthOut	Integer	No	50	<p>The maximum outbound public bandwidth. Unit: Mbit/s.</p> <ul style="list-style-type: none"> • Valid values if you set InternetChargeType to PayByBandwidth: 0 to 100. If you leave this parameter empty, a value of 0 is used. • Valid values if you set InternetChargeType to PayByTraffic: 0 to 100. If you leave this parameter empty, an error is reported.

Parameter	Type	Required	Example	Description
SystemDisk.Category	String	No	cloud_ssd	<p>The category of the system disk. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk • cloud_efficiency: ultra disk • cloud_ssd: standard SSD • ephemeral_ssd: local standard SSD • cloud_essd: enhanced SSD (ESSD) <p>You cannot specify the SystemDisk.Category and SystemDiskCategories.N parameters at the same time. If you do not specify the SystemDisk.Category and SystemDiskCategories.N parameters, the default value of SystemDisk.Category is used. For non-I/O optimized instances of Generation I instance types, the default value is cloud. For instances of other types, the default value is cloud_efficiency.</p>
SystemDisk.Size	Integer	No	100	<p>The size of the system disk. Unit: GiB. Valid values: 20 to 500.</p> <p>The value of SystemDisk.Size must be greater than or equal to max{20, ImageSize}.</p> <p>Default value: max{40, ImageSize}.</p>
SystemDisk.DiskName	String	No	cloud_ssdSystem	<p>The name of the system disk. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with http:// or https:// . The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).</p>
SystemDisk.Description	String	No	Test system disk.	<p>The description of the system disk. The description must be 2 to 256 characters in length, and cannot start with http:// or https:// .</p>
SystemDisk.AutoSnapshotPolicyId	String	No	sp-bp12m37ccmxvbmj5***	<p>The ID of the automatic snapshot policy that you want to apply to the system disk.</p>

Parameter	Type	Required	Example	Description
SystemDisk.PerformanceLevel	String	No	PL0	<p>The performance level of the system disk that is an ESSD. Valid values:</p> <ul style="list-style-type: none"> • PL0: A single ESSD can deliver up to 10,000 random read/write IOPS. • PL1: A single ESSD can deliver up to 50,000 random read/write IOPS. • PL2: A single ESSD can deliver up to 100,000 random read/write IOPS. • PL3: A single ESSD can deliver up to 1,000,000 random read/write IOPS. <p>Default value: PL0.</p> <div style="background-color: #e1f5fe; padding: 10px; border-radius: 5px;"> ? Note For more information about how to select ESSD performance levels, see ESSDs. </div>
ScalingConfigurationName	String	No	scalingconfig***	<p>The name of the scaling configuration. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.</p> <p>The name of the scaling configuration must be unique within a scaling group in a region. If you do not specify this parameter, the value of ScalingConfigurationId is used.</p>
LoadBalancerWeight	Integer	No	50	<p>The weight of the ECS instance as a backend server of the associated Server Load Balancer (SLB) instance. Valid values: 1 to 100.</p> <p>Default value: 50.</p>

Parameter	Type	Required	Example	Description
Tags	String	No	{"key1":"value1", "key2":"value2", ... "key5":"value5"}	<p>The tags of the ECS instance. The tags must be specified in the key-value pair format. You can specify up to 20 tags. Take note of the following items when you specify tag keys and tag values:</p> <ul style="list-style-type: none">• A tag key can be up to 64 characters in length. The key cannot start with acs: or aliyun, and cannot contain http:// or https:// . You cannot specify an empty string as a tag key.• A tag value can be up to 128 characters in length. The value cannot start with acs: or aliyun, and cannot contain http:// or https:// . You can specify an empty string as a tag value.
UserData	String	No	echo hello ecs!	The user data of the ECS instance. The data must be encoded in Base64. The maximum size of the raw data before encoding is 16 KB.
KeyPairName	String	No	KeyPairTest	<p>The name of the key pair that is used to log on to an ECS instance.</p> <ul style="list-style-type: none">• Windows instances do not use this parameter.• By default, the username and password authentication method is disabled for Linux instances.
RamRoleName	String	No	ramrole***	The name of the RAM role that is associated with the ECS instance. The name is provided and maintained by Resource Access Management (RAM). You can call the ListRoles operation to query the available RAM roles. For information about how to create a RAM role, see CreateRole .

Parameter	Type	Required	Example	Description
SecurityEnhancementStrategy	String	No	Active	<p>Specifies whether to enable security hardening. Valid values:</p> <ul style="list-style-type: none"> • Active: enables security hardening. This value is available only for public images. • Deactive: disables security hardening. This value is available for all image types.
InstanceName	String	No	instance****	The name of the ECS instance that is automatically created based on the scaling configuration.
HostName	String	No	host****	<p>The hostname of the ECS instance. The hostname cannot start or end with a period (.) or a hyphen (-). The hostname cannot contain consecutive periods (.) or hyphens (-). Naming conventions for different types of instances:</p> <ul style="list-style-type: none"> • Windows instances: The hostname must be 2 to 15 characters in length, and can contain letters, digits, and hyphens (-). The hostname cannot contain periods (.). It cannot contain only digits. • Other instances such as Linux instances: The hostname must be 2 to 64 characters in length. You can use periods (.) to separate a hostname into multiple segments. Each segment can contain letters, digits, and hyphens (-).

Parameter	Type	Required	Example	Description
SpotStrategy	String	No	NoSpot	<p>The preemption policy that you want to apply to pay-as-you-go instances and preemptible instances. Valid values:</p> <ul style="list-style-type: none">• NoSpot: This policy applies to regular pay-as-you-go instances.• SpotWithPriceLimit: This policy applies to preemptible instances for which a maximum hourly price is specified.• SpotAsPriceGo: This policy applies to preemptible instances for which the market price at the time of purchase is used as the bid price. <p>Default value: NoSpot.</p>
PasswordInherit	Boolean	No	false	<p>Specifies whether to use the password preset in the image. Before you use this parameter, make sure that a password is configured for the image. Valid values:</p> <ul style="list-style-type: none">• true• false

Parameter	Type	Required	Example	Description
				<p>The password that is used to log on to the ECS instance. The password must be 8 to 30 characters in length, and contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The following special characters are supported:</p> <pre>() ` ~ ! @ # \$ % ^ & * - _ + = \ { } [] : ; ' < > , . ? /</pre>
Password	String	No	123abc****	<p>The password of a Windows instance cannot start with a forward slash (/).</p> <div style="background-color: #e1f5fe; padding: 10px;"> <p>? Note If you specify the Password parameter, we recommend that you use HTTPS to send requests for security purposes.</p> </div>
ResourceGroupId	String	No	rg-resource****	The ID of the resource group to which the ECS instance belongs.
HpcClusterId	String	No	hpc-clusterid	The ID of the Elastic High Performance Computing (E-HPC) cluster to which the ECS instance belongs.
InstanceDescription	String	No	Test instance.	The description of the ECS instance. The description must be 2 to 256 characters in length, and cannot start with http:// or https:// .

Parameter	Type	Required	Example	Description
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can contain only ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence .
Ipv6AddressCount	Integer	No	1	The number of randomly generated IPv6 addresses that you want to allocate to the elastic network interface (ENI).
CreditSpecification	String	No	Standard	<p>The performance mode of the burstable instance. Valid values:</p> <ul style="list-style-type: none">• Standard: standard mode• Unlimited: unlimited mode <p>For more information, see the "Performance modes" section in Overview.</p>
ImageFamily	String	No	hangzhou-daily-update	The name of the image family. You can specify this parameter to obtain the latest custom images that are available in the specified image family. The images are used to create ECS instances. If you specify the ImageId parameter, you cannot specify the ImageFamily parameter.
ZoneId	String	No	cn-hangzhou-g	The ID of the zone where the ECS instance resides.

Parameter	Type	Required	Example	Description
DedicatedHostId	String	No	dh-bp67acfmxazb4p****	<p>The ID of the dedicated host on which you want to create the ECS instance. You cannot create preemptible instances on dedicated hosts. If you specify the DedicatedHostId parameter, the SpotStrategy and SpotPriceLimit parameters are ignored.</p> <p>You can call the DescribeDedicatedHosts operation to query the IDs of dedicated hosts.</p>
Affinity	String	No	default	<p>Specifies whether to associate the instance on a dedicated host with the dedicated host. Valid values:</p> <ul style="list-style-type: none"> • default: does not associate the instance with the dedicated host. If you start an instance that was in economical mode when the original dedicated host of the instance has insufficient resources, the instance is deployed to another dedicated host in the automatic deployment resource pool. • host: associates the instance with the dedicated host. If you start an instance that was in economical mode, the instance remains on the original dedicated host. If the original dedicated host has insufficient resources, the instance fails to start. <p>Default value: default.</p>
Tenancy	String	No	default	<p>Specifies whether to create the instance on a dedicated host. Valid values:</p> <ul style="list-style-type: none"> • default: creates the instance on a non-dedicated host. • host: creates the instance on a dedicated host. If you do not specify the DedicatedHostId parameter, the system selects a dedicated host for the instance that you created. <p>Default value: default.</p>

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

Parameter	Type	Required	Example	Description
SpotInterruptionBehavior	String	No	Terminate	The interruption mode of the preemptible instance. Default value: Terminate. Set the value to Terminate, which specifies that the instance is released.
SystemDisk.Encrypted	Boolean	No	false	Specifies whether to encrypt the system disk. Valid values: <ul style="list-style-type: none">• true• false Default value: false.
SystemDisk.KMSKeyId	String	No	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the Key Management Service (KMS) key that is used to encrypt the system disk.
SystemDisk.EncryptAlgorithm	String	No	AES-256	The algorithm that is used to encrypt the system disk. Valid values: <ul style="list-style-type: none">• AES-256.• SM4-128. Default value: AES-256.
InstanceTypes.N	String	No	ecs.g6.large	Instance type N of ECS instances. If you specify this parameter, the InstanceType parameter is ignored. You can specify up to 10 instance types for a scaling configuration. Valid values of N: 1 to 10. A smaller value of N specifies a higher priority. Auto Scaling creates instances based on the priorities of the instance types. If Auto Scaling cannot create instances of the instance type that has the highest priority, it creates instances of the instance type that has the next highest priority.

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

Parameter	Type	Required	Example	Description
InstanceTypeOverride.N.WeightedCapacity	Integer	No	4	<p>If you want to specify the capacity of instance types in the scaling configuration, you must specify the InstanceTypeOverride.N.InstanceType and InstanceTypeOverride.N.WeightedCapacity parameters. The two parameters have a one-to-one correspondence. The value of N in the two parameters must be the same.</p> <p>This parameter specifies the weight of the instance type. The weight specifies the capacity of a single instance of the specified instance type in the scaling group. A higher weight specifies that a smaller number of instances of the specified instance type are required to meet the expected capacity.</p> <p>Performance metrics such as the number of vCPUs and the memory size of each instance type may vary. You can specify different weights for different instance types based on your business requirements.</p> <p>Example:</p> <ul style="list-style-type: none"> • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 <p>To meet the expected capacity, Auto Scaling creates two ecs.c5.xlarge instances.</p> <div style="background-color: #e0f2ff; padding: 10px;"> <p>? Note The capacity of the scaling group cannot exceed the sum of the maximum capacity that is specified by MaxSize and the maximum weight of the instance type.</p> </div> <p>Valid values of WeightedCapacity in InstanceTypeOverride.N.WeightedCapacity: 1 to 500.</p>

Parameter	Type	Required	Example	Description
DataDisk.N.PerformanceLevel	String	No	PL1	<p>The performance level of data disk N that is an ESSD. If you set the DataDisk.N.Category parameter to cloud_essd, the N value must be the same as the N value in DataDisk.N.Category. Valid values:</p> <ul style="list-style-type: none">• PL0: A single ESSD can deliver up to 10,000 random read/write IOPS.• PL1: A single ESSD can deliver up to 50,000 random read/write IOPS.• PL2: A single ESSD can deliver up to 100,000 random read/write IOPS.• PL3: A single ESSD can deliver up to 1,000,000 random read/write IOPS. <p>Default value: PL1.</p> <div style="background-color: #e0f2ff; padding: 10px; border-radius: 5px;"><p>? Note For more information about how to select ESSD performance levels, see ESSDs.</p></div>
DataDisk.N.AutoSnapshotPolicyId	String	No	sp-bp19nq9enxqkomib***	The ID of the automatic snapshot policy that you want to apply to data disk N. Valid values of N: 1 to 16.
DataDisk.N.Encrypted	String	No	false	<p>Specifies whether to encrypt data disk N. Valid values of N: 1 to 16. Valid values:</p> <ul style="list-style-type: none">• true• false <p>Default value: false.</p>
DataDisk.N.Description	String	No	Test data disk.	The description of data disk N. Valid values of N: 1 to 16. The description must be 2 to 256 characters in length, and cannot start with http:// or https:// .

Parameter	Type	Required	Example	Description
DataDisk.N.SnapshotId	String	No	s-280s7****	<p>The ID of the snapshot that you want to use to create data disk N. Valid values of N: 1 to 16. If you specify this parameter, the DataDisk.N.Size parameter is ignored. The size of the disk is the same as the size of the specified snapshot.</p> <p>If you specify a snapshot that is created on or before July 15, 2013, the operation fails and the system returns InvalidSnapshot.TooOld.</p>
DataDisk.N.Categories.N	String	No	cloud	<p>Category N of data disk N. The first N in DataDisk.N.Categories.N represents the serial number of the data disk. Valid values of N: 1 to 16. The second N in DataDisk.N.Categories.N represents the serial number of the data disk category. Valid values of N: 1 to 4. You can specify up to four data disk categories in a scaling configuration. A smaller value of N specifies a higher priority. If Auto Scaling cannot create data disks of the disk category that has the highest priority, it creates disks of the disk category that has the next highest priority. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: SSD. • cloud_essd: ESSD. <div style="background-color: #e0f2ff; padding: 10px;"> ? Note You cannot specify the DataDisk.N.Categories.N and DataDisk.N.Category parameters at the same time. </div>

Parameter	Type	Required	Example	Description
DataDisk.N.Size	Integer	No	100	<p>The size of data disk N. Unit: GiB. Valid values of N: 1 to 16.</p> <ul style="list-style-type: none">• Valid values if you set DataDisk.N.Category to cloud: 5 to 2000.• Valid values if you set DataDisk.N.Category to cloud_efficiency: 20 to 32768.• Valid values if you set DataDisk.N.Category to cloud_ssd: 20 to 32768.• Valid values if you set DataDisk.N.Category to cloud_essd: 20 to 32768 .• Valid values if you set DataDisk.N.Category to ephemeral_ssd: 5 to 800. <p>The value of DataDisk.N.Size must be greater than or equal to the size of the snapshot that is specified by SnapshotId.</p>
DataDisk.N.Device	String	No	/dev/xvdb	The mount target of data disk N. Valid values of N: 1 to 16. If you do not specify this parameter, the system automatically assigns a mount target when you create the ECS instance. The name of the mount target ranges from /dev/xvdb to /dev/xvdz.
DataDisk.N.DiskName	String	No	cloud_ssdData	The name of data disk N. Valid values of N: 1 to 16. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with http:// or https:// . The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).

Parameter	Type	Required	Example	Description
DataDisk.N.Category	String	No	cloud_ssd	<p>The category of data disk N. Valid values of N: 1 to 16. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • ephemeral_ssd: local standard SSD. • cloud_essd: ESSD. <p>You cannot specify the DataDisk.N.Category and DataDisk.N.Categories.N parameters at the same time. If you do not specify the DataDisk.N.Category and DataDisk.N.Categories.N parameters, the default value of DataDisk.N.Category is:</p> <ul style="list-style-type: none"> • cloud_efficiency for I/O optimized instances. • cloud for non-I/O optimized instances.
DataDisk.N.DeleteWithInstance	Boolean	No	true	<p>Specifies whether to release data disk N if the instance to which data disk N is attached is released. Valid values of N: 1 to 16. Valid values:</p> <ul style="list-style-type: none"> • true: If the instance is released, the data disk is released. • false: If the instance is released, the data disk is retained. <p>This parameter is available only for independent disks whose DataDisk.N.Category parameter is set to cloud, cloud_efficiency, cloud_ssd, or cloud_essd. If you specify this parameter for other disks, an error is reported.</p> <p>Default value: true.</p>
DataDisk.N.KMSKeyId	String	No	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the KMS key of data disk N. Valid values of N: 1 to 16.

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 is available only if you set SpotStrategy to SpotWithPriceLimit.
SpotPriceLimit.N.InstanceType	String	No	ecs.g6.large	The instance type of preemptible instance N. Valid values of N: 1 to 10. This parameter is available only if you set SpotStrategy to SpotWithPriceLimit.
SecurityGroupIds.N	String	No	sg-bp18kz60mefs*** *	The ID of security group N with which you want to associate the ECS instance. The valid values of N vary based on the maximum number of security groups with which an instance can be associated. For more information, see the "Security group limits" section in Limits .

 **Note** You cannot specify the SecurityGroupId and SecurityGroupIds.N parameters at the same time.

Parameter	Type	Required	Example	Description
InstancePatternInfo.N.Cores	Integer	No	2	<p>The number of vCPUs that are specified for an instance type in intelligent configuration mode. This parameter is used to filter the available instance types that meet the specified criteria. For more information, see Instance families. Valid values of N: 1 to 10.</p> <p>Take note of the following items:</p> <ul style="list-style-type: none">• The InstancePatternInfo.N parameter is available only for scaling groups that reside in VPCs.• You must specify the InstancePatternInfo.N, InstancePatternInfo.N.Cores, and InstancePatternInfo.N.Memory parameters at the same time.• If you specify an instance type in the InstanceType or InstanceTypes.N parameter, Auto Scaling preferentially uses the instance type that is specified by the InstanceType or InstanceTypes.N parameter for scale-out activities. If insufficient instances of the specified instance type are available, Auto Scaling uses the instance types that are specified by the InstancePatternInfo.N parameter based on the unit price in ascending order.

Parameter	Type	Required	Example	Description
InstancePatternIn fo.N.InstanceFami lyLevel	String	No	EnterpriseLevel	<p>The instance family level in intelligent configuration mode. This parameter is used to filter the available instance types that meet the specified criteria. Valid values:</p> <ul style="list-style-type: none">• EntryLevel: shared instance type. Instances of this level are more cost-effective, but do not provide stable computing performance. Instances of this level are suitable for business scenarios in which the CPU utilization is low. For more information, see Shared instance families.• EnterpriseLevel: Instances of this level provide stable performance and dedicated resources, and are suitable for business scenarios that require high stability. For more information, see Instance families.• CreditEntryLevel: This value is available only for burstable instances. CPU credits are used to ensure computing performance. Instances of this level are suitable for scenarios in which the CPU utilization is low but may fluctuate in specific cases. For more information, see What are burstable instances? <p>Valid values of N: 1 to 10.</p>

Parameter	Type	Required	Example	Description
InstancePatternInfo.N.MaxPrice	Float	No	2	<p>The maximum hourly price for a pay-as-you-go instance or a preemptible instance in intelligent configuration mode. This parameter is used to filter the available instance types that meet the specified criteria. Valid values of N: 1 to 10.</p> <div style="background-color: #e0f2ff; padding: 10px; border-radius: 5px;"><p>? Note If you set SpotStrategy to SpotWithPriceLimit, you must specify this parameter. Otherwise, you do not need to specify this parameter.</p></div>
InstancePatternInfo.N.Memory	Float	No	4	<p>The memory size that is specified for an instance type in intelligent configuration mode. Unit: GiB. This parameter is used to filter the available instance types that meet the specified criteria. For more information, see Instance families. Valid values of N: 1 to 10.</p>

Parameter	Type	Required	Example	Description
SystemDiskCategories.N	String	No	cloud	<p>Category N of the system disk. Valid values of N: 1 to 4. You can specify up to four system disk categories in a scaling configuration. A smaller value of N specifies a higher priority. If Auto Scaling cannot create system disks of the disk category that has the highest priority, it creates disks of the disk category that has the next highest priority. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk • cloud_efficiency: ultra disk • cloud_ssd: standard SSD • cloud_essd: ESSD <div style="background-color: #e0f2ff; padding: 10px; margin-top: 10px;"> ? Note You cannot specify the SystemDiskCategories.N and SystemDisk.Category parameters at the same time. </div>

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action/CreateScalingConfiguration
&ScalingGroupId=asg-bp14wlu85wrpchm0*****
&ImageId=centos6u5_64_20G_aliaegeis****.vhd
&ImageName=image****
&InstanceType=ecs.g6.large
&Cpu=2
&Memory=16
&DeploymentSetId=ds-bp1frxuzdg87zh4pz****
&SecurityGroupId=sg-280ih****
&IoOptimized=optimized
&InternetChargeType=PayByTraffic
&InternetMaxBandwidthIn=100
&InternetMaxBandwidthOut=50
```

```
&SystemDisk.Category=cloud_ssd
&SystemDisk.Size=100
&SystemDisk.DiskName=cloud_ssdSystem
&SystemDisk.Description=Test system disk.
&SystemDisk.AutoSnapshotPolicyId=sp-bp12m37ccmxvbm15****
&SystemDisk.PerformanceLevel=PL0
&ScalingConfigurationName=scalingconfig****
&LoadBalancerWeight=50
&Tags={"key1":"value1","key2":"value2", ... "key5":"value5"}
&UserData=echo hello ecs!
&KeyPairName=KeyPairTest
&RamRoleName=ramrole****
&SecurityEnhancementStrategy=Active
&InstanceName=instance****
&HostName=host****
&SpotStrategy=NoSpot
&PasswordInherit=false
&Password=123abc****
&ResourceGroupId=rg-resource****
&HpcClusterId=hpc-clusterid
&InstanceDescription=Test instance.
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&Ipv6AddressCount=1
&CreditSpecification=Standard
&ImageFamily=hangzhou-daily-update
&ZoneId=cn-hangzhou-g
&DedicatedHostId=dh-bp67acfmxazb4p****
&Affinity=default
&Tenancy=default
&PrivatePoolOptions.MatchCriteria=Open
&PrivatePoolOptions.Id=eap-bp67acfmxazb4****
&SpotDuration=1
&SpotInterruptionBehavior=Terminate
&InstanceTypes=["ecs.g6.large"]
&InstanceTypeOverride=[{"InstanceType":"ecs.c5.xlarge","WeightedCapacity":4}]
&DataDisk=[{"PerformanceLevel":"PL1","AutoSnapshotPolicyId":"sp-bp19nq9enxqkomib****","Encrypted":false,"Description":"Test data disk.","SnapshotId":"s-280s7****","Categories":["cloud"],"Size":100,"Device":"/dev/xvdb","DiskName":"cloud_ssdData","Category":"cloud_ssd","DeleteWithInstance":true,"KMSKeyId":"0e478b7a-4262-4802-b8cb-00d3fb40****"}]
&SpotPriceLimit=[{"PriceLimit":0.5,"InstanceType":"ecs.g6.large"}]
&SecurityGroupIds=["sg-bp18kz60mefs****"]
&InstancePatternInfo=[{"Cores":2,"InstanceFamilyLevel":"EnterpriseLevel","MaxPrice":2.0,"Memory":4.0}]
&SystemDiskCategories=["cloud"]
&SystemDisk.Encrypted=false
&SystemDisk.KMSKeyId=0e478b7a-4262-4802-b8cb-00d3fb40****
&SystemDisk.EncryptAlgorithm=AES-256
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<CreateScalingConfigurationResponse>
  <ScalingConfigurationId>asc-bp1ffogfdauy0nu5****</ScalingConfigurationId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</CreateScalingConfigurationResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "ScalingConfigurationId" : "asc-bp1ffogfdauy0nu5****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InstanceType.Mismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	The error message returned because the instance type in the specified scaling configuration is different from the instance type in the current scaling configuration.
404	InvalidDataDiskSnapshotId.NotFound	Snapshot "XXX" does not exist.	The error message returned because the specified snapshot does not exist.
400	InvalidDataDiskSnapshotId.SizeTypeNotSupported	The capacity of snapshot "XXX" exceeds the size limit of the specified disk category.	The error message returned because the size of the specified snapshot exceeds the maximum size that is allowed for the specified disk.

HTTP status code	Error code	Error message	Description
403	InvalidDevice.InUse	Device "XXX" has been occupied.	The error message returned because the mount target of the data disk is occupied.
400	InvalidImageId.InstanceTypeMismatch	The specified image does not support the specified instance type.	The error message returned because the specified image does not support the specified instance type.
404	InvalidImageId.NotFound	The specified image does not exist.	The error message returned because the specified image does not exist within the Alibaba Cloud account.
400	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	The error message returned because the name of the specified key pair does not exist.
400	InvalidNetworkType.ForRAMRole	RAMRole can't be used For classic instance.	The error message returned because the network type of the instance is classic. The classic network does not support RamRoleName.
400	InvalidParameter	The specified value of parameter KeyPairName is not valid.	The error message returned because the OS of the specified instance is Windows. Windows instances do not support KeyPairName.

HTTP status code	Error code	Error message	Description
400	InvalidParameter.Conflict	The value of parameter SystemDisk.Category and parameter DataDisk.N.Category are conflict.	The error message returned because the specified system disk category conflicts with the data disk category.
400	InvalidRamRole.NotFound	The specified RamRoleName does not exist.	The error message returned because the value that you specified for RamRoleName is invalid.
400	InvalidScalingConfigurationName.Duplicate	The specified value of parameter ScalingConfigurationName is duplicated.	The error message returned because the specified scaling configuration name already exists.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Alibaba Cloud account.
400	InvalidSecurityGroupId.IncorrectNetworkType	The network type of specified security Group does not support this action.	The error message returned because the network type of the specified security group is different from the network type of the scaling group.
404	InvalidSecurityGroupId.NotFound	The specified security group does not exist.	The error message returned because the specified security group does not exist within the Alibaba Cloud 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 because 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 because the specified snapshot is created on or before July 15, 2013 and the request is rejected.
403	InvalidSystemDiskCategory.ValueUnauthorized	The system disk category is not authorized.	The error message returned because you are not authorized to create an ephemeral system disk.
400	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded.	The error message returned because the specified user data is not encoded in Base64.
400	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	The error message returned because the size of user data exceeds the upper limit.
403	QuotaExceeded.EphemeralDiskSize	Ephemeral disk size quota exceeded.	The error message returned because the total capacity of mounted ephemeral disks is larger than 2 TiB (2,048 GiB).
400	QuotaExceeded.ScalingConfiguration	Scaling configuration quota exceeded in the specified scaling group.	The error message returned because the maximum number of scaling configurations has been reached.

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

10.2. DescribeScalingConfigurations

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

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

Response parameters

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

Parameter	Type	Example	Description
Affinity	String	default	<p>Indicates whether an instance on a dedicated host is associated with the dedicated host. Valid values:</p> <ul style="list-style-type: none">• default: The instance is not associated with the dedicated host. When you start an instance that was stopped in economical mode, the instance is automatically deployed to another dedicated host in the automatic deployment resource pool if the available resources of the original dedicated host are insufficient.• host: The instance is associated with the dedicated host. When you start an instance that was stopped in economical mode, the instance remains on the original dedicated host. If the original dedicated host has insufficient resources, the instance cannot be started.
Cpu	Integer	2	<p>The number of vCPUs.</p> <p>You can specify the number of vCPUs and the memory size to determine the range of instance types. For example, you can set Cpu to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. If you specify the Cpu and Memory parameters, Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price.</p> <div style="background-color: #e0f2f1; padding: 10px; border-radius: 5px;"><p>? Note This instance type range is available only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.</p></div>
CreationTime	String	2014-08-14T10:58Z	The time when the scaling configuration was created.

Parameter	Type	Example	Description
CreditSpecification	String	Standard	<p>The performance mode of the burstable instance. Valid values:</p> <ul style="list-style-type: none"> • Standard: the standard mode. For more information, see the "Standard mode" section of the Burstable instances topic. • Unlimited: the unlimited mode. For more information, see the "Unlimited mode" section of the Burstable instances topic.
DataDisks	Array of DataDisk		Information about the data disks.
DataDisk			
AutoSnapshotPolicyId	String	sp-bp19nq9enxqkomib****	The ID of the automatic snapshot policy applied to the data disk.
Categories	List	cloud_essd	<p>The categories of the data disks. The values are sorted based on their priorities. The first value has the highest priority. If Auto Scaling cannot create instances that use the disk category of the highest priority, it creates instances that uses the disk category of the next highest priority. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • cloud_essd: enhanced SSD (ESSD).
Category	String	cloud	<p>The category of the data disk. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • ephemeral_ssd: local SSD. • cloud_essd: ESSD.

Parameter	Type	Example	Description
DeleteWithInstance	Boolean	true	<p>Indicates whether the data disk is released when the instance to which the data disk is attached is released. Valid values:</p> <ul style="list-style-type: none"> • true: The data disk is released when the instance is released. • false: The data disk is retained when the instance is released.
Description	String	FinanceDept	The description of the data disk.
Device	String	/dev/xvdb	The mount target of the data disk.
DiskName	String	cloud_ssdData	The name of the data disk.
Encrypted	String	false	<p>Indicates whether the data disk is encrypted. Valid values:</p> <ul style="list-style-type: none"> • true • false <p>Default value: false.</p>
KMSKeyId	String	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the Key Management Service (KMS) key used by the data disk.
PerformanceLevel	String	PL1	The performance level of the data disk that is an ESSD.
Size	Integer	200	<p>The size of the data disk. Unit: GiB.</p> <ul style="list-style-type: none"> • Valid values if Category is set to cloud: 5 to 2000. • Valid values if Category is set to cloud_efficiency: 20 to 32768. • Valid values if Category is set to cloud_ssd: 20 to 32768. • Valid values if Category is set to cloud_essd: 20 to 32768. • Valid values if Category is set to ephemeral_ssd: 5 to 800.
SnapshotId	String	s-23f2i****	The ID of the snapshot used to create the data disk.

Parameter	Type	Example	Description
DedicatedHostId	String	dh-bp67acfmxazb4p***	The ID of the dedicated host on which the Elastic Compute Service (ECS) instance runs. You cannot create preemptible instances on dedicated hosts. If you specify the DedicatedHostId parameter, the SpotStrategy and SpotPriceLimit parameters are ignored. You can call the DescribeDedicatedHosts operation to query the IDs of dedicated hosts.
DeploymentSetId	String	ds-bp1frxuzdg87zh4p***	The ID of the deployment set to which the ECS instance belongs.
HostName	String	LocalHost	The hostname of the ECS instance.
HpcClusterId	String	hpc-clus****	The ID of the Elastic High Performance Computing (E-HPC) cluster to which the ECS instance belongs.
ImageFamily	String	hangzhou-daily-update	The name of the image family. If you specify this parameter, the latest custom images that are available in the specified image family are returned. You can use the images to create ECS instances. If you specify the ImageId parameter, you cannot specify the ImageFamily parameter.
ImageId	String	centos6u5_64_20G_aliaegeis_2014****.vhd	The ID of the image that is used by Auto Scaling to automatically create ECS instances.
ImageName	String	centos6u5_64_20G_aliaegeis_20140703.vhd	The name of the image file.
InstanceDescription	String	FinanceDept	The description of the ECS instance.
InstanceGeneration	String	ecs-3	The generation of the ECS instance.
InstanceName	String	instance****	The name of the ECS instance.

Parameter	Type	Example	Description
InstancePatternInfos	Array of InstancePatternInfo		Information about the intelligent configuration settings, which determines the available instance types.
InstancePatternInfo			
Cores	Integer	2	The number of vCPUs that are supported by the instance type.
InstanceFamilyLevel	String	EnterpriseLevel	<p>The level of the instance family.</p> <ul style="list-style-type: none">• EntryLevel: shared instance type. Instances of this level are most cost-effective, but may not always provide stable computing performance. Instances of this level are suitable for business scenarios in which the CPU utilization is low. For more information, see Shared instance families.• EnterpriseLevel: Instances of this level provide stable performance and dedicated resources, and are suitable for business scenarios that require high stability. For more information, see Instance family.• CreditEntryLevel: This value is available only for burstable instances. CPU credits are used to ensure computing performance. Instances of this level are suitable for scenarios in which the CPU utilization is low but may fluctuate in specific cases. For more information, see What are burstable instances?
MaxPrice	Float	2	The maximum hourly price for pay-as-you-go instances or preemptible instances.
Memory	Float	4	The memory size that is supported by the instance type. Unit: GiB.
InstanceType	String	ecs.g6.large	The instance type of the ECS instance.
InstanceTypes	List	ecs.g6.large	The list of ECS instance types.

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

Parameter	Type	Example	Description
LifecycleState	String	Active	<p>The lifecycle status of the scaling configuration in the scaling group. Valid values:</p> <ul style="list-style-type: none">• Active: The scaling configuration is active in the scaling group. Auto Scaling uses the active scaling configuration to automatically create ECS instances.• Inactive: The scaling configuration is inactive in the scaling group. Auto Scaling does not use the inactive scaling configuration to automatically create ECS instances. Inactive scaling configurations are retained in the scaling group.
LoadBalancerWeight	Integer	1	The weight of the ECS instance as a backend server. Valid values: 1 to 100.
Memory	Integer	16	<p>The memory size. Unit: GiB.</p> <p>You can specify the number of vCPUs and the memory size to determine the range of instance types. For example, you can set Cpu to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. If you specify the Cpu and Memory parameters, Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price.</p> <div style="background-color: #e0f2f1; padding: 5px;"><p>Note This instance type range is available only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.</p></div>
PasswordInherit	Boolean	true	Indicates whether the password preset in the image is used.

Parameter	Type	Example	Description
PrivatePoolOptions.Id	String	eap-bp67acfmxazb4****	The ID of the private pool. The ID of a private pool is the same as the ID of the elasticity assurance or capacity reservation for which the private pool is generated.
PrivatePoolOptions.MatchCriteria	String	Open	<p>The type of the private pool. A private pool is generated after an elasticity assurance or a capacity reservation takes effect. You can select a private pool when you create instances. Valid values:</p> <ul style="list-style-type: none"> • Open: open private pool. The system selects a matching open private pool to create instances. If no matching open private pools are available, the resources in the public pool are used. • Target: specified private pool. The system uses the capacity of a specified private pool to create instances. If the specified private pool is unavailable, the instances cannot be created. • None: no private pool. The capacity of private pools is not used.
RamRoleName	String	ramrole****	The name of the RAM role that is associated with the ECS instance. The name is provided and maintained by Resource Access Management (RAM). You can call the ListRoles operation to query the available RAM roles. For more information about how to create a RAM role, see CreateRole .
ResourceGroupId	String	rg-aeakzn2ou7xo****	The ID of the resource group to which the ECS instance belongs.
ScalingConfigurationId	String	asc-bp1ezrfgoyn5kijl****	The ID of the scaling configuration.
ScalingConfigurationName	String	scalingconfj****	The name of the scaling configuration.
ScalingGroupId	String	asg-bp17pelvl720x3v7**	The ID of the scaling group in which the scaling configuration is created.

Parameter	Type	Example	Description
SchedulerOptions	Struct		<p>? Note This parameter is in invitational preview and unavailable for public use.</p>
ManagedPrivateSpaceId	String	testManagedPrivateSpaceId	<p>? Note This parameter is in invitational preview and unavailable for public use.</p>
SecurityEnhancementStrategy	String	Active	<p>Indicates whether security hardening is enabled. Valid values:</p> <ul style="list-style-type: none">• Active: Security hardening is enabled. This value is available only for public images.• Deactive: Security hardening is disabled. This value is available for all types of images.
SecurityGroupId	String	sg-bp18kz60mefs***	The ID of the security group with which the ECS instance is associated. The ECS instances that are associated with the same security group can access each other.
SecurityGroupIds	List	sg-bp18kz60mefs***	The IDs of the security groups with which the ECS instance is associated. The ECS instances that are associated with the same security group can access each other.
SpotDuration	Integer	1	The retention period of the preemptible instance. Unit: hours.
SpotInterruptoinBehavior	String	Terminate	The interruption mode of the preemptible instance.
SpotPriceLimit	Array of SpotPriceModel		Information about the preemptible instance.
SpotPriceModel			
InstanceType	String	ecs.g6.large	The instance type of the preemptible instance.

Parameter	Type	Example	Description
PriceLimit	Float	0.125	The price limit of the preemptible instance.
SpotStrategy	String	NoSpot	<p>The preemption policy that is applied to pay-as-you-go instances and preemptible instances. Valid values:</p> <ul style="list-style-type: none"> • NoSpot: The instance is created as a pay-as-you-go instance. • SpotWithPriceLimit: The instance is a preemptible instance with a user-defined maximum hourly price. • SpotAsPriceGo: The instance is a preemptible instance for which the market price at the time of purchase is used as the bid price.
SystemDisk.EncryptAlgorithm	String	AES-256	<p>The algorithm that is used to encrypt the system disk. Valid values:</p> <ul style="list-style-type: none"> • AES-256 • SM4-128 <p>Default value: AES-256.</p>
SystemDisk.Encrypted	Boolean	false	<p>Indicates whether the system disk is encrypted. Valid values:</p> <ul style="list-style-type: none"> • true • false <p>Default value: false.</p>
SystemDisk.KMSKeyId	String	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the KMS key that is used to encrypt the system disk.
SystemDiskAutoSnapshotPolicyId	String	sp-bp12m37ccmxvbm5****	The ID of the automatic snapshot policy that is applied to the system disk.

Parameter	Type	Example	Description
SystemDiskCategories	List	cloud	<p>The categories of system disks. The values are sorted based on their priorities. The first value has the highest priority. If Auto Scaling cannot create instances that use the disk category of the highest priority, it creates instances that use the disk category of the next highest priority. Valid values:</p> <ul style="list-style-type: none">• cloud: basic disk• cloud_efficiency: ultra disk• cloud_ssd: standard SSD• cloud_essd: ESSD
SystemDiskCategory	String	cloud	<p>The category of the system disk. Valid values:</p> <ul style="list-style-type: none">• cloud: basic disk• cloud_efficiency: ultra disk• cloud_ssd: standard SSD• ephemeral_ssd: local SSD• cloud_essd: ESSD
SystemDiskDescription	String	Test system disk.	The description of the system disk.
SystemDiskName	String	cloud_ssd_Test	The name of the system disk.
SystemDiskPerformanceLevel	String	PL1	The performance level of the system disk that is an ESSD.
SystemDiskSize	Integer	100	The size of the system disk.
Tags	Array of Tag		Information about tags.
Tag			
Key	String	binary	<p>The tag key of the instance.</p> <p>The tag key cannot be an empty string. The tag key can be up to 128 characters in length and cannot start with acs: or aliyun . It cannot contain http:// or https:// .</p>

Parameter	Type	Example	Description
Value	String	alterTable	The tag value of the instance. The tag value can be an empty string. The tag value can be up to 128 characters in length and cannot start with <code>acs:</code> . It cannot contain <code>http://</code> or <code>https://</code> .
Tenancy	String	default	Indicates whether the instance is created on a dedicated host. Valid values: <ul style="list-style-type: none">• <code>default</code>: The instance is not created on a dedicated host.• <code>host</code>: The instance is created on a dedicated host. If you do not specify the <code>DedicatedHostId</code> parameter, the system selects a dedicated host for the instance that you created.
UserData	String	echo hello ecs!	The user data of the ECS instance.
WeightedCapacities	List	4	The weight of the instance type. This parameter indicates the capacity of a single instance of this instance type in the scaling group. A higher weight indicates that a smaller number of instances of this instance type are required to meet the expected capacity.
ZoneId	String	cn-hangzhou-g	The ID of the zone in which the instance is created. You can call the DescribeZones operation to query the most recent zone list.
TotalCount	Integer	1	The total number of scaling configurations.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DescribeScalingConfigurations
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
<DescribeScalingConfigurationsResponse>
  <RequestId>804F240A-8D3E-40A1-BD68-6B333DEA2CA8</RequestId>
  <TotalCount>1</TotalCount>
  <PageNumber>1</PageNumber>
  <PageSize>50</PageSize>
  <ScalingConfigurations>
    <ScalingConfiguration>
      <CreationTime>2014-08-14T10:58Z</CreationTime>
      <ImageId>centos6u5_64_20G_aliaegis_2014****.vhd</ImageId>
      <InstanceType>ecs.g6.large</InstanceType>
      <InternetChargeType>PayByTraffic</InternetChargeType>
      <InternetMaxBandwidthIn>200</InternetMaxBandwidthIn>
      <InternetMaxBandwidthOut>0</InternetMaxBandwidthOut>
      <LifecycleState>Active</LifecycleState>
      <ScalingConfigurationId>asc-bp1ezrfgoyn5kijl****</ScalingConfigurationId>
      <ScalingConfigurationName>scalingconfig****</ScalingConfigurationName>
      <ScalingGroupId>asg-bp17pelvl720x3v7****</ScalingGroupId>
      <SecurityGroupId>sg-bp18kz60mefs****</SecurityGroupId>
      <SystemDiskCategory>cloud</SystemDiskCategory>
      <DataDisks>
        <DataDisk>
          <Size>200</Size>
          <Category>cloud</Category>
          <SnapshotId>s-280s7****</SnapshotId>
          <Device>/dev/xvdb</Device>
        </DataDisk>
      </DataDisks>
    </ScalingConfiguration>
  </ScalingConfigurations>
</DescribeScalingConfigurationsResponse>
```

JSON format

```
{  
    "DescribeScalingConfigurationsResponse": {  
        "RequestId": "804F240A-8D3E-40A1-BD68-6B333DEA2CA8",  
        "TotalCount": 1,  
        "PageNumber": 1,  
        "PageSize": 50,  
        "ScalingConfigurations": {  
            "ScalingConfiguration": {  
                "CreationTime": "2014-08-14T10:58Z",  
                "ImageId": "centos6u5_64_20G_aliaegeis_2014****.vhf",  
                "InstanceType": "ecs.g6.large",  
                "InternetChargeType": "PayByTraffic",  
                "InternetMaxBandwidthIn": 200,  
                "InternetMaxBandwidthOut": 0,  
                "LifecycleState": "Active",  
                "ScalingConfigurationId": "asc-bp1ezrfgoyn5kijl****",  
                "ScalingConfigurationName": "scalingconfig****",  
                "ScalingGroupId": "asg-bp17pelv1720x3v7****",  
                "SecurityGroupId": "sg-bp18kz60mefs****",  
                "SystemDiskCategory": "cloud",  
                "DataDisks": {  
                    "DataDisk": {  
                        "Size": 200,  
                        "Category": "cloud",  
                        "SnapshotId": "s-280s7****",  
                        "Device": "/dev/xvdb"  
                    }  
                }  
            }  
        }  
    }  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

10.3. ModifyScalingConfiguration

Modifies a scaling configuration.

Description

If you want to change the name of a scaling configuration in a scaling group, make sure that the new name is unique within 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
Action	String	Yes	ModifyScalingConfiguration	The operation that you want to perform. Set the value to ModifyScalingConfiguration .
ScalingConfigurationId	String	Yes	asc-bp16har3jppj6fjbx***	The ID of the scaling configuration that you want to modify.
IoOptimized	String	No	none	<p>Specifies whether the instance that you want to create is an I/O optimized instance. Valid values:</p> <ul style="list-style-type: none"> • none: non-I/O optimized instance. • optimized: I/O optimized instance.
SpotStrategy	String	No	NoSpot	<p>The preemption policy that you want to apply to pay-as-you-go instances and preemptible instances. Valid values:</p> <ul style="list-style-type: none"> • NoSpot: This policy applies to regular pay-as-you-go instances. • SpotWithPriceLimit: This policy applies to the preemptible instance for which a maximum hourly price is specified. • SpotAsPriceGo: This policy applies to the preemptible instance for which the market price at the time of purchase is used as the bid price.
ScalingConfigurationName	String	No	test-modify	<p>The name of the scaling configuration. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or a digit.</p> <p>The name of the scaling configuration must be unique within a scaling group in a region. If you do not specify this parameter, the value of ScalingConfigurationId is used.</p>
InstanceName	String	No	inst****	The name of the Elastic Compute Service (ECS) instance that is automatically created based on the scaling configuration.

Parameter	Type	Required	Example	Description
HostName	String	No	hos***	<p>The hostname of the ECS instance. The hostname cannot start or end with a period (.) or a hyphen (-). The hostname cannot contain consecutive periods (.) or hyphens (-). Naming conventions for different types of instances:</p> <ul style="list-style-type: none"> Windows instances: The hostname must be 2 to 15 characters in length, and can contain letters, digits, and hyphens (-). The hostname cannot contain periods (.). It cannot contain only digits. Other instances such as Linux instances: The hostname must be 2 to 64 characters in length. You can use periods (.) to separate a hostname into multiple segments. Each segment can contain letters, digits, and hyphens (-).
ImageId	String	No	centos6u5_64_20 G_aliaegeis_2014** **.vhd	<p>The ID of the image that is used by Auto Scaling to automatically create instances.</p> <div style="background-color: #e1f5fe; padding: 10px;"> ? Note If the image that is specified in the scaling configuration contains a system disk and data disks, the data that is stored in the data disks is cleared after you modify the image. </div>
ImageName	String	No	suse11sp3_64_20 G_aliaegeis_2015** **.vhd	<p>The name of the image. The name must be unique within a region. If you specify the ImageId parameter, the ImageName parameter is ignored.</p> <p>Alibaba Cloud Marketplace images cannot be specified by using ImageName.</p>

Parameter	Type	Required	Example	Description
Cpu	Integer	No	2	<p>The number of vCPUs.</p> <p>You can specify the number of vCPUs and memory size to determine the range of instance types. For example, you can set CPU to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. Auto Scaling determines a set of available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest unit price.</p> <div style="background-color: #e1f5fe; padding: 10px; border-radius: 5px;"><p>? Note You can specify CPU and memory specifications to create ECS instances only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.</p></div>
Memory	Integer	No	16	<p>The memory size.</p> <p>You can specify the number of vCPUs and memory size to determine the range of instance types. For example, you can set CPU to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. Auto Scaling determines a set of available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest unit price.</p> <div style="background-color: #e1f5fe; padding: 10px; border-radius: 5px;"><p>? Note You can specify CPU and memory specifications to create ECS instances only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.</p></div>

Parameter	Type	Required	Example	Description
InternetChargeType	String	No	PayByBandwidth	<p>The billing method for network usage. Valid values:</p> <ul style="list-style-type: none">• PayByBandwidth: You are charged for the maximum available bandwidth that is specified by InternetMaxBandwidthOut.• PayByTraffic: You are charged for the actual traffic used. InternetMaxBandwidthOut specifies only the maximum available bandwidth.
InternetMaxBandwidthOut	Integer	No	50	<p>The maximum outbound public bandwidth. Unit: Mbit/s. Valid values:</p> <ul style="list-style-type: none">• Valid values if you set InternetChargeType to PayByBandwidth: 0 to 100. If you leave this parameter empty, a value of 0 is used.• Valid values if you set InternetChargeType to PayByTraffic: 0 to 100. If you leave this parameter empty, an error is reported.

Parameter	Type	Required	Example	Description
SystemDisk.Category	String	No	cloud_efficiency	<p>The category of the system disk. Valid values:</p> <ul style="list-style-type: none">• cloud: basic disk• cloud_efficiency: ultra disk• cloud_ssd: standard SSD• cloud_essd: enhanced SSD (ESSD)• ephemeral_ssd: local standard SSD <p>You cannot specify the SystemDisk.Category and SystemDiskCategories.N parameters at the same time. If you do not specify the SystemDisk.Category and SystemDiskCategories.N parameters, the default value of SystemDisk.Category is used. The default value for non-I/O optimized instances of Generation I instance types is cloud. The default value for instances of other types is cloud_efficiency.</p>
SystemDisk.Size	Integer	No	50	<p>The size of the system disk. Unit: GiB. Valid values: 20 to 500.</p> <p>The value of SystemDisk.Size must be greater than or equal to max[20, ImageSize].</p>
SystemDisk.DiskName	String	No	cloud_ssdSystem	<p>The name of the system disk. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with http:// or https://. The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).</p>

Parameter	Type	Required	Example	Description
SystemDisk.Description	String	No	Test system disk.	The description of the system disk. The description must be 2 to 256 characters in length. The description can contain letters but cannot start with <code>http://</code> or <code>https://</code> .
SystemDisk.AutoSnapshotPolicyId	String	No	sp-bp12m37ccmxvbm5****	The ID of the automatic snapshot policy that you want to apply to the system disk.
SystemDisk.PerformanceLevel	String	No	PL0	<p>The performance level of the system disk that is an ESSD. Valid values:</p> <ul style="list-style-type: none"> • PL0: A single ESSD can deliver up to 10,000 random read/write IOPS. • PL1: A single ESSD can deliver up to 50,000 random read/write IOPS. • PL2: A single ESSD can deliver up to 100,000 random read/write IOPS. • PL3: A single ESSD can deliver up to 1,000,000 random read/write IOPS. <div style="background-color: #e1f5fe; padding: 10px; border-radius: 5px;"> ? Note For more information about how to select ESSD performance levels, see ESSDs. </div>
LoadBalancerWeight	Integer	No	50	The weight of the ECS instance as a backend server. Valid values: 1 to 100.
UserData	String	No	echo hello ecs!	The user data of the ECS instance. The data must be encoded in Base64. The maximum size of the raw data before encoding is 16 KB.
KeyPairName	String	No	KeyPair_Name	<p>The name of the key pair that is used to log on to an ECS instance.</p> <ul style="list-style-type: none"> • Windows instances do not use this parameter. • By default, the username and password authentication method is disabled for Linux instances.

Parameter	Type	Required	Example	Description
RamRoleName	String	No	RamRoleTest	The name of the RAM role that is associated with the ECS instance. The name is provided and maintained by Resource Access Management (RAM). You can call the ListRoles operation to query the available RAM roles. For more information about how to create a RAM role, see CreateRole .
PasswordInherit	Boolean	No	false	Specifies whether to use the password preset in the image. Before you use this parameter, make sure that a password is configured for the image.
Tags	String	No	{"key1":"value1", "key2":"value2", ... "key5":"value5"}	<p>The tags of the ECS instance. The tags must be specified in the key-value pair format. You can specify up to 20 tags. The following rules apply to tag keys and tag values:</p> <ul style="list-style-type: none">• A tag key can be up to 64 characters in length. The key cannot start with <code>acs:</code> or <code>aliyun</code>, and cannot contain <code>http://</code> or <code>https://</code>. You cannot specify an empty string as a tag key.• A tag value can be up to 128 characters in length. The value cannot start with <code>acs:</code> or <code>aliyun</code>, and cannot contain <code>http://</code> or <code>https://</code>. You can specify an empty string as a tag value.
DeploymentSetId	String	No	ds-bp13v7bjnj9gis***	The ID of the deployment set to which the ECS instance belongs.
SecurityGroupId	String	No	sg-F876F***	The ID of the security group to which the ECS instance belongs. The ECS instances that belong to the same security group can access each other.

Parameter	Type	Required	Example	Description
Override	Boolean	No	true	<p>Specifies whether to overwrite existing data. Valid values:</p> <ul style="list-style-type: none"> • true: overwrites existing data. • false: does not overwrite existing data.
ResourceGroupId	String	No	abcd1234abcd*** *	The ID of the resource group to which the ECS instance belongs.
HpcClusterId	String	No	hpc-clusterid	The ID of the Elastic High Performance Computing (E-HPC) cluster to which the ECS instance belongs.
InstanceDescription	String	No	Test instance.	<p>The description of the ECS instance. The description must be 2 to 256 characters in length. The description can contain letters but cannot start with <code>http://</code> or <code>https://</code>.</p>
Ipv6AddressCount	Integer	No	1	The number of randomly generated IPv6 addresses that you want to allocate to the elastic network interface (ENI).
CreditSpecification	String	No	Standard	<p>The performance mode of the burstable instance. Valid values:</p> <ul style="list-style-type: none"> • Standard: the standard mode. For more information, see the "Standard mode" section in the Burstable instances topic. • Unlimited: the unlimited mode. For more information, see the "Unlimited mode" section in the Burstable instances topic.
ImageFamily	String	No	hangzhou-daily-update	The name of the image family. You can specify this parameter to obtain the latest custom images that are available in the specified image family. You can use the images to create instances. If you specify the <code>ImageId</code> parameter, you cannot specify the <code>ImageFamily</code> parameter.

Parameter	Type	Required	Example	Description
Zoneld	String	No	cn-hangzhou-g	The ID of the zone where the ECS instance resides.
DedicatedHostId	String	No	dh-bp67acfmxazb4p****	<p>The ID of the dedicated host on which the ECS instance is created. Preemptible instances cannot be created on dedicated hosts. If you specify the DedicatedHostId parameter, the SpotStrategy and SpotPriceLimit parameters are ignored.</p> <p>You can call the DescribeDedicatedHosts operation to query dedicated host IDs.</p>
Affinity	String	No	default	<p>Specifies whether to associate the instance on a dedicated host with the dedicated host. Valid values:</p> <ul style="list-style-type: none">• default: does not associate the instance with the dedicated host. If you restart an instance that was in economical mode when the original dedicated host of the instance has insufficient resources, the instance is deployed to another dedicated host in the automatic deployment resource pool.• host: associates the instance with the dedicated host. If you restart an instance that was in economical mode, the instance remains on the original dedicated host. If the original dedicated host has insufficient resources, the instance fails to start.
Tenancy	String	No	default	<p>Specifies whether to create the instance on a dedicated host. Valid values:</p> <ul style="list-style-type: none">• default: creates the instance on a non-dedicated host.• host: creates the instance on a dedicated host. If you do not specify the DedicatedHostId parameter, the system automatically selects a dedicated host for the instance that you created.

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

Parameter	Type	Required	Example	Description
SpotInterruptionBehavior	String	No	Terminate	The interruption mode of the preemptible instance. Default value: Terminate. Set the value to Terminate. In this case, the instance is directly released.
SystemDisk.Encrypted	Boolean	No	false	Specifies whether to encrypt the system disk. Valid values: <ul style="list-style-type: none">• true: encrypts the system disk.• false: does not encrypt the system disk. Default value: false.
SystemDisk.KMSKeyId	String	No	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the Key Management Service (KMS) key that is used to encrypt the system disk.
SystemDisk.EncryptAlgorithm	String	No	AES-256	The encryption algorithm that is used to encrypt the system disk. Valid values: <ul style="list-style-type: none">• AES-256• SM4-128 Default value: AES-256.
DataDisk.N.PerformanceLevel	String	No	PL1	The performance level of data disk N that is an ESSD. If you set the DataDisk.N.Category parameter to cloud_essd, the N value must be the same as the N value in DataDisk.N.Category. Valid values: <ul style="list-style-type: none">• PL0: A single ESSD can deliver up to 10,000 random read/write IOPS.• PL1: A single ESSD can deliver up to 50,000 random read/write IOPS.• PL2: A single ESSD can deliver up to 100,000 random read/write IOPS.• PL3: A single ESSD can deliver up to 1,000,000 random read/write IOPS. <div style="background-color: #e1f5fe; padding: 10px; border-radius: 5px;"><p>? Note For more information about how to select ESSD performance levels, see ESSDs.</p></div>

Parameter	Type	Required	Example	Description
DataDisk.N.AutoScaleSnapshotPolicyId	String	No	sp-bp19nq9enxqkomib****	The ID of the automatic snapshot policy that you want to apply to data disk N. Valid values of N: 1 to 16.
DataDisk.N.Encrypted	String	No	false	Specifies whether to encrypt data disk N. Valid values of N: 1 to 16. Valid values: <ul style="list-style-type: none">• true: encrypts data disk N.• false: does not encrypt data disk N.
DataDisk.N.Description	String	No	Test data disk.	The description of data disk N. Valid values of N: 1 to 16. The description must be 2 to 256 characters in length. The description can contain letters but cannot start with <code>http://</code> or <code>https://</code> .
DataDisk.N.SnapshotId	String	No	s-snapshot****	The ID of the snapshot that you want to use to create data disk N. Valid values of N: 1 to 16. If you specify this parameter, the DataDisk.N.Size parameter is ignored. The size of the disk is the same as the size of the specified snapshot. If you specify a snapshot that is created on or before July 15, 2013, the operation fails and the system returns InvalidSnapshot.TooOld.

Parameter	Type	Required	Example	Description
DataDisk.N.Categories.N	String	No	cloud	<p>Category N of data disk N. The first N in DataDisk.N.Categories.N represents the serial number of the data disk. Valid values of N: 1 to 16. The second N in DataDisk.N.Categories.N represents the serial number of the data disk category. Valid values of N: 1 to 4. You can specify up to four data disk categories in a scaling configuration. A smaller value of N specifies a higher priority. If Auto Scaling cannot create data disks of the disk category that has the highest priority, it creates disks of the disk category that has the next highest priority. Valid values:</p> <ul style="list-style-type: none">• cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true.• cloud_efficiency: ultra disk.• cloud_ssd: standard SSD.• cloud_essd: ESSD. <div style="background-color: #e0f2ff; padding: 10px; margin-top: 10px;"><p>? Note You cannot specify the DataDisk.N.Categories.N and DataDisk.N.Category parameters at the same time.</p></div>
DataDisk.N.Size	Integer	No	100	<p>The size of data disk N. Unit: GiB. Valid values of N: 1 to 16. Valid values:</p> <ul style="list-style-type: none">• Valid values if you set DataDisk.N.Category to cloud: 5 to 2000.• Valid values if you set DataDisk.N.Category to cloud_efficiency: 20 to 32768.• Valid values if you set DataDisk.N.Category to cloud_essd: 20 to 32768.• Valid values if you set DataDisk.N.Category to ephemeral_ssd: 5 to 800. <p>The value of DataDisk.N.Size must be greater than or equal to the size of the snapshot that is specified by SnapshotId.</p>

Parameter	Type	Required	Example	Description
DataDisk.N.Device	String	No	/dev/xvdb	The mount target of data disk N. Valid values of N: 1 to 16. If you do not specify this parameter, the system automatically assigns a mount target when you create the ECS instance. The name of the mount target ranges from /dev/xvdb to /dev/xvdz.
DataDisk.N.DiskName	String	No	cloud_ssdData	The name of data disk N. Valid values of N: 1 to 16. The name must be 2 to 128 characters in length. The name must start with a letter but cannot start with http:// or https:// . The name can contain letters, digits, colons (:), underscores (_), and hyphens (-).
DataDisk.N.Category	String	No	cloud_ssd	<p>The category of data disk N. Valid values of N: 1 to 16. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk. The DeleteWithInstance attribute of a basic disk that is created together with the instance is set to true. • cloud_efficiency: ultra disk. • cloud_ssd: standard SSD. • cloud_essd: ESSD. • ephemeral_ssd: local SSD. <p>You cannot specify the DataDisk.N.Category and DataDisk.N.Categories.N parameters at the same time. If you do not specify the DataDisk.N.Category and DataDisk.N.Categories.N parameters, the default value of DataDisk.N.Category is:</p> <ul style="list-style-type: none"> • cloud_efficiency for an I/O optimized instance. • cloud for a non-I/O optimized instance.

Parameter	Type	Required	Example	Description
DataDisk.N.DeleteWithInstance	Boolean	No	true	<p>Specifies whether to release data disk N if the instance to which data disk N is attached is released. Valid values of N: 1 to 16. Valid values:</p> <ul style="list-style-type: none">• true: If the instance is released, the data disk is released.• false: If the instance is released, the data disk is retained. <p>This parameter is available only for independent disks whose DataDisk.N.Category parameter is set to cloud, cloud_efficiency, cloud_ssd, or cloud_essd. If you specify this parameter for other disks, an error is reported.</p>
DataDisk.N.KMSKeyId	String	No	0e478b7a-4262-4802-b8cb-00d3fb40****	The ID of the KMS key of data disk N. Valid values of N: 1 to 16.
SpotPriceLimit.N.PriceLimit	Float	No	0.125	The price limit of preemptible instance N. Valid values of N: 1 to 10. This parameter is available only if you set the SpotStrategy parameter to SpotWithPriceLimit.
SpotPriceLimit.N.InstanceType	String	No	ecs.g6.large	The instance type of preemptible instance N. Valid values of N: 1 to 10. This parameter is available only if you set SpotStrategy to SpotWithPriceLimit.

Parameter	Type	Required	Example	Description
InstanceTypes.N	String	No	ecs.g6.large	<p>Instance type N of ECS instances. If you specify this parameter, the InstanceType parameter is ignored. You can specify up to 10 instance types for a scaling configuration. Valid values of N: 1 to 10.</p> <p>A smaller value of N specifies a higher priority. Auto Scaling creates instances based on the priorities of the instance types. If Auto Scaling cannot create instances of the instance type that has the highest priority, it creates instances of the instance type that has the next highest priority.</p>
InstanceTypeOverride.N.InstanceType	String	No	ecs.c5.xlarge	<p>If you want to specify the capacity of instance types in the scaling configuration, you must specify the InstanceTypeOverride.N.InstanceType and InstanceTypeOverride.N.WeightedCapacity parameters.</p> <p>This parameter is used to specify the instance type. You can specify N values for this parameter. You can use this parameter together with the InstanceTypeOverride.N.WeightedCapacity parameter to specify weights for multiple instance types. Valid values of N: 1 to 10.</p> <p>Note You cannot specify the InstanceTypeOverride.N.InstanceType and InstanceTypes parameters at the same time.</p> <p>For information about valid values of InstanceType in InstanceTypeOverride.N.InstanceType, see Instance families.</p>

Parameter	Type	Required	Example	Description
InstanceTypeOverride.N.WeightedCapacity	Integer	No	4	<p>If you want to specify the capacity of instance types in the scaling configuration, you must specify the InstanceTypeOverride.N.InstanceType and InstanceTypeOverride.N.WeightedCapacity parameters. The two parameters have a one-to-one correspondence. The value of N in the two parameters must be the same.</p> <p>This parameter specifies the weight of the instance type. The weight specifies the capacity of a single instance of the specified instance type in the scaling group. A higher weight specifies that a smaller number of instances of the specified instance type are required to meet the expected capacity.</p> <p>Performance metrics such as the number of vCPUs and the memory size of each instance type may vary. You can specify different weights for different instance types based on your business requirements.</p> <p>Example:</p> <ul style="list-style-type: none"> • Current capacity: 0 • Expected capacity: 6 • Capacity of ecs.c5.xlarge: 4 <p>To meet the expected capacity, Auto Scaling creates two ecs.c5.xlarge instances.</p> <div style="background-color: #e0f2f1; padding: 10px;"> <p>? Note The capacity of the scaling group cannot exceed the sum of the maximum capacity that is specified by MaxSize and the maximum weight of the instance type.</p> </div> <p>Valid values of WeightedCapacity in InstanceTypeOverride.N.WeightedCapacity: 1 to 500.</p>
SecurityGroupIds.N	String	No	sg-bp18kz60mefs*** *	The ID of the security group.

Parameter	Type	Required	Example	Description
InstancePatternInfo.N.Cores	Integer	No	2	<p>The number of vCPUs that are specified for an instance type in the intelligent configuration mode. This parameter is used to filter the available instance types that meet the specified criteria. For more information, see Instance families. Valid values of N: 1 to 10.</p> <p>Take note of the following items:</p> <ul style="list-style-type: none">• The InstancePatternInfo.N parameter is available only for scaling groups that reside in virtual private clouds (VPCs).• You must specify the InstancePatternInfo.N, InstancePatternInfo.N.Cores, and InstancePatternInfo.N.Memory parameters at the same time.• If you specify an instance type in the InstanceType or InstanceTypes.N parameter, Auto Scaling preferentially uses the instance type that is specified by the InstanceType or InstanceTypes.N parameter for scale-out activities. If the specified instance type has insufficient inventory, Auto Scaling uses the instance types that are specified by the InstancePatternInfo.N parameter based on the unit price in ascending order.

Parameter	Type	Required	Example	Description
InstancePatternInfo.N.InstanceFamilyLevel	String	No	EnterpriseLevel	<p>The instance family level in intelligent configuration mode. This parameter is used to filter the available instance types that meet the specified criteria. Valid values:</p> <ul style="list-style-type: none">• EntryLevel: shared instance type. Instances of this level are cost-effective, but do not provide stable computing performance. Instances of this level are suitable for business scenarios in which the CPU utilization is low. For more information, see Shared instance families.• EnterpriseLevel: Instances of this level provide stable performance and dedicated resources, and are suitable for business scenarios that require high stability. For more information, see Instance families.• CreditEntryLevel: This value is available only for burstable instances. CPU credits are used to ensure computing performance. This level is suitable for scenarios in which the CPU utilization is low but may fluctuate in specific cases. For more information, see What are burstable instances? <p>Valid values of N: 1 to 10.</p>
InstancePatternInfo.N.MaxPrice	Float	No	2	<p>The maximum hourly price for a pay-as-you-go instance or a preemptible instance in intelligent configuration mode. This parameter is used to filter the available instance types that meet the specified criteria. Valid values of N: 1 to 10.</p> <div style="background-color: #e0f2ff; padding: 10px; border-radius: 5px;"><p>? Note If you set SpotStrategy to SpotWithPriceLimit, you must specify this parameter. Otherwise, you do not need to specify this parameter.</p></div>

Parameter	Type	Required	Example	Description
InstancePatternInfo.N.Memory	Float	No	4	The memory size that is specified for an instance type in intelligent configuration mode. Unit: GiB. This parameter is used to filter the available instance types that meet the specified criteria. For more information, see Instance families . Valid values of N: 1 to 10.
SystemDiskCategories.N	String	No	cloud	<p>Category N of the system disk. Valid values of N: 1 to 4. You can specify up to four system disk categories in a scaling configuration. A smaller value of N specifies a higher priority. If Auto Scaling cannot create system disks of the disk category that has the highest priority, it creates disks of the disk category that has the next highest priority. Valid values:</p> <ul style="list-style-type: none"> • cloud: basic disk • cloud_efficiency: ultra disk • cloud_ssd: standard SSD • cloud_essd: ESSD <div style="background-color: #e1f5fe; padding: 10px;"> ? Note You cannot specify the SystemDiskCategories.N and SystemDisk.Category parameters at the same time. </div>

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=ModifyScalingConfiguration
&ScalingConfigurationId=asc-bp16har3jpj6fjbx****
&IoOptimized=none
&SpotStrategy=NoSpot
&ScalingConfigurationName=test-modify
&InstanceId=inst****
&HostName=hos****
```

```
&ImageId=centos6u5_64_20G_aliaegeis_2014****.vhd
&ImageName=suse11sp3_64_20G_aliaegeis_2015****.vhd
&Cpu=2
&Memory=16
&InternetChargeType=PayByBandwidth
&InternetMaxBandwidthOut=50
&SystemDisk.Category=cloud_efficiency
&SystemDisk.Size=50
&SystemDisk.DiskName=cloud_ssdSystem
&SystemDisk.Description=Test system disk.
&SystemDisk.AutoSnapshotPolicyId=sp-bp12m37ccmxvbm15****
&SystemDisk.PerformanceLevel=PL0
&LoadBalancerWeight=50
&UserData=echo hello ecs!
&KeyPairName=KeyPair_Name
&RamRoleName=RamRoleTest
&PasswordInherit=false
&Tags={"key1":"value1","key2":"value2", ... "key5":"value5"}
&DeploymentSetId=ds-bp13v7bjnj9gis****
&SecurityGroupId=sg-F876F****
&Override=true
&ResourceGroupId=abcd1234abcd****
&HpcClusterId=hpc-clusterid
&InstanceDescription=Test instance.
&Ipv6AddressCount=1
&CreditSpecification=Standard
&ImageFamily=hangzhou-daily-update
&ZoneId=cn-hangzhou-g
&DedicatedHostId=dh-bp67acfmxazb4p****
&Affinity=default
&Tenancy=default
&PrivatePoolOptions.MatchCriteria=Open
&PrivatePoolOptions.Id=eap-bp67acfmxazb4****
&SpotDuration=1
&SpotInterruptionBehavior=Terminate
&DataDisk=[{"PerformanceLevel":"PL1","AutoSnapshotPolicyId":"sp-bp19nq9enxqkomib****","Encrypted":"false","Description":"Test data disk.","SnapshotId":"s-snapshot****","Categories": ["cloud"], "Size":100,"Device":"/dev/xvdb","DiskName": "cloud_ssdData", "Category": "cloud_ssd", "DeleteWithInstance":true, "KMSKeyId": "0e478b7a-4262-4802-b8cb-00d3fb40****"}]
&SpotPriceLimit=[{"PriceLimit":0.125,"InstanceType":"ecs.g6.large"}]
&InstanceTypes=["ecs.g6.large"]
&InstanceTypeOverride=[{"InstanceType":"ecs.c5.xlarge","WeightedCapacity":4}]
&SecurityGroupIds=["sg-bp18kz60mefs****"]
&InstancePatternInfo=[{"Cores":2,"InstanceFamilyLevel":"EnterpriseLevel","MaxPrice":2.0,"Memory":4.0}]
&SystemDiskCategories=["cloud"]
&SystemDisk.Encrypted=false
&SystemDisk.KMSKeyId=0e478b7a-4262-4802-b8cb-00d3fb40****
&SystemDisk.EncryptAlgorithm=AES-256
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<ModifyScalingConfigurationResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</ModifyScalingConfigurationResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because you are not authorized to perform the operation.
404	InvalidDataDiskSnapshotId.NotFound	Snapshot "XXX" does not exist.	The error message returned because the specified snapshot does not exist.
400	InvalidDataDiskSnapshotId.SizeNotSupported	The capacity of snapshot "XXX" exceeds the size limit of the specified disk category.	The error message returned because the size of the specified snapshot exceeds the maximum size allowed for the specified disk.
404	InvalidImageId.NotFound	The specified image does not exist.	The error message returned because the specified image does not exist.

HTTP status code	Error code	Error message	Description
400	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	The error message returned because the name of the specified key pair does not exist.
400	InvalidNetworkType.ForRAMRole	RAMRole can't be used For classic instance.	The error message returned because the network type of the instance is classic. The classic network does not support RamRoleName.
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned because the value that you specified for a parameter is invalid.
400	InvalidScalingConfigurationName.Duplicate	The specified value of parameter is duplicated.	The error message returned because 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 because the network type of the specified security group is different from the network type of the scaling group.
400	InvalidSecurityGroupId.VPCMismatch	The specified security group and the specified virtual switch are not in the same VPC.	The error message returned because the specified security group and vSwitch are not in the same VPC.
400	InvalidTags.KeyValue	The specified tags key/value cannot be empty.	The error message returned because no value is specified for the Tags parameter.

HTTP status code	Error code	Error message	Description
400	InvalidTags.ListSize	The specified tags list size cannot be more than "20".	The error message returned because the maximum number of tags that can be specified for the instance is reached.
400	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded.	The error message returned because the specified user data is not encoded in Base64.
400	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	The error message returned because the size of user data exceeds the upper limit.

10.4. DeleteScalingConfiguration

You can call this operation to delete a scaling configuration.

Description

The scaling configuration cannot be deleted in one of the following conditions:

- The scaling configuration in the scaling group is in the Active state.
- The scaling group contains ECS instances that were created based on the scaling configuration.

Debugging

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

Request parameters

Parameter	Type	Required	Example	Description
Action	String	Yes	DeleteScalingConfiguration	The operation that you want to perform. Set the value to DeleteScalingConfiguration.
ScalingConfigurationId	String	Yes	asc-bp1bx8mzur534edp***	The ID of the scaling configuration to be 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

```
https://ess.aliyuncs.com/?Action=DeleteScalingConfiguration  
&ScalingConfigurationId=asc-bp1bx8mzur534edp***  
&<Common request parameters>
```

Sample success responses

XML format

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

JSON format

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

Error codes

For a list of error codes, visit the [API Error Center](#).

HttpCode	Error code	Error message	Description
404	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	The error message returned because the specified scaling configuration does not exist in the current account.

HttpCode	Error code	Error message	Description
400	IncorrectScalingConfigurationLifecycleState	The current lifecycle state of specified scaling configuration does not support this action.	The error message returned because the specified scaling configuration is not in the Inactive state.
400	InstanceInUse	You cannot delete a scaling configuration or scaling group while there is an instance associated with it.	The error message returned because the scaling group contains ECS instances that were created based on the specified scaling configuration.

10.5. CreateEciScalingConfiguration

Creates a scaling configuration for a scaling group that contains elastic container instances.

Description

A scaling configuration is a template that is used to create elastic container instances during scale-out activities.

You can specify the Cpu and Memory parameters to determine the range of instance types. If you specify the Cpu and Memory parameters, Auto Scaling determines the available instance types based on factors such as I/O optimization requirements and zones. Auto Scaling preferentially creates elastic container instances of the instance type that is provided at the lowest price. This method is available only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration.

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
Action	String	Yes	CreateEciScalingConfiguration	The operation that you want to perform. Set the value to <code>CreateEciScalingConfiguration</code> .

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	asg-bp14wlu85wrpchm0****	The ID of the scaling group in which you want to create the scaling configuration.
ScalingConfigurationName	String	Yes	scalingconfig****	<p>The name of the scaling configuration. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.</p> <p>The name of the scaling configuration must be unique within a scaling group in a region. If you do not specify this parameter, the value of ScalingConfigurationId is used.</p>
SecurityGroupId	String	Yes	sg-uf66jeqopgqa9hdn****	<p>The ID of the security group with which you want to associate the elastic container instance. Elastic container instances that are associated with the same security group can access each other.</p> <p>If you do not specify a security group, the system uses the default security group in the region that you selected. Make sure that the inbound rules of the security group contain the protocols and the port numbers of the containers that you want to expose. If you do not have a default security group in the region, the system creates a default security group, and then adds the declared container protocols and port numbers to the inbound rules of the security group.</p>
ContainerGroupName	String	Yes	nginx-test	The name of the elastic container instance.

Parameter	Type	Required	Example	Description
RestartPolicy	String	No	Always	<p>The restart policy of the elastic container instance. Valid values:</p> <ul style="list-style-type: none"> • Always: always restarts the elastic container instance. • Never: never restarts the elastic container instance. • OnFailure: restarts the elastic container instance upon failures. <p>Default value: Always.</p>
Cpu	Float	No	1.0	The number of vCPUs.
Memory	Float	No	2.0	The size of the memory. Unit: GiB.
ResourceGroupId	String	No	rg-uf66jeqopgqa9hdn***	The ID of the resource group.
DnsPolicy	String	No	Default	The Domain Name System (DNS) policy.
ImageSnapshotId	String	No	imc-2zebxkiifuyzlhlt**	The ID of the image cache.
RamRoleName	String	No	RamTestRole	<p>The name of the instance RAM role. You can use an instance RAM role to access both elastic container instances and Elastic Compute Service (ECS) instances. For more information, see Use an instance RAM role by calling API operations.</p>
TerminationGracePeriodSeconds	Long	No	60	The buffer time in which the program handles operations before the program stops.
AutoMatchImageCache	Boolean	No	false	<p>Specifies whether to automatically match the image cache. Valid values:</p> <ul style="list-style-type: none"> • true • false <p>Default value: false.</p>

Parameter	Type	Required	Example	Description
Ipv6AddressCount	Integer	No	1	The number of IPv6 addresses.
ActiveDeadlineSeconds	Long	No	1000	The validity period. Unit: seconds.
SpotStrategy	String	No	SpotPriceLimit	<p>The preemption policy for the instance. Valid values:</p> <ul style="list-style-type: none"> • NoSpot: The instance is created as a pay-as-you-go instance. • SpotWithPriceLimit: The instance is a preemptible instance with a user-defined maximum hourly price. • SpotAsPriceGo: The instance is a preemptible instance for which the market price at the time of purchase is used as the bid price. <p>Default value: NoSpot.</p>
SpotPriceLimit	Float	No	0.025	<p>The maximum hourly price of the preemptible elastic container instance. The value can be accurate to three decimal places.</p> <p>If you set SpotStrategy to SpotWithPriceLimit, you must specify the SpotPriceLimit parameter.</p>
AutoCreateEip	Boolean	No	true	Specifies whether to automatically create an elastic IP address (EIP), and then bind the EIP to the elastic container instance.
EipBandwidth	Integer	No	5	The bandwidth of the EIP. Default value: 5 Mbit/s.
HostName	String	No	test	The hostname of the elastic container instance.
IngressBandwidth	Long	No	1024000	The maximum inbound bandwidth. Unit: bit/s.
EgressBandwidth	Long	No	1024000	The maximum outbound bandwidth. Unit: bit/s.

Parameter	Type	Required	Example	Description
CpuOptionsCore	Integer	No	2	The number of physical CPU cores. You can specify this parameter for only some instance types. For more information, see Customize CPU options .
CpuOptionsThreadsPerCore	Integer	No	2	The number of threads per core. You can specify this parameter for only some instance types. If you set this parameter to 1, Hyper-Threading is disabled. For more information, see Customize CPU options .
EphemeralStorage	Integer	No	20	The size of the temporary storage space. By default, an enhanced SSD (ESSD) of the PL1 level is used. Unit: GiB.
LoadBalancerWeight	Integer	No	50	The weight of the elastic container instance as a backend server of the associated Server Load Balancer (SLB) instance. Valid values: 1 to 100. Default value: 50.
Tag.N.Key	String	No	version	The key of tag N. Valid values of N: 1 to 20. The tag key cannot be an empty string. The tag key can be up to 128 characters in length and cannot contain <code>http://</code> or <code>https://</code> . It cannot start with <code>acs:</code> or <code>aliyun</code> .
Tag.N.Value	String	No	3	The value of tag N. Valid values of N: 1 to 20. The tag value can be an empty string. The tag value can be up to 128 characters in length and cannot contain <code>http://</code> or <code>https://</code> . It cannot start with <code>acs:</code> .
ImageRegistryCredential.N.Password	String	No	yourpaasword	The password that is used to access image repository N.

Parameter	Type	Required	Example	Description
ImageRegistryCredential.N.Server	String	No	registry-vpc.cn-shanghai.aliyuncs.com	The address of image repository N.
ImageRegistryCredential.N.UserName	String	No	yourusername	The username that is used to access image repository N.
Container.N.ReadinessProbe.TimeoutSeconds	Integer	No	5	The timeout period for the readiness probe. Unit: seconds. Default value: 1. Minimum value: 1.
Container.N.ReadinessProbe.SuccessThreshold	Integer	No	1	The minimum number of consecutive successes for the readiness probe to be considered successful after having failed. Default value: 1. Set the value to 1.
Container.N.SecurityContext.Capability.Add.N	String	No	NET_ADMIN	<p>The permissions that are granted to processes in container N. Only NET_ADMIN and NET_RAW are supported.</p> <div style="background-color: #e0f2ff; padding: 5px;"> ? Note If you want to use NET_RAW, submit a ticket. </div>
Container.N.ReadinessProbe.TcpSocket.Port	Integer	No	8000	The port detected by Transmission Control Protocol (TCP) sockets when you use TCP sockets to perform readiness probes.
Container.N.ReadinessProbe.HttpGet.Scheme	String	No	HTTP	<p>The protocol type of HTTP GET requests when you use HTTP requests for readiness probes. Valid values:</p> <ul style="list-style-type: none"> • HTTP • HTTPS
Container.N.LivenessProbe.PeriodSeconds	Integer	No	5	The interval at which the liveness probe is performed. Unit: seconds. Default value: 10. Minimum value: 1.
Container.N.Port.N.Protocol	String	No	TCP	<p>The protocol type. Valid values:</p> <ul style="list-style-type: none"> • TCP • UDP

Parameter	Type	Required	Example	Description
Container.N.Port.N.Port	Integer	No	80	The port number. Valid values: 1 to 65535.
Container.N.EnvironmentVar.N.Key	String	No	PATH	The name of environment variable N. Specify the name in the [0-9a-zA-Z] format. The name can be 1 to 128 characters in length, and can contain underscores (_). It cannot start with a digit.
Container.N.EnvironmentVar.N.Value	String	No	/usr/local/bin	The value of environment variable N. The value must be 0 to 256 characters in length.
Container.N.LivenessProbe.TcpSocket.Port	Integer	No	8000	The port detected by TCP sockets when you use TCP sockets to perform liveness probes.
Container.N.Tty	Boolean	No	false	Specifies whether to enable interaction. Valid values: <ul style="list-style-type: none">• true• false Default value: false. If the Command parameter is set to /bin/bash, set this parameter to true.
Container.N.WorkingDir	String	No	/usr/local/	The working directory of container N.
Container.N.LivenessProbe.HttpGet.Scheme	String	No	HTTP	The protocol type of HTTP GET requests when you use HTTP requests for liveness probes. Valid values: <ul style="list-style-type: none">• HTTP• HTTPS
Container.N.ReadinessProbe.HttpGet.Port	Integer	No	8080	The port to which HTTP GET requests are sent when you use HTTP requests to perform readiness probes.

Parameter	Type	Required	Example	Description
Container.N.Arg.N	String	No	100	Argument N that corresponds to the command that you run to start container N. You can specify up to 10 arguments.
Container.N.Gpu	Integer	No	1	The number of GPUs that you want to allocate to container N.
Container.N.ReadinessProbe.InitialDelaySeconds	Integer	No	3	The number of seconds after container N has started before readiness probes are initiated.
Container.N.Stdin	Boolean	No	false	Specifies whether container N allocates buffer resources to standard input streams when the container runs. If you do not specify this parameter, an end-of-file (EOF) error may occur. Default value: false.
Container.N.Memory	Float	No	0.5	The memory size of container N. Unit: GiB.
Container.N.Name	String	No	nginx	The image name of container N.
Container.N.Image	String	No	registry-vpc.cn-hangzhou.aliyuncs.com/eci_open/nginx:latest	The image of container N.
Container.N.LivenessProbe.InitialDelaySeconds	Integer	No	5	The number of seconds after container N has started before liveness probes are initiated.

Parameter	Type	Required	Example	Description
Container.N.VolumeMount.N.MountPropagation	String	No	None	<p>The mount propagation setting of volume N. Mount propagation allows volumes that are mounted on one container to be shared with other containers in the same pod, or even with other pods on the same node. Valid values:</p> <ul style="list-style-type: none"> • None: The volume mount does not receive subsequent mounts that are mounted to this volume or its subdirectories. • HostToContainer: The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. • Bidirectional: This value is similar to HostToContainer. The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. In addition, all volume mounts that are created by container N are propagated back to the instance and to all containers of all pods that use the same volume. <p>Default value: None.</p>
Container.N.VolumeMount.N.MountPath	String	No	/pod/data	<p>The directory on which volume N is mounted.</p> <div style="background-color: #e0f2ff; padding: 10px;"> ? Note Data in this directory is overwritten by the data on the volume. </div>
Container.N.VolumeMount.N.Read Only	Boolean	No	false	<p>Specifies whether volume N is read-only. Valid values:</p> <ul style="list-style-type: none"> • true • false <p>Default value: false.</p>
Container.N.VolumeMount.N.Name	String	No	default-volume1	The name of volume N. The value of this parameter is the same as the value of the Volumne.N.Name parameter.

Parameter	Type	Required	Example	Description
Container.N.VolumeMount.N.SubPath	String	No	data2/	The subdirectory of volume N.
Container.N.LivenessProbe.FailureThreshold	Integer	No	3	<p>The minimum number of consecutive failures for the liveness probe to be considered failed after having been successful.</p> <p>Default value: 3.</p>
Container.N.ReadinessProbe.Exec.Command.N	String	No	cat /tmp/healthy	Command N that you want to run in container N when you use the CLI to perform readiness probes.
Container.N.ReadinessProbe.FailureThreshold	Integer	No	3	The minimum number of consecutive failures for the readiness probe to be considered failed after having been successful. Default value: 3.
Container.N.ImagePullPolicy	String	No	Always	The policy to pull an image.
Container.N.StdinOnce	Boolean	No	false	<p>Specifies whether standard input streams are disconnected after a client is disconnected. If Container.N.StdinOnce is set to true, standard input streams remain connected during multiple sessions.</p> <p>When Container.N.StdinOnce is set to true, standard input streams are connected after container N is started, and remain idle until a client is connected to receive data. After the client is disconnected, streams are also disconnected, and remain disconnected until the container is started again.</p>
Container.N.Cpu	Float	No	0.25	The number of CPU cores in container N.
Container.N.LivenessProbe.HttpGet.Port	Integer	No	8888	The port to which HTTP GET requests are sent when you use HTTP requests to perform liveness probes.

Parameter	Type	Required	Example	Description
Container.N.LivenessProbe.HttpGet.Path	String	No	/healthyz	The path to which HTTP GET requests are sent when you use HTTP requests to perform liveness probes.
Container.N.LivenessProbe.SuccessThreshold	Integer	No	1	The minimum number of consecutive successes for the liveness probe to be considered successful after having failed. Default value: 1. Set the value to 1.
Container.N.ReadinessProbe.PeriodSeconds	Integer	No	3	The interval at which the readiness probe is performed. Unit: seconds. Default value: 10. Minimum value: 1.
Container.N.LivenessProbe.TimeoutSeconds	Integer	No	1	The timeout period for the liveness probe. Unit: seconds. Default value: 1. Minimum value: 1.
Container.N.Command.N	String	No	sleep	Command N that you want to run to start container N. You can specify up to 20 commands. Each command contains a maximum of 256 characters.
Container.N.SecurityContext.RunAsUser	Long	No	1000	The ID of the user that runs container N.
Container.N.ReadinessProbe.HttpGet.Path	String	No	/healthz	The path to which HTTP GET requests are sent when you use HTTP requests to perform readiness probes.
Container.N.LivenessProbe.Exec.Command.N	String	No	cat/tmp/healthy	Command N that you want to run in container N when you use the CLI to perform liveness probes.
Volume.N.Type	String	No	ConfigFileVolume	The type of volume N.
Volume.N.DiskVolume.DiskSize	Integer	No	15	The storage size of the disk as a volume. Unit: GiB.
Volume.N.NFSVolume.Path	String	No	/share	The path to the Network File System (NFS) volume.

Parameter	Type	Required	Example	Description
Volume.N.FlexVolume.FsType	String	No	ext4	The file system type of FlexVolume. The default value is determined by the script of FlexVolume.
Volume.N.DiskVolume.FsType	String	No	xfs	The type of the file system of the disk as a volume.
Volume.N.HostPathVolume.Type	String	No	Directory	The type of HostPathVolume. Examples: File, Directory, and Socket.
Volume.N.NFSVolume.ReadOnly	Boolean	No	false	Specifies whether the NFS volume is read-only. Valid values: <ul style="list-style-type: none">• true• false Default value: false.
Volume.N.HostPathVolume.Path	String	No	/xx/xx/path	The absolute path to HostPathVolume.
Volume.N.FlexVolume.Options	String	No	{"volumeId": "d-2zehdahrwoa7sr9****", "performanceLevel": "PL2"}	The FlexVolume options. Each option is a key-value pair in a JSON string. For example, when you use FlexVolume to mount a disk, the format of Options is <code>{"volumeId": "d-2zehdahrwoa7sr9****", "performanceLevel": "PL2"}</code> .
Volume.N.FlexVolume.Driver	String	No	flexvolume	The name of the FlexVolume driver.
Volume.N.ConfigFileVolumeDefaultMode	Integer	No	0644	The default permissions on ConfigFileVolume.
Volume.N.NFSVolume.Server	String	No	3f9cd4a596-naw76.cn-shanghai.nas.aliyuncs.com	The endpoint of the NFS server.
Volume.N.DiskVolume.DiskId	String	No	d-xx	The ID of the disk as a volume.
Volume.N.Name	String	No	default-volume1	The name of volume N.

Parameter	Type	Required	Example	Description
Volume.N.EmptyDirVolume.Medium	String	No	memory	The storage medium of EmptyDirVolume. If you leave this parameter empty, the file system that backs the node is used as the storage medium. If you set this parameter to memory, the memory is used as the storage medium.
Volume.N.ConfigFileVolumeConfigFileToPath.N.Path	String	No	PATH	The name of environment variable N. The name must be 1 to 128 characters in length. Specify the name in the [0-9a-zA-Z] format. The name can contain underscores and cannot start with a digit.
Volume.N.ConfigFileVolumeConfigFileToPath.N.Mode	Integer	No	0644	The permissions on the ConfigFileVolume directory.
Volume.N.ConfigFileVolumeConfigFileToPath.N.Content	String	No	bG1bWk=	The content of the configuration file (32 KB).
InitContainer.N.InitContainerEnvironmentVar.N.Key	String	No	Path	The name of environment variable N. The name must be 1 to 128 characters in length. Specify the name in the [0-9a-zA-Z] format. The name can contain underscores and cannot start with a digit.
InitContainer.N.InitContainerEnvironmentVar.N.Value	String	No	/usr/bin/	The value of environment variable N. The value must be 0 to 256 characters in length.
InitContainer.N.SecurityContext.Capability.Add.N	String	No	NET_ADMIN	The permissions that are granted to the processes in container N. Only NET_ADMIN and NET_RAW are supported.
InitContainer.N.Image	String	No	nginx	The image of container N.

Parameter	Type	Required	Example	Description
InitContainer.N.InitContainerVolumeMount.N.MountPropagation	String	No	None	<p>The mount propagation setting of volume N. Mount propagation allows the sharing of volumes that are mounted on one container to other containers in the same pod, or even to other pods on the same node. Valid values:</p> <ul style="list-style-type: none"> • None: The volume mount does not receive subsequent mounts that are mounted to this volume or its subdirectories. • HostToContainer: The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. • Bidirectional: This value is similar to HostToContainer. The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. In addition, all volume mounts that are created by the container are propagated back to the instance and to all containers of all pods that use the same volume.
InitContainer.N.InitContainerVolumeMount.N.MountPath	String	No	/usr/share/	The directory to which volume N is mounted. Data in this directory is overwritten by the data on the volume.
InitContainer.N.InitContainerVolumeMount.N.ReadOnly	Boolean	No	false	<p>Specifies whether the mount path is read-only. Default value: false.</p>
InitContainer.N.InitContainerVolumeMount.N.Name	String	No	test-empty	The name of volume N that you want to mount.
InitContainer.N.InitContainerVolumeMount.N.SubPath	String	No	/usr/sub/	The subdirectory of volume N. The elastic container instance can mount different directories of the same volume to different subdirectories of container N.
InitContainer.N.ImagePullPolicy	String	No	Always	The policy to pull an image.

Parameter	Type	Required	Example	Description
InitContainer.N.Cpu	Float	No	0.5	The number of vCPUs.
InitContainer.N.WorkingDir	String	No	/usr/local	The working directory of container N.
InitContainer.N.Command.N	String	No	sleep	Command N that you want to run to start container N.
InitContainer.N.Arg.N	String	No	10	Argument N that you want to use to start container N.
InitContainer.N.SecurityContext.RunAsUser	Long	No	587	The ID of the user that runs container N.
InitContainer.N.Gpu	Integer	No	1	The number of GPUs that you want to allocate to container N.
InitContainer.N.Memory	Float	No	1.0	The memory size of container N. Unit: GiB.
InitContainer.N.Name	String	No	test-init	The name of container N.
DnsConfigNameServer.N	String	No	172.10.*.**	The IP addresses of DNS server N.
DnsConfigSearch.N	String	No	svc.local.kubernetes	The lookup domains of DNS server N.
DnsConfigOption.N.Value	String	No	value	The value of option N.
DnsConfigOption.N.Name	String	No	name	The name of option N.
HostAliase.N.Ip	String	No	1.1.1.1	The IP address of host N that you want to add.
HostAliase.N.Hostname.N	String	No	hehe.com	The name of host N that you want to add.
SecurityContextSysctl.N.Value	String	No	65536	The variable value of security context N in which the elastic container instance runs.

Parameter	Type	Required	Example	Description
SecurityContextSystl.N.Name	String	No	kernel.msgmax	The name of security context N in which the elastic container instance runs.
NtpServer.N	String	No	ntp.cloud.aliyuncs.com	The domain name of Network Time Protocol (NTP) server N.
AcrRegistryInfo.N.Domain.N	String	No	*****_****_registry.cn-beijing.cr.aliyuncs.com	Endpoint N of Container Registry Enterprise Edition instance N. By default, all endpoints of the Container Registry Enterprise Edition instance are displayed. You can specify one or more endpoints. Separate multiple endpoints with commas (,).
AcrRegistryInfo.N.InstanceName	String	No	acr-test	The name of Container Registry Enterprise Edition instance N.
AcrRegistryInfo.N.InstanceId	String	No	cri-nwj395hgf6f3***	The ID of Container Registry Enterprise Edition instance N.
AcrRegistryInfo.N.RegionId	String	No	cn-beijing	The region ID of Container Registry Enterprise Edition instance N.

Response parameters

Parameter	Type	Example	Description
ScalingConfigurationId	String	eci-uf6fonnghi50u374***	The ID of the elastic container instance.
RequestId	String	89945DD3-9072-47D0-A318-353284CFC7B3	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action/CreateEciScalingConfiguration
&ScalingGroupId=asg-bp14wlu85wrpchm0****
&ScalingConfigurationName=scalingconfig****
&SecurityGroupId=sg-uf66jeqopgqa9hdn****
&ContainerGroupName=nginx-test
&RestartPolicy=Always
&Cpu=1.0
```

```
&CPU=1.0
&Memory=2.0
&ResourceGroupId=rg-uf66jeqopgqa9hdn****
&DnsPolicy=Default
&ImageSnapshotId=imc-2zebxkiifuyzzlhl****
&RamRoleName=RamTestRole
&TerminationGracePeriodSeconds=60
&AutoMatchImageCache=false
&Ipv6AddressCount=1
&ActiveDeadlineSeconds=1000
&SpotStrategy=SpotPriceLimit
&SpotPriceLimit=0.025
&AutoCreateEip=true
&EipBandwidth=5
&HostName=test
&IngressBandwidth=1024000
&EgressBandwidth=1024000
&CpuOptionsCore=2
&CpuOptionsThreadsPerCore=2
&EphemeralStorage=20
&LoadBalancerWeight=50
&Tag=[{"Key": "version", "Value": "3"}]
&ImageRegistryCredential=[{"Password": "yourpaasword", "Server": "registry-vpc.cn-shanghai.aliyuncs.com", "UserName": "yourusername"}]
&Container=[{"ReadinessProbe.TimeoutSeconds": 5, "ReadinessProbe.SuccessThreshold": 1, "SecurityContext.Capability.Add": ["NET_ADMIN"], "ReadinessProbe.TcpSocket.Port": 8000, "ReadinessProbe.HttpGet.Scheme": "HTTP", "LivenessProbe.PeriodSeconds": 5, "Port": [{"Protocol": "TCP", "Port": 80}], "EnvironmentVar": [{"Key": "PATH", "Value": "/usr/local/bin"}], "LivenessProbe.TcpSocket.Port": 8000, "Tty": false, "WorkingDir": "/usr/local/", "LivenessProbe.HttpGet.Scheme": "HTTP", "ReadinessProbe.HttpGet.Port": 8080, "Arg": ["100"], "Gpu": 1, "ReadinessProbe.InitialDelaySeconds": 3, "Stdin": false, "Memory": 0.5, "Name": "nginx", "Image": "registry-vpc.cn-hangzhou.aliyuncs.com/eci-open/nginx:latest", "LivenessProbe.InitialDelaySeconds": 5, "VolumeMount": [{"MountPropagation": "None", "MountPath": "/pod/data", "ReadOnly": false, "Name": "default-volume1", "SubPath": "data2"}], "LivenessProbe.FailureThreshold": 3, "ReadinessProbe.Exec.Command": ["cat /tmp/healthy"], "ReadinessProbe.FailureThreshold": 3, "ImagePullPolicy": "Always", "StdinOnce": false, "Cpu": 0.25, "LivenessProbe.HttpGet.Port": 8888, "LivenessProbe.HttpGet.Path": "/healthz", "LivenessProbe.SuccessThreshold": 1, "ReadinessProbe.PeriodSeconds": 3, "LivenessProbe.TimeoutSeconds": 1, "Command": ["sleep"], "SecurityContext.RunAsUser": 1000, "ReadinessProbe.HttpGet.Path": "/healthz", "LivenessProbe.Exec.Command": ["cat /tmp/healthy"]}]
&Volume=[{"Type": "ConfigFileVolume", "DiskVolume.DiskSize": 15, "NFSVolume.Path": "/share", "FileVolume.FsType": "ext4", "DiskVolume.FsType": "xfs", "HostPathVolume.Type": "Directory", "NFSVolume.ReadOnly": false, "HostPathVolume.Path": "/xx/xx/path", "FlexVolume.Options": "{\"volumeId\": \"d-2zehdahrwoa7srg****\", \"performanceLevel\": \"PL2\"}", "FlexVolume.Driver": "flexvolume", "ConfigFileVolumeDefaultMode": 644, "NFSVolume.Server": "3f9cd4a596-naw76.cn-shanghai.nas.aliyuncs.com", "DiskVolume.DiskId": "d-xx", "Name": "default-volume1", "EmptyDirVolume.Medium": "memory", "ConfigFileVolumeConfigFileToPath": [{"Path": "PATH", "Mode": 644, "Content": "bGlibWk="}]}, {"InitContainer": [{"InitContainerEnvironmentVar": [{"Key": "Path", "Value": "/usr/bin/"}]}, {"SecurityContext.Capability.Add": ["NET_ADMIN"], "Image": "nginx", "InitContainerVolumeMount": [{"MountPropagation": "None", "MountPath": "/usr/share/", "ReadOnly": false, "Name": "test-empty", "SubPath": "/usr/sub/"}], "ImagePullPolicy": "Always", "Cpu": 0.5, "WorkingDir": "/usr/local", "Command": ["sleep"], "Arg": ["10"], "SecurityContext.RunAsUser": 587, "Gpu": 1, "Memory": 1.0, "Name": "test-init"}]}
&DnsConfigNameServer=["172.10.*.*"]
&DnsConfigSearch=["svc.local.kubernetes"]
&DnsConfigOption=[{"Value": "value", "Name": "name"}]
```

```
&HostAliase=[{"Ip":"1.1.1.1","Hostname":["hehe.com"]}]  
&SecurityContextSysctl=[{"Value":"65536","Name":"kernel.msgmax"}]  
&NtpServer=["ntp.cloud.aliyuncs.com"]  
&AcrRegistryInfo=[{"Domain":["*****-****-registry.cn-beijing.cr.aliyuncs.com"],"InstanceName":"acr-test","InstanceId":"cri-nwj395hgf6f3****","RegionId":"cn-beijing"}]  
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK  
Content-Type:application/xml  
<CreateEciScalingConfigurationResponse>  
    <ScalingConfigurationId>eci-uf6fonnghi50u374****</ScalingConfigurationId>  
    <RequestId>89945DD3-9072-47D0-A318-353284CFC7B3</RequestId>  
</CreateEciScalingConfigurationResponse>
```

JSON format

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
    "ScalingConfigurationId" : "eci-uf6fonnghi50u374****",  
    "RequestId" : "89945DD3-9072-47D0-A318-353284CFC7B3"  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InstanceType.Mismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	The error message returned because the instance type in the specified scaling configuration is different from the instance type in the current scaling configuration.

HTTP status code	Error code	Error message	Description
404	InvalidDataDiskSnapshotId.NotFound	Snapshot "XXX" does not exist.	The error message returned because the specified snapshot does not exist.
400	InvalidDataDiskSnapshotId.SizeNotSupported	The capacity of snapshot "XXX" exceeds the size limit of the specified disk category.	The error message returned because the size of the specified snapshot exceeds the maximum size allowed for the specified disk.
403	InvalidDevice.InUse	Device "XXX" has been occupied.	The error message returned because the mount target of the data disk is occupied.
400	InvalidImageId.InstanceTypeMismatch	The specified image does not support the specified instance type.	The error message returned because the specified image does not support the specified instance type.
404	InvalidImageId.NotFound	The specified image does not exist.	The error message returned because the specified image does not exist within the Alibaba Cloud account.
400	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	The error message returned because the name of the specified key pair does not exist.
400	InvalidNetworkType.ForRAMRole	RAMRole can't be used For classic instance.	The error message returned because the network type of the instance is classic network. The classic network does not support RamRoleName.

HTTP status code	Error code	Error message	Description
400	InvalidParameter	The specified value of parameter KeyPairName is not valid.	The error message returned because the OS of the specified instance is Windows. Windows instances do not support KeyPairName.
400	InvalidParameter.Conflict	The value of parameter SystemDisk.Category and parameter DataDisk.N.Category are conflict.	The error message returned because the specified system disk category conflicts with the data disk category.
400	InvalidRamRole.NotFound	The specified RamRoleName does not exist.	The error message returned because the value that you specified for the RamRoleName parameter is invalid.
400	InvalidScalingConfigurationName.Duplicate	The specified value of parameter ScalingConfigurationName is duplicated.	The error message returned because the specified scaling configuration name already exists.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Alibaba Cloud account.
400	InvalidSecurityGroupId.IncorrectNetworkType	The network type of specified security Group does not support this action.	The error message returned because the network type of the specified security group is different from the network type of the scaling group.

HTTP status code	Error code	Error message	Description
404	InvalidSecurityGroupId.NotFound	The specified security group does not exist.	The error message returned because the specified security group does not exist within the Alibaba Cloud account.
400	InvalidSecurityGroupId.VPCMismatch	The specified security group and the specified virtual switch are not in the same VPC.	The error message returned because the specified security group and vSwitch are not in the same VPC.
403	InvalidSnapshot.TooOld	This operation is denied because the specified snapshot is created before 2013-07-15.	The error message returned because the snapshot was created on or before July 15, 2013. Therefore, the request is rejected.
403	InvalidSystemDiskCategory.ValueUnauthorized	The system disk category is not authorized.	The error message returned because you are not authorized to create an ephemeral system disk.
400	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded.	The error message returned because the specified user data is not encoded in Base64.
400	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	The error message returned because the user data size exceeds the upper limit.

HTTP status code	Error code	Error message	Description
403	QuotaExceeded.EphemeralDiskSize	Ephemeral disk size quota exceeded.	The error message returned because the total capacity of mounted ephemeral disks is larger than 2 TiB (2,048 GiB).
400	QuotaExceeded.ScalingConfiguration	Scaling configuration quota exceeded in the specified scaling group.	The error message returned because the maximum number of scaling configurations has been reached.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the maximum number of elastic container instances that can be associated with the specified security group has been reached.

10.6.

DescribeEciScalingConfigurations

Queries the scaling configurations of a scaling group that contains elastic container instances.

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
Action	String	Yes	DescribeEciScalingConfigurations	The operation that you want to perform. Set the value to DescribeEciScalingConfigurations .

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group to which the scaling configuration belongs.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1.
PageSize	Integer	No	50	The number of entries to return on each page. Maximum value: 50. Default value: 10.
ScalingGroupId	String	No	asg-bp17pelvl720x3v7****	The ID of the scaling group. You can use the ID to query all scaling configurations in the scaling group.
ScalingConfigurationId.N	String	No	asc-bp17pelvl720x5ub****	The ID of scaling configuration N that you want to query. Valid values of N: 1 to 10. The IDs of active and inactive scaling configurations are displayed in the query results. You can differentiate between active and inactive scaling configurations based on the value of the LifecycleState parameter.
ScalingConfigurationName.N	String	No	scalingcon****	The name of scaling configuration N that you want to query. Valid values of N: 1 to 10. The names of inactive scaling configurations are not displayed in the query results, and no error is reported.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
TotalCount	Integer	1	The total number of scaling configurations.
ScalingConfigurations	Array of ScalingConfiguration		Details of the scaling configurations.
CreationTime	String	2014-08-14T10:58Z	The time when the scaling configuration was created.
ScalingConfigurationName	String	scalingconfi****	The name of the scaling configuration.
SecurityGroupId	String	sg-bp18kz60mefs****	The ID of the security group with which the elastic container instance is associated. Elastic container instances that are associated with the same security group can access each other.
ActiveDeadlineSeconds	Integer	1000	The validity period. Unit: seconds.
SpotStrategy	String	NoSpot	<p>The bidding policy for the instance. Valid values:</p> <ul style="list-style-type: none"> • NoSpot: The instance is created as a regular pay-as-you-go instance. • SpotWithPriceLimit: The instance is created as a preemptible instance with a user-defined maximum hourly price. • SpotAsPriceGo: The instance is created as a preemptible instance whose price is based on the market price. <p>Default value: NoSpot.</p>
AutoCreateEip	Boolean	true	Indicates whether an elastic IP address (EIP) is automatically created, and then bound to the elastic container instance.
ScalingGroupId	String	asg-bp17pelvl720x3v7**	The ID of the scaling group in which the scaling configuration is created.

Parameter	Type	Example	Description
EphemeralStorage	Integer	20	The size of the temporary storage space. Unit: GiB.
Ipv6AddressCount	Integer	1	The number of IPv6 addresses.
EipBandwidth	Integer	5	The bandwidth of the EIP. Default value: 5 Mbit/s.
ImageSnapshotId	String	imc-2zebxkiifuyzzlh****	The ID of the image cache.
Memory	Float	16	<p>The size of the memory.</p> <p>You can specify the number of vCPUs and the memory size to determine the range of instance types. For example, you can set Cpu to 2 and Memory to 16 to specify instance types that have 2 vCPUs and 16 GiB of memory. If you specify the Cpu and Memory parameters, Auto Scaling determines available instance types based on factors such as I/O optimization requirements and zones. Then, Auto Scaling preferentially creates instances of the instance type that is provided at the lowest price.</p> <div style="background-color: #e0f2f1; padding: 10px; border-radius: 5px;"> ? Note This instance type range is available only if Scaling Policy is set to Cost Optimization Policy and no instance type is specified in the scaling configuration. </div>
RestartPolicy	String	Never	<p>The restart policy of the elastic container instance. Valid values:</p> <ul style="list-style-type: none"> • Never: never restarts the elastic container instance. • Always: always restarts the elastic container instance. • OnFailure: restarts the elastic container instance upon failures.
SpotPriceLimit	Float	0.025	<p>The maximum hourly price for the preemptible elastic container instance.</p> <p>This parameter is returned only if SpotStrategy is set to SpotWithPriceLimit.</p>

Parameter	Type	Example	Description
IngressBandwidth	Long	1024000	The maximum inbound bandwidth. Unit: bit/s.
CpuOptionsThreadsPerCore	Integer	2	The number of threads per core. You can specify this parameter for only some instance types. If you set this parameter to 1, Hyper-Threading is disabled. For more information, see Customize CPU options .
TerminationGracePeriodSeconds	Integer	60	The buffer time in which the program handles operations before the program stops.
LoadBalancerWeight	Integer	1	The weight of the elastic container instance as a backend server. Valid values: 1 to 100.
DnsPolicy	String	Default	The Domain Name System (DNS) policy.
HostName	String	['hehe.com', 'haha.com']	The name of the host.
EgressBandwidth	Long	1024000	The maximum outbound bandwidth. Unit: bit/s.
RegionId	String	cn-hangzhou	The ID of the region where the elastic container instance resides.
RamRoleName	String	ram:PassRole	The name of the instance RAM role. You can use an instance RAM role to access both elastic container instances and Elastic Compute Service (ECS) instances. For more information, see Use the instance RAM role by calling APIs .
Cpu	Float	2.0	The number of vCPUs that are allocated to the elastic container instance.
AutoMatchImageCache	Boolean	false	Indicates whether the image cache is automatically matched. Default value: false.

Parameter	Type	Example	Description
ResourceGroupId	String	rg-8db03793gfrz****	The ID of the resource group.
ContainerGroupName	String	test	The name of the elastic container instance.
ScalingConfigurationId	String	asc-bp1ezrfgoyn5kijl****	The ID of the scaling configuration.
CpuOptionsCore	Integer	2	The number of physical CPU cores. You can specify this parameter for only some instance types. For more information, see Customize CPU options .
Containers	Array of Container		The containers in the elastic container instance.
ReadinessProbeHttpGetPath	String	/usr/local	The path to which HTTP GET requests were sent.
ReadinessProbeSuccessThreshold	Integer	1	The minimum number of consecutive successes for the readiness probe to be considered successful after having failed. Default value: 1. Valid value: 1.
LivenessProbePeriodSeconds	Integer	5	The interval at which the liveness probe is performed. Default value: 10. Minimum value: 1. Unit: seconds.
LivenessProbeTcpSocketPort	Integer	80	The port number of TcpSocket.
LivenessProbeHttpGetScheme	String	HTTP	The protocol type of HTTP GET requests when you use HTTP requests for liveness probes. Valid values: <ul style="list-style-type: none"> • HTTP • HTTPS
WorkingDir	String	/usr/local/nginx	The working directory of the container.

Parameter	Type	Example	Description
Tty	Boolean	false	<p>Indicates whether interaction is enabled. Valid values:</p> <ul style="list-style-type: none">• true• false <p>If the value of the Command parameter is /bin/bash, the value of this parameter is true.</p> <p>Default value: false.</p>
LivenessProbeHttpGetPort	Integer	80	The port to which HTTP GET requests were sent.
Gpu	Integer	1	The number of GPUs.
ReadinessProbeInitialDelaySeconds	Integer	5	The number of seconds after the container has started before readiness probes are initiated.
Stdin	Boolean	true	<p>Indicates whether the container allocates buffer resources to standard input streams when the container runs. If this parameter is not specified, an end-of-file (EOF) error may occur.</p> <p>Default value: false.</p>
Memory	Float	2.0	The memory size of the container.
Name	String	nginx	The name of the container.
LivenessProbeHttpGetPath	String	/usr/nginx/	The path to which HTTP GET requests were sent.
Image	String	mysql	The image of the container.
ReadinessProbeTcpSocketPort	Integer	8888	The port number of TcpSocket.

Parameter	Type	Example	Description
ReadinessProbe.HttpGetScheme	String	HTTP	The protocol type of HTTP GET requests when you use HTTP requests for readiness probes. Valid values: <ul style="list-style-type: none">• HTTP• HTTPS
ReadinessProbe.HttpGetPort	Integer	/usr/nginx/	The path to which HTTP GET requests were sent.
LivenessProbe.SuccessThreshold	Integer	1	The minimum number of consecutive successes for the liveness probe to be considered successful after having failed. Default value: 1. Valid value: 1.
ReadinessProbe.TimeoutSeconds	Integer	5	The timeout period of the readiness probe. Default value: 1. Minimum value: 1. Unit: seconds.
SecurityContext.RunAsUser	Long	1000	The ID of the user that runs the entry point of the container process.
LivenessProbe.InitialDelaySeconds	Integer	10	The number of seconds after the container has started before liveness probes are initiated.
ReadinessProbe.PeriodSeconds	Integer	1	The interval at which the readiness probe is performed. Default value: 10. Minimum value: 1. Unit: seconds.
ImagePullPolicy	String	Always	The policy to pull an image. Valid values: <ul style="list-style-type: none">• Always: Image pulling is performed each time.• IfNotPresent: Image pulling is performed as needed. On-premises images are preferentially used. If no on-premises images are available, image pulling is performed.• Never: On-premises images are always used. Image pulling is not performed.

Parameter	Type	Example	Description
StdinOnce	Boolean	true	<p>Indicates whether standard input streams are disconnected after a client is disconnected. If Stdin is set to true, standard input streams remain connected during multiple sessions.</p> <p>When Container.N.StdinOnce is set to true, standard input streams are connected after the container is started and remain idle until a client is connected to receive data. After the client is disconnected, streams are also disconnected and remain in the disconnected state until the container is started again.</p>
Cpu	Float	2.0	The number of vCPUs that are allocated to the container.
LivenessProbeTimeoutSeconds	Integer	10	The timeout period of the liveness probe. Default value: 1. Minimum value: 1. Unit: seconds.
ReadinessProbeFailureThreshold	Integer	3	<p>The minimum number of consecutive failures for the readiness probe to be considered failed after having been successful.</p> <p>Default value: 3.</p>
LivenessProbeFailureThreshold	Integer	3	<p>The minimum number of consecutive failures for the liveness probe to be considered failed after having been successful.</p> <p>Default value: 3.</p>
Ports	Array of Port		The exposed ports and protocols of the container.
Port	Integer	8888	The port number. Valid values: 1 to 65535.
Protocol	String	TCP	<p>The protocol type. Valid values:</p> <ul style="list-style-type: none">• TCP• UDP
VolumeMounts	Array of VolumeMount		The volumes that are mounted on the container.

Parameter	Type	Example	Description
ReadOnly	Boolean	false	Indicates whether the volume is read-only. Default value: false.
SubPath	String	data2/	The subdirectory of the volume.
Name	String	default-volume1	The name of the volume. The value of this parameter is the same as the value of the Volumne.N.Name parameter.
MountPropagation	String	None	<p>The mount propagation setting of the volume. Mount propagation allows the sharing of volumes that are mounted on one container with other containers in the same pod, or even with other pods on the same node. Valid values:</p> <ul style="list-style-type: none"> • None: The volume mount does not receive subsequent mounts that are mounted to this volume or its subdirectories. • HostToContainer: The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. • Bidirectional: This value is similar to HostToContainer. The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. In addition, all volume mounts that are created by the container are propagated back to the instance and to all containers of all pods that use the same volume. <p>Default value: None.</p>
MountPath	String	/pod/data	<p>The directory on which the data volume is mounted.</p> <div style="background-color: #e0f2f1; padding: 5px;"> ? Note Data in this directory is overwritten by the data on the volume. </div>
EnvironmentVars	Array of EnvironmentVar		Information about environment variables.

Parameter	Type	Example	Description
Key	String	PATH	The name of the environment variable.
Value	String	/usr/bin/	The value of the environment variable.
Commands	Array of String	sleep	The commands that you run to start the container. You can specify up to 20 commands. Each command contains a maximum of 256 characters.
LivenessProbeExecCommands	Array of String	cat/tmp/healthy	The commands that you run in the container when you used the CLI to perform liveness probes.
Args	Array of String	100	The arguments that correspond to the commands that you run to start the container. You can specify up to 10 arguments.
ReadinessProbeExecCommands	Array of String	cat/tmp/healthy	The commands that you run in the container when you used the CLI to perform readiness probes.
SecurityContextCapabilityAdds	Array of String	NET_ADMIN	<p>The permissions that are granted to the processes in the container. Only NET_ADMIN and NET_RAW are supported.</p> <div style="background-color: #e0f2f1; padding: 5px;"> ? Note If you want to use NET_RAW, submit a ticket. </div>
InitContainers	Array of InitContainer		The init containers.
SecurityContextReadOnlyRootFilesystem	Boolean	true	Indicates whether the root file system is read-only. Valid value: true.
ImagePullPolicy	String	Always	The policy to pull an image.
WorkingDir	String	/usr/local	The working directory.
Cpu	Float	0.5	The number of vCPUs.

Parameter	Type	Example	Description
Image	String	nginx	The image of the container.
Gpu	Integer	1	The number of GPUs that are allocated to the container.
SecurityContext.RunAsUser	String	587	The ID of the user that runs the container.
Memory	Float	1.0	The size of the memory.
Name	String	test-init	The name of the container.
InitContainerEnv ironmentVars	Array of InitContainerEnv ironmentVar		Information about environment variables.
Key	String	PATH	The name of the environment variable.
Value	String	/usr/local/bin	The value of the environment variable.
InitContainerVol umeMounts	Array of InitContainerVol umeMount		The volumes that are mounted on the container.
ReadOnly	Boolean	false	Indicates whether the mount path is read-only. Default value: false.
SubPath	String	/usr/sub/	The subdirectory of the volume. The container group can mount different directories of the same volume to different subdirectories of containers.
Name	String	test-empty	The name of the volume.

Parameter	Type	Example	Description
MountPropagation	String	None	<p>The mount propagation setting of the volume. Mount propagation allows the sharing of volumes that are mounted on one container to other containers in the same pod, or even to other pods on the same node. Valid values:</p> <ul style="list-style-type: none">• None: The volume mount does not receive subsequent mounts that are mounted to this volume or its subdirectories.• HostToContainer: The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories.• Bidirectional: This value is similar to HostToContainer. The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. In addition, all volume mounts that are created by the container are propagated back to the instance and to all containers of all pods that use the same volume. <p>Default value: None.</p>
MountPath	String	/usr/share/	The directory to which the volume is mounted. Data in this directory is overwritten by the data on the volume.
SecurityContextCapabilityAdds	Array of String	NET_ADMIN	<p>The permissions that are granted to the processes in the container. Only NET_ADMIN and NET_RAW are supported.</p> <div style="background-color: #e0f2ff; padding: 10px; border-radius: 5px;"><p>? Note If you want to use NET_RAW, submit a ticket.</p></div>
InitContainerCommands	Array of String	sleep	The commands that you run to start the container.
InitContainerArgs	Array of String	10	The arguments that you use to start the container.

Parameter	Type	Example	Description
Tags	Array of Tag		The tags of the elastic container instance. The tags are specified in the key-value pair format.
Key	String	version	The tag key.
Value	String	3	The tag value.
Volumes	Array of Volume		Information about volumes.
Type	String	EmptyDirVolume	<p>The type of the volume. Valid values:</p> <ul style="list-style-type: none"> • EmptyDirVolume • NFSVolume • ConfigFileVolume • FlexVolume
DiskVolumeDiskId	String	15	The storage size of the disk as a volume. Unit: GiB.
DiskVolumeFsType	String	xfs	The type of the file system of the disk as a volume.
EmptyDirVolumeMedium	String	memory	The storage medium of EmptyDirVolume. If this parameter is left empty, the file system that backs the node is used as the storage medium. If this parameter is set to memory, the memory is used as the storage medium.
NFSVolumePath	String	/share	The path to the Network File System (NFS) volume.
DiskVolumeDiskSize	Integer	15	The storage size of the disk as a volume. Unit: GiB.
NFSVolumeReadOnly	Boolean	false	<p>Indicates whether the NFS volume is read-only.</p> <p>Default value: false.</p>

Parameter	Type	Example	Description
FlexVolumeFsType	String	ext4	The type of the file system of FlexVolume. The default value is determined by the script of FlexVolume.
ConfigFileVolumeDefaultMode	Integer	0644	The default permissions on ConfigFileVolume.
FlexVolumeDriver	String	flexvolume	The name of the FlexVolume driver.
FlexVolumeOptions	String	{"volumeId": "d-2zehdahrwoa7srg****", "performanceLevel": "PL2"}	The FlexVolume options. Each option is a key-value pair in a JSON string. For example, when you use FlexVolume to mount a disk, the format of Options is <code>{"volumeId": "d-2zehdahrwoa7srg****", "performanceLevel": "PL2"} .</code>
NFSVolumeServer	String	3f9cd4a596-naw76.cn-shanghai.nas.aliyuncs.com	The endpoint of the NFS server.
Name	String	default-volume1	The name of the volume.
ConfigFileVolumeConfigFileToPaths	Array of ConfigFileVolumeConfigFileToPath		The paths to configuration files.
Path	String	/usr/bin/	The relative file path.
Mode	Integer	0644	The permissions on the ConfigFileVolume directory.
Content	String	bG1bwk=	The content of the configuration file (32 KB).
ImageRegistryCredentials	Array of ImageRegistryCredential		Information about the image repository.
Password	String	yourpaasword	The password that is used to access the image repository.

Parameter	Type	Example	Description
Server	String	registry-vpc.cn-shanghai.aliyuncs.com	The address of the image repository.
UserName	String	yourusername	The username that is used to access the image repository.
DnsConfigOptions	Array of DnsConfigOption		The options. Each option is a name-value pair. The value in the name-value pair is optional.
Name	String	name	The variable name of the option.
Value	String	value	The variable value of the option.
HostAliases	Array of HostAlias		The hostnames and IP addresses of a container that are added to the hosts file of the elastic container instance.
Ip	String	192.0.XX.XX	The IP address that is added.
Hostnames	Array of String	['hehe.com', 'haha.com']	The hostnames that are added.
SecurityContext.SysCtls	Array of SecurityContext.SysCtl		The system information of the security context in which the elastic container instance runs.
Name	String	kernel.msgmax	The name of the security context in which the elastic container instance runs.
Value	String	65536	The variable value of the security context in which the elastic container instance runs.
AcrRegistryInfo.s	Array of AcrRegistryInfo		Information about the Container Registry Enterprise Edition instance.
Instanceid	String	cri-nwj395hgf6f3****	The ID of the Container Registry Enterprise Edition instance.

Parameter	Type	Example	Description
InstanceName	String	acr-test	The name of the Container Registry Enterprise Edition instance.
RegionId	String	cn-hangzhou	The ID of the region where the Container Registry Enterprise Edition instance resides.
Domains	Array of String	*****_****_ registry.cn-beijing.cr.aliyuncs.com	The endpoints of the Container Registry Enterprise Edition instance. By default, all endpoints of the Container Registry Enterprise Edition instance are displayed. You can specify one or more endpoints. Separate multiple endpoints with commas (,).
DnsConfigNameservers	Array of String	172.10.*.**	The IP addresses of the DNS server.
DnsConfigSearches	Array of String	svc.local.kubernetes	The lookup domains of the DNS server.
NtpServers	Array of String	ntp.cloud.aliyuncs.com	The domain name of the Network Time Protocol (NTP) server.
LifecycleState	String	Active	<p>The status of the scaling configuration in the scaling group. Valid values:</p> <ul style="list-style-type: none">• Active: The scaling configuration is active in the scaling group. Auto Scaling uses the active scaling configuration to automatically create elastic container instances.• Inactive: The scaling configuration is inactive in the scaling group. Auto Scaling does not use the inactive scaling configuration to automatically create elastic container instances. Inactive scaling configurations are retained in the scaling group.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DescribeEciScalingConfigurations  
&RegionId=cn-qingdao  
&PageNumber=1  
&PageSize=50  
&ScalingGroupId=asg-bp17pelvl720x3v7****  
&ScalingConfigurationId=[ "asc-bp17pelvl720x5ub****" ]  
&ScalingConfigurationName=[ "scalingcon****" ]  
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK  
Content-Type:application/xml  
<DescribeEciScalingConfigurationsResponse>  
    <PageNumber>1</PageNumber>  
    <PageSize>50</PageSize>  
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>  
    <TotalCount>1</TotalCount>  
    <ScalingConfigurations>  
        <CreationTime>2014-08-14T10:58Z</CreationTime>  
        <ScalingConfigurationName>scalingconfi****</ScalingConfigurationName>  
        <SecurityGroupId>sg-bp18kz60mefs****</SecurityGroupId>  
        <ActiveDeadlineSeconds>1000</ActiveDeadlineSeconds>  
        <SpotStrategy>NoSpot</SpotStrategy>  
        <AutoCreateEip>true</AutoCreateEip>  
        <ScalingGroupId>asg-bp17pelvl720x3v7****</ScalingGroupId>  
        <EphemeralStorage>20</EphemeralStorage>  
        <Ipv6AddressCount>1</Ipv6AddressCount>  
        <EipBandwidth>5</EipBandwidth>  
        <ImageSnapshotId>imc-2zebxkiifuyzzlh1****</ImageSnapshotId>  
        <Memory>16</Memory>  
        <RestartPolicy>Never</RestartPolicy>  
        <SpotPriceLimit>0.025</SpotPriceLimit>  
        <IngressBandwidth>1024000</IngressBandwidth>  
        <CpuOptionsThreadsPerCore>2</CpuOptionsThreadsPerCore>  
        <TerminationGracePeriodSeconds>60</TerminationGracePeriodSeconds>  
        <LoadBalancerWeight>1</LoadBalancerWeight>  
        <DnsPolicy>Default</DnsPolicy>  
        <HostName>['hehe.com', 'haha.com']</HostName>  
        <EgressBandwidth>1024000</EgressBandwidth>  
        <RegionId>cn-hangzhou</RegionId>  
        <RamRoleName>ram:PassRole</RamRoleName>  
        <Cpu>2</Cpu>  
        <AutoMatchImageCache>false</AutoMatchImageCache>  
        <ResourceGroupId>rg-8db03793gfrz****</ResourceGroupId>  
        <ContainerGroupName>test</ContainerGroupName>  
        <ScalingConfigurationId>asc-bplezrfgoyn5kijl****</ScalingConfigurationId>  
        <CpuOptionsCore>2</CpuOptionsCore>  
        <Containers>  
            <ReadinessProbeHttpGetPath>/usr/local</ReadinessProbeHttpGetPath>  
            <ReadinessProbeSuccessThreshold>1</ReadinessProbeSuccessThreshold>  
            <LivenessProbePeriodSeconds>5</LivenessProbePeriodSeconds>
```

```
<LivenessProbeTcpSocketPort>80</LivenessProbeTcpSocketPort>
<LivenessProbeHttpGetScheme>HTTP</LivenessProbeHttpGetScheme>
<WorkingDir>/usr/local/nginx</WorkingDir>
<Tty>false</Tty>
<LivenessProbeHttpGetPort>80</LivenessProbeHttpGetPort>
<Gpu>1</Gpu>
<ReadinessProbeInitialDelaySeconds>5</ReadinessProbeInitialDelaySeconds>
<Stdin>true</Stdin>
<Memory>2</Memory>
<Name>nginx</Name>
<LivenessProbeHttpGetPath>/usr/nginx/</LivenessProbeHttpGetPath>
<Image>mysql</Image>
<ReadinessProbeTcpSocketPort>8888</ReadinessProbeTcpSocketPort>
<ReadinessProbeHttpGetScheme>HTTP</ReadinessProbeHttpGetScheme>
<LivenessProbeSuccessThreshold>1</LivenessProbeSuccessThreshold>
<ReadinessProbeTimeoutSeconds>5</ReadinessProbeTimeoutSeconds>
<SecurityContextRunAsUser>1000</SecurityContextRunAsUser>
<LivenessProbeInitialDelaySeconds>10</LivenessProbeInitialDelaySeconds>
<ReadinessProbePeriodSeconds>1</ReadinessProbePeriodSeconds>
<ImagePullPolicy>Always</ImagePullPolicy>
<StdinOnce>true</StdinOnce>
<Cpu>2</Cpu>
<LivenessProbeTimeoutSeconds>10</LivenessProbeTimeoutSeconds>
<ReadinessProbeFailureThreshold>3</ReadinessProbeFailureThreshold>
<LivenessProbeFailureThreshold>3</LivenessProbeFailureThreshold>
<Ports>
    <Port>8888</Port>
    <Protocol>TCP</Protocol>
</Ports>
<VolumeMounts>
    <ReadOnly>false</ReadOnly>
    <SubPath>data2</SubPath>
    <Name>default-volume1</Name>
    <MountPropagation>None</MountPropagation>
    <MountPath>/pod/data</MountPath>
</VolumeMounts>
<EnvironmentVars>
    <Key>PATH</Key>
    <Value>/usr/bin/</Value>
</EnvironmentVars>
<Commands>sleep</Commands>
<LivenessProbeExecCommands>cat /tmp/healthy</LivenessProbeExecCommands>
<Args>100</Args>
<ReadinessProbeExecCommands>cat /tmp/healthy</ReadinessProbeExecCommands>
<SecurityContextCapabilityAddrs>NET_ADMIN</SecurityContextCapabilityAddrs>
</Containers>
<InitContainers>
    <SecurityContextReadOnlyRootFilesystem>true</SecurityContextReadOnlyRootFilesys
tem>
    <ImagePullPolicy>Always</ImagePullPolicy>
    <WorkingDir>/usr/local</WorkingDir>
    <Cpu>0.5</Cpu>
    <Image>nginx</Image>
    <Gpu>1</Gpu>
```

```
<SecurityContextRunAsUser>587</SecurityContextRunAsUser>
<Memory>1</Memory>
<Name>test-init</Name>
<InitContainerEnvironmentVars>
    <Key>PATH</Key>
    <Value>/usr/local/bin</Value>
</InitContainerEnvironmentVars>
<InitContainerVolumeMounts>
    <ReadOnly>false</ReadOnly>
    <SubPath>/usr/sub/</SubPath>
    <Name>test-empty</Name>
    <MountPropagation>None</MountPropagation>
    <MountPath>/usr/share/</MountPath>
</InitContainerVolumeMounts>
<SecurityContextCapabilityAdds>NET_ADMIN</SecurityContextCapabilityAdds>
<InitContainerCommands>sleep</InitContainerCommands>
<InitContainerArgs>10</InitContainerArgs>
</InitContainers>
<Tags>
    <Key>version</Key>
    <Value>3</Value>
</Tags>
<Volumes>
    <Type>EmptyDirVolume</Type>
    <DiskVolumeDiskId>15</DiskVolumeDiskId>
    <DiskVolumeFsType>xfs</DiskVolumeFsType>
    <EmptyDirVolumeMedium>memory</EmptyDirVolumeMedium>
    <NFSServerPath>/share</NFSServerPath>
    <DiskVolumeDiskSize>15</DiskVolumeDiskSize>
    <NFSServerReadOnly>false</NFSServerReadOnly>
    <FlexVolumeFsType>ext4</FlexVolumeFsType>
    <ConfigFileVolumeDefaultMode>644</ConfigFileVolumeDefaultMode>
    <FlexVolumeDriver>flexvolume</FlexVolumeDriver>
    <FlexVolumeOptions>{"volumeId": "d-2zehdahrwoa7srg****", "performanceLevel": "PL2"}</FlexVolumeOptions>
    <NFSServer>3f9cd4a596-naw76.cn-shanghai.nas.aliyuncs.com</NFSServer>
>
    <Name>default-volume1</Name>
    <ConfigFileVolumeConfigFileToPaths>
        <Path>/usr/bin/</Path>
        <Mode>644</Mode>
        <Content>bG1lbWk=</Content>
    </ConfigFileVolumeConfigFileToPaths>
</Volumes>
<ImageRegistryCredentials>
    <Password>yourpaasword</Password>
    <Server>registry-vpc.cn-shanghai.aliyuncs.com</Server>
    <UserName>yourusername</UserName>
</ImageRegistryCredentials>
<DnsConfigOptions>
    <Name>name</Name>
    <Value>value</Value>
</DnsConfigOptions>
<HostAliases>
```

```
<Ip>192.0.XX.XX</Ip>
<Hostnames>['hehe.com', 'haha.com']</Hostnames>
</HostAliases>
<SecurityContextSysCtls>
    <Name>kernel.msgmax</Name>
    <Value>65536</Value>
</SecurityContextSysCtls>
<AcrRegistryInfos>
    <InstanceId>cri-nwj395hgf6f3****</InstanceId>
    <InstanceName>acr-test</InstanceName>
    <RegionId>cn-hangzhou</RegionId>
    <Domains>*****-*****-registry.cn-beijing.cr.aliyuncs.com</Domains>
</AcrRegistryInfos>
<DnsConfigNameServers>172.10.*.**</DnsConfigNameServers>
<DnsConfigSearches>svc.local.kubernetes</DnsConfigSearches>
<NtpServers>ntp.cloud.aliyuncs.com</NtpServers>
<LifecycleState>Active</LifecycleState>
</ScalingConfigurations>
</DescribeEciScalingConfigurationsResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "PageNumber" : 1,
    "PageSize" : 50,
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "TotalCount" : 1,
    "ScalingConfigurations" : [ {
        "CreationTime" : "2014-08-14T10:58Z",
        "ScalingConfigurationName" : "scalingconfi****",
        "SecurityGroupId" : "sg-bp18kz60mefs****",
        "ActiveDeadlineSeconds" : 1000,
        "SpotStrategy" : "NoSpot",
        "AutoCreateEip" : true,
        "ScalingGroupId" : "asg-bp17pelvl720x3v7****",
        "EphemeralStorage" : 20,
        "Ipv6AddressCount" : 1,
        "EipBandwidth" : 5,
        "ImageSnapshotId" : "imc-2zebxkiifuyzzlhl****",
        "Memory" : 16,
        "RestartPolicy" : "Never",
        "SpotPriceLimit" : 0.025,
        "IngressBandwidth" : 1024000,
        "CpuOptionsThreadsPerCore" : 2,
        "TerminationGracePeriodSeconds" : 60,
        "LoadBalancerWeight" : 1,
        "DnsPolicy" : "Default",
        "HostName" : "[ 'hehe.com', 'haha.com' ]",
        "EgressBandwidth" : 1024000,
        "RegionId" : "cn-hangzhou",
        "RamRoleName" : "ram:PassRole",
        "Cpu" : 2,
    } ]
```

```
"AutoMatchImageCache" : false,
"ResourceGroupId" : "rg-8db03793gfrz****",
"ContainerGroupName" : "test",
"ScalingConfigurationId" : "asc-bplezrfgoyn5kijl****",
"CpuOptionsCore" : 2,
"Containers" : [ {
    "ReadinessProbeHttpGetPath" : "/usr/local",
    "ReadinessProbeSuccessThreshold" : 1,
    "LivenessProbePeriodSeconds" : 5,
    "LivenessProbeTcpSocketPort" : 80,
    "LivenessProbeHttpGetScheme" : "HTTP",
    "WorkingDir" : "/usr/local/nginx",
    "Tty" : false,
    "LivenessProbeHttpGetPort" : 80,
    "Gpu" : 1,
    "ReadinessProbeInitialDelaySeconds" : 5,
    "Stdin" : true,
    "Memory" : 2,
    "Name" : "nginx",
    "LivenessProbeHttpGetPath" : "/usr/nginx/",
    "Image" : "mysql",
    "ReadinessProbeTcpSocketPort" : 8888,
    "ReadinessProbeHttpGetScheme" : "HTTP",
    "LivenessProbeSuccessThreshold" : 1,
    "ReadinessProbeTimeoutSeconds" : 5,
    "SecurityContextRunAsUser" : 1000,
    "LivenessProbeInitialDelaySeconds" : 10,
    "ReadinessProbePeriodSeconds" : 1,
    "ImagePullPolicy" : "Always",
    "StdinOnce" : true,
    "Cpu" : 2,
    "LivenessProbeTimeoutSeconds" : 10,
    "ReadinessProbeFailureThreshold" : 3,
    "LivenessProbeFailureThreshold" : 3,
    "Ports" : [ {
        "Port" : 8888,
        "Protocol" : "TCP"
    }],
    "VolumeMounts" : [ {
        "ReadOnly" : false,
        "SubPath" : "data2/",
        "Name" : "default-volume1",
        "MountPropagation" : "None",
        "MountPath" : "/pod/data"
    }],
    "EnvironmentVars" : [ {
        "Key" : "PATH",
        "Value" : "/usr/bin/"
    }],
    "Commands" : [ "sleep" ],
    "LivenessProbeExecCommands" : [ "cat/tmp/healthy" ],
    "Args" : [ "100" ],
    "ReadinessProbeExecCommands" : [ "cat/tmp/healthy" ],
    "SecurityContextCapabilityAdds" : [ "NET_ADMIN" ]
}
```

```
    } ],
  "InitContainers" : [ {
    "SecurityContextReadOnlyRootFilesystem" : true,
    "ImagePullPolicy" : "Always",
    "WorkingDir" : "/usr/local",
    "Cpu" : 0.5,
    "Image" : "nginx",
    "Gpu" : 1,
    "SecurityContextRunAsUser" : "587",
    "Memory" : 1,
    "Name" : "test-init",
    "InitContainerEnvironmentVars" : [ {
      "Key" : "PATH",
      "Value" : "/usr/local/bin"
    }],
    "InitContainerVolumeMounts" : [ {
      "ReadOnly" : false,
      "SubPath" : "/usr/sub/",
      "Name" : "test-empty",
      "MountPropagation" : "None",
      "MountPath" : "/usr/share/"
    }],
    "SecurityContextCapabilityAdds" : [ "NET_ADMIN" ],
    "InitContainerCommands" : [ "sleep" ],
    "InitContainerArgs" : [ "10" ]
  }],
  "Tags" : [ {
    "Key" : "version",
    "Value" : "3"
  }],
  "Volumes" : [ {
    "Type" : "EmptyDirVolume",
    "DiskVolumeDiskId" : "15",
    "DiskVolumeFsType" : "xfs",
    "EmptyDirVolumeMedium" : "memory",
    "NFSVolumePath" : "/share",
    "DiskVolumeDiskSize" : 15,
    "NFSVolumeReadOnly" : false,
    "FlexVolumeFsType" : "ext4",
    "ConfigFileVolumeDefaultMode" : 644,
    "FlexVolumeDriver" : "flexvolume",
    "FlexVolumeOptions" : "{\"volumeId\":\"d-2zehdahrwoa7srg****\", \"performanceLevel\": \"PL2\"}",
    "NFSVolumeServer" : "3f9cd4a596-naw76.cn-shanghai.nas.aliyuncs.com",
    "Name" : "default-volume1",
    "ConfigFileVolumeConfigFileToPaths" : [ {
      "Path" : "/usr/bin/",
      "Mode" : 644,
      "Content" : "bGl1bWk="
    }]
  }],
  "ImageRegistryCredentials" : [ {
    "Password" : "yourpaasword",
    "Server" : "registry-vpc.cn-shanghai.aliyuncs.com",
  }]
}
```

```
        "UserName" : "yourusername"
    } ],
    "DnsConfigOptions" : [ {
        "Name" : "name",
        "Value" : "value"
    } ],
    "HostAliases" : [ {
        "Ip" : "192.0.XX.XX",
        "Hostnames" : [ "'hehe.com', 'haha.com'" ]
    } ],
    "SecurityContextSysCtls" : [ {
        "Name" : "kernel.msgmax",
        "Value" : "65536"
    } ],
    "AcrRegistryInfos" : [ {
        "InstanceId" : "cri-nwj395hgf6f3****",
        "InstanceName" : "acr-test",
        "RegionId" : "cn-hangzhou",
        "Domains" : [ "*****-****-registry.cn-beijing.cr.aliyuncs.com" ]
    } ],
    "DnsConfigNameServers" : [ "172.10.*.*" ],
    "DnsConfigSearches" : [ "svc.local.kubernetes" ],
    "NtpServers" : [ "ntp.cloud.aliyuncs.com" ],
    "LifecycleState" : "Active"
} ]
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

10.7. ModifyEciScalingConfiguration

Modifies a scaling configuration of a scaling group that contains elastic container instances.

Description

If you want to change the name of a scaling configuration in a scaling group, make sure that the new name is unique within 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
-----------	------	----------	---------	-------------

Parameter	Type	Required	Example	Description
Action	String	Yes	ModifyEciScalingConfiguration	The operation that you want to perform. Set the value to ModifyEciScalingConfiguration .
ScalingConfigurationId	String	Yes	asc-bp16har3jppj6fjbx***	The ID of the scaling configuration that you want to modify.
ScalingConfigurationName	String	Yes	test-modify	<p>The name of the scaling configuration. The name must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). The name must start with a letter or a digit.</p> <p>The name of the scaling configuration must be unique within a scaling group in a region. If you do not specify this parameter, the value of the <code>ScalingConfigurationId</code> parameter is used.</p>
SecurityGroupId	String	Yes	sg-uf66jeqopgqa9hdn***	<p>The ID of the security group with which you want to associate the elastic container instance. Elastic container instances that are associated with the same security group can access each other.</p> <p>If you do not specify a security group, the system uses the default security group in the region that you selected. Make sure that the inbound rules of the security group contain the protocols and the port numbers of the containers that you want to expose. If you do not have a default security group in the region, the system creates a default security group, and then adds the declared container protocols and port numbers to the inbound rules of the security group.</p>

Parameter	Type	Required	Example	Description
ContainerGroupName	String	Yes	nginx-test	<p>The name of the elastic container instance.</p> <ul style="list-style-type: none"> • The name must be 2 to 128 characters in length. • The name can contain only lowercase letters, digits, and hyphens (-). It cannot start or end with a hyphen (-).
RestartPolicy	String	No	Always	<p>The restart policy of the elastic container instance. Valid values:</p> <ul style="list-style-type: none"> • Always: always restarts the elastic container instance. • Never: never restarts the elastic container instance. • OnFailure: restarts the elastic container instance upon failures. <p>Default value: Always.</p>
Cpu	Float	No	1.0	The number of vCPUs.
Memory	Float	No	2.0	The size of the memory. Unit: GiB.
ResourceGroupId	String	No	rg-uf66jeqopgqa9hdn***	The ID of the resource group.
DnsPolicy	String	No	Default	The Domain Name System (DNS) policy.
ImageSnapshotId	String	No	imc-2zebxkiifuyzlhll**	The ID of the image cache.
RamRoleName	String	No	RamTestRole	<p>The name of the instance RAM role. You can use an instance RAM role to access both elastic container instances and Elastic Compute Service (ECS) instances. For more information, see Use an instance RAM role by calling API operations.</p>

Parameter	Type	Required	Example	Description
TerminationGracePeriodSeconds	Long	No	60	The buffer time in which the program handles operations before the program stops.
AutoMatchImageCache	Boolean	No	false	Specifies whether to automatically match the image cache. Default value: false.
Ipv6AddressCount	Integer	No	1	The number of IPv6 addresses.
ActiveDeadlineSeconds	Long	No	1000	The validity period. Unit: seconds.
SpotStrategy	String	No	SpotPriceLimit	<p>The bidding policy for the elastic container instance. Valid values:</p> <ul style="list-style-type: none"> • NoSpot: The instance is created as a pay-as-you-go instance. • SpotWithPriceLimit: The instance is a preemptible instance with a user-defined maximum hourly price. • SpotAsPriceGo: The instance is a preemptible instance for which the market price at the time of purchase is used as the bid price. <p>Default value: NoSpot.</p>
SpotPriceLimit	Float	No	0.025	<p>The maximum hourly price of the preemptible elastic container instance. The value can be accurate to three decimal places.</p> <p>If you set SpotStrategy to SpotWithPriceLimit, you must specify the SpotPriceLimit parameter.</p>
AutoCreateEip	Boolean	No	true	Specifies whether to automatically create an elastic IP address (EIP), and then bind the EIP to the elastic container instance.
EipBandwidth	Integer	No	5	<p>The bandwidth of the EIP.</p> <p>Default value: 5 Mbit/s.</p>

Parameter	Type	Required	Example	Description
HostName	String	No	test	The hostname of the elastic container instance.
IngressBandwidth	Long	No	1024000	The maximum inbound bandwidth. Unit: bit/s.
EgressBandwidth	Long	No	1024000	The maximum outbound bandwidth. Unit: bit/s.
CpuOptionsCore	Integer	No	2	The number of physical CPU cores. You can specify this parameter for only some instance types. For more information, see Customize CPU options .
CpuOptionsThreadsPerCore	Integer	No	2	The number of threads per core. You can specify this parameter for only some instance types. If you set this parameter to 1, Hyper-Threading is disabled. For more information, see Customize CPU options .
EphemeralStorage	Integer	No	20	The size of the temporary storage space. By default, an enhanced SSD (ESSD) of the PL1 level is used. Unit: GiB.
LoadBalancerWeight	Integer	No	50	The weight of the elastic container instance as a backend server. Valid values: 1 to 100.
Tag.N.Key	String	No	version	The key of tag N.
Tag.N.Value	String	No	3	The value of tag N.
ImageRegistryCredential.N.Password	String	No	yourpaasword	The password that is used to access image repository N.
ImageRegistryCredential.N.Server	String	No	registry-vpc.cn-shanghai.aliyuncs.com	The address of image repository N.

Parameter	Type	Required	Example	Description
ImageRegistryCredential.N.UserName	String	No	yourusername	The username that is used to access image repository N.
Container.N.ReadinessProbe.TimeoutSeconds	Integer	No	1	The timeout period for the readiness probe. Unit: seconds. Default value: 1. Minimum value: 1.
Container.N.ReadinessProbe.SuccessThreshold	Integer	No	1	The minimum number of consecutive successes for the readiness probe to be considered successful after having failed. Default value: 1. Valid value: 1.
Container.N.SecurityContext.Capability.Add.N	String	No	NET_ADMIN	The permissions that are granted to processes in container N. Only NET_ADMIN and NET_RAW are supported. Note If you want to use NET_RAW, submit a ticket.
Container.N.ReadinessProbe.TcpSocket.Port	Integer	No	8000	The port detected by Transmission Control Protocol (TCP) sockets when you use TCP sockets to perform readiness probes.
Container.N.ReadinessProbe.HttpGet.Scheme	String	No	HTTP	The protocol type of HTTP GET requests when you use HTTP requests for readiness probes. Valid values: <ul style="list-style-type: none">• HTTP• HTTPS
Container.N.LivenessProbe.PeriodSeconds	Integer	No	5	The interval at which the liveness probe is performed. Unit: seconds. Default value: 10. Minimum value: 1.
Container.N.Port.N.Protocol	String	No	TCP	The protocol type. Valid values: <ul style="list-style-type: none">• TCP• UDP
Container.N.Port.N.Port	Integer	No	80	The port number. Valid values: 1 to 65535.

Parameter	Type	Required	Example	Description
Container.N.SecurityContext.ReadOnlyRootFilesystem	Boolean	No	true	Specifies whether the root file system on which the container runs is read-only. Valid value: true.
Container.N.EnvironmentVar.N.Key	String	No	PATH	The name of environment variable N. Specify the name in the [0-9a-zA-Z] format. The name can be 1 to 128 characters in length, and can contain underscores (_). It cannot start with a digit.
Container.N.EnvironmentVar.N.Value	String	No	/usr/local/bin	The value of environment variable N. The value must be 0 to 256 characters in length.
Container.N.LivenessProbe.TcpSocket.Port	Integer	No	8000	The port detected by TCP sockets when you use TCP sockets to perform liveness probes.
Container.N.Tty	Boolean	No	false	Specifies whether to enable interaction. Default value: false. If the Command parameter is set to /bin/bash, set this parameter to true.
Container.N.WorkingDir	String	No	/usr/local/	The working directory of container N.
Container.N.LivenessProbe.HttpGet.Scheme	String	No	HTTP	The protocol type of HTTP GET requests when you use HTTP requests for liveness probes. Valid values: <ul style="list-style-type: none">• HTTP• HTTPS
Container.N.ReadinessProbe.HttpGet.Port	Integer	No	8080	The port to which HTTP GET requests are sent when you use HTTP requests to perform readiness probes.
Container.N.Arg.N	String	No	100	Argument N that corresponds to the command that you run to start container N. You can specify up to 10 arguments.

Parameter	Type	Required	Example	Description
Container.N.Gpu	Integer	No	1	The number of GPUs that you want to allocate to container N.
Container.N.ReadinessProbe.InitialDelaySeconds	Integer	No	3	The number of seconds after container N has started before readiness probes are initiated.
Container.N.Stdin	Boolean	No	false	Specifies whether container N allocates buffer resources to standard input streams when the container runs. If you do not specify this parameter, an end-of-file (EOF) error may occur. Default value: false.
Container.N.Memory	Float	No	0.5	The memory size of container N. Unit: GiB.
Container.N.Name	String	No	nginx	The image name of container N.
Container.N.Image	String	No	registry-vpc.cn-hangzhou.aliyuncs.com/eci_open/nginx:latest	The image of container N.
Container.N.LivenessProbe.InitialDelaySeconds	Integer	No	5	The number of seconds after container N has started before liveness probes are initiated.

Parameter	Type	Required	Example	Description
Container.N.VolumeMount.N.MountPropagation	String	No	None	<p>The mount propagation setting of volume N. Mount propagation allows volumes that are mounted on one container to be shared with other containers in the same pod, or even with other pods on the same node. Valid values:</p> <ul style="list-style-type: none"> • None: The volume mount does not receive subsequent mounts that are mounted to this volume or its subdirectories. • HostToContainer: The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. • Bidirectional: This value is similar to HostToContainer. The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. In addition, all volume mounts that are created by container N are propagated back to the instance and to all containers of all pods that use the same volume. <p>Default value: None.</p>
Container.N.VolumeMount.N.MountPath	String	No	/pod/data	<p>The directory on which volume N is mounted.</p> <div style="background-color: #e0f2ff; padding: 10px;"> 💡 Notice Data in this directory is overwritten by the data on the volume. </div>
Container.N.VolumeMount.N.Read Only	Boolean	No	false	<p>Specifies whether volume N is read-only.</p> <p>Default value: false.</p>
Container.N.VolumeMount.N.Name	String	No	default-volume1	The name of volume N. The value of this parameter is the same as the value of the Volumne.N.Name parameter.
Container.N.VolumeMount.N.SubPath	String	No	data2/	The subdirectory of volume N.

Parameter	Type	Required	Example	Description
Container.N.LivenessProbe.FailureThreshold	Integer	No	3	The minimum number of consecutive failures for the liveness probe to be considered failed after having been successful. Default value: 3.
Container.N.ReadinessProbe.Exec.Command.N	String	No	cat/tmp/healthy	Command N that you want to run in container N when you use the CLI to perform readiness probes.
Container.N.ReadinessProbe.FailureThreshold	Integer	No	3	The minimum number of consecutive failures for the readiness probe to be considered failed after having been successful. Default value: 3.
Container.N.ImagePullPolicy	String	No	Always	The policy to pull an image. Valid values: <ul style="list-style-type: none">• Always: Image pulling is performed each time.• IfNotPresent: Image pulling is performed as needed. On-premises images are preferentially used. If no on-premises images are available, image pulling is performed.• Never: On-premises images are always used. Image pulling is not performed.

Parameter	Type	Required	Example	Description
Container.N.StdinOnce	Boolean	No	false	<p>Specifies whether standard input streams are disconnected after a client is disconnected. If Container.N.StdinOnce is set to true, standard input streams remain connected during multiple sessions.</p> <p>When Container.N.StdinOnce is set to true, standard input streams are connected after container N is started, and remain idle until a client is connected to receive data. After the client is disconnected, streams are also disconnected, and remain disconnected until the container is started again.</p>
Container.N.Cpu	Float	No	0.25	The number of CPU cores that you want to allocate to container N.
Container.N.LivenessProbe.HttpGet.Port	Integer	No	8888	The port to which HTTP GET requests are sent when you use HTTP requests to perform liveness probes.
Container.N.LivenessProbe.HttpGet.Path	String	No	/healthyz	The path to which HTTP GET requests are sent when you use HTTP requests to perform liveness probes.
Container.N.LivenessProbe.SuccessThreshold	Integer	No	1	The minimum number of consecutive successes for the liveness probe to be considered successful after having failed. Default value: 1. Valid value: 1.
Container.N.ReadinessProbe.PeriodSeconds	Integer	No	3	The interval at which the readiness probe is performed. Unit: seconds. Default value: 10. Minimum value: 1.
Container.N.LivenessProbe.TimeoutSeconds	Integer	No	1	The timeout period for the liveness probe. Unit: seconds. Default value: 1. Minimum value: 1.

Parameter	Type	Required	Example	Description
Container.N.Command.N	String	No	sleep	Command N that you want to run to start container N. You can specify up to 20 commands. Each command contains a maximum of 256 characters.
Container.N.SecurityContext.RunAsUser	Long	No	1000	The ID of the user that runs container N.
Container.N.ReadinessProbe.HttpGet.Path	String	No	/healthz	The path to which HTTP GET requests are sent when you use HTTP requests to perform readiness probes.
Container.N.LivenessProbe.Exec.Command.N	String	No	cat /tmp/healthy	Command N that you want to run in container N when you use the CLI to perform liveness probes.
Volume.N.Type	String	No	EmptyDirVolume	The type of volume N. Valid values: <ul style="list-style-type: none">• EmptyDirVolume• NFSVolume• ConfigFileVolume• FlexVolume
Volume.N.DiskVolume.DiskSize	Integer	No	15	The storage size of the disk as a volume. Unit: GiB.
Volume.N.NFSVolume.Path	String	No	/share	The path to the Network File System (NFS) volume.
Volume.N.FlexVolume.FsType	String	No	ext4	The file system type of FlexVolume. The default value is determined by the script of FlexVolume.
Volume.N.DiskVolume.FsType	String	No	xfs	The type of the file system of the disk as a volume.
Volume.N.HostPathVolume.Type	String	No	Directory	The type of HostPathVolume. Examples: File, Directory, and Socket.

Parameter	Type	Required	Example	Description
Volume.N.NFSVolume.ReadOnly	Boolean	No	false	Specifies whether the NFS volume is read-only. Default value: false.
Volume.N.HostPathVolume.Path	String	No	/xx/xx/name	The absolute path to HostPathVolume.
Volume.N.FlexVolume.Options	String	No	{"volumeId": "d-2zehdahrwoa7srg****", "performanceLevel": "PL2"}	The FlexVolume options. Each option is a key-value pair in a JSON string. For example, when you use FlexVolume to mount a disk, the format of Options is {"volumeId": "d-2zehdahrwoa7srg****", "performanceLevel": "PL2"}
Volume.N.FlexVolume.Driver	String	No	flexvolume	The name of the FlexVolume driver.
Volume.N.ConfigFileVolumeDefaultMode	Integer	No	0644	The default permissions on ConfigFileVolume.
Volume.N.NFSVolume.Server	String	No	3f9cd4a596-naw76.cn-shanghai.nas.aliyuncs.com	The endpoint of the NFS server.
Volume.N.DiskVolume.DiskId	String	No	d-xx	The ID of the disk as a volume.
Volume.N.Name	String	No	default-volume1	The name of volume N.
Volume.N.EmptyDirVolume.Medium	String	No	memory	The storage medium of EmptyDirVolume. If you leave this parameter empty, the file system that backs the node is used as the storage medium. If you set this parameter to memory, the memory is used as the storage medium.
Volume.N.ConfigFileVolumeConfigFileToPath.N.Path	String	No	/usr/bin/	The relative path to the configuration file.
Volume.N.ConfigFileVolumeConfigFileToPath.N.Mode	Integer	No	0644	The permissions on the ConfigFileVolume directory.

Parameter	Type	Required	Example	Description
Volume.N.ConfigFileVolumeConfigFileToPath.N.Content	String	No	bG1bWk=	The content of the configuration file (32 KB).
InitContainer.N.InitContainerEnvironmentVar.N.Key	String	No	Path	The name of environment variable N. The name must be 1 to 128 characters in length. Specify the name in the [0-9a-zA-Z] format. The name can contain underscores (_) and cannot start with a digit.
InitContainer.N.InitContainerEnvironmentVar.N.Value	String	No	/usr/bin/	The value of environment variable N. The value must be 0 to 256 characters in length.
InitContainer.N.SecurityContext.Capability.Add.N	String	No	NET_ADMIN	The permissions that are granted to the processes in container N. Only NET_ADMIN and NET_RAW are supported. ? Note If you want to use NET_RAW, submit a ticket.
InitContainer.N.Image	String	No	nginx	The image of container N.

Parameter	Type	Required	Example	Description
InitContainer.N.InitContainerVolumeMount.N.MountPropagation	String	No	None	<p>The mount propagation setting of volume N. Mount propagation allows the sharing of volumes that are mounted on one container to other containers in the same pod, or even to other pods on the same node. Valid values:</p> <ul style="list-style-type: none"> • None: The volume mount does not receive subsequent mounts that are mounted to this volume or its subdirectories. • HostToContainer: The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. • Bidirectional: This value is similar to HostToContainer. The volume mount receives all subsequent mounts that are mounted to this volume or its subdirectories. In addition, all volume mounts that are created by container N are propagated back to the instance and to all containers of all pods that use the same volume. <p>Default value: None.</p>
InitContainer.N.InitContainerVolumeMount.N.MountPath	String	No	/usr/share/	<p>The directory on which volume N is mounted.</p> <div style="background-color: #e0f2ff; padding: 10px;"> 💡 Notice Data in this directory is overwritten by the data on the volume. </div>
InitContainer.N.InitContainerVolumeMount.N.ReadOnly	Boolean	No	false	<p>Specifies whether the mount path is read-only.</p> <p>Default value: false.</p>
InitContainer.N.InitContainerVolumeMount.N.Name	String	No	test-empty	The name of volume N that you want to mount.

Parameter	Type	Required	Example	Description
InitContainer.N.InitContainerVolumeMount.N.SubPath	String	No	/usr/sub/	The subdirectory of volume N. The elastic container instance can mount different directories of the same volume to different subdirectories of container N.
InitContainer.N.ImagePullPolicy	String	No	Always	The policy to pull an image.
InitContainer.N.Cpu	Float	No	0.5	The number of vCPUs.
InitContainer.N.WorkingDir	String	No	/usr/local	The working directory of container N.
InitContainer.N.Command.N	String	No	sleep	Command N that you want to run to start container N.
InitContainer.N.Arg.N	String	No	10	Argument N that you want to use to start container N.
InitContainer.N.SecurityContext.RunAsUser	Long	No	587	The ID of the user that runs container N.
InitContainer.N.Gpu	Integer	No	1	The number of GPUs that you want to allocate to container N.
InitContainer.N.Memory	Float	No	1.0	The memory size of container N. Unit: GiB.
InitContainer.N.Name	String	No	test-init	The name of container N.
DnsConfigNameServer.N	String	No	172.10.*.**	The IP addresses of DNS server N.
DnsConfigSearch.N	String	No	svc.local.kubernetes	The lookup domains of DNS server N.
DnsConfigOption.N.Value	String	No	value	The value of option N.
DnsConfigOption.N.Name	String	No	name	The name of option N.

Parameter	Type	Required	Example	Description
HostAliase.N.ip	String	No	192.0.XX.XX	The IP address of host N that you want to add.
HostAliase.N.Hostname.N	String	No	['hehe.com', 'haha.com']	The name of host N that you want to add.
SecurityContextSystcl.N.Value	String	No	65536	The variable value of security context N in which the elastic container instance runs.
SecurityContextSystcl.N.Name	String	No	kernel.msgmax	The name of security context N in which the elastic container instance runs.
NtpServer.N	String	No	ntp.cloud.aliyuncs.com	The domain name of Network Time Protocol (NTP) server N.
AcrRegistryInfo.N.Domain.N	String	No	*****_****_registry.cn-beijing.cr.aliyuncs.com	Endpoint N of Container Registry Enterprise Edition instance N. By default, all endpoints of instance N are displayed. You can specify one or more endpoints. Separate multiple endpoints with commas (,).
AcrRegistryInfo.N.InstanceName	String	No	acr-test	The name of Container Registry Enterprise Edition instance N.
AcrRegistryInfo.N.InstanceId	String	No	cri-nwj395hgf6f3****	The ID of Container Registry Enterprise Edition instance N.
AcrRegistryInfo.N.RegionId	String	No	cn-hangzhou	The region ID of Container Registry Enterprise Edition instance N.

Response parameters

Parameter	Type	Example	Description
RequestId	String	89945DD3-9072-47D0-A318-353284CFC7B3	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=ModifyEciScalingConfiguration
&ScalingConfigurationId=asc-bp16har3jpj6fjbx****
&ScalingConfigurationName=test-modify
&SecurityGroupId=sg-uf66jeqopgqa9hdn****
&ContainerGroupName=nginx-test
&RestartPolicy=Always
&Cpu=1.0
&Memory=2.0
&ResourceGroupId=rg-uf66jeqopgqa9hdn****
&DnsPolicy=Default
&ImageSnapshotId=imc-2zebxkiifuyzzlhl****
&RamRoleName=RamTestRole
&TerminationGracePeriodSeconds=60
&AutoMatchImageCache=false
&Ipv6AddressCount=1
&ActiveDeadlineSeconds=1000
&SpotStrategy=SpotPriceLimit
&SpotPriceLimit=0.025
&AutoCreateEip=true
&EipBandwidth=5
&HostName=test
&IngressBandwidth=1024000
&EgressBandwidth=1024000
&CpuOptionsCore=2
&CpuOptionsThreadsPerCore=2
&EphemeralStorage=20
&LoadBalancerWeight=50
&Tag=[{"Key":"version","Value":"3"}]
&ImageRegistryCredential=[{"Password":"yourpaasword","Server":"registry-vpc.cn-shanghai.aliyuncs.com","UserName":"yourusername"}]
&Container=[{"ReadinessProbe.TimeoutSeconds":1,"ReadinessProbe.SuccessThreshold":1,"SecurityContext.Capability.Add":["NET_ADMIN"],"ReadinessProbe.TcpSocket.Port":8000,"ReadinessProbe.HttpGet.Scheme":"HTTP","LivenessProbe.PeriodSeconds":5,"Port": [{"Protocol":"TCP","Port":80}], "SecurityContext.ReadOnlyRootFilesystem":true,"EnvironmentVar":[{"Key":"PATH","Value":"/usr/local/bin"}],"LivenessProbe.TcpSocket.Port":8000,"Tty":false,"WorkingDir":"/usr/local/","LivenessProbe.HttpGet.Scheme":"HTTP","ReadinessProbe.HttpGet.Port":8080,"Arg":["100"],"GPU":1,"ReadinessProbe.InitialDelaySeconds":3,"Stdin":false,"Memory":0.5,"Name":"nginx","Image":"registry-vpc.cn-hangzhou.aliyuncs.com/eci_open/nginx:latest","LivenessProbe.InitialDelaySeconds":5,"VolumeMount": [{"MountPropagation":"None","MountPath":"/pod/data","ReadOnly":false,"Name":"default-volume1","SubPath":"data2/"}],"LivenessProbe.FailureThreshold":3,"ReadinessProbe.Exec.Command": ["cat /tmp/healthy"], "ReadinessProbe.FailureThreshold":3,"ImagePullPolicy":"Always","StdinOnce":false,"Cpu":0.25,"LivenessProbe.HttpGet.Port":8888,"LivenessProbe.HttpGet.Path":"/healthz","LivenessProbe.SuccessThreshold":1,"ReadinessProbe.PeriodSeconds":3,"LivenessProbe.TimeoutSeconds":1,"Command": ["sleep"], "SecurityContext.RunAsUser":1000,"ReadinessProbe.HttpGet.Path":"/healthz","LivenessProbe.Exec.Command": ["cat /tmp/healthy"]}]
&Volume=[{"Type":"EmptyDirVolume","DiskVolume.DiskSize":15,"NFSVolume.Path":"/share","FlexVolume.FsType":"ext4","DiskVolume.FsType":"xfs","HostPathVolume.Type":"Directory","NFSVolume.ReadOnly":false,"HostPathVolume.Path":"/xx/xx/name","FlexVolume.Options":"{\\"volumeId\\":\\"d-2zehdahrwoa7srg****\\",\\"performanceLevel\\": \"PL2\\\"}","FlexVolume.Driver":"flexvolume","ConfigFileVolumeDefaultMode":644,"NFSVolume.Server":"3f9cd4a596-naw76.cn-shanghai.nas.aliyuncs.com","DiskVolume.DiskId":"d-xx","Name":"default-volume1","EmptyDirVolume.Medium":"memory"}]
```

```

    ], "ConfigFileVolumeConfigFileToPath": [{"Path":"/usr/bin/","Mode":644,"Content":"bG1lbWk="}]}
  }
  &InitContainer=[{"InitContainerEnvironmentVar":[{"Key":"Path","Value":"/usr/bin/"}],"SecurityContext.Capability.Add":["NET_ADMIN"],"Image":"nginx","InitContainerVolumeMount":[{"MountPropagation":"None","MountPath":"/usr/share/","ReadOnly":false,"Name":"test-empty","SubPath":"/usr/sub/"}],"ImagePullPolicy":"Always","Cpu":0.5,"WorkingDir":"/usr/local","Command":["sleep"],"Arg":["10"]}, "SecurityContext.RunAsUser":587,"Gpu":1,"Memory":1.0,"Name":"test-init"}]
  &DnsConfigNameServer=["172.10.*.*"]
  &DnsConfigSearch=["svc.local.kubernetes"]
  &DnsConfigOption=[{"Value":"value","Name":"name"}]
  &HostAliase=[{"Ip":"192.0.XX.XX","Hostname":["[hehe.com', 'haha.com']]"}]
  &SecurityContextSysctl=[{"Value":65536,"Name":kernel.msgmax}]
  &NtpServer=["ntp.cloud.aliyuncs.com"]
  &AcrRegistryInfo=[{"Domain":["*****-****-registry.cn-beijing.cr.aliyuncs.com"],"InstanceName":"acr-test","InstanceId":"cri-nwj395hgf6f3****","RegionId":cn-hangzhou}]

  &<Common request parameters>

```

Sample success responses

XML format

```

HTTP/1.1 200 OK
Content-Type:application/xml
<ModifyEcisScalingConfigurationResponse>
  <RequestId>89945DD3-9072-47D0-A318-353284CFC7B3</RequestId>
</ModifyEcisScalingConfigurationResponse>

```

JSON format

```

HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "89945DD3-9072-47D0-A318-353284CFC7B3"
}

```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because you are not authorized to call the operation.

HTTP status code	Error code	Error message	Description
404	InvalidDataDiskSnapshotId.NotFound	Snapshot "XXX" does not exist.	The error message returned because the specified snapshot does not exist.
400	InvalidDataDiskSnapshotId.SizeNotSupported	The capacity of snapshot "XXX" exceeds the size limit of the specified disk category.	The error message returned because the size of the specified snapshot exceeds the maximum size allowed for the specified disk.
404	InvalidImageId.NotFound	The specified image does not exist.	The error message returned because the specified image does not exist.
400	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	The error message returned because the name of the specified key pair does not exist.
400	InvalidNetworkType.ForRAMRole	RAMRole can't be used For classic instance.	The error message returned because the network type of the instance is classic network. The classic network does not support RamRoleName.
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned because the value that you specified for a parameter is invalid.
400	InvalidScalingConfigurationName.Duplicate	The specified value of parameter is duplicated.	The error message returned because the specified scaling configuration name already exists.

HTTP status code	Error code	Error message	Description
400	InvalidSecurityGroupId.IncorrectNetworkType	The network type of specified Security Group does not support this action.	The error message returned because the network type of the specified security group is different from the network type of the scaling group.
400	InvalidSecurityGroupId.VPCMismatch	The specified security group and the specified virtual switch are not in the same VPC.	The error message returned because the specified security group and vSwitch are not in the same VPC.
400	InvalidTags.KeyValue	The specified tags key/value cannot be empty.	The error message returned because no value is specified for the Tags parameter.
400	InvalidTags.ListSize	The specified tags list size cannot be more than "20".	The error message returned because the maximum number of tags that can be specified for the elastic container instance has been reached.
400	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded.	The error message returned because the specified user data is not encoded in Base64.
400	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	The error message returned because the user data size exceeds the upper limit.

10.8. DeleteEciScalingConfiguration

Deletes a scaling configuration that is used to create elastic container instances

Description

You cannot delete a scaling configuration that is used to create elastic container instances in the following scenarios:

- The scaling configuration is in the Active state.
- The scaling group contains elastic container instances that are created based on the scaling configuration.

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
Action	String	Yes	DeleteEciScalingConfiguration	The operation that you want to perform. Set the value to DeleteEciScalingConfiguration .
ScalingConfigurationId	String	Yes	asc-bp1bx8mzur534edp***	The ID of the scaling configuration that you want to delete.
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.

Response parameters

Parameter	Type	Example	Description
RequestId	String	45D5B0AD-3B00-4A9B-9911-6D5303B06712	The ID of the request. This parameter is returned regardless of whether the request is successful.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DeleteEciScalingConfiguration  
&ScalingConfigurationId=asc-bp1bx8mzur534edp***  
&RegionId=cn-qingdao  
&<Common request parameters>
```

Sample responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DeleteEciScalingConfigurationResponse>
  <RequestId>45D5B0AD-3B00-4A9B-9911-6D5303B06712</RequestId>
</DeleteEciScalingConfigurationResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "45D5B0AD-3B00-4A9B-9911-6D5303B06712"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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

11. Scaling rule

11.1. CreateScalingRule

Creates a scaling rule.

Description

A scaling rule defines a specific scaling activity, such as adding or removing N instances. If the number of Elastic Compute Service (ECS) instances in a scaling group is less than the minimum number allowed or greater than the maximum number allowed after a scaling rule is executed, Auto Scaling adjusts the number of ECS instances to add or remove. This way, the number of ECS instances can be maintained within the specified range after the scaling rule is executed. The number of ECS instances that is specified in the scaling rule remains unchanged. Examples:

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

Before you call this operation, take note of the following items:

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

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
Action	String	Yes	CreateScalingRule	The operation that you want to perform. Set the value to CreateScalingRule .
ScalingGroupId	String	Yes	asg-bp1ffogfdauy0jw0***	The ID of the scaling group to which the scaling rule belongs.
ScalingRuleName	String	No	scalingrule***	<p>The name of the scaling rule. It must be 2 to 64 characters in length, and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or a digit. The name of a scaling rule must be unique in the scaling group to which the scaling rule belongs within an Alibaba Cloud account.</p> <p>If you do not specify this parameter, the ScalingRuleId value is used.</p>
Cooldown	Integer	No	60	<p>The cooldown time of the scaling rule. This parameter is available only if you set the ScalingRuleType parameter to SimpleScalingRule. Valid values: 0 to 86400. Unit: seconds.</p> <p>By default, this parameter is left empty.</p>
MinAdjustmentMagnitude	Integer	No	1	The minimum number of instances that need to be scaled when the adjustment type is PercentChangeInCapacity. This parameter is available only if you set the ScalingRuleType parameter to SimpleScalingRule or StepScalingRule.

Parameter	Type	Required	Example	Description
AdjustmentType	String	No	QuantityChangeInCapacity	<p>The adjustment method of the scaling rule. This parameter is required only if you set the ScalingRuleType parameter to SimpleScalingRule or StepScalingRule. Valid values:</p> <ul style="list-style-type: none"> • QuantityChangeInCapacity: adds the specified number of ECS instances to or removes the specified number of ECS instances from the scaling group. • PercentChangeInCapacity: adds the specified percentage of ECS instances to or removes the specified percentage of ECS instances from the scaling group. • TotalCapacity: adjusts the number of ECS instances in the scaling group to a specified number.
AdjustmentValue	Integer	No	100	<p>The number of instances that need to be scaled based on the scaling rule. This parameter is required only if you set the ScalingRuleType parameter to SimpleScalingRule or StepScalingRule. The number of ECS instances that need to be scaled in a single scaling activity cannot exceed 1,000. Valid values based on the AdjustmentType value:</p> <ul style="list-style-type: none"> • -1000 to 1000 if you set the AdjustmentType parameter to QuantityChangeInCapacity. • -100 to 10000 if you set the AdjustmentType parameter to PercentChangeInCapacity. • 0 to 2000 if you set the AdjustmentType parameter to TotalCapacity.

Parameter	Type	Required	Example	Description
ScalingRuleType	String	No	SimpleScalingRule	<p>The type of the scaling rule. Valid values:</p> <ul style="list-style-type: none"> • SimpleScalingRule: scales the number of ECS instances based on the AdjustmentType and AdjustmentValue values. • TargetTrackingScalingRule: calculates the number of ECS instances that need to be scaled and attempts to maintain the value of a predefined metric close to the TargetValue value. • StepScalingRule: scales ECS instances in steps based on the specified thresholds and metric values. • PredictiveScalingRule: uses machine learning to analyze historical monitoring data of the scaling group and predicts the future values of metrics. Then, Auto Scaling automatically creates scheduled tasks to specify the boundary values for the scaling group. <p>Default value: SimpleScalingRule.</p>
EstimatedInstanceWarmup	Integer	No	300	<p>The warmup period of an instance. This parameter is available only if you set the ScalingRuleType parameter to TargetTrackingScalingRule or PredictiveScalingRule. Auto Scaling adds ECS instances that are in the warmup state to a scaling group but does not report monitoring data to CloudMonitor during the warmup period.</p> <div style="background-color: #e0f2ff; padding: 10px; margin-top: 10px;"> Note Auto Scaling calculates the number of ECS instances that need to be scaled. ECS instances in the warmup state are not counted towards the current capacity of the scaling group. </div> <p>Valid values: 0 to 86400. Unit: seconds.</p> <p>Default value: 300.</p>

Parameter	Type	Required	Example	Description
MetricName	String	No	CpuUtilization	<p>The predefined metric that you want to monitor. This parameter is required only if you set the <code>ScalingRuleType</code> parameter to <code>TargetTrackingScalingRule</code> or <code>PredictiveScalingRule</code>.</p> <p>Valid values if you set the <code>ScalingRuleType</code> parameter to <code>TargetTrackingScalingRule</code>:</p> <ul style="list-style-type: none"> • <code>CpuUtilization</code>: the average CPU utilization • <code>ClassicInternetRx</code>: the average inbound Internet traffic over the classic network • <code>ClassicInternetTx</code>: the average outbound Internet traffic over the classic network • <code>VpcInternetRx</code>: the average inbound Internet traffic over the virtual private cloud (VPC) • <code>VpcInternetTx</code>: the average outbound Internet traffic over the VPC • <code>IntranetRx</code>: the average inbound traffic over the internal network • <code>IntranetTx</code>: the average outbound traffic over the internal network <p>Valid values if you set the <code>ScalingRuleType</code> parameter to <code>PredictiveScalingRule</code>:</p> <ul style="list-style-type: none"> • <code>CpuUtilization</code>: the average CPU utilization • <code>IntranetRx</code>: the average inbound traffic over the internal network • <code>IntranetTx</code>: the average outbound traffic over the internal network
TargetValue	Float	No	0.125	The target value. This parameter is required only if you set <code>ScalingRuleType</code> to <code>TargetTrackingScalingRule</code> or <code>PredictiveScalingRule</code> . The value must be greater than 0 and can have up to three decimal places.

Parameter	Type	Required	Example	Description
DisableScaleIn	Boolean	No	false	<p>Specifies whether to disable scale-in. This parameter is available only if you set the ScalingRuleType parameter to TargetTrackingScalingRule.</p> <p>Default value: false.</p>
ScaleInEvaluationCount	Integer	No	15	<p>The number of consecutive times that the event-triggered task created for scale-in activities meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule.</p> <p>Default value: 15.</p>
ScaleOutEvaluationCount	Integer	No	3	<p>The number of consecutive times that the event-triggered task created for scale-out activities meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and then associated with the target tracking scaling rule.</p> <p>Default value: 3.</p>
PredictiveScalingMode	String	No	PredictAndScale	<p>The mode of the predictive scaling rule. Valid values:</p> <ul style="list-style-type: none"> • PredictAndScale: provides predictions and creates prediction tasks. • PredictOnly: provides predictions but does not create prediction tasks. <p>Default value: PredictAndScale.</p>

Parameter	Type	Required	Example	Description
PredictiveValueBehavior	String	No	MaxOverridePredictiveValue	<p>The action to take on the predicted maximum value. Valid values:</p> <ul style="list-style-type: none"> • MaxOverridePredictiveValue: uses the initial maximum capacity as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMax: uses the predicted value as the maximum value for prediction tasks if the predicted value is greater than the initial maximum capacity. • PredictiveValueOverrideMaxWithBuffer: increases the predicted value by a percentage that is specified by the PredictiveValueBuffer parameter, and uses the increased value as the maximum value for prediction tasks if the predicted value increased by this percentage is greater than the initial maximum capacity. <p>Default value: MaxOverridePredictiveValue.</p>
PredictiveValueBuffer	Integer	No	50	<p>The percentage of the increment to the predicted value when the PredictiveValueBehavior parameter is set to PredictiveValueOverrideMaxWithBuffer. If the predicted value increased by this percentage is greater than the initial maximum capacity, the increased value is used as the maximum value for prediction tasks. Valid values: 0 to 100.</p> <p>Default value: 0.</p>
PredictiveTaskBufferTime	Integer	No	30	<p>The amount of buffer time before the prediction task is executed. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed on the hour. You can specify a buffer time for resource preparation before prediction tasks are executed. Valid values: 0 to 60. Unit: minutes.</p> <p>Default value: 0.</p>

Parameter	Type	Required	Example	Description
InitialMaxSize	Integer	No	100	The maximum number of ECS instances in the scaling group. This parameter is used together with the PredictiveValueBehavior parameter. The default value of this parameter is the value of MaxSize.
RegionId	String	No	cn-hangzhou	The ID of the region where the scaling group resides.
StepAdjustment.N.MetricIntervalUpperBound	Float	No	5.0	The upper limit value specified in a step adjustment. This parameter is available only if you set ScalingRuleType to StepScalingRule. Valid values: -9.99999E18 to 9.99999E18.
StepAdjustment.N.ScalingAdjustment	Integer	No	1	The number of ECS instances that need to be scaled in a step adjustment. This parameter is available only if you set the ScalingRuleType parameter to StepScalingRule.
StepAdjustment.N.MetricIntervalLowerBound	Float	No	1.0	The lower limit value specified in a step adjustment. This parameter is available only if you set ScalingRuleType to StepScalingRule. Valid values: -9.99999E18 to 9.99999E18.

Response parameters

Parameter	Type	Example	Description
ScalingRuleAri	String	ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp1dvirgwkoowxk7***	The unique identifier of the scaling rule.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Parameter	Type	Example	Description
ScalingRuleId	String	asr-bp1dvirgwkoowxk7***	The ID of the scaling rule, which is generated by the system and is globally unique.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action/CreateScalingRule
&ScalingGroupId=asg-bp1ffogfdauy0jw0****
&ScalingRuleName=scalingrule****
&Cooldown=60
&MinAdjustmentMagnitude=1
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=100
&ScalingRuleType=SimpleScalingRule
&EstimatedInstanceWarmup=300
&MetricName=CpuUtilization
&TargetValue=0.125
&DisableScaleIn=false
&ScaleInEvaluationCount=15
&ScaleOutEvaluationCount=3
&PredictiveScalingMode=PredictAndScale
&PredictiveValueBehavior=MaxOverridePredictiveValue
&PredictiveValueBuffer=50
&PredictiveTaskBufferTime=30
&InitialMaxSize=100
&StepAdjustment=[{"MetricIntervalUpperBound":5.0,"ScalingAdjustment":1,"MetricIntervalLowerBound":1.0}]
&RegionId=cn-hangzhou
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<CreateScalingRuleResponse>
  <ScalingRuleAri>ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bpldvirgwkoowxk7****</ScalingRuleAri>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
  <ScalingRuleId>asr-bp1dvirgwkoowxk7****</ScalingRuleId>
</CreateScalingRuleResponse>
```

JSON format

```

HTTP/1.1 200 OK
Content-Type:application/json
{
  "ScalingRuleAri" : "ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bpldvirgwkoo
wxk7****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
  "ScalingRuleId" : "asr-bpldvirgwkoowxk7****"
}

```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.N otFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Alibaba Cloud account.
400	InvalidScalingRuleName. Duplicate	The specified value of parameter <parameter name> is duplicated.	The error message returned because the specified scaling rule name already exists.
400	QuotaExceeded.Scali ng Rule	Scaling rule quota exceeded in the specified scaling group.	The error message returned because the maximum number of scaling rules in the scaling group has been reached.
400	TargetTrackingScalingR ule.UnsupportedMetric	Specific metric is not supported for target tracking scaling rule.	The error message returned because the target tracking scaling rules do not support the specified metric.
400	TargetTrackingScalingR ule.DuplicateMetric	Only one TargetTrackingScaling rule for a given metric specification is allowed.	The error message returned because the metric is already specified for a target tracking scaling rule in the scaling group.

HTTP status code	Error code	Error message	Description
400	InvalidMinAdjustmentMagnitudeMismatchAdjustmentType	MinAdjustmentMagnitude is not supported by the specified adjustment type.	The error message returned because the MinAdjustmentMagnitude parameter does not support the adjustment method that you specify for the scaling rule.
400	InvalidStepAdjustments. MultipleNullUpperBound	At most one StepAdjustment may have an unspecified upper bound.	The error message returned because a step adjustment that does not have an 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 that does not have a 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, and step adjustments that do not have a lower limit value are unavailable.
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, and step adjustments that do not have an upper limit value are unavailable.
400	InvalidStepAdjustments. Gap	StepAdjustment intervals can not have gaps between them.	The error message returned because the specified ranges of step adjustments have gaps.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.Overlap	StepAdjustment intervals can not overlap.	The error message returned because the specified ranges of step adjustments overlap.
400	InvalidStepAdjustments.LowerGtUpper	LowerBound must be less than the UpperBound for StepAdjustment :%s.	The error message returned because the lower limit 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 can not be left unspecified.	The error message returned because an upper limit value or a lower limit value must be specified for a step adjustment.
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 been reached.
400	StepBeyondPermitRange	Specific parameter "%s" beyond permit range.	The error message returned because the upper limit value or lower limit value specified for a step adjustment is invalid.

11.2. 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 specify a scaling group ID to query all scaling rules in the scaling group. You can also specify the scaling rule ID, name, unique identifier, and type in the request parameters as filter conditions.

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
Action	String	Yes	DescribeScalingRules	The operation that you want to perform. Set the value to DescribeScalingRules.
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group.
PageNumber	Integer	No	1	The number of the page to return. Pages start from page 1. Default value: 1
PageSize	Integer	No	50	The number of entries to return on each page. Valid values: 1 to 50. Default value: 10
ScalingGroupId	String	No	asg-bp1ffogfdauy0jw0****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
ScalingRuleType	String	No	SimpleScalingRule	<p>The type of the scaling rule. Valid values:</p> <ul style="list-style-type: none"> • SimpleScalingRule: simple scaling rules, which scale the number of ECS instances based on the values of AdjustmentType and AdjustmentValue. • TargetTrackingScalingRule: target tracking scaling rules, which calculate the number of ECS instances to be scaled and try to keep the value of a predefined metric close to the value of TargetValue. • StepScalingRule: step scaling rules, which scale the number of ECS instances in steps based on specified thresholds and metric values. • PredictiveScalingRule: predictive scaling rules, which use machine learning to analyze historical monitoring data of the scaling group and predict the future values of metrics. Then, the predictive scaling rules automatically create scheduled tasks to set the boundary values for the scaling group.
ShowAlarmRules	Boolean	No	false	<p>Specifies whether to return Cloud Monitor event-triggered tasks associated with scaling rules. Valid values:</p> <ul style="list-style-type: none"> • true: The Cloud Monitor event-triggered tasks associated with a scaling rule are returned. • false: The Cloud Monitor event-triggered tasks associated with a scaling rule are not returned. <p>Default value: false</p>
ScalingRuleId.N	RepeatList	No	asr-bp1dvirgwkoowxk7***	The ID of scaling rule N to be queried. Valid values of N: 1 to 10.
ScalingRuleName.N	RepeatList	No	scalingrule***	The name of scaling rule N to be queried. Valid values of N: 1 to 10.

Parameter	Type	Required	Example	Description
ScalingRuleAri.N	RepeatList	No	ari:acs:ess:cn-hangzhou:140692647406****:scalngrule/asr-bp1dvirgwkoowxk7****	The unique identifier of scaling rule N to be queried. Valid values of N: 1 to 10.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
ScalingRules	Array of ScalingRule		Details about the scaling rules.
ScalingRule			
AdjustmentType	String	QuantityChangeInCapacity	<p>The adjustment mode of the scaling rule. Valid values:</p> <ul style="list-style-type: none"> • QuantityChangeInCapacity: adds or removes the specified number of ECS instances to or from the scaling group. • PercentChangeInCapacity: adds or removes the specified percentage of ECS instances to or from the scaling group. • TotalCapacity: adjusts the number of ECS instances in the scaling group to the specified number.
AdjustmentValue	Integer	1	The adjustment value specified for the scaling rule.
Alarms	Array of Alarm		The Cloud Monitor event-triggered tasks associated with scaling rules. The Cloud Monitor event-triggered tasks associated with scaling rules are returned only when the ShowAlarmRules parameter is set to true. Otherwise, an empty list is returned.

Parameter	Type	Example	Description
Alarm			
AlarmTaskId	String	asg-bp18p2yfxow2dloq* *** _1f9458d1-70e1-4bee-8c7f-7a47695b***	The ID of the event-triggered task associated with the scaling rule.
AlarmTaskName	String	alarmtask***	The name of the event-triggered task associated with the scaling rule.
ComparisonOperator	String	>=	<p>The comparison operator between the metric value and the threshold for the event-triggered task associated with the scaling rule. It indicates the relationship in which the metric value and the threshold can meet the condition. Valid values:</p> <ul style="list-style-type: none"> • >=: The metric value is greater than or equal to the threshold. • <=: The metric value is less than or equal to the threshold. • >: The metric value is greater than the threshold. • <: The metric value is less than the threshold.
Dimensions	Array of Dimension		The dimensions of the event-triggered task associated with the scaling rule.
Dimension			
DimensionKey	String	scaling_group	<p>The key values of dimensions associated with a metric. Valid values:</p> <ul style="list-style-type: none"> • 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.
EvaluationCount	Integer	3	The number of consecutive times when the event-triggered task associated with a scaling rule meets the threshold expression to trigger an alert.

Parameter	Type	Example	Description
MetricName	String	CpuUtilization	The name of the metric of the event-triggered task associated with the scaling rule.
Statistics	String	Average	The statistical method of the event-triggered task associated with the scaling rule. Valid values: <ul style="list-style-type: none">• Average• Maximum• Minimum
Threshold	Float	50	The alert threshold for the event-triggered task associated with the scaling rule.
Cooldown	Integer	20	The cooldown period of the scaling rule. This parameter takes effect only when ScalingRuleType is set to SimpleScalingRule. Valid values: 0 to 86400. Unit: seconds.
DisableScaleIn	Boolean	true	Indicates whether scale-in is disabled. This parameter takes effect only when ScalingRuleType is set to TargetTrackingScalingRule. Valid values: <ul style="list-style-type: none">• 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	<p>The predefined metric to monitor. This parameter is required and takes effect only when ScalingRuleType is set to TargetTrackingScalingRule or PredictiveScalingRule.</p> <p>Valid values for target tracking scaling rules:</p> <ul style="list-style-type: none"> • CpuUtilization: the average CPU utilization • ClassicInternetRx: the average inbound public traffic over the classic network • ClassicInternetTx: the average outbound public traffic over the classic network • VpcInternetRx: the average inbound public traffic over a VPC • VpcInternetTx: the average outbound public traffic over a VPC • IntranetRx: the average inbound internal traffic • IntranetTx: the average outbound internal traffic <p>Valid values for predictive scaling rules:</p> <ul style="list-style-type: none"> • CpuUtilization: the average CPU utilization • IntranetRx: the average inbound internal traffic • IntranetTx: the average outbound internal traffic
MinAdjustmentMagnitude	Integer	1	The minimum number of ECS instances that can be scaled by using a scaling rule. This parameter takes effect only when ScalingRuleType is set to SimpleScalingRule or StepScalingRule and AdjustmentType is set to PercentChangeInCapacity.
MinSize	Integer	1	The minimum number of ECS instances in the scaling group.
PredictiveScalingMode	String	PredictAndScale	The mode of the predictive scaling rule. Valid values: <ul style="list-style-type: none"> • PredictAndScale: produces predictions and creates prediction tasks. • PredictOnly: produces predictions but does not create prediction tasks.

Parameter	Type	Example	Description
PredictiveTaskBufferTime	Integer	30	The amount of buffer time ahead of when the prediction task is executed. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed on the hour. You can 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.
PredictiveValueBehavior	String	MaxOverridePredictiveValue	<p>The action to take on the predicted maximum value. Valid values:</p> <ul style="list-style-type: none"> • 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 increased value is used as the maximum value for prediction tasks.
PredictiveValueBuffer	Integer	50	The ratio of the increment to the predicted value when PredictiveValueBehavior is set to PredictiveValueOverrideMaxWithBuffer. If the predicted value increased by this ratio is greater than the initial maximum capacity, the increased value is used as the maximum value for prediction tasks. Valid values: 0 to 100.

Parameter	Type	Example	Description
ScaleInEvaluationCount	Integer	15	The number of consecutive times when the event-triggered task created for scale-in events meets the threshold conditions to trigger an alert. After a target tracking scaling rule is created, an event-triggered task is automatically created and associated with the target tracking scaling rule.
ScaleOutEvaluationCount	Integer	3	The number of consecutive times when the event-triggered task created for scale-out events meets the threshold conditions to trigger an alert. After a target tracking scaling rule is created, an event-triggered task is automatically created and associated with the target tracking scaling rule.
ScalingGroupId	String	asg-bp1ffogfdauy0jw0**	The ID of the scaling group.
ScalingRuleArn	String	ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp1dvirgwkoowxk7***	The unique identifier of the scaling rule.
ScalingRuleId	String	asr-bp1dvirgwkoowxk7***	The ID of the scaling rule.
ScalingRuleName	String	scalingrule****	The name of the scaling rule.

Parameter	Type	Example	Description
ScalingRuleType	String	SimpleScalingRule	<p>The type of the scaling rule. Valid values:</p> <ul style="list-style-type: none"> • SimpleScalingRule: simple scaling rules, which scale the number of ECS instances based on the values of AdjustmentType and AdjustmentValue. • TargetTrackingScalingRule: target tracking scaling rules, which calculate the number of ECS instances to be scaled and try to keep the value of a predefined metric close to the value of TargetValue. • StepScalingRule: step scaling rules, which scale ECS instances in steps based on specified thresholds and metric values. • PredictiveScalingRule: predictive scaling rules, which use machine learning to analyze historical monitoring data of the scaling group and predict the future values of metrics. Then, the predictive scaling rules automatically create scheduled tasks to set the boundary values for the scaling group.
StepAdjustments	Array of StepAdjustment		The step adjustments of the step scaling rule.
StepAdjustment			
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

```
https://ess.aliyuncs.com/?Action=DescribeScalingRules  
&RegionId=cn-qingdao  
&ScalingGroupId=asg-bplffogfdauy0jw0****  
&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>asg-bplffogfdauy0jw0****</ScalingGroupId>  
            <ScalingRuleAri>ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-  
bpldvirgwkoowxk7****</ScalingRuleAri>  
            <ScalingRuleId>asr-bpldvirgwkoowxk7****</ScalingRuleId>  
  
            <ScalingRuleName>scalingrule****</ScalingRuleName>  
        </ScalingRule>  
    </ScalingRules>  
    <TotalCount>1</TotalCount>  
    <RequestId>3306A40D-3412-4101-9F19-5F81E3055DAD</RequestId>  
</DescribeScalingRulesResponse>
```

JSON format

```
{  
    "RequestId": "B583BFEF-A779-427A-9B74-262DDD249702",  
    "TotalCount": 1,  
    "PageNumber": 1,  
    "PageSize": 10,  
    "ScalingRules": {  
        "ScalingRule": [  
            {  
                "ScalingRuleId": "asr-bp1dvirgwkoowxk7****",  
                "ScalingRuleAri": "ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr  
-bp1dvirgwkoowxk7****",  
                "Cooldown": 500,  
                "ScalingGroupId": "asg-bp1ffogfdauy0jw0****",  
                "AdjustmentType": "TotalCapacity",  
                "ScalingRuleName": "scalingrule****",  
                "AdjustmentValue": 5  
            }  
        ]  
    }  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

11.3. ModifyScalingRule

Modifies a scaling rule.

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
Action	String	Yes	ModifyScalingRule	The operation that you want to perform. Set the value to ModifyScalingRule .
ScalingRuleId	String	Yes	asr-bp1dvirgwkoowxk7****	The ID of the scaling rule that you want to modify.

Parameter	Type	Required	Example	Description
ScalingRuleName	String	No	scalingrule****	The name of the scaling rule. It must be 2 to 64 characters in length and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or a digit. The name of a scaling rule must be unique within the scaling group to which it belongs for an Alibaba Cloud account.
Cooldown	Integer	No	60	<p>The cooldown time of the scaling rule. This parameter is valid only when the ScalingRuleType parameter is set to SimpleScalingRule.</p> <p>Valid values: 0 to 86,400. Unit: seconds.</p>
MinAdjustmentMagnitude	Integer	No	1	The minimum value to scale by when the adjustment type is PercentChangeInCapacity. This parameter is valid only when the ScalingRuleType parameter is set to SimpleScalingRule or StepScalingRule and the AdjustmentType parameter is set to PercentChangeInCapacity.

Parameter	Type	Required	Example	Description
AdjustmentType	String	No	QuantityChangeInCapacity	<p>The adjustment method of the scaling rule. This parameter is valid and required only when the ScalingRuleType parameter is set to SimpleScalingRule or StepScalingRule. Valid values:</p> <ul style="list-style-type: none">• QuantityChangeInCapacity: adds or removes the specified number of ECS instances to or from the scaling group.• PercentChangeInCapacity: adds or removes the specified percentage of ECS instances to or from the scaling group.• TotalCapacity: adjusts the number of ECS instances in the scaling group to the specified number.
AdjustmentValue	Integer	No	100	<p>The adjustment value specified for the scaling rule. This parameter is valid and required only when the ScalingRuleType parameter is set to SimpleScalingRule or StepScalingRule. The number of ECS instances to be scaled in a single scaling activity cannot exceed 1,000.</p> <ul style="list-style-type: none">• Valid values when the AdjustmentType parameter is set to QuantityChangeInCapacity: -1000 to 1000• Valid values when the AdjustmentType parameter is set to PercentChangeInCapacity: -100 to 10000• Valid values when the AdjustmentType parameter is set to TotalCapacity: 0 to 2000

Parameter	Type	Required	Example	Description
EstimatedInstancesWarmup	Integer	No	60	<p>The warmup period of an ECS instance. This parameter is valid only when the ScalingRuleType parameter is set to TargetTrackingScalingRule or PredictiveScalingRule. Auto Scaling adds ECS instances that are in the warmup state to a scaling group but does not report monitoring data to CloudMonitor during the warmup period.</p> <p>? Note Auto Scaling calculates the number of ECS instances to be scaled. ECS instances in the warmup state are not counted towards the current capacity of the scaling group.</p> <p>Valid values: 0 to 86,400. Unit: seconds.</p>

Parameter	Type	Required	Example	Description
MetricName	String	No	CpuUtilization	<p>The predefined metric that you want to monitor. This parameter is valid and required only when the <code>ScalingRuleType</code> parameter is set to <code>TargetTrackingScalingRule</code> or <code>PredictiveScalingRule</code>.</p> <p>Valid values when the <code>ScalingRuleType</code> parameter is set to <code>TargetTrackingScalingRule</code>:</p> <ul style="list-style-type: none"> • <code>CpuUtilization</code>: the average CPU utilization • <code>ClassicInternetRx</code>: the average inbound public traffic volume over the classic network • <code>ClassicInternetTx</code>: the average outbound public traffic volume over the classic network • <code>VpcInternetRx</code>: the average inbound public traffic volume over a VPC • <code>VpcInternetTx</code>: the average outbound public traffic volume over a VPC • <code>IntranetRx</code>: the average inbound internal traffic volume • <code>IntranetTx</code>: the average outbound internal traffic volume <p>Valid values when the <code>ScalingRuleType</code> parameter is set to <code>PredictiveScalingRule</code>:</p> <ul style="list-style-type: none"> • <code>CpuUtilization</code>: the average CPU utilization • <code>IntranetRx</code>: the average inbound internal traffic volume • <code>IntranetTx</code>: the average outbound internal traffic volume
TargetValue	Float	No	0.125	The target value. This parameter is valid only when the <code>ScalingRuleType</code> parameter is set to <code>TargetTrackingScalingRule</code> or <code>PredictiveScalingRule</code> . The value must be greater than 0 and can have a maximum of three decimal places.

Parameter	Type	Required	Example	Description
DisableScaleIn	Boolean	No	true	Specifies whether to disable scale-in. This parameter is valid only when the ScalingRuleType parameter is set to TargetTrackingScalingRule.
ScaleInEvaluationCount	Integer	No	15	The number of consecutive times that the event-triggered task created for scale-in events meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and associated with the target tracking scaling rule.
ScaleOutEvaluationCount	Integer	No	3	The number of consecutive times that the event-triggered task created for scale-in events meets the threshold conditions before an alert is triggered. After a target tracking scaling rule is created, an event-triggered task is automatically created and associated with the target tracking scaling rule.
PredictiveScalingMode	String	No	PredictAndScale	<p>The mode of the predictive scaling rule. Valid values:</p> <ul style="list-style-type: none"> • PredictAndScale: produces predictions and creates prediction tasks. • PredictOnly: produces predictions but does not create prediction tasks.

Parameter	Type	Required	Example	Description
PredictiveValueBehavior	String	No	MaxOverridePredictiveValue	<p>The action you want to take on the predicted maximum value. Valid values:</p> <ul style="list-style-type: none"> • 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 percentage which is specified by the PredictiveValueBuffer parameter, and uses the increased value as the maximum value for prediction tasks if the predicted value increased by this percentage is greater than the initial maximum capacity.
PredictiveValueBuffer	Integer	No	50	<p>The percentage of the increment to the predicted value when the PredictiveValueBehavior parameter is set to PredictiveValueOverrideMaxWithBuffer. If the predicted value increased by this percentage is greater than the initial maximum capacity, the increased value is used as the maximum value for prediction tasks. Valid values: 0 to 100.</p>
PredictiveTaskBufferTime	Integer	No	30	<p>The amount of buffer time specified before the prediction task is executed. By default, all scheduled tasks that are automatically created for a predictive scaling rule are executed on the hour. You can set a buffer time before prediction tasks are executed to allow resources to be prepared in advance. Valid values: 0 to 60.</p>

Parameter	Type	Required	Example	Description
InitialMaxSize	Integer	No	100	The maximum number of ECS instances in the scaling group. This parameter is used together with the PredictiveValueBehavior parameter.
StepAdjustment.N.MetricIntervalUpperBound	Float	No	5.0	The upper limit value specified in a step adjustment. This parameter is valid only when the ScalingRuleType parameter is set to StepScalingRule. Valid values: -9.99999E18 to 9.99999E18.
StepAdjustment.N.ScalingAdjustment	Integer	No	1	The specified number of ECS instances to be scaled in a step adjustment. This parameter is valid only when the ScalingRuleType parameter is set to StepScalingRule.
StepAdjustment.N.MetricIntervalLowerBound	Float	No	1.0	The lower limit value specified in a step adjustment. This parameter is valid only when the ScalingRuleType parameter is set to StepScalingRule. Valid values: -9.99999E18 to 9.99999E18.

Response parameters

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

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=ModifyScalingRule
&ScalingRuleId=asg-bp1ffogfdauy0jw0****
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=-10
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<ModifyScalingRuleResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</ModifyScalingRuleResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
404	InvalidScalingRuleId.NotFound	The specified scaling rule does not exist.	The error message returned because the specified scaling rule does not exist 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 maximum number of scaling rules in the specified scaling group has been reached.
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 without an 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 without a 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 without a lower limit value does not exist.

HTTP status code	Error code	Error message	Description
400	InvalidStepAdjustments.NoNullUpperBound	There must be a StepAdjustment with an unspecified upper bound when one StepAdjustment has a positive upper bound.	The error message returned because the upper limit value of a step adjustment is positive but a different step adjustment without an upper limit value does not exist.
400	InvalidStepAdjustments.Gap	StepAdjustment intervals can not have gaps between them.	The error message returned because the specified ranges of step adjustments have gaps between them.
400	InvalidStepAdjustments.Overlap	StepAdjustment intervals can not overlap.	The error message returned because the specified ranges of step adjustments overlap.
400	InvalidStepAdjustments.LowerGtUpper	LowerBound must be less than the UpperBound for StepAdjustment :%s.	The error message returned because the lower limit 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 can not be left unspecified.	The error message returned because the upper limit value or the lower limit value for a step adjustment must be 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 been reached.

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 invalid.

11.4. DeleteScalingRule

Deletes a scaling rule.

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
Action	String	Yes	DeleteScalingRule	The operation that you want to perform. Set the value to DeleteScalingRule .
ScalingRuleId	String	Yes	asr-bp163l21e07uhnyt****	The ID of the scaling rule that you want to delete.
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DeleteScalingRule
&ScalingRuleId=asr-bp163121e07uhnyt****
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

[XML](#) **format**

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DeleteScalingRuleResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DeleteScalingRuleResponse>
```

[JSON](#) **format**

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
404	InvalidScalingRuleId.NotFound	The specified scaling rule does not exist.	The error message returned because the specified scaling rule does not exist in the current account.

12. Trigger task

12.1. ExecuteScalingRule

You can call this operation to execute a scaling rule.

Description

Before you call this operation, make sure that the following conditions are met:

- The scaling group is in the Active state.
- The scaling group does not have scaling activities in progress.

If no scaling activity is being executed in the scaling group, this operation can trigger scaling activities immediately without waiting for the cooldown time to expire.

If the call is successful, Auto Scaling has accepted the request. However, this does not mean that the scaling activity will succeed. You can determine the status of a scaling activity based on the return value of the ScalingActivityId parameter.

If the number of ECS instances to be added will leave the number of instances in the scaling group greater than MaxSize, Auto Scaling adds the proper number of ECS instances to maintain the number of ECS instances in the scaling group equal to the MaxSize value.

If the number of ECS instances to be removed will leave the number of instances in the scaling group less than MinSize, Auto Scaling removes the proper number of ECS instances to maintain the number of ECS instances in the scaling group equal to the MinSize value.

A limited number of ECS instances can be adjusted at a time. For more information, see the description about the AdjustmentValue parameter in [CreateScalingRule](#).

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
Action	String	Yes	ExecuteScalingRule	The operation that you want to perform. Set the value to ExecuteScalingRule.

Parameter	Type	Required	Example	Description
ScalingRuleAri	String	Yes	ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp1dvirgwkoowxk7***	<p>The unique identifier of the scaling rule.</p> <p>Note You can call the ExecuteScalingRule operation to execute only simple scaling rules and step scaling rules. To execute a step scaling rule, you must specify both BreachThreshold and MetricValue.</p>
ClientToken	String	No	123e4567-e89b-12d3-a456-426655440000	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence.</p>
BreachThreshold	Float	No	1.0	<p>The threshold specified when the step scaling rule is executed. Valid values: -9.999999E18~9.999999E18.</p>
MetricValue	Float	No	1.0	<p>The metric value specified when the step scaling rule is executed. Valid values: -9.999999E18~9.999999E18.</p>

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3C83E	The ID of the request.
ScalingActivityId	String	asa-bp13o672yeautiil****	The ID of the scaling activity.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=ExecuteScalingRule
&ScalingRuleAri=ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp1dvirgwkoowxk7**
*&
&<Common request parameters>
```

Sample success responses

XML **format**

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

JSON **format**

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

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
404	InvalidScalingRuleAri.No tFound	The specified scaling rule Ari does not exist.	The error message returned because 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 because Auto Scaling is not authorized to call the specified operation.
400	IncorrectScalingGroupSt atus	The current status of the specified scaling group does not support this action.	The error message returned because the scaling group to which the specified scaling rule belongs is not in the Active state.

HTTP status code	Error code	Error message	Description
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because the scaling group to which the specified scaling rule belongs has a scaling activity in progress.
400	InsufficientBalance	Your account does not have enough balance.	The error message returned because the balance in your account is insufficient. Add funds to your account and try again.
400	QuotaExceed.Instance	Living instance quota exceeded.	The error message returned because the maximum number of ECS instances has been reached.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified Server Load Balancer (SLB) instance associated with the scaling group to which the specified scaling rule belongs is not in the Active state.
400	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	The error message returned because health check is not enabled for 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 because the network type of the ECS instance associated with 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 associated with the specified SLB instance is not in the same VPC as the vSwitch specified by VSwitchID.
400	IncorrectDBInstanceState	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance of the scaling group to which the specified scaling rule belongs is not in the Running state.
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because for the scaling group to which the specified scaling rule belongs, the maximum number of IP addresses in the whitelist of the associated ApsaraDB RDS instance has been reached.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the maximum number of ECS instances in the specified security group has been reached.

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

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

12.2. ScaleWithAdjustment

Scales instances in a scaling group based on the specified adjustment policy.

Description

Different from the ExecuteScalingRule operation, the ScaleWithAdjustment operation does not require a scaling rule that is created in advance.

Before you call the ScaleWithAdjustment operation, take note of the following items:

- The following conditions must be met:
 - The scaling group is in the Active state.
 - No scaling activities in the scaling group are in progress.
 - The Expected Number of Instances feature is disabled for the scaling group.
- If no scaling activities in the scaling group are in progress, you can call the ScaleWithAdjustment operation to scale instances in the scaling group, without the need to wait for the cooldown time to end.
- If the addition of a specified number of Elastic Compute Service (ECS) instances to a scaling group causes the total number of ECS instances in the scaling group to exceed the maximum number of instances allowed, Auto Scaling adds only some ECS instances to ensure that the total number of

- instances is equal to the maximum number of instances allowed.
- If the removal of a specified number of ECS instances from a scaling group causes the total number of ECS instances in the scaling group to drop below the minimum number of instances allowed, Auto Scaling removes only some ECS instances to ensure that the total number of instances is equal to the minimum number of instances allowed.

If the call is successful, Auto Scaling accepted the request. However, this does not ensure a successful scaling activity. You can obtain the status of a scaling activity based on the return value of the `ScalingActivityId` 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
ScalingGroupId	String	Yes	asg-j6c1o397427hyjdcgczx	The ID of the scaling group.
AdjustmentType	String	Yes	QuantityChangeInCapacity	<p>The type of the adjustment policy. Valid values:</p> <ul style="list-style-type: none">QuantityChangeInCapacity: adds the specified number of ECS instances to or removes the specified number of ECS instances from the scaling group.PercentChangeInCapacity: adds the specified percentage of ECS instances to or removes the specified percentage of ECS instances from the scaling group.TotalCapacity: adjusts the number of ECS instances in the scaling group to the specified number.

Parameter	Type	Required	Example	Description
AdjustmentValue	Integer	Yes	100	<p>The number of instances in each adjustment. The number of ECS instances in each adjustment cannot exceed 1,000.</p> <ul style="list-style-type: none"> • Valid values if you set the AdjustmentType parameter to QuantityChangeInCapacity: -1000 to 1000. • Valid values if you set the AdjustmentType parameter to PercentChangeInCapacity: -100 to 10000. • Valid values if you set the AdjustmentType parameter to TotalCapacity: 0 to 2000.
MinAdjustmentMagnitude	Integer	No	1	The minimum number of instances allowed in each adjustment. This parameter takes effect only if you set the AdjustmentType parameter to PercentChangeInCapacity.
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests. The token can only contain ASCII characters and cannot exceed 64 characters in length.
Action	String	Yes	ScaleWithAdjustment	The operation that you want to perform. Set the value to ScaleWithAdjustment .

Response parameters

Parameter	Type	Example	Description
ScalingActivityId	String	asa-bp175o6f6ego3r2j**	The ID of the scaling activity.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3C83E	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?ScalingGroupId=asg-j6c1o397427hyjdcgczx  
&AdjustmentType=QuantityChangeInCapacity  
&AdjustmentValue=100  
&MinAdjustmentMagnitude=1  
&ClientToken=123e4567-e89b-12d3-a456-42665544****  
&Action=ScaleWithAdjustment  
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK  
Content-Type:application/xml  
<ScaleWithAdjustmentResponse>  
    <ScalingActivityId>asa-bp175o6f6ego3r2j****</ScalingActivityId>  
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>  
</ScaleWithAdjustmentResponse>
```

JSON format

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
    "ScalingActivityId" : "asa-bp175o6f6ego3r2j****",  
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
404	InvalidScalingRuleAri.No tFound	The specified scaling rule Ari does not exist.	The error message returned because the specified scaling rule does not exist within the Alibaba Cloud account.

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the operation.
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the scaling group to which the specified scaling rule belongs is not in the Active state.
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because the scaling group to which the specified scaling rule belongs has a scaling activity in progress.
400	InsufficientBalance	Your account does not have enough balance.	The error message returned because the account balance is insufficient.
400	QuotaExceed.Instance	Living instance quota exceeded.	The error message returned because the maximum number of ECS instances has been reached.

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

HTTP status code	Error code	Error message	Description
400	IncorrectDBInstanceState	The current status of DB instance "XXX" does not support this action.	The error message returned because the ApsaraDB RDS instance that is associated with the scaling group to which the specified scaling rule belongs is not in the Running state.
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because the maximum number of IP addresses in the whitelist that manages access to the ApsaraDB RDS instance associated with the scaling group to which the specified scaling rule belongs has been reached.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the maximum number of ECS instances associated with the specified security group has been reached.
400	IncorrectCapacity.NoChange	To execute the specified scaling rule, the total capacity will not change.	The error message returned because no changes are made to the number of instances in the scaling group 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 because the maximum number of ECS instances that can be scaled has been reached.
400	QuotaExceeded.AfterpayInstance	Living afterpay instance quota exceeded.	The error message returned because the maximum number of pay-as-you-go ECS instances has been reached.
400	ResourceNotAvailable.ECS	The specified region or zone does not offer the specified disk or instance category.	The error message returned because ECS instances of the specified instance type or ECS instances that use the specified disk category cannot be created in the specified region.
400	ScalingRule.InvalidScalingRuleType	Specific scaling rule type: %s can not be executed.	The error message returned because the scaling rule of the specified type cannot be executed.
400	InvalidStepAdjustments.NoStepFound	No adjustment step found for a metric value of: %s.	The error message returned because no matching adjustment steps are found for the specified metric.
400	MissingParameter.MetricValue	Metric value must be specified for StepScalingRule.	The error message returned because no metric value is specified for the step scaling rule.

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

12.3. AttachInstances

Manually adds Elastic Compute Service (ECS) instances or elastic container instances to a scaling group.

Description

Before you call this operation, make sure that the following conditions are met:

- The scaling group is in the Active state.
- No scaling activities in the scaling group are in progress.

The ECS instances and the elastic container instances that you want to add to a scaling group must meet the following requirements:

- The instances reside in the same region as the scaling group.
- The instances are in the Running state.
- The instances are not added to other scaling groups.
- The instances use the subscription or pay-as-you-go billing method, or are preemptible instances.
- If the VswitchID parameter is specified for a scaling group, the instances that are in the classic network or those that are not in the same virtual private cloud (VPC) as the specified vSwitch cannot be added to the scaling group.
- If the VswitchID parameter is not specified for a scaling group, the instances that are in VPCs cannot be added to the scaling group.

If no scaling activities in the specified scaling group are in progress, the operation can trigger scaling activities even before the cooldown time expires.

A successful call indicates that Auto Scaling accepts the request. However, the scaling activity may still fail. You can obtain the status of a scaling activity by using the value of the ScalingActivityId parameter in the response.

If the sum of the number of instances that you want to add and the number of existing instances in the scaling group is greater than the MaxSize value, the call fails.

Instances that are manually added by calling the AttachInstances operation 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
Action	String	Yes	AttachInstances	The operation that you want to perform. Set the value to AttachInstances .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dl0q***	The ID of the scaling group.
Entrusted	Boolean	No	false	<p>Specifies whether the scaling group manages the lifecycle of instances that are manually added to the scaling group. Valid values:</p> <ul style="list-style-type: none">• true: The scaling group manages the lifecycle of instances that are manually added in a similar manner in which the scaling group manages the lifecycle of automatically created instances. When the instances are removed from the scaling group, the instances are automatically released. However, if you call the <code>DetachInstances</code> operation to remove the instances from the scaling group, the instances are not released.• false: The scaling group does not manage the lifecycle of instances that are manually added. After the instances are removed from the scaling group, the instances are not released. <p>Note This parameter is unavailable for subscription instances.</p> <p>Default value: false.</p>

Parameter	Type	Required	Example	Description
LifecycleHook	Boolean	No	false	<p>Specifies whether to trigger a lifecycle hook for the scaling group to which instances are being added. Valid values:</p> <ul style="list-style-type: none"> • true • false <p>Default value: false.</p>
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.
InstanceId.N	String	No	i-28wt4****	<p>The ID of instance N that you want to add. Valid values of N: 1 to 20.</p> <p>Examples:</p> <ul style="list-style-type: none"> • The ID of the ECS instance that you want to add is <code>i-28wt4****</code>. • The ID of the elastic container instance that you want to add is <code>eci-bp17gw49eu09yiwm****</code>.
LoadBalancerWeight.N	Integer	No	50	<p>The weight of instance N as a backend server of the Server Load Balancer (SLB) instance that is associated with the scaling group to which you want to add instance N. Valid values of N: 1 to 20. Valid values of this parameter: 1 to 100.</p> <p>Default value: 50.</p>

Response parameters

Parameter	Type	Example	Description
ScalingActivityId	String	asa-bp1crxor24s28xf1***	The ID of the scaling activity.

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=AttachInstances
&ScalingGroupId=asg-bp18p2yfxow2dlog****
&Entrusted=false
&InstanceId=["i-28wt4****"]
&LoadBalancerWeight=[50]
&LifecycleHook=false
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<AttachInstancesResponse>
  <ScalingActivityId>asa-bplcrxor24s28xf1****</ScalingActivityId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</AttachInstancesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "ScalingActivityId" : "asa-bplcrxor24s28xf1****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description

HTTP status code	Error code	Error message	Description
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Alibaba Cloud account.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because Auto Scaling is not authorized to call the specified operation.
400	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	The error message returned because the specified scaling group is not in the Active state.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the specified instance does not exist within the Alibaba Cloud account.
400	InvalidInstanceId.RegionMismatch	Instance "XXX" and the specified scaling group are not in the same Region.	The error message returned because the specified 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 instance type of the specified instance is different from the instance type in the active scaling configuration.

HTTP status code	Error code	Error message	Description
400	IncorrectInstanceState	The current status of instance "XXX" does not support this action.	The error message returned because the specified instance is not in the Running state.
400	InvalidInstanceId.NetworkTypeMismatch	The network type of instance "XXX" does not support this action.	The error message returned because the network type of the specified 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 instance that you want to add 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 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 scaling activities are in progress in the specified scaling group.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified SLB instance that is associated with the scaling group is not in the Active 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 the health check feature is disabled for the SLB instance that is associated with the specified scaling group.
400	InvalidLoadBalancerId.InvalidInstanceNetworkType	The network type of the instance in specified load balancer does not support this action.	The error message returned because the network type of an instance that is attached to the specified SLB instance is different from the network type of the scaling group.
400	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in specified load balancer are not in the same VPC.	The error message returned because an instance that is attached to the specified SLB instance is not in the same VPC as the vSwitch that is specified by the VSwitchID parameter.
400	IncorrectDBInstanceState	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance that is associated with the scaling group is not in the Running state.

HTTP status code	Error code	Error message	Description
400	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance "XXX".	The error message returned because the maximum number of IP addresses in the whitelist that manages access to the ApsaraDB RDS instance that is associated with the specified scaling group has been reached.
400	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	The error message returned because the maximum number of instances that are associated with the specified security group has been reached.
400	IncorrectCapacity.MaxValue	To attach the instances, the total capacity will be greater than the MaxSize.	The error message returned because the total number of instances as indicated by Total Capacity exceeds the specified value of MaxSize after you add instances to the specified scaling group.

12.4. RemoveInstances

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

Description

- Before you call this operation, make sure that the following requirements are met:
 - The scaling group is in the Active state.
 - No scaling activity is in progress within the scaling group.

If no scaling activity is in progress within the scaling group, you can call the operation even within the cooldown period.

- If an ECS instance is automatically created by Auto Scaling, or if an ECS instance is manually added to a scaling group and managed by the scaling group, the ECS instance enters the economical mode or is released when the instance is removed from the scaling group.
- If an ECS instance is manually added to a scaling group and is not managed by the scaling group, the

- ECS instance is not stopped or released when the instance is removed from the scaling group.
- If the difference between the number of existing ECS instances specified by TotalCapacity and the number of ECS instances that you specified to remove is less than the value of MinSize, the call fails. A successful call indicates that Auto Scaling accepts the request. However, the scaling activity may still fail. You can obtain the status of a scaling activity based on the value of the ScalingActivityId parameter in the response.

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
Action	String	Yes	RemoveInstances	The operation that you want to perform. Set the value to RemoveInstances .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dl0q***	The ID of the scaling group.

Parameter	Type	Required	Example	Description
RemovePolicy	String	No	release	<p>The action that is performed when ECS instances are removed. Valid values:</p> <ul style="list-style-type: none">• recycle: The ECS instances enter the economical mode. <p>Note This setting takes effect only if you set ScalingPolicy to recycle.</p> <ul style="list-style-type: none">• release: The ECS instances are released. <p>The ScalingPolicy parameter that you specify when you call the CreateScalingGroup operation specifies the reclaim mode of the scaling group. The RemovePolicy parameter that you specify when you call the RemoveInstances operation specifies the action to be performed on ECS instances when the ECS instances are removed. Examples:</p> <ul style="list-style-type: none">• If you set ScalingPolicy and RemovePolicy to recycle, the ECS instances enter the economical mode when they are removed.• If you set ScalingPolicy to recycle and RemovePolicy to release, the ECS instances are released when they are removed.• If you set ScalingPolicy to release and RemovePolicy to recycle, the ECS instances are released when they are removed.• If you set ScalingPolicy and RemovePolicy to release, the ECS instances are released when they are removed. <p>Default value: release.</p>

Parameter	Type	Required	Example	Description
DecreaseDesiredCapacity	Boolean	No	true	<p>Specifies whether to adjust the expected number of ECS instances in the scaling group. Valid values:</p> <ul style="list-style-type: none"> • true: After ECS instances are removed from a scaling group, the expected number of ECS instances in the scaling group decreases. • false: After ECS instances are removed from a scaling group, the expected number of ECS instances in the scaling group remains unchanged. <p>Default value: true.</p>
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.
InstanceId.N	String	Yes	i-28wt4****	The ID of ECS instance N that you want to remove. Valid values of N: 1 to 20.

Response parameters

Parameter	Type	Example	Description
ScalingActivityId	String	asa-bp175o6f6ego3r2j**	The ID of the scaling activity.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=RemoveInstances
&ScalingGroupId=asg-bp18p2yfxow2dlog****
&RemovePolicy=release
&DecreaseDesiredCapacity=true
&InstanceId=["i-28wt4****"]
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<RemoveInstancesResponse>
  <ScalingActivityId>asa-bp175o6f6ego3r2j****</ScalingActivityId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</RemoveInstancesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "ScalingActivityId" : "asa-bp175o6f6ego3r2j****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [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 within the Alibaba Cloud account.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the specified ECS instance does not exist in the scaling group.
400	InvalidParameter	The specified group does not support the specified RemovePolicy.	The error message returned because the specified scaling group does not support the RemovePolicy setting.
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 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 scaling group is not in the Active state.
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because a scaling activity is in progress within the scaling group.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the Server Load Balancer (SLB) instance that is associated with the scaling group to which the specified scaling rule applies is not in the Active state.
400	IncorrectDBInstanceState	The current status of DB instance "XXX" does not support this action.	The error message returned because the ApsaraDB RDS instance that is associated with the scaling group to which the specified scaling rule applies is not in the Running state.
400	IncorrectCapacity.MinSize	To remove the instances, the total capacity will be lesser than the MinSize.	The error message returned because the difference between the number of existing ECS instances specified by TotalCapacity and the number of ECS instances that you specified to remove is less than the value of MinSize.

12.5. DetachInstances

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

Description

After ECS instances or elastic container instances are removed from a scaling group, you can call the [AttachInstances](#) operation to add the ECS instances or elastic container instances that are removed from the scaling group to other scaling groups.

After you remove an ECS instance or elastic container instance by calling the [DetachInstances](#) operation, the instance is not stopped or released.

Before you call this operation, make sure that the following conditions are met:

- The specified scaling group is enabled.
- No scaling activities in the specified scaling group are in progress.

If no scaling activities in the specified scaling group are in progress, the operation can immediately trigger scaling activities without the need to wait for the cooldown time to expire.

A successful call indicates only that Auto Scaling accepts the request. However, the scaling activity may still fail. You can obtain the status of a scaling activity based on the return value of the `ScalingActivityId` parameter.

The number of ECS instances or elastic container instances in a scaling group after you remove a specific number of instances from the scaling group must be equal to or greater than the value of the `MinSize` parameter. Otherwise, an error is reported when you call the [DetachInstances](#) operation.

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
Action	String	Yes	DetachInstances	The operation that you want to perform. Set the value to DetachInstances .
ScalingGroupId	String	Yes	asg-bp1igpak5ft1flyp****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
DecreaseDesiredCapacity	Boolean	No	true	<p>Specifies whether to adjust the expected number of instances in the scaling group. Valid values:</p> <ul style="list-style-type: none"> • true: After a specific number of instances are removed from the scaling group, the expected number of instances in the scaling group decreases. • false: After a specific number of instances are removed from the scaling group, the expected number of instances in the scaling group remains unchanged. <p>Default value: true.</p>
DetachOption	String	No	both	<p>Specifies whether to remove the instances from the default server groups and vServer groups of the Server Load Balancer (SLB) instance that is associated with the scaling group, and whether to remove the IP addresses of the instances from the whitelist that manages access to the ApsaraDB RDS instance that is associated with the scaling group.</p> <p>If you set this parameter to both, the instances are removed from the default sever groups and vServer groups of the associated SLB instance, and the IP addresses of the instances are removed from the whitelist that manages access to the associated ApsaraDB RDS instance.</p>
LifecycleHook	Boolean	No	false	<p>Specifies whether to trigger the lifecycle hook for the scaling group when you remove instances from the scaling group. Valid values:</p> <ul style="list-style-type: none"> • true • false <p>Default value: false.</p>

Parameter	Type	Required	Example	Description
InstanceId.N	String	Yes	i-bp109k5j3dum1ce6****	<p>The ID of ECS instance N or elastic container instance N that you want to remove from the scaling group. Valid values of N: 1 to 20.</p> <p>Examples:</p> <ul style="list-style-type: none"> The ID of the ECS instance that you want to remove is <code>i-bp109k5j3dum1ce6****</code>. The ID of the elastic container instance that you want to remove is <code>eci-bp17gw49eu09yiwm****</code>.

Response parameters

Parameter	Type	Example	Description
ScalingActivityId	String	asa-bp1gbswjhjr8tko**	The ID of the scaling activity.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DetachInstances
&ScalingGroupId=asg-bp1igpak5ft1flyp****
&DecreaseDesiredCapacity=true
&DetachOption=both
&InstanceId=["i-bp109k5j3dum1ce6****"]
&LifecycleHook=false
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DetachInstancesResponse>
    <ScalingActivityId>asa-bp1gbswjhjr8tko****</ScalingActivityId>
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DetachInstancesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "ScalingActivityId" : "asa-bplgbswjhjrw8tko****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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 disabled.
400	ScalingActivityInProgress	You cannot delete a scaling group or launch a new scaling activity while there is a scaling activity in progress for the specified scaling group.	The error message returned because a scaling activity in the specified scaling group is in progress.
400	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	The error message returned because the specified SLB instance is not in the Active state.
400	IncorrectDBInstanceState	The current status of DB instance "XXX" does not support this action.	The error message returned because the specified ApsaraDB RDS instance is not in the Running state.
400	IncorrectCapacity.MinSize	To remove the instances, the total capacity will be lesser than the MinSize.	The error message returned because the number of instances in the scaling group is less than the value of the MinSize parameter if the specified number of instances are removed.

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because you are not authorized to call this operation.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the specified instance does not exist within the Alibaba Cloud account.

13. Scheduled tasks

13.1. CreateScheduledTask

You can call this operation to create a scheduled task.

Description

When a scheduled task fails to be triggered because the scaling group is executing a scaling activity or the scaling group is disabled, the scheduled task will be retried for the duration of the launch expiration time. After the duration of the launch expiration time, a scheduled task that has failed to be triggered will be abandoned.

If multiple tasks are scheduled at similar points in time to execute scaling rules in the same group, the earliest task will execute its scaling activity first. Other tasks will make attempts to execute their rules during their launch expiration time, but only one scaling activity can be run in a scaling group at a time. If the previous scaling activity is completed while another scheduled task is making triggering attempts during its launch expiration time, Auto Scaling executes the scaling rule that is triggered by the scheduled task and then triggers the corresponding scaling activity.

A scheduled task supports the following scaling methods:

- You can set the `ScheduledAction` parameter to specify the scaling rule to be executed.
- You can set the `ScalingGroupId` parameter to specify the number of instances in a scaling group.



Note You cannot set `ScheduledAction` and `ScalingGroupId` at the same time.

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
Action	String	Yes	CreateScheduledTask	The operation that you want to perform. Set the value to <code>CreateScheduledTask</code> .
RegionId	String	Yes	cn-qingdao	The region ID of the scheduled task.

Parameter	Type	Required	Example	Description
ScheduledTaskName	String	No	scheduled****	<p>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 within the same region for the same Alibaba Cloud account.</p> <p>The default value is the value of the ScheduledTaskId parameter.</p>
Description	String	No	Test scheduled task.	<p>The description of the scheduled task. The description must be 2 to 200 characters in length.</p>
ScheduledAction	String	No	ari:acs:ess:cn-hangzhou:140692647****:scalingrule/asr-bp12tcnol686y1ik****	<p>The operation to be executed when the scheduled task is triggered. Specify the unique identifier of the scaling rule. After the ScheduledAction parameter is specified, the scaling method of the scheduled task is to select an existing scaling rule.</p> <div style="background-color: #e1f5fe; padding: 5px; margin-top: 10px;"> ? Note You cannot set ScheduledAction and ScalingGroupId at the same time. </div>
RecurrenceEndTime	String	No	2014-08-17T16:55Z	<p>The end time of the scheduled task to be repeated. Specify the time in the ISO 8601 standard in the YYYY-MM-DDThh:mmZ format.</p> <p>The time must be in UTC. You cannot enter a time point later than 365 days from the scheduled task creation.</p>

Parameter	Type	Required	Example	Description
LaunchTime	String	No	2014-08-17T16:52Z	<p>The time at which the scheduled task is triggered. Specify the time in the ISO 8601 standard in the YYYY-MM-DDThh:mmZ format. The time must be in UTC. You cannot enter a time point later than 90 days from the scheduled task creation.</p> <p>If the RecurrenceType parameter is specified, the task is executed each day at the time specified by LaunchTime.</p> <p>If the RecurrenceType parameter is not specified, the task is executed only once at the time specified by LaunchTime.</p>
RecurrenceType	String	No	Daily	<p>The interval that a scheduled task is repeated at. Valid values:</p> <ul style="list-style-type: none"> • Daily: The scheduled task is executed once every 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 based on the specified cron expression. <p>You must specify RecurrenceType and RecurrenceValue at the same time.</p>

Parameter	Type	Required	Example	Description
RecurrenceValue	String	No	1	<p>The recurrence value of the scheduled task to be repeated.</p> <ul style="list-style-type: none"> • Daily: If you set RecurrenceType to Daily, you can specify only one value. Valid values: 1 to 31. • Weekly: If you set RecurrenceType to Weekly, you can specify multiple values and separate them with commas (,). For example, the values of Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday are 0, 1, 2, 3, 4, 5, and 6. • Monthly: If you set RecurrenceType to Monthly, you can specify two values in the A-B format. Valid values: 1 to 31. B must be greater than or equal to A. • Cron: If you set RecurrenceType to Cron, you can specify a cron expression. 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 letters. <p>You must specify RecurrenceType and RecurrenceValue at the same time.</p>
TaskEnabled	Boolean	No	true	<p>Specifies whether to start the scheduled task.</p> <ul style="list-style-type: none"> • true: starts the scheduled task. • false: stops the scheduled task. <p>Default value: true.</p>
LaunchExpirationTime	Integer	No	600	<p>The time period during which the failed scheduled task is retried. Unit: seconds. Valid values: 0 to 21600.</p> <p>Default value: 600.</p>

Parameter	Type	Required	Example	Description
MinValue	Integer	No	0	The minimum number of instances in a scaling group when the scaling method of the scheduled task is to specify the number of instances in a scaling group.
MaxValue	Integer	No	10	The maximum number of instances in a scaling group when the scaling method of the scheduled task is to specify the number of instances in a scaling group.
DesiredCapacity	Integer	No	10	The expected number of instances in a scaling group when the scaling method of the scheduled task is to specify the number of instances in a scaling group. ? Note You must specify the DesiredCapacity parameter when you create the scaling group.
ScalingGroupId	String	No	asg-bp18p2yfxow2dl0q****	The ID of the scaling group whose number of instances is modified when the scheduled task is triggered. After the ScalingGroupId parameter is specified, the scaling method of the scheduled task is to specify the number of instance in a scaling group. You must specify at least one of the MinValue, MaxValue, and DesiredCapacity parameters. ? Note You cannot set ScheduledAction and ScalingGroupId at the same time.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3****	The ID of the request.
ScheduledTaskId	String	edRtShc57WGxdt8TlPbr****	The globally unique ID of the scheduled task generated by the system.

Examples

Sample requests

```
https://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-hangzhou:140692647****:scalingrule/asr-bp12tcnol686ylik****  
&<Common request parameters>
```

Sample success responses

XML format

```
<CreateScheduledTaskResponse>  
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>  
    <ScheduledTaskId>edRtShc57WXdt8TlPbr****</ScheduledTaskId>  
</CreateScheduledTaskResponse>
```

JSON format

```
{  
    "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3****",  
    "ScheduledTaskId": "edRtShc57WXdt8TlPbr****"  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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

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

13.2. ModifyScheduledTask

Modifies a scheduled task.

Description

A scheduled task supports two scaling methods.

- You can use the `ScheduledAction` parameter to specify the scaling rule that you want to execute.
- You can use the `ScalingGroupId` parameter to specify the number of instances in a scaling group.

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

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
Action	String	Yes	ModifyScheduledTask	The operation that you want to perform. Set the value to <code>ModifyScheduledTask</code> .

Parameter	Type	Required	Example	Description
ScheduledTaskId	String	Yes	edRtShc57WGXd8TlPbr****	The ID of the scheduled task.
ScheduledTaskName	String	No	scheduled****	The name of the scheduled task. The name must be 2 to 64 characters in length, and can contain letters, digits, underscores (_), hyphens (-), and periods (.). It must start with a letter or a digit. The name of the scheduled task must be unique in the region and within the Alibaba Cloud account.
Description	String	No	Test scheduled task.	The description of the scheduled task. The description must be 2 to 200 characters in length.
ScheduledAction	String	No	ari:acs:ess:cn-hangzhou:14069264****:scalingrule/asr-bp12tcnol686y1ik****	The scaling rule that you want to execute when the scheduled task is triggered. Specify the unique identifier of the scaling rule. If you specify the ScheduledAction parameter, the scheduled task selects an existing scaling rule to perform scaling.
RecurrenceEndTime	String	No	2014-08-20T16:55Z	The end time of the scheduled task. Specify the time in the ISO 8601 standard in the YYYY-MM-DDThh:mmZ format. The time must be in UTC. You cannot enter a time point later than 365 days from the scheduled task modification.

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

Parameter	Type	Required	Example	Description
LaunchTime	String	No	2014-08-18T10:52Z	<p>The point in time at which the scheduled task is triggered. Specify the time in the ISO 8601 standard in the YYYY-MM-DDThh:mmZ format. The time must be in UTC. You cannot enter a time point later than 90 days from the scheduled task modification.</p> <ul style="list-style-type: none">• If you specify the RecurrenceType parameter, the task is repeatedly executed at the time that is specified by LaunchTime.• If you do not specify the RecurrenceType parameter, the task is executed only once at the time that is specified by LaunchTime.
RecurrenceType	String	No	Daily	<p>The interval at which a scheduled task is repeated. Valid values:</p> <ul style="list-style-type: none">• Daily: The scheduled task is executed once every 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 based on the specified cron expression. <p>After you modify the scheduled task, the values that you specify for the RecurrenceType and RecurrenceValue parameters must be valid at the same time.</p>

Parameter	Type	Required	Example	Description
RecurrenceValue	String	No	2	<p>The number of recurrences of the scheduled task.</p> <ul style="list-style-type: none"> • Daily: If you set RecurrenceType to Daily, you can specify only one value. Valid values: 1 to 31. • Weekly: If you set RecurrenceType to Weekly, you can specify multiple values and separate them with commas (,). For example, the values that correspond to Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday are 0, 1, 2, 3, 4, 5, and 6. • Monthly: If you set RecurrenceType to Monthly, you can specify two values in the A-B format. Valid values of A and B: 1 to 31. B must be greater than or equal to A. • Cron: If you set RecurrenceType to Cron, you can specify a cron expression. A cron expression is written in UTC time and consists of five fields: minute, hour, day (date), month, and week. The expression can contain wildcard characters, including commas (,), question marks (?), hyphens (-), asterisks (*), number signs (#), forward slashes (/), and the letters L and W. <p>After you modify the scheduled task, the values that you specify for the RecurrenceType and RecurrenceValue parameters must be valid at the same time.</p>
TaskEnabled	Boolean	No	true	<p>Specifies whether the scheduled task is enabled. Valid values:</p> <ul style="list-style-type: none"> • true • false
LaunchExpirationTime	Integer	No	600	The period of time during which the failed scheduled task is retried. Unit: seconds. Valid values: 0 to 21600.

Parameter	Type	Required	Example	Description
MinValue	Integer	No	0	The minimum number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.
MaxValue	Integer	No	10	The maximum number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.
DesiredCapacity	Integer	No	10	The expected number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group. ? Note You must specify the DesiredCapacity parameter when you create the scaling group.
ScalingGroupId	String	No	asg-bp18p2yfxow2dl0q***	The ID of the scaling group whose number of instances is changed when the scheduled task is triggered. If you specify the ScalingGroupId parameter, the scheduled task specifies a number of instances in the scaling group to perform scaling. You must specify at least one of the MinValue, MaxValue, and DesiredCapacity parameters when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group. ? Note You cannot specify the ScheduledAction and ScalingGroupId parameters at the same time.
RegionId	String	No	cn-hangzhou	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=ModifyScheduledTask
&ScheduledTaskId=edRtShc57WGXd8T1Pbr****
&ScheduledTaskName=scheduled****
&Description=Test scheduled task.
&ScheduledAction=ari:acs:ess:cn-hangzhou:14069264****:scalingrule/asr-bp12tcnol686ylik****
&RecurrenceEndTime=2014-08-20T16:55Z
&LaunchTime=2014-08-18T10:52Z
&RecurrenceType=Daily
&RecurrenceValue=2
&TaskEnabled=true
&LaunchExpirationTime=600
&MinValue=0
&MaxValue=10
&DesiredCapacity=10
&ScalingGroupId=asg-bp18p2yfxow2dloq****
&RegionId=cn-hangzhou
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<ModifyScheduledTaskResponse>
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>
</ModifyScheduledTaskResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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

13.3. DescribeScheduledTasks

Queries scheduled tasks.

Description

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

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
Action	String	Yes	DescribeScheduledTasks	The operation that you want to perform. Set the value to <code>DescribeScheduledTasks</code> .

Parameter	Type	Required	Example	Description
RegionId	String	Yes	cn-qingdao	The region ID of the scaling group to which a scheduled task belongs.
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. Valid values: 1 to 50. Default value: 10.
ScheduledAction.N	RepeatList	No	ari:acs:ess:cn-hangzhou:1406926474****:scaling rule/asr-bp1id5rhu8edp7kd****	The operation that you want to perform when scheduled task N is triggered. Valid values of N: 1 to 20.
ScheduledTaskId.N	RepeatList	No	edRtShc57WGXdtsTIPbr****	The ID of scheduled task N that you want to query. Valid values of N: 1 to 20.
ScheduledTaskName.N	RepeatList	No	scheduled****	The name of scheduled task N that you want to query. Valid values of N: 1 to 20.
ScalingGroupId	String	No	asg-bp1bo5tca4m56nap****	The ID of the scaling group in which you want to execute the scheduled task.

Response parameters

Parameter	Type	Example	Description
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3C83E	The ID of the request.

Parameter	Type	Example	Description
ScheduledTasks	Array of ScheduledTask		The collection of scheduled tasks.
ScheduledTask			
Description	String	Test scheduled task.	The description of the scheduled task.
DesiredCapacity	Integer	10	The expected number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.
LaunchExpirationTime	Integer	600	The time period during which the failed scheduled task is retried. Unit: seconds. Valid values: 0 to 21600.
LaunchTime	String	2014-08-18T10:52Z	The point in time at which the scheduled task is triggered.
MaxValue	Integer	10	The maximum number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.
MinValue	Integer	0	The minimum number of instances in the scaling group when the Scaling Method parameter of the scheduled task is set to Configure Number of Instances in Scaling Group.
RecurrenceEndTime	String	2014-08-20T16:55Z	The end time of the recurrence of the scheduled task.
RecurrenceType	String	Daily	The interval at which the scheduled task recurs.
RecurrenceValue	String	1	The number of recurrences of the scheduled task.

Parameter	Type	Example	Description
ScalingGroupId	String	asg-bp1bo5tca4m56nap****	The ID of the scaling group in which you want to execute the scheduled task.
ScheduledAction	String	ari:acs:ess:cn-hangzhou:1406926474****:scalingrule/asr-bp1id5rhu8edp7kd**	The scaling rule that you want to use to execute the scheduled task. This parameter is returned only if a scaling rule is specified for the scheduled task.
ScheduledTaskId	String	edRtShc57WGXd8TlPbr****	The ID of the scheduled task.
ScheduledTaskName	String	scheduled****	The name of the scheduled task.
TaskEnabled	Boolean	true	Specifies whether the scheduled task is enabled.
TotalCount	Integer	1	The total number of scheduled tasks.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=DescribeScheduledTasks
&RegionId=cn-qingdao
&ScheduledTaskId.1=edRtShc57WGXd8TlPbr****
&PageSize=50
&<Common request parameters>
```

Sample success responses

XML format

```
<DescribeScheduledTasksResponse>
  <ScheduledTasks>
    <ScheduledTask>
      <LaunchExpirationTime>600</LaunchExpirationTime>
      <RecurrenceType>Daily</RecurrenceType>
      <RecurrenceEndTime>2020-03-02T00:00Z</RecurrenceEndTime>
      <ScalingGroupId>asg-bp1bo5tca4m56nap****</ScalingGroupId>
      <TaskEnabled>false</TaskEnabled>
      <ScheduledTaskName>scheduled****</ScheduledTaskName>
      <MaxValue>1</MaxValue>
      <LaunchTime>2020-03-01T00:00Z</LaunchTime>
      <ScheduledTaskId>cv0bJbAXpipdbFI1NbM****</ScheduledTaskId>
      <RecurrenceValue>31</RecurrenceValue>
    </ScheduledTask>
    <ScheduledTask>
      <LaunchExpirationTime>600</LaunchExpirationTime>
      <MinValue>1</MinValue>
      <RecurrenceType>Daily</RecurrenceType>
      <RecurrenceEndTime>2020-02-27T00:00Z</RecurrenceEndTime>
      <ScalingGroupId>asg-bp1bo5tca4m56nap****</ScalingGroupId>
      <TaskEnabled>true</TaskEnabled>
      <ScheduledTaskName>scheduled****</ScheduledTaskName>
      <LaunchTime>2020-02-26T00:00Z</LaunchTime>
      <ScheduledTaskId>cd3G3pesE65NcSTmkld1****</ScheduledTaskId>
      <RecurrenceValue>31</RecurrenceValue>
    </ScheduledTask>
  </ScheduledTasks>
  <PageNumber>1</PageNumber>
  <TotalCount>34</TotalCount>
  <PageSize>10</PageSize>
  <RequestId>4CF33369-7318-4BD5-BE1C-A776BA2C6627</RequestId>
</DescribeScheduledTasksResponse>
```

JSON format

```
{  
    "ScheduledTasks": [  
        {"  
            "LaunchExpirationTime": 600,  
            "RecurrenceType": "Daily",  
            "RecurrenceEndTime": "2020-03-02T00:00Z",  
            "ScalingGroupId": "asg-bp1bo5tca4m56nap****",  
            "TaskEnabled": false,  
            "ScheduledTaskName": "scheduled****",  
            "MaxValue": 1,  
            "LaunchTime": "2020-03-01T00:00Z",  
            "ScheduledTaskId": "cV0bJbAXpipdbFI1NbM****",  
            "RecurrenceValue": "31"  
        },  
        {"  
            "LaunchExpirationTime": 600,  
            "MinValue": 1,  
            "RecurrenceType": "Daily",  
            "RecurrenceEndTime": "2020-02-27T00:00Z",  
            "ScalingGroupId": "asg-bp1bo5tca4m56nap****",  
            "TaskEnabled": true,  
            "ScheduledTaskName": "scheduled****",  
            "LaunchTime": "2020-02-26T00:00Z",  
            "ScheduledTaskId": "cD3G3pesE65NcSTMkld1****",  
            "RecurrenceValue": "31"  
        }  
    ],  
    "PageNumber": 1,  
    "TotalCount": 34,  
    "PageSize": 10,  
    "RequestId": "4CF33369-7318-4BD5-BE1C-A776BA2C6627"  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

13.4. DeleteScheduledTask

Deletes a scheduled task.

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
Action	String	Yes	DeleteScheduledTask	The operation that you want to perform. Set the value to DeleteScheduledTask .
ScheduledTaskId	String	Yes	edRtShc57WGxdt8TlPbr****	The ID of the scheduled task. An ID is a globally unique identifier (GUID) that is generated by the system for a scheduled task.
RegionId	String	No	cn-qingdao	The ID of the region where the scheduled task is created.

Response parameters

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

Examples

Sample request

```
http(s)://ess.aliyuncs.com/?Action=DeleteScheduledTask
&ScheduledTaskId=edRtShc57WGxdt8TlPbr****
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DeleteScheduledTaskResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DeleteScheduledTaskResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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

14. Alarm tasks

14.1. CreateAlarm

Creates an event-triggered task.

Description

If you set MetricType to custom, you must first report your custom metrics to CloudMonitor before you can create event-triggered tasks by using custom metrics. For more information, see [Custom monitoring event-triggered tasks](#).

When you create an event-triggered task, you must configure the MetricName, Dimension.N.DimensionKey, and Dimension.N.DimensionValue parameters to specify the range of aggregated statistics for the metrics of the scaling group. For example, you can specify the user_id and scaling_group dimensions for an event-triggered task to collect monitoring data about all the Elastic Compute Service (ECS) instances in a specific scaling group within a specific account. The following table describes the metrics that are available when MetricType is set to system.

 Note

- The user_id and scaling_group dimensions are automatically populated. You need only to configure the device and state dimensions. For more information, see the Dimension.N.DimensionKey and Dimension.N.DimensionValue parameters in the "Request parameters" section of this topic.
- If you set MetricType to custom, the available metrics vary based on the custom metrics that you have.

Collection source	Metric	Description	Dimension	Applicable network
ECS instance	CpuUtilization	The CPU utilization. Unit: %.	user_id and scaling_group	Virtual private cloud (VPC) and classic network
ECS instance	IntranetTx	The outbound traffic over the internal network. Unit: KB/min.	user_id and scaling_group	VPC and classic network
ECS instance	IntranetRx	The inbound traffic over the internal network. Unit: KB/min.	user_id and scaling_group	VPC and classic network

Collection source	Metric	Description	Dimension	Applicable network
ECS instance	VpcInternetTx	The outbound traffic over the Internet from the VPC. Unit: KB/min.	user_id and scaling_group	VPC
ECS instance	VpcInternetRx	The inbound traffic over the Internet to the VPC. Unit: KB/min.	user_id and scaling_group	VPC
ECS instance	ClassicInternetTx	The outbound traffic over the Internet from the classic network. Unit: KB/min.	user_id and scaling_group	Classic network
ECS instance	ClassicInternetRx	The inbound traffic over the Internet to the classic network. Unit: KB/min.	user_id and scaling_group	Classic network
ECS instance	SystemDiskReadBps	The number of bytes read from the system disk per second.	user_id and scaling_group	VPC and classic network
ECS instance	SystemDiskWriteBps	The number of bytes written to the system disk per second.	user_id and scaling_group	VPC and classic network
ECS instance	SystemDiskReadOps	The number of read operations on the system disk per second.	user_id and scaling_group	VPC and classic network

Collection source	Metric	Description	Dimension	Applicable network
ECS instance	SystemDiskWriteOps	The number of write operations on the system disk per second.	user_id and scaling_group	VPC and classic network
CloudMonitor agent	CpuUtilizationAgent	The CPU utilization. Unit: %.	user_id and scaling_group	VPC and classic network
CloudMonitor agent	GpuUtilizationAgent	The GPU utilization. Unit: %.	user_id and scaling_group	VPC
CloudMonitor agent	GpuMemoryFreeUtilizationAgent	The percentage of idle GPU memory.	user_id and scaling_group	VPC
CloudMonitor agent	GpuMemoryUtilizationAgent	The GPU memory utilization. Unit: %.	user_id and scaling_group	VPC
CloudMonitor agent	MemoryUtilization	The memory utilization. Unit: %.	user_id and scaling_group	VPC and classic network
CloudMonitor agent	LoadAverage	The average system load.	user_id and scaling_group	VPC and classic network
CloudMonitor agent	TcpConnection	The total number of TCP connections.	user_id, scaling_group, and state	VPC and classic network
CloudMonitor agent	PackagesNetOut	The number of packets that are sent by the network interface controller (NIC).	user_id, scaling_group, and device	VPC and classic network
CloudMonitor agent	PackagesNetIn	The number of packets that are received by the NIC.	user_id, scaling_group, and device	VPC and classic network

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
Action	String	Yes	CreateAlarm	The operation that you want to perform. Set the value to CreateAlarm .
RegionId	String	Yes	cn-hangzhou	The ID of the region where the scaling group resides.
Name	String	No	TestAlarmTask	The name of the event-triggered task.
Description	String	No	Test alarm task.	The description of the event-triggered task.
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dl0q***	The ID of the scaling group with which the event-triggered task is associated.

Parameter	Type	Required	Example	Description
MetricName	String	Yes	CpuUtilization	<p>The name of the metric. The valid values vary based on the metric type.</p> <ul style="list-style-type: none">• If you set MetricType to custom, the valid values are the metrics that you have.• If you set MetricType to system, the valid values are: <ul style="list-style-type: none">◦ CpuUtilization◦ IntranetTx◦ IntranetRx◦ VpcInternetTx◦ VpcInternetRx◦ ClassicInternetTx◦ ClassicInternetRx◦ SystemDiskReadBps◦ SystemDiskWriteBps◦ SystemDiskReadOps◦ SystemDiskWriteOps◦ CpuUtilizationAgent◦ GpuUtilizationAgent◦ GpuMemoryFreeUtilizationAgent◦ GpuMemoryUtilizationAgent◦ MemoryUtilization◦ LoadAverage◦ TcpConnection◦ PackagesNetOut◦ PackagesNetIn <p>For more information, see Description in this topic.</p>
MetricType	String	No	system	<p>The type of the metric. Valid values:</p> <ul style="list-style-type: none">• system: system metrics of CloudMonitor• custom: custom metrics that are published to CloudMonitor

Parameter	Type	Required	Example	Description
Period	Integer	No	300	<p>The period during which the statistical value of the metric is collected. Unit: seconds. Valid values:</p> <ul style="list-style-type: none"> • 60 • 120 • 300 • 900 <p>Default value: 300.</p>
Statistics	String	No	Average	<p>The method that is used to collect statistics for the metric. Valid values:</p> <ul style="list-style-type: none"> • Average • Minimum • Maximum <p>Default value: Average.</p>
Threshold	Float	Yes	80	<p>The threshold of the metric. If the threshold is reached N times, a scaling rule is triggered.</p>
ComparisonOperator	String	No	>=	<p>The operator that is used to compare the metric value and the threshold.</p> <ul style="list-style-type: none"> • Valid value if the metric value is greater than or equal to the threshold: >= • Valid value if the metric value is less than or equal to the threshold: <= • Valid value if the metric value is greater than the threshold: > • Valid value if the metric value is less than the threshold: < <p>Default value: >=.</p>
EvaluationCount	Integer	No	3	<p>The number of times that the threshold must be reached to trigger an alert and execute a scaling rule. For example, if you set this parameter to 3, the average CPU utilization must reach 80% three times to trigger an alert.</p> <p>Default value: 3.</p>

Parameter	Type	Required	Example	Description
GroupId	Integer	No	4055401	<p>The ID of the application group to which the custom metric belongs. This parameter is required only if you set MetricType to custom.</p>
Effective	String	No	TZ=+00 * * 1-2 * * ?	<p>Specifies the effective period of the even-triggered task. By default, the event-triggered task is effective all the time.</p> <p>This parameter follows the cron expression format. The default format is <code>x x x x x ?</code>. In the format:</p> <ul style="list-style-type: none"> • X: a placeholder for a field, which represents seconds, minutes, hours, days, and months in sequence. X can be a definite value or a special character that has logical meaning. For information about the valid values of X, see Cron expression. • ?: No value is specified. <div style="background-color: #e0f2ff; padding: 10px;"> <p>? Note By default, the value of this parameter is specified in UTC+8. You can specify the time zone before a cron expression in the <code>TZ=+yy</code> format. yy indicates the time zone. For example, <code>TZ=+00 * * 1-2 * * ?</code> specifies that the event-triggered task is effective between 01:00 and 02:59 (UTC+0) every day.</p> </div> <p>Sample values:</p> <ul style="list-style-type: none"> • <code>* * * * * ?</code> : The event-triggered task is effective all the time. • <code>* * 17-18 * * ?</code> : The event-triggered task is effective between 17:00 and 18:59 (UTC+8) every day. • <code>TZ=+00 * * 1-2 * * ?</code> : The event-triggered task is effective between 01:00 and 02:59 (UTC+0) every day.

Parameter	Type	Required	Example	Description
AlarmAction.N	String	No	ari:acs:ess:cn-hangzhou:1406926****:scalingrule/asr-bp163l21e07uhn***	The unique identifier of scaling rule N that is associated with the event-triggered task.
Dimension.N.DimensionKey	String	No	device	<p>The key of dimension N that is associated with the metric. The valid values vary based on the metric type.</p> <ul style="list-style-type: none">• If you set MetricType to custom, you can specify the valid values as you want.• If you set MetricType to system, the valid values are:<ul style="list-style-type: none">◦ user_id: the ID of your account◦ scaling_group: the scaling group that is monitored◦ device: the type of the NIC◦ state: the status of the TCP connection

Parameter	Type	Required	Example	Description
Dimension.N.DimensionValue	String	No	eth0	<p>The value of dimension N that is associated with the metric. The valid values vary based on the value of Dimension.N.DimensionKey.</p> <ul style="list-style-type: none"> If you set Dimension.N.DimensionKey to a custom value or if you set MetricType to custom, you can specify the valid values as you want. If the value of Dimension.N.DimensionKey is specified by the system or if you set MetricType to system, the valid values are: <ul style="list-style-type: none"> user_id: The system specifies the value. scaling_group: The system specifies the value. device: You can specify eth0 or eth1. <ul style="list-style-type: none"> For instances of the classic network type, eth0 specifies the internal NIC. Only one eth0 NIC exists on each instance of the VPC type. For instances of the classic network type, eth1 specifies the public NIC. state: You can specify TCP_TOTAL or ESTABLISHED. <ul style="list-style-type: none"> TCP_TOTAL specifies the total number of TCP connections. ESTABLISHED specifies the number of TCP connections that are established.

Response parameters

Parameter	Type	Example	Description
-----------	------	---------	-------------

Parameter	Type	Example	Description
AlarmTaskId	String	asg-bp1hvbnmk10vll5**_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3****	The ID of the request.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action/CreateAlarm
&RegionId=cn-hangzhou
&Name=TestAlarmTask
&Description=Test alarm task.
&ScalingGroupId=asg-bp18p2yfxow2dloq****
&MetricName=CpuUtilization
&MetricType=system
&Period=300
&Statistics=Average
&Threshold=80.0
&ComparisonOperator=>=
&EvaluationCount=3
&GroupId=4055401
&Effective=TZ=+00 * * 1-2 * * ?
&AlarmAction=["ari:acs:ess:cn-hangzhou:1406926****:scalingrule/asr-bp163121e07uhn****"]
&Dimension=[{"DimensionKey": "device", "DimensionValue": "eth0"}]
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<CreateAlarmResponse>
    <AlarmTaskId>asg-bp1hvbnmk10vll5**_f95ce797-dc2e-4bad-9618-14fee7d1****</AlarmTaskId>
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>
</CreateAlarmResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "AlarmTaskId" : "asg-bplhvbnmk110v115****_f95ce797-dc2e-4bad-9618-14fee7d1****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
404	InvalidParameter	The specified value of parameter "%s" is not valid.	The error message returned because the value that you specified for the "%s" parameter is invalid.

14.2. DescribeAlarms

You can call this operation to query one or more event-triggered tasks.

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
Action	String	Yes	DescribeAlarms	The operation that you want to perform. Set the value to DescribeAlarms.
RegionId	String	Yes	cn-qingdao	The region ID of the event-triggered task.
ScalingGroupId	String	No	asg-bp18p2yfxow2dl0q****	The ID of the scaling group to be monitored by the event-triggered task.

Parameter	Type	Required	Example	Description
AlarmTaskId	String	No	asg-bp1hvbnmk10vll5****_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.
State	String	No	OK	<p>The status of the event-triggered task. Valid values:</p> <ul style="list-style-type: none"> ALARM: One or more alerts are triggered because the specified alert condition is met. OK: No alert is triggered because the specified alert condition is not met. INSUFFICIENT_DATA: Auto Scaling cannot determine whether the specified alert condition is met because data is insufficient.
IsEnable	Boolean	No	true	<p>Specifies whether the event-triggered task is enabled. Valid values:</p> <ul style="list-style-type: none"> true: The event-triggered task is enabled. false: The event-triggered task is disabled.
MetricType	String	No	system	<p>The type of the metric. Valid values:</p> <ul style="list-style-type: none"> system: the built-in metric of Cloud Monitor. custom: the custom metric that is used to report relevant data of ECS instances to Cloud Monitor.
PageSize	Integer	No	10	<p>The number of entries to return on each page. Maximum value: 50. Default: 10.</p>
PageNumber	Integer	No	1	<p>The number of the page to return. Pages start from page 1. Default value: 1.</p>

Response parameters

Parameter	Type	Example	Description
AlarmList	Array of Alarm		The list of event-triggered tasks.
Alarm			
AlarmActions	List	ari:acs:ess:cn-hangzhou:1406926474****:scalingrule/asr-bp163l21e07uh****	The list of unique identifiers of the scaling rules that are associated with the event-triggered tasks.
AlarmTaskId	String	asg-bp1hvbnmk10vll5**_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.
ComparisonOperator	String	>=	<p>The operator for comparison between the collected metric value and the threshold. This parameter is used to specify the relationship between the metric value and the threshold that meets the alert condition. Valid values:</p> <ul style="list-style-type: none"> • >=: The metric value is greater than or equal to the threshold. • <=: The metric value is less than or equal to the threshold. • >: The metric value is greater than the threshold. • <: The metric value is less than the threshold.
Description	String	Test alarm task.	The description of the event-triggered task.
Dimensions	Array of Dimension		The dimensions that are associated with the metric.
Dimension			

Parameter	Type	Example	Description
DimensionKey	String	device	<p>The key of the dimension that is associated with the metric. Valid values:</p> <ul style="list-style-type: none"> • user_id: the ID of your account. • scaling_group: the scaling group that is monitored. • device: the type of the network interface controller (NIC). • state: indicates whether the total number of TCP connections or the number of TCP connections that have been established and are available is counted.
DimensionValue	String	eth0	<p>The value of the dimension that is associated with the metric. The value of this parameter depends on the DimensionKey parameter.</p> <p>When the value of DimensionKey is user_id, the value of this parameter is set by the system.</p> <p>When the value of DimensionKey is scaling_group, the value of this parameter is set by the system.</p> <p>When the value of DimensionKey is device, the valid values of this parameter are:</p> <ul style="list-style-type: none"> • eth0: the internal network NIC for classic network-type ECS instances, and the only NIC for VPC-type ECS instances. • eth1: the public network NIC for classic network-type ECS instances. <p>When the value of DimensionKey is state, the valid values of this parameter are:</p> <ul style="list-style-type: none"> • TCP_TOTAL: indicates that the total number of TCP connections is counted. • ESTABLISHED: indicates that the number of TCP connections that have been established and are available is counted.
Effective	String	Test	<p> Note This parameter is in invitational preview and not available.</p>

Parameter	Type	Example	Description
Enable	Boolean	true	<p>Indicates whether the event-triggered task is enabled. Valid values:</p> <ul style="list-style-type: none">• true: The event-triggered task is enabled.• false: The event-triggered task is disabled.
EvaluationCount	Integer	3	The number of times that the alert condition must be met to trigger an alert and execute scaling rules. For example, if this parameter is set to 3, an alert is triggered when the average CPU utilization is greater than or equal to 80% for three times.

Parameter	Type	Example	Description
MetricName	String	CpuUtilization	<p>The name of the metric. Valid values:</p> <ul style="list-style-type: none"> • CpuUtilization: the CPU utilization in percentage. • ClassicInternetRx: the inbound traffic over the public network to the classic network. Unit: KB/min. • ClassicInternetTx: the outbound traffic over the public network from the classic network. Unit: KB/min. • VpcInternetRx: the inbound traffic over the public network to the VPC. Unit: KB/min. • VpcInternetTx: the outbound traffic over the public network from the VPC. Unit: KB/min. • IntranetRx: the inbound traffic over the internal network. Unit: KB/min. • IntranetTx: the outbound traffic over the internal network. Unit: KB/min. • LoadAverage: the average system load. • MemoryUtilization: the memory usage in percentage. • SystemDiskReadBps: the number of bytes read from the system disk per second. • SystemDiskWriteBps: the number of bytes written to the system disk per second. • SystemDiskReadOps: the number of times that the system disk is read per second. • SystemDiskWriteOps: the number of times that the system disk is written per second. • PackagesNetIn: the number of packets that are received by the NIC per second. • PackagesNetOut: the number of packets that are sent by the NIC per second. • TcpConnection: the number of TCP connections. <p>For more information, see the Description section of this topic.</p>
MetricType	String	system	<p>The type of the metric. Valid values:</p> <ul style="list-style-type: none"> • system: the built-in metric of Cloud Monitor. • custom: the custom metric that is used to report relevant data of ECS instances to Cloud Monitor.

Parameter	Type	Example	Description
Name	String	TestAlarmTask	The name of the event-triggered task.
Period	Integer	300	<p>The period for collecting the statistical value of the metric. Unit: seconds. Valid values:</p> <ul style="list-style-type: none"> • 60 • 120 • 300 • 900
ScalingGroupId	String	asg-bp18p2yfxow2dloq***	The ID of the scaling group that is monitored by the event-triggered task.
State	String	ALARM	<p>The status of the event-triggered task. Valid values:</p> <ul style="list-style-type: none"> • ALARM: One or more alerts are triggered because the specified alert condition is met. • OK: No alert is triggered because the specified alert condition is not met. • INSUFFICIENT_DATA: Auto Scaling cannot determine whether the specified alert condition is met because data is insufficient.
Statistics	String	Average	<p>The type of the statistical value that is collected for the metric. Valid values:</p> <ul style="list-style-type: none"> • Average: the average value. • Minimum: the minimum value. • Maximum: the maximum value.
Threshold	Float	80	<p>The threshold of the metric. The alert condition of the event-triggered task specifies the relationship between the statistical value and the threshold of the metric. When the alert condition is met for the specified number of times, Auto Scaling triggers an alert and executes the scaling rules specified in the event-triggered task.</p>
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	10	The number of entries returned per page.

Parameter	Type	Example	Description
RequestId	String	871C7C53-34A4-45AA-8C14-4B72FA6A****	The ID of the request.
TotalCount	Integer	2	The total number of event-triggered tasks.

Examples

Sample requests

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

Sample success responses

XML format

```
<DescribeAlarmsResponse>
<PageNumber>1</PageNumber>
<TotalCount>2</TotalCount>
<PageSize>10</PageSize>
<AlarmList>
<Alarm>
<Period>300</Period>
<Statistics>Average</Statistics>
<MetricType>system</MetricType>
<EvaluationCount>3</EvaluationCount>
<Name>asg-bp1hvbnmk110v115xtt2_f95ce797-dc2e-4bad-9618-14fee7d1****</Name>
<AlarmTaskId>asg-bp1hvbnmk110v115xtt2_f95ce797-dc2e-4bad-9618-14fee7d1****</AlarmTask
Id>
<MetricName>MemoryUtilization</MetricName>
<Threshold>55</Threshold>
<State>INSUFFICIENT_DATA</State>
<Enable>false</Enable>
<ScalingGroupId>asg-bp1hvbnmk110v115****</ScalingGroupId>
<Dimensions>
<Dimension>
<DimensionValue>asg-bp1hvbnmk110v115****</DimensionValue>
<DimensionKey>scaling_group</DimensionKey>
</Dimension>
<Dimension>
<DimensionValue>140692647406****</DimensionValue>
<DimensionKey>userId</DimensionKey>
</Dimension>
</Dimensions>
<ComparisonOperator>>=</ComparisonOperator>
</Alarm>
<Alarm>
<Period>60</Period>
<Description>DO NOT EDIT OR DELETE. For TargetTrackingScalingRule scalingRuleAri:ari:a
```

```

cs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp13jvr9iummzvgk****, scalingRuleName:*
***-target-tracking-scaling-rule</Description>
<Statistics>Average</Statistics>
<MetricType>system</MetricType>
<EvaluationCount>15</EvaluationCount>
<AlarmActions>
    <AlarmAction>ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp13jvr9iummz
vgk****</AlarmAction>
</AlarmActions>
<Name>TargetTracking-asr-bp13jvr9iummzvgk****-AlarmLow-3f4730b9-c275-4527-b057-847fa8
83****</Name>
<AlarmTaskId>asg-bp18p2yfxow2dloq****_48efb36a-c0b8-4173-b699-bd37c430****</AlarmTask
Id>
<MetricName>CpuUtilization</MetricName>
<Threshold>56</Threshold>
<State>ALARM</State>
<Enable>true</Enable>
<ScalingGroupId>asg-bp18p2yfxow2dloq****</ScalingGroupId>
<Dimensions>
    <Dimension>
        <DimensionValue>asg-bp18p2yfxow2dloq****</DimensionValue>
        <DimensionKey>scaling_group</DimensionKey>
    </Dimension>
    <Dimension>
        <DimensionValue>140692647406****</DimensionValue>
        <DimensionKey>userId</DimensionKey>
    </Dimension>
</Dimensions>
<ComparisonOperator><</ComparisonOperator>
</Alarm>
</AlarmList>
<RequestId>871C7C53-34A4-45AA-8C14-4B72FA6A****</RequestId>
</DescribeAlarmsResponse>

```

JSON format

```
{
    "PageNumber": 1,
    "TotalCount": 2,
    "PageSize": 10,
    "AlarmList": {
        "Alarm": [
            {
                "Period": 300,
                "Statistics": "Average",
                "MetricType": "system",
                "EvaluationCount": 3,
                "Name": "asg-bp1hvbnnmk110v115xtt2_f95ce797-dc2e-4bad-9618-14fee7d1****",
                "AlarmTaskId": "asg-bp1hvbnnmk110v115xtt2_f95ce797-dc2e-4bad-9618-14fee7d1**"
            },
            {
                "MetricName": "MemoryUtilization",
                "Threshold": 55,
                "State": "INSUFFICIENT_DATA",
                "Enable": false
            }
        ]
    }
}
```

```
        "Enabled": true,
        "ScalingGroupId": "asg-bp1hvbnnmk110v115****",
        "Dimensions": {
            "Dimension": [
                {
                    "DimensionValue": "asg-bp1hvbnnmk110v115****",
                    "DimensionKey": "scaling_group"
                },
                {
                    "DimensionValue": "140692647406****",
                    "DimensionKey": "userId"
                }
            ]
        },
        "ComparisonOperator": ">="
    },
    {
        "Period": 60,
        "Description": "DO NOT EDIT OR DELETE. For TargetTrackingScalingRule scaling RuleAri:ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp13jvr9iummzvgk****, scalingRuleName:****-target-tracking-scaling-rule",
        "Statistics": "Average",
        "MetricType": "system",
        "EvaluationCount": 15,
        "AlarmActions": [
            "AlarmAction": [
                "ari:acs:ess:cn-hangzhou:140692647406****:scalingrule/asr-bp13jvr9iummzvgk****"
            ]
        ],
        "Name": "TargetTracking-asr-bp13jvr9iummzvgk****-AlarmLow-3f4730b9-c275-4527-b057-847fa883****",
        "AlarmTaskId": "asg-bp18p2yfxow2dloq****_48efb36a-c0b8-4173-b699-bd37c430**",
        "MetricName": "CpuUtilization",
        "Threshold": 56,
        "State": "ALARM",
        "Enable": true,
        "ScalingGroupId": "asg-bp18p2yfxow2dloq****",
        "Dimensions": {
            "Dimension": [
                {
                    "DimensionValue": "asg-bp18p2yfxow2dloq****",
                    "DimensionKey": "scaling_group"
                },
                {
                    "DimensionValue": "140692647406****",
                    "DimensionKey": "userId"
                }
            ]
        },
        "ComparisonOperator": "<"
    }
]
}.
```

```
        "RequestId": "871C7C53-34A4-45AA-8C14-4B72FA6A****"  
    }
```

Error codes

For a list of error codes, visit the [API Error Center](#).

14.3. ModifyAlarm

Modifies an event-triggered task.

Description

If you set MetricType to custom, you must first publish the custom metric to CloudMonitor before you can create an event-triggered task for the metric. For more information, see [Custom monitoring event-triggered tasks](#).

When you create an event-triggered task, you must configure the MetricName, Dimension.N.DimensionKey, and Dimension.N.DimensionValue parameters to specify the range of aggregated statistics about metrics of the scaling group that you specified. For example, if user_id and scaling_group are specified, the statistics about the metrics of all the Elastic Compute Service (ECS) instances in the specified scaling group within the specified account is aggregated. The following table describes the metrics that are available when MetricType is set to system.

(?) Note

- The user_id and scaling_group dimensions are automatically populated. You need only to configure the device and state dimensions. For more information, see the Dimension.N.DimensionKey and Dimension.N.DimensionValue parameters in the "Request parameters" section of this topic.
- If you set MetricType to custom, the available metrics vary based on the custom metrics that you have.

Collection source	Metric	Description	Dimension	Applicable network
ECS instances	CpuUtilization	The CPU utilization. Unit: %.	user_id and scaling_group	Virtual Private Cloud (VPC) and classic network
ECS instances	IntranetTx	The outbound traffic over the internal network. Unit: KB/min.	user_id and scaling_group	VPC and classic network

Collection source	Metric	Description	Dimension	Applicable network
ECS instances	IntranetRx	The inbound traffic over the internal network. Unit: KB/min.	user_id and scaling_group	VPC and classic network
ECS instances	VpcInternetTx	The outbound traffic over the Internet from the VPC. Unit: KB/min.	user_id and scaling_group	VPC
ECS instances	VpcInternetRx	The inbound traffic over the Internet to the VPC. Unit: KB/min.	user_id and scaling_group	VPC
ECS instances	ClassicInternetTx	The outbound traffic over the Internet from the classic network. Unit: KB/min.	user_id and scaling_group	Classic network
ECS instances	ClassicInternetRx	The inbound traffic over the Internet to the classic network. Unit: KB/min.	user_id and scaling_group	Classic network
ECS instances	SystemDiskReadBps	The number of bytes read from the system disk per second.	user_id and scaling_group	VPC and classic network
ECS instances	SystemDiskWriteBps	The number of bytes written to the system disk per second.	user_id and scaling_group	VPC and classic network
ECS instances	SystemDiskReadOps	The number of read operations on the system disk per second.	user_id and scaling_group	VPC and classic network

Collection source	Metric	Description	Dimension	Applicable network
ECS instances	SystemDiskWriteOps	The number of write operations on the system disk per second.	user_id and scaling_group	VPC and classic network
CloudMonitor agents	CpuUtilizationAgent	The CPU utilization. Unit: %.	user_id and scaling_group	VPC and classic network
CloudMonitor agents	GpuUtilizationAgent	The GPU utilization. Unit: %.	user_id and scaling_group	VPC
CloudMonitor agents	GpuMemoryFreeUtilizationAgent	The percentage of idle GPU memory.	user_id and scaling_group	VPC
CloudMonitor agents	GpuMemoryUtilizationAgent	The GPU memory utilization. Unit: %.	user_id and scaling_group	VPC
CloudMonitor agents	MemoryUtilization	The memory utilization. Unit: %.	user_id and scaling_group	VPC and classic network
CloudMonitor agents	LoadAverage	The average system load.	user_id and scaling_group	VPC and classic network
CloudMonitor agents	TcpConnection	The total number of TCP connections.	user_id, scaling_group, and state	VPC and classic network
CloudMonitor agents	PackagesNetOut	The number of packets that are sent by the network interface controller (NIC) per second.	user_id, scaling_group, and device	VPC and classic network

Collection source	Metric	Description	Dimension	Applicable network
CloudMonitor agents	PackagesNetIn	The number of packets that are received by the NIC per second.	user_id, scaling_group, and device	VPC and classic network

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
Action	String	Yes	ModifyAlarm	The operation that you want to perform. Set the value to ModifyAlarm.
AlarmTaskId	String	Yes	asg-bp1hvbnmkl10vll5****_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.
RegionId	String	Yes	cn-qingdao	The ID of the region to which the event-triggered task belongs.
Name	String	No	alarmtask****	The name of the event-triggered task.
Description	String	No	Test alarm task.	The description of the event-triggered task.
AlarmAction.N	RepeatList	No	ari:acs:ess:cn-hangzhou:140692647****:scalingrule/asr-bp163l21e07uhn***	The unique identifier of scaling rule N that is associated with the event-triggered task.

Parameter	Type	Required	Example	Description
MetricName	String	No	MemoryUtilization	<p>The name of the metric. The valid values vary based on the metric type.</p> <ul style="list-style-type: none"> If you set MetricType to custom, the valid values are the metrics that you have. If you set MetricType to system, the following values are valid: <ul style="list-style-type: none"> CpuUtilization IntranetTx IntranetRx VpcInternetTx VpcInternetRx ClassicInternetTx ClassicInternetRx SystemDiskReadBps SystemDiskWriteBps SystemDiskReadOps SystemDiskWriteOps CpuUtilizationAgent GpuUtilizationAgent GpuMemoryFreeUtilizationAgent GpuMemoryUtilizationAgent MemoryUtilization LoadAverage TcpConnection PackagesNetOut PackagesNetIn <p>For more information, see Description in this topic.</p>
MetricType	String	No	system	<p>The type of the metric. Valid values:</p> <ul style="list-style-type: none"> system: the system metrics of CloudMonitor custom: the custom metrics that are published to CloudMonitor

Parameter	Type	Required	Example	Description
Period	Integer	No	300	<p>The period of time during which statistics about the metric is collected. Unit: seconds. Valid values:</p> <ul style="list-style-type: none"> • 60 • 120 • 300 • 900
Statistics	String	No	Average	<p>The method that is used to collect statistics about the metric. Valid values:</p> <ul style="list-style-type: none"> • Average • Minimum • Maximum
Threshold	Float	No	80	The threshold of the metric. If the threshold is reached, a scaling rule is triggered.
ComparisonOperator	String	No	>=	<p>The operator that is used to compare the metric value and the threshold.</p> <ul style="list-style-type: none"> • Valid value if the metric value is greater than or equal to the threshold: >= • Valid value if the metric value is less than or equal to the threshold: <= • Valid value if the metric value is greater than the threshold: > • Valid value if the metric value is less than the threshold: <
EvaluationCount	Integer	No	3	The number of times that the threshold must be reached to trigger an alert and execute scaling rules. For example, if this parameter is set to 3, an alert is triggered when the average CPU utilization is greater than or equal to 80% for three times.
GroupId	Integer	No	4055401	The ID of the application group to which the custom metric belongs. This parameter is required only if you set MetricType to custom.

Parameter	Type	Required	Example	Description
Dimension.N.DimensionKey	String	No	device	<p>The key of dimension N that is associated with the metric. The valid values vary based on the metric type.</p> <ul style="list-style-type: none">• If you set MetricType to custom, you can specify the valid values as you want.• If you set MetricType to system, the following values are valid:<ul style="list-style-type: none">◦ user_id: the ID of your account.◦ scaling_group: the scaling group that is monitored◦ device: the type of the NIC◦ state: the status of the TCP connection.

Parameter	Type	Required	Example	Description
Dimension.N.DimensionValue	String	No	eth0	<p>The value of dimension N that is associated with the metric. The valid values vary based on the value of Dimension.N.DimensionKey.</p> <ul style="list-style-type: none">• If the value of Dimension.N.DimensionKey is a custom value or if you set MetricType to custom, you can specify the valid values as you want.• If the value of Dimension.N.DimensionKey is specified by the system or if you set MetricType to system, the following values are valid:<ul style="list-style-type: none">◦ user_id: The system automatically fills in the value.◦ scaling_group: The system automatically fills in the value.◦ device: You can specify eth0 or eth1.<ul style="list-style-type: none">▪ For instances of the classic network type, eth0 specifies the internal NIC. Only one eth0 NIC exists on each instance of the VPC type.▪ For instances of the classic network type, eth1 specifies the public NIC.◦ state: You can specify TCP_TOTAL or ESTABLISHED.<ul style="list-style-type: none">▪ TCP_TOTAL specifies the total number of TCP connections.▪ ESTABLISHED specifies the number of TCP connections that are established.

Parameter	Type	Required	Example	Description
Effective	String	No	TZ=+00 * * 1-2 * * ?	<p>The effective period of the event-triggered task.</p> <p>This parameter follows the Cron expression format. The default format is <code>X X X X X ?</code>. In the format:</p> <ul style="list-style-type: none"> • X is a placeholder for a field, which represents seconds, minutes, hours, day of month, and months in sequence. X can be a definite value or a special character that has logical meaning. For information about the valid values of X, see Cron expression. • ? indicates that no value is specified. <div style="background-color: #e0f2ff; padding: 10px;"> <p>? Note By default, the value of this parameter is specified in UTC+8. You can specify the time zone before a Cron expression in the <code>TZ=+yy</code> format. yy indicates the time zone. For example, <code>TZ=+00 * * 1-2 * * ?</code> indicates that the event-triggered task is effective between 01:00 and 02:59 (UTC+0) every day.</p> </div> <p>Sample values:</p> <ul style="list-style-type: none"> • <code>* * * * * ?</code> : The event-triggered task is effective all the time. • <code>* * 17-18 * * ?</code> : The event-triggered task is effective between 17:00 and 18:59 (UTC+8) every day. • <code>TZ=+00 * * 1-2 * * ?</code> : The event-triggered task is effective between 01:00 and 02:59 (UTC+0) every day.

Response parameters

Parameter	Type	Example	Description
AlarmTaskId	String	asg-bp1hvbnmk110vll5**_83948190-acdd-483f-98f7-b77f4778****	The ID of the event-triggered task.
RequestId	String	BACACF83-7070-4953-A8FD-D81F89F1****	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=ModifyAlarm
&RegionId=cn-qingdao
&AlarmTaskId=asg-bp1hvbnmk110vll5**_83948190-acdd-483f-98f7-b77f4778****
&MetricName=MemoryUtilization
&<Common request parameters>
```

Sample success responses

XML format

```
<ModifyAlarmResponse>
  <AlarmTaskId>asg-bp1hvbnmk110vll5**_83948190-acdd-483f-98f7-b77f4778****</AlarmTaskId>
  <RequestId>BACACF83-7070-4953-A8FD-D81F89F1****</RequestId>
</ModifyAlarmResponse>
```

JSON format

```
{
  "AlarmTaskId": "asg-bp1hvbnmk110vll5**_83948190-acdd-483f-98f7-b77f4778****",
  "RequestId": "BACACF83-7070-4953-A8FD-D81F89F1****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

14.4. EnableAlarm

You can call this operation to enable an event-triggered task.

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
Action	String	Yes	EnableAlarm	The operation that you want to perform. Set the value to EnableAlarm.
AlarmTaskId	String	Yes	asg-bp1hvbnmk110vll5****_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.
RegionId	String	Yes	cn-qingdao	The region ID of the event-triggered task.

Response parameters

Parameter	Type	Example	Description
RequestId	String	688B18B8-FB1E-42EB-A1ED-7F55B090****	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=EnableAlarm
&RegionId=cn-qingdao
&AlarmTaskId=asg-bp1hvbnmk110vll5****_f95ce797-dc2e-4bad-9618-14fee7d1****
&<Common request parameters>
```

Sample success responses

XML format

```
<EnableAlarmResponse>
  <RequestId>688B18B8-FB1E-42EB-A1ED-7F55B090****</RequestId>
</EnableAlarmResponse>
```

JSON format

```
{
  "RequestId": "688B18B8-FB1E-42EB-A1ED-7F55B090****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

14.5. DisableAlarm

You can call this operation to disable an event-triggered task.

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
Action	String	Yes	DisableAlarm	The operation that you want to perform. Set the value to DisableAlarm.
AlarmTaskId	String	Yes	asg-bp1hvbnmk10vll5****_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.
RegionId	String	Yes	cn-qingdao	The region ID of the event-triggered task.

Response parameters

Parameter	Type	Example	Description
RequestId	String	086EFCD4-C76F-4DC6-9EE9-0D9B71E****	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=DisableAlarm
&RegionId=cn-qingdao
&AlarmTaskId=asg-bp1hvbnmk10vll5****_f95ce797-dc2e-4bad-9618-14fee7d1****
&<Common request parameters>
```

Sample success responses

XML format

```
<DisableAlarmResponse>
<RequestId>086EFCD4-C76F-4DC6-9EE9-0D9B711E****</RequestId>
</DisableAlarmResponse>
```

JSON format

```
{
  "RequestId": "086EFCD4-C76F-4DC6-9EE9-0D9B711E****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

14.6. DeleteAlarm

You can call this operation to delete an event-triggered task.

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
Action	String	Yes	DeleteAlarm	The operation that you want to perform. Set the value to DeleteAlarm.
AlarmTaskId	String	Yes	asg-bp1hvbnmk10vll5****_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.
RegionId	String	Yes	cn-qingdao	The region ID of the event-triggered task.

Response parameters

Parameter	Type	Example	Description
AlarmTaskId	String	asg-bp1hvbnmk10vll5**_f95ce797-dc2e-4bad-9618-14fee7d1****	The ID of the event-triggered task.

Parameter	Type	Example	Description
RequestId	String	6EF9BFEE-FE07-4627-B8FB-14326FB9****	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=DeleteAlarm
&RegionId=cn-qingdao
&AlarmTaskId=asg-bp1hvbnmk110v115****_f95ce797-dc2e-4bad-9618-14fee7d1****
&<Common request parameters>
```

Sample success responses

XML format

```
<DeleteAlarmResponse>
    <AlarmTaskId>asg-bp1hvbnmk110v115****_f95ce797-dc2e-4bad-9618-14fee7d1****</AlarmTaskId>
    <RequestId>6EF9BFEE-FE07-4627-B8FB-14326FB9****</RequestId>
</DeleteAlarmResponse>
```

JSON format

```
{
    "AlarmTaskId": "asg-bp1hvbnmk110v115****_f95ce797-dc2e-4bad-9618-14fee7d1****",
    "RequestId": "6EF9BFEE-FE07-4627-B8FB-14326FB9****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

15.Lifecycle Hook

15.1. CreateLifecycleHook

You can call this operation to create one or more lifecycle hooks for a scaling group.

Description

You can create a maximum of six lifecycle hooks for each scaling group. After a lifecycle hook is created for a scaling group, ECS instances in the scaling group are put into the wait state during scaling activities. You can use the HeartbeatTimeout parameter to specify the timeout period of the lifecycle hook. While an ECS instance is in the wait state, you can perform operations such as downloading data and initializing the configurations of the instance.

During a scale-out event, 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 ECS instances will be added to a specified SLB backend server group. During a scale-in event, ECS instances enter the wait state after they are removed from the SLB backend server group. When the wait state ends, their IP addresses are removed from the whitelist of the ApsaraDB for RDS instance.

We recommend that you use an Alibaba Cloud Message Service (MNS) queue or topic to create notifications about lifecycle hooks. Then, you can learn when ECS instances are started or released.

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
Action	String	Yes	CreateLifecycleHook	The operation that you want to perform. Set the value to CreateLifecycleHook.
LifecycleTransition	String	Yes	SCALE_OUT	The type of scaling activities to which the lifecycle hook applies. Valid values: <ul style="list-style-type: none">• SCALE_OUT: scale-out events of the scaling group• SCALE_IN: scale-in events of the scaling group
ScalingGroupId	String	Yes	asg-bp1eyv4qn8ssgv43***	The ID of the scaling group.

Parameter	Type	Required	Example	Description
LifecycleHookName	String	No	lifecyclehook****	<p>The name of the lifecycle hook. Each lifecycle hook name must be unique within a scaling group. 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.</p> <p>The default value is the value of the LifecycleHookId parameter.</p>
DefaultResult	String	No	CONTINUE	<p>The action that the scaling group takes when the lifecycle hook times out. Valid values:</p> <ul style="list-style-type: none"> • CONTINUE: The scaling group continues to respond to a 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. <p>If a 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 times out and continue with the action specified by the DefaultResult parameter.</p> <p>Default value: CONTINUE.</p>

Parameter	Type	Required	Example	Description
HeartbeatTimeout	Integer	No	600	<p>The wait period before the lifecycle hook times out. When the lifecycle hook times out, the scaling group performs the default action. Valid values: 30 to 21600. Unit: seconds.</p> <p>After you create a lifecycle hook, you can call the RecordLifecycleActionHeartbeat operation to extend the timeout period and keep the instance in the wait state. You can also call the CompleteLifecycleAction operation to terminate the wait state of a scaling activity.</p> <p>Default value: 600.</p>
NotificationMetadata	String	No	Test lifecycle hook.	<p>The fixed string to include when Auto Scaling sends a notification about the wait state of a scaling activity. The parameter value cannot exceed 4,096 characters in length.</p> <p>Auto Scaling sends the specified NotificationMetadata parameter value along with the notification message so that you can categorize your notifications. The NotificationMetadata parameter is valid only after you set the NotificationArn parameter.</p>

Parameter	Type	Required	Example	Description
NotificationArn	String	No	acs:mns:cn-beijing:161456884340****:queue/modifyLifecycleH0****	<p>The Alibaba Cloud Resource Name (ARN) of the notification object that Auto Scaling uses to notify you when ECS instances are put into the wait state by the lifecycle hook. If you do not set this parameter, Auto Scaling does not send you notifications. If you set this parameter, Auto Scaling sends you notifications of the following types:</p> <ul style="list-style-type: none"> • MNS queue. The format of the parameter value is acs:mns:{region-id}:{account-id}:queue/{queuename}. • MNS topic. The format of the parameter value is acs:mns:{region-id}:{account-id}:topic/{topicname}. • Operation Orchestration Service (OOS) template. The format of the parameter value is acs:oos:{region-id}:{account-id}:template/{templatename}. <p>The variables in the preceding parameter formats have the following meanings:</p> <ul style="list-style-type: none"> • region-id: the region ID of the scaling group • account-id: the ID of Alibaba Cloud account • queuename: the name of the MNS queue • topicname: the name of the MNS topic • templatename: the name of the OOS template

Response parameters

Parameter	Type	Example	Description
LifecycleHookId	String	ash-bp1at9ufhmcf9cmy***	The ID of the lifecycle hook.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action/CreateLifecycleHook
&ScalingGroupId=asg-bp1eyv4qn8ssgv43****
&LifecycleHookName=lifecyclehook****
&LifecycleTransition=SCALE_OUT
&NotificationArn=acs:mns:cn-beijing:161456884340****:queue/modifyLifecycleHo****
&NotificationMetadata=Test lifecycle hook.
&<Common request parameters>
```

Sample success responses

XML format

```
<CreateLifecycleHookResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
  <LifecycleHookId>ash-bp1at9ufhmcf9cmy***</LifecycleHookId>
</CreateLifecycleHookResponse>
```

JSON format

```
{
  "RequestId": "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
  "LifecycleHookId": "ash-bp1at9ufhmcf9cmy***"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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

HTTP status code	Error code	Error message	Description
400	InvalidNotificationArn	The specified parameter notificationArn is invalid.	The error message returned because the specified NotificationArn parameter is invalid.
400	UnsupportedNotificationType.CurrentRegion	The notificationType is not supported in the special region which scalingGroup belongs to.	The error message returned because the notification type is not supported in the region to which the scaling group belongs.
400	QueueNotExist	The specified queue does not exist.	The error message returned because the specified MNS queue does not exist.
400	TopicNotExist	The specified topic does not exist.	The error message returned because the specified MNS topic does not exist.
400	InvalidLifecycleHookName.Duplicate	The specified value of parameter lifecycleHookName is duplicated.	The error message returned because the specified lifecycle hook name already exists.
400	QuotaExceeded.LifecycleHook	Lifecycle hook quota exceeded in the specified scaling group.	The error message returned because the number of lifecycle hooks that can be created for a scaling group has exceeded the maximum value of six.

15.2. ModifyLifecycleHook

Modifies a lifecycle hook.

Description

You can use one of the following methods to specify the lifecycle hook that you want to modify:

- Specify the lifecycle hook ID by using the LifecycleHookId parameter. When you use this method, the ScalingGroupId and LifecycleHookName parameters are ignored.

- Specify the scaling group ID by using the `ScalingGroupId` parameter and the lifecycle hook name by using the `LifecycleHookName` 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
Action	String	Yes	ModifyLifecycleHook	The operation that you want to perform. Set the value to <code>ModifyLifecycleHook</code> .
LifecycleHookId	String	No	ash-bp1fxuqyi98w0ai b****	The ID of the lifecycle hook that you want to modify.
ScalingGroupId	String	No	asg-bp18p2yfxow2dl oq***	The ID of the scaling group to which the lifecycle hook belongs.
LifecycleHookName	String	No	test_SCALE_IN	The name of the lifecycle hook that you want to modify.

Parameter	Type	Required	Example	Description
DefaultResult	String	No	CONTINUE	<p>The operation that is performed when the lifecycle hook times out. Valid values:</p> <ul style="list-style-type: none"> • CONTINUE: Auto Scaling continues to respond to scale-in or scale-out requests. • ABANDON: Auto Scaling releases Elastic Compute Service (ECS) instances that are created during scale-out activities, or removes ECS instances from the scaling group during scale-in activities. <p>If multiple lifecycle hooks in a scaling group are triggered during scale-in activities and you set the DefaultResult parameter to ABANDON for the lifecycle hook that you want to modify, Auto Scaling immediately starts to perform the action after the lifecycle hook that you want to modify times out. As a result, other lifecycle hooks time out ahead of schedule. In other cases, Auto Scaling performs the action only after all lifecycle hooks time out.</p>
HeartbeatTimeout	Integer	No	600	<p>The period of time before the lifecycle hook times out. When the lifecycle hook times out, Auto Scaling performs the default action. Valid values: 30 to 21600. Unit: seconds.</p> <p>You can call the RecordLifecycleActionHeartbeat operation to prolong the period of time before the lifecycle hook times out. You can also call the CompleteLifecycleAction operation to end the lifecycle hook ahead of schedule.</p>
LifecycleTransition	String	No	SCALE_IN	<p>The type of scaling activities to which the lifecycle hook applies. Valid values:</p> <ul style="list-style-type: none"> • SCALE_OUT • SCALE_IN

Parameter	Type	Required	Example	Description
NotificationMetadata	String	No	Test	<p>The fixed string that is included in a notification when Auto Scaling sends the notification which indicates that the scaling activity is pending. The parameter value cannot exceed 4,096 characters in length.</p> <p>Auto Scaling sends the specified NotificationMetadata parameter value together with the notification. This helps you categorize your notifications. The NotificationMetadata parameter takes effect only after you specify the NotificationArn parameter.</p>
NotificationArn	String	No	acs:mns:cn-beijing:161456884340****:queue/modifyLifecycleH0****	<p>The Alibaba Cloud Resource Name (ARN) of the notified party.</p> <ul style="list-style-type: none"> If the notified party is a Message Service (MNS) queue, the value format of this parameter is <code>acs:mns:{region-id}:{account-id}:queue/{queuename}</code>. If the notified party is an MNS topic, the value format of this parameter is <code>acs:mns:{region-id}:{account-id}:topic/{topicname}</code>. If the notified party is an Operation Orchestration Service (OOS) template, the value format of this parameter is <code>acs:oos:{region-id}:{account-id}:template/{templatename}</code>. <p>The variables in the preceding formats have the following meanings:</p> <ul style="list-style-type: none"> region-id: the ID of the region where the scaling group resides. account-id: the ID of the Alibaba Cloud account. queuename: the name of the MNS queue. topicname: the name of the MNS topic. templatename: the name of the OOS template.
RegionId	String	No	cn-beijing	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=ModifyLifecycleHook
&LifecycleHookId=ash-bp1fxuqyi98w0ai*****  

&ScalingGroupId=asg-bp18p2yfxow2dloq*****  

&LifecycleHookName=test_SCALE_IN  

&DefaultResult=CONTINUE  

&HeartbeatTimeout=600  

&LifecycleTransition=SCALE_IN  

&NotificationMetadata=Test  

&NotificationArn=acs:mns:cn-beijing:161456884340****:queue/modifyLifecycleHo****  

&RegionId=cn-beijing  

&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<ModifyLifecycleHookResponse>
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</ModifyLifecycleHookResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned because the value that you specified for a parameter is invalid.
400	InvalidLifecycleHookId.NotExist	The specified lifecycleHookId not exist.	The error message returned because the lifecycle hook ID does not exist.
400	InvalidLifecycleHookName.NotExist	The specified lifecycleHookName you provided not exist.	The error message returned because the lifecycle hook name does not exist.
400	InvalidNotificationArn	The specified parameter NotificationArn is invalid.	The error message returned because the value that you specified for the NotificationArn parameter is invalid.
400	UnsupportedNotificationType.CurrentRegion	The NotificationType is not supported in the special region which scalingGroup belongs to.	The error message returned because 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 because the MNS queue does not exist.
400	TopicNotExist	The specified topic does not exist.	The error message returned because the MNS topic does not exist.

15.3. DescribeLifecycleHooks

Queries lifecycle hooks.

Description

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

- Specify a list of lifecycle hook IDs by using the LifecycleHookId.N parameter. In this case, you do not need to specify the ScalingGroupId and LifecycleHookName parameters.
- Specify the scaling group ID by using the ScalingGroupId parameter.
- Specify the scaling group ID by using the ScalingGroupId parameter and the lifecycle hook name by using the LifecycleHookName 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
Action	String	Yes	DescribeLifecycleHooks	The operation that you want to perform. Set the value to DescribeLifecycleHooks .
ScalingGroupId	String	No	asg-bp1igpak5ft1flyp****	The ID of the scaling group.
LifecycleHookName	String	No	lifecyclehook****	The name of the lifecycle hook.
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. Valid values: 1 to 50. Default value: 50.
RegionId	String	No	cn-beijing	The ID of the region to which the scaling group belongs.
LifecycleHookId.N	String	No	ash-bp1at9ufhmcf9cmy****	The ID of the lifecycle hook.

Response parameters

Parameter	Type	Example	Description
-----------	------	---------	-------------

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.
PageNumber	Integer	1	The page number of the returned page.
PageSize	Integer	50	The number of entries returned per page.
TotalCount	Integer	1	The total number of lifecycle hooks.
LifecycleHooks	Array of LifecycleHook		The details of the lifecycle hooks.
LifecycleHook			
DefaultResult	String	CONTINUE	The subsequent action to take when the lifecycle hook times out.
LifecycleHookId	String	ash-bp19d1032y9kij96****	The ID of the lifecycle hook.
LifecycleHookName	String	lifecyclehook****	The name of the lifecycle hook.
LifecycleTransition	String	SCALE_OUT	The type of the scaling activity to which the lifecycle hook applies.
NotificationMetadata	String	Test Lifecycle Hook.	The fixed string that is included when Auto Scaling sends a notification that a scaling activity is in the pending state.

Parameter	Type	Example	Description
NotificationArn	String	acs:ess:cn-beijing:161456884340****:null/null	<p>The Alibaba Cloud Resource Name (ARN) of the party to which Auto Scaling sends notifications when the lifecycle hook takes effect. The following list describes the value formats of this parameter:</p> <ul style="list-style-type: none"> If you did not create an event notification, the value format of this parameter is acs:ess:{region-id}:{account-id}:null/null. If the notified party is a Message Service (MNS) queue, the value format of this parameter is acs:mns:{region-id}:{account-id}:queue/{queuename}. If the notified party is an MNS topic, the value format of this parameter is acs:mns:{region-id}:{account-id}:topic/{topicname}. If the notified party is an Operation Orchestration Service (OOS) template, the value format of this parameter is acs:oos:{region-id}:{account-id}:template/{templatename}. <p>The variables in the preceding formats have the following meanings:</p> <ul style="list-style-type: none"> region-id: the ID of the region where the scaling group resides. account-id: the ID of the Alibaba Cloud account. queuename: the name of the MNS queue. topicname: the name of the MNS topic. templatename: the name of the OOS template.
HeartbeatTime out	Integer	60	The period of time during which the lifecycle hook takes effect. Auto Scaling performs the default action when the lifecycle hook times out.
ScalingGroupId	String	asg-bp1igpak5ft1flyp****	The ID of the scaling group.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DescribeLifecycleHooks
&ScalingGroupId=asg-bpligpdk5ft1flyp****
&LifecycleHookName=lifecyclehook****
&PageNumber=1
&PageSize=50
&LifecycleHookId=["ash-bplat9ufhmcf9cmy****"]
&RegionId=cn-beijing
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DescribeLifecycleHooksResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
  <PageNumber>1</PageNumber>
  <PageSize>50</PageSize>
  <TotalCount>1</TotalCount>
  <LifecycleHooks>
    <DefaultResult>CONTINUE</DefaultResult>
    <LifecycleHookId>ash-bp19d1032y9kij96****</LifecycleHookId>
    <LifecycleHookName>lifecyclehook****</LifecycleHookName>
    <LifecycleTransition>SCALE_OUT</LifecycleTransition>
    <NotificationMetadata>Test Lifecycle Hook.</NotificationMetadata>
    <NotificationArn>acs:ess:cn-beijing:161456884340****:null/null</NotificationArn>
    <HeartbeatTimeout>60</HeartbeatTimeout>
    <ScalingGroupId>asg-bpligpdk5ft1flyp****</ScalingGroupId>
  </LifecycleHooks>
</DescribeLifecycleHooksResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
  "PageNumber" : 1,
  "PageSize" : 50,
  "TotalCount" : 1,
  "LifecycleHooks" : [ {
    "DefaultResult" : "CONTINUE",
    "LifecycleHookId" : "ash-bp19d1032y9kij96****",
    "LifecycleHookName" : "lifecyclehook****",
    "LifecycleTransition" : "SCALE_OUT",
    "NotificationMetadata" : "Test Lifecycle Hook.",
    "NotificationArn" : "acs:ess:cn-beijing:161456884340****:null/null",
    "HeartbeatTimeout" : 60,
    "ScalingGroupId" : "asg-bpligpdk5ft1flyp****"
  } ]
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned because the value that you specified for a parameter is invalid.

15.4. RecordLifecycleActionHeartbeat

Extends the timeout period during which an Elastic Compute Service (ECS) instance is in the pending state after a lifecycle hook is triggered.

Description

You can call the operation up to 20 times to extend the timeout period of a lifecycle hook. However, the total length of the timeout period cannot exceed 6 hours because an ECS instance cannot be in the pending state for more than 6 hours.

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
Action	String	Yes	RecordLifecycleActionHeartbeat	The operation that you want to perform. Set the value to RecordLifecycleActionHeartbeat .
lifecycleHookId	String	Yes	ash-bp1fxuqyi98w0ai b****	The ID of the lifecycle hook.
lifecycleActionToken	String	Yes	F324B880-900E-4968-85DD-81691113****	The token that indicates the pending state of a scaling activity. You can obtain this token by using a Message Service (MNS) queue or an MNS topic that is specified for the lifecycle hook.

Parameter	Type	Required	Example	Description
heartbeatTimeout	Integer	No	600	<p>The timeout period of the lifecycle hook. When the lifecycle hook times out, Auto Scaling performs the default action. Unit: seconds. Valid values: 30 to 21600.</p> <p>After you create a lifecycle hook, you can call the RecordLifecycleActionHeartbeat operation to extend the timeout period of the lifecycle hook. You can also call the CompleteLifecycleAction operation to end the timeout period of the lifecycle hook in advance.</p> <p>Default value: 600.</p>
RegionId	String	No	cn-hangzhou	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=RecordLifecycleActionHeartbeat
&lifecycleHookId=ash-bp1fxuqy198w0aib****
&lifecycleActionToken=F324B880-900E-4968-85DD-81691113****
&heartbeatTimeout=600
&RegionId=cn-hangzhou
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<RecordLifecycleActionHeartbeatResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</RecordLifecycleActionHeartbeatResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is not valid.	The error message returned because the value that you specified for a 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 because the value specified for the lifecycleActionToken parameter does not match the value specified for the LifecycleHookId parameter.
400	LifecycleAction.TimeExceeded	The specified parameter heartbeatTime exceed lifecycleAction max suspend time.	The error message returned because the timeout period during which the ECS instance is in the pending state has exceeded 6 hours.
400	LifecycleAction.RecordTimesExceeded	The specified lifecycleAction exceed lifecycleAction max record times.	The error message returned because the operation is called more than 20 times.

15.5. DescribeLifecycleActions

Queries the lifecycle actions of a scaling activity.

Description

If a scaling activity whose type is the same as the type of a lifecycle hook occurs, the lifecycle hook triggers a lifecycle action. The status of a lifecycle action can be:

- Pending: Elastic Compute Service (ECS) instances in the scaling group enter the Pending state.
- Timeout: The lifecycle hook times out. ECS instances in the scaling group are automatically removed from the Pending state.
- Completed: ECS instances are manually removed from the Pending state.

If you do not specify the subsequent actions when you create a lifecycle hook, such as triggering the execution of a specific Operation Orchestration Service (OOS) template after the ECS instances are removed from the Pending state, you can call this operation to obtain the token of the lifecycle action that corresponds to the scaling activity. This helps you specify the subsequent actions.

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
Action	String	Yes	DescribeLifecycleActions	The operation that you want to perform. Set the value to DescribeLifecycleActions .
ScalingActivityId	String	Yes	asa-bp17mug9t0pegagw****	The ID of the scaling activity.
LifecycleActionStatus	String	No	Pending	The status of the lifecycle action. Valid values: <ul style="list-style-type: none">• Pending: ECS instances enter the Pending state.• Timeout: The lifecycle hook times out. ECS instances are automatically removed from the Pending state.• Completed: ECS instances are manually removed from the Pending state.

Parameter	Type	Required	Example	Description
NextToken	String	No	AAAAAcSz4VTb1Nq****	<p>The token that is used to specify the lifecycle action from which the query starts.</p> <p>For example, after the first 10 lifecycle actions were queried, the query starts from the 11th lifecycle action. Set this parameter to the NextToken value that is returned in the previous API call. If you do not specify this parameter, the query starts from the beginning.</p>
MaxResults	Integer	No	10	<p>The maximum number of entries to return on each page. Valid values: 1 to 50.</p> <p>Default value: 10.</p>
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.

Response parameters

Parameter	Type	Example	Description
NextToken	String	AAAAAcSz4VTb1Nq* ***	The query token returned in this call.
RequestId	String	42A742EB-FCF3- 459E-9C62- E107048C17E3	The ID of the request.
TotalCount	Integer	3	The total number of queried lifecycle actions.
MaxResults	Integer	3	The maximum number of entries returned per page.
LifecycleActions	Array of LifecycleAction		The lifecycle actions that correspond to each lifecycle hook.
LifecycleAction			

Parameter	Type	Example	Description
LifecycleHookId	String	ash-bp18uoft0deax0f7**	The ID of the lifecycle hook.
LifecycleAction Token	String	9C2E9DA7-F794-449A-ACF6-CEE24444F7BB	The token of the lifecycle action.
LifecycleAction Status	String	Pending	The status of the lifecycle action.
LifecycleActionResult	String	CONTINUE	<p>The action that is performed after the lifecycle action triggered by the lifecycle hook is complete. Valid values:</p> <ul style="list-style-type: none"> CONTINUE: Auto Scaling continues to respond to scale-out requests and adds ECS instances to the scaling group, or continues to respond to scale-in requests and removes ECS instances from the scaling group. ABANDON: Auto Scaling stops scale-out activities and releases the created ECS instances, or continues to respond to scale-in requests and removes ECS instances from the scaling group.
InstanceIds	Array of String	["i-bp11m3fzlqrgk5vh****", "i-bp11m3fzlqrgk5vh****"]	The IDs of the ECS instances to which the lifecycle hook applies.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DescribeLifecycleActions
&ScalingActivityId=asa-bp17mug9t0pegagw****
&LifecycleActionStatus=Pending
&NextToken=AAAAAcSz4VTb1Nq****
&MaxResults=10
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DescribeLifecycleActionsResponse>
  <NextToken>AAAAAcSz4VTb1Nq****</NextToken>
  <RequestId>42A742EB-FCF3-459E-9C62-E107048C17E3</RequestId>
  <TotalCount>3</TotalCount>
  <MaxResults>3</MaxResults>
  <LifecycleActions>
    <LifecycleHookId>ash-bp18uoft0deax0f7****</LifecycleHookId>
    <LifecycleActionToken>9C2E9DA7-F794-449A-ACF6-CEE24444F7BB</LifecycleActionToken>
    <LifecycleActionStatus>Pending</LifecycleActionStatus>
    <LifecycleActionResult>CONTINUE</LifecycleActionResult>
    <InstanceIds>["i-bp11m3fz1qrgk5vh****", "i-bp11m3fz1qrgk5vh****"]</InstanceIds>
  </LifecycleActions>
</DescribeLifecycleActionsResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "NextToken" : "AAAAAcSz4VTb1Nq****",
  "RequestId" : "42A742EB-FCF3-459E-9C62-E107048C17E3",
  "TotalCount" : 3,
  "MaxResults" : 3,
  "LifecycleActions" : [ {
    "LifecycleHookId" : "ash-bp18uoft0deax0f7****",
    "LifecycleActionToken" : "9C2E9DA7-F794-449A-ACF6-CEE24444F7BB",
    "LifecycleActionStatus" : "Pending",
    "LifecycleActionResult" : "CONTINUE",
    "InstanceIds" : [ "\"i-bp11m3fz1qrgk5vh****\", \"i-bp11m3fz1qrgk5vh****\""] }
  ]
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidParameter	The specified value of parameter "ScalingActivityId" is not valid.	The error message returned because the value that you specified for the ScalingActivityId parameter is invalid.

HTTP status code	Error code	Error message	Description
400	InvalidParameter	The specified value of parameter "MaxResults" is not valid.	The error message returned because the value that you specified for the MaxResults parameter is invalid.
400	InvalidParameter	The specified value of parameter "LifecycleActionResult" is not valid.	The error message returned because the value that you specified for the LifecycleActionResult parameter is invalid.

15.6. CompleteLifecycleAction

Removes a scaling activity from the pending state in advance.

Description

After you remove a scaling activity from the pending state, you can set the LifecycleActionResult parameter to CONTINUE to continue the scaling activity. If you want to skip the scaling activity, you can set the LifecycleActionResult parameter to ABANDON.

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
Action	String	Yes	CompleteLifecycleAction	The operation that you want to perform. Set the value to CompleteLifecycleAction .
LifecycleHookId	String	Yes	ash-bp14g3ee6bt3sc98***	The ID of the lifecycle hook.
LifecycleActionToken	String	Yes	aaaa-bbbbbcccc-ddddd	The token that indicates the pending state of a scaling activity. You can obtain this token by using a Message Service (MNS) queue or an MNS topic that is specified for the lifecycle hook.

Parameter	Type	Required	Example	Description
LifecycleActionResult	String	No	CONTINUE	<p>The action that you want Auto Scaling to perform after the lifecycle hook times out. Valid values:</p> <ul style="list-style-type: none">• CONTINUE: Auto Scaling continues to add Elastic Compute Service (ECS) instances to the scaling group, or continues to remove ECS instances from the scaling group.• ABANDON: Auto Scaling stops adding ECS instances to the scaling group and releases the ECS instances, or continues to respond to scale-in requests and remove ECS instances from the scaling group. <p>Default value: CONTINUE.</p> <p>If multiple lifecycle hooks exist in a scaling group and are triggered at the same time, the following rules apply:</p> <ul style="list-style-type: none">• If you set the LifecycleActionResult parameter to ABANDON for the lifecycle hook that is applied to a scale-in activity, Auto Scaling immediately removes ECS instances from the scaling group after the lifecycle hook expires, without the need to wait for the last lifecycle hook to expire.• If you set the LifecycleActionResult parameter to CONTINUE for the lifecycle hook that is applied to a scale-in or scale-out activity, Auto Scaling performs the next action until the last lifecycle hook in the scaling group times out. The action that Auto Scaling performs varies based on the value that you specify for the LifecycleActionResult parameter of the last lifecycle hook.

Parameter	Type	Required	Example	Description
ClientToken	String	No	123e4567-e89b-12d3-a456-42665544****	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests.</p> <p>The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence.</p>
RegionId	String	No	cn-qingdao	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=CompleteLifecycleAction
&LifecycleHookId=ash-bp14g3ee6bt3sc98****
&LifecycleActionResult=CONTINUE
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&RegionId=cn-qingdao
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<CompleteLifecycleActionResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</CompleteLifecycleActionResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidParamter	The specified value of parameter is invalid.	The error message returned because the value that you specified for a 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 because the value specified for the LifecycleActionToken parameter does not match the value specified for the LifecycleHookId parameter.

15.7. DeleteLifecycleHook

Deletes a lifecycle hook.

Description

If you delete a lifecycle hook that is in effect in a scaling group, instances are removed from the pending state in advance. You can use one of the following methods to specify the lifecycle hooks that you want to delete:

- Specify the lifecycle hook ID by using the `LifecycleHookId` parameter. In this case, the `ScalingGroupId` parameter and the `LifecycleHookName` parameter are ignored.
- Specify the scaling group ID by using the `ScalingGroupId` parameter and specify the lifecycle hook name by using the `LifecycleHookName` 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
Action	String	Yes	DeleteLifecycleHook	The operation that you want to perform. Set the value to DeleteLifecycleHook .
LifecycleHookId	String	No	ash-bp14g3ee6bt3sc98****	The ID of the lifecycle hook.
ScalingGroupId	String	No	asg-bp18p2yfxow2dl0q****	The ID of the scaling group.
LifecycleHookName	String	No	lifecyclehook****	The name of the lifecycle hook.
RegionId	String	No	cn-hangzhou	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DeleteLifecycleHook
&LifecycleHookId=ash-bp14g3ee6bt3sc98****
&ScalingGroupId=asg-bp18p2yfxow2dl0q****
&LifecycleHookName=lifecyclehook****
&RegionId=cn-hangzhou
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DeleteLifecycleHookResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</DeleteLifecycleHookResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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

16. Event notification

16.1. CreateNotificationConfiguration

Creates an event notification.

Description

You can configure CloudMonitor system events, Message Service (MNS) queues, or MNS topics to receive notifications. When a specified type of scaling activity or resource change occurs in a scaling group, Auto Scaling notifies CloudMonitor or MNS.

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
Action	String	Yes	CreateNotificationConfiguration	The operation that you want to perform. Set the value to CreateNotificationConfiguration .
ScalingGroupId	String	Yes	asg-bp1igpak5ft1flyp****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
NotificationArn	String	Yes	acs:mns:cn-beijing:161456884340****:queue/modifyLifecycleH0****	<p>The Alibaba Cloud Resource Name (ARN) of the notified party. The following list describes the value formats of this parameter:</p> <ul style="list-style-type: none"> If the notified party is CloudMonitor, the value format of this parameter is acs:ess:{region-id}:{account-id}:cloudmonitor. If the notified party is an MNS queue, the value format of this parameter is acs:mns:{region-id}:{account-id}:queue/{queuename}. If the notified party is an MNS topic, the value format of this parameter is acs:mns:{region-id}:{account-id}:topic/{topicname}. <p>The variables in the preceding formats have the following meanings:</p> <ul style="list-style-type: none"> region-id: the region ID of the scaling group. account-id: the ID of the Alibaba Cloud account. queuename: the name of the MNS queue. topicname: the name of the MNS topic.
RegionId	String	No	cn-beijing	The ID of the region where the scaling group resides.
NotificationType.N	String	Yes	AUTOSCALING:SCALE_OUT_SUCCESS	<p>The type of notification N for scaling activities or resource changes. Valid values of N: 1 to 8. Specify multiple values in the repeated list form.</p> <p>You can call the DescribeNotificationTypes operation to query the values of this parameter.</p>

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action/CreateNotificationConfiguration  
&ScalingGroupId=asg-bp1igpak5ft1flyp****  
&NotificationArn=acs:mns:cn-beijing:161456884340****:queue/modifyLifecycleHo****  
&NotificationType=[ "AUTOSCALING:SCALE_OUT_SUCCESS" ]  
&RegionId=cn-beijing  
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK  
Content-Type:application/xml  
<CreateNotificationConfigurationResponse>  
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>  
</CreateNotificationConfigurationResponse>
```

JSON format

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidNotificationArn	The specified parameter notificationArn is invalid.	The error message returned because the value that you specified for the NotificationArn parameter is invalid.

HTTP status code	Error code	Error message	Description
400	InvalidNotificationTypes	The specified notification type is invalid.	The error message returned because the value that you specified for the NotificationType.N parameter is invalid.
400	NotificationConfigurationExist	The specified notification configuration already exist for the scalingGroup.	The error message returned because the specified event notification already exists in the scaling group.
400	NotificationConfigurationQuotaExceed.ForScalingGroup	NotificationConfiguration num exceed for the specified scalingGroup.	The error message returned because the maximum number of notifications allowed for the scaling group has been reached.
400	QueueNotExist	The specified queue queuename does not exist.	The error message returned because the specified MNS queue does not exist.
400	TopicNotExist	The specified topic topicname does not exist.	The error message returned because the specified MNS topic does not exist.
400	UnsupportedNotificationType.CurrentRegion	The NotificationType is not supported in the special region which scaling group belongs to.	The error message returned because the specified 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 because the specified scaling group does not exist.

16.2. ModifyNotificationConfiguration

Modifies the configurations of an event notification.

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
Action	String	Yes	ModifyNotificationConfiguration	The operation that you want to perform. Set the value to ModifyNotificationConfiguration .
ScalingGroupId	String	Yes	asg-bp1igpak5ft1flyp****	The ID of the scaling group.
NotificationArn	String	Yes	acs:ess:cn-beijing:161456884340****:cloudmonitor	<p>The Alibaba Cloud Resource Name (ARN) of the notified party. The following list describes the value formats of this parameter:</p> <ul style="list-style-type: none">If the notified party is CloudMonitor, the value format of this parameter is <code>acs:ess:{region-id}:{account-id}:cloudmonitor</code>.If the notified party is a Message Service (MNS) queue, the value format of this parameter is <code>acs:mns:{region-id}:{account-id}:queue/{queuename}</code>.If the notified party is an MNS topic, the value format of this parameter is <code>acs:mns:{region-id}:{account-id}:topic/{topicname}</code>. <p>The variables in the preceding formats have the following meanings:</p> <ul style="list-style-type: none">region-id: the ID of the region where the scaling group resides.account-id: the ID of the Alibaba Cloud account.queuename: the name of the MNS queue.topicname: the name of the MNS topic.

Parameter	Type	Required	Example	Description
RegionId	String	No	cn-beijing	The ID of the region where the scaling group resides.
NotificationType.N	String	Yes	AUTOSCALING:SCALE_OUT_SUCCESS	The type of notification N for scaling activities or resource changes. Valid values of N: 1 to 8. Specify multiple values in the repeated list form. You can call the DescribeNotificationTypes operation to query the value of this parameter.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=ModifyNotificationConfiguration
&ScalingGroupId=asg-bpligpak5ft1flyp****
&NotificationArn=acs:ess:cn-beijing:161456884340****:cloudmonitor
&NotificationType=["AUTOSCALING:SCALE_OUT_SUCCESS"]
&RegionId=cn-beijing
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<ModifyNotificationConfigurationResponse>
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</ModifyNotificationConfigurationResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidNotificationTypes	The specified NotificationType is invalid.	The error message returned because the value that you specified for the NotificationType.N parameter is invalid.
400	NotificationConfigurationNotExist	The specified notification configuration not exist for the scaling group.	The error message returned because the specified notification type for scaling activities and resource changes does not exist in the scaling group.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist.

16.3.

DescribeNotificationConfigurations

Queries event notifications.

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
Action	String	Yes	DescribeNotificationConfigurations	The operation that you want to perform. Set the value to DescribeNotificationConfigurations .

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	asg-bp1igpak5ft1flyp****	The ID of the scaling group.
RegionId	String	No	cn-beijing	The ID of the region where the scaling group resides.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3C83E	The ID of the request.
NotificationConfigurationModes	Array of NotificationConfigurationMode l		The details of the notifications for scaling activities or resource changes.
NotificationConfigurationMode l			
NotificationArn	String	acs:mns:cn-beijing:161456884340****:topic/modifyLifeCycleHo****	<p>The Alibaba Cloud Resource Name (ARN) of the notified party. The following list describes the value formats of this parameter:</p> <ul style="list-style-type: none"> If the notified party is CloudMonitor, the value format is acs:ess:{region-id}:{account-id}:cloudmonitor. If the notified party is a Message Service (MNS) queue, the value format is acs:mns:{region-id}:{account-id}:queue/{queuename}. If the notified party is an MNS topic, the value format is acs:mns:{region-id}:{account-id}:topic/{topicname}. <p>The variables in the preceding formats have the following meanings:</p> <ul style="list-style-type: none"> region-id: the region ID of the scaling group. account-id: the ID of the Alibaba Cloud account. queuename: the name of the MNS queue. topicname: the name of the MNS topic.

Parameter	Type	Example	Description
ScalingGroupId	String	asg-bp1igpak5ft1flyp**** *	The ID of the scaling group.
NotificationTypes	Array of String	AUTOSCALING:SCALE_IN_SUCCESS	The types of notifications for scaling activities or resource changes.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=DescribeNotificationConfigurations
&ScalingGroupId=asg-bp1igpak5ft1flyp****
&RegionId=cn-beijing
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<DescribeNotificationConfigurationsResponse>
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
    <NotificationConfigurationModels>
        <NotificationArn>acs:mns:cn-beijing:161456884340****:topic/modifyLifecycleHo****</NotificationArn>
        <ScalingGroupId>asg-bp1igpak5ft1flyp****</ScalingGroupId>
        <NotificationTypes>AUTOSCALING:SCALE_IN_SUCCESS</NotificationTypes>
    </NotificationConfigurationModels>
</DescribeNotificationConfigurationsResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E",
    "NotificationConfigurationModels" : [ {
        "NotificationArn" : "acs:mns:cn-beijing:161456884340****:topic/modifyLifecycleHo****",
        "ScalingGroupId" : "asg-bp1igpak5ft1flyp****",
        "NotificationTypes" : [ "AUTOSCALING:SCALE_IN_SUCCESS" ]
    } ]
}
```

Error codes

For a list of error codes, visit the [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.

16.4. DeleteNotificationConfiguration

Deletes a notification.

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
Action	String	Yes	DeleteNotificationConfiguration	The operation that you want to perform. Set the value to DeleteNotificationConfiguration .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dloq****	The ID of the scaling group.

Parameter	Type	Required	Example	Description
NotificationArn	String	Yes	acs:ess:cn-beijing:161456884340****:cloudmonitor	<p>The Alibaba Cloud Resource Name (ARN) of the notified party. The following list describes the value formats of this parameter:</p> <ul style="list-style-type: none"> If the notified party is CloudMonitor, the value format of this parameter is acs:ess:{region-id}:{account-id}:cloudmonitor. If the notified party is a Message Service (MNS) queue, the value format of this parameter is acs:mns:{region-id}:{account-id}:queue/{queuename}. If the notified party is an MNS topic, the value format of this parameter is acs:mns:{region-id}:{account-id}:topic/{topicname}. <p>The variables in the preceding formats have the following meanings:</p> <ul style="list-style-type: none"> region-id: the region ID of the scaling group. account-id: the ID of the Alibaba Cloud account. queuename: the name of the MNS queue. topicname: the name of the MNS topic.
RegionId	String	No	cn-beijing	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample request

```
http(s)://ess.aliyuncs.com/?Action=DeleteNotificationConfiguration  
&ScalingGroupId=asg-bp18p2yfxow2dlog****  
&NotificationArn=acs:ess:cn-beijing:161456884340****:cloudmonitor  
&RegionId=cn-beijing  
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK  
Content-Type:application/xml  
<DeleteNotificationConfigurationResponse>  
    <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>  
</DeleteNotificationConfigurationResponse>
```

JSON format

```
HTTP/1.1 200 OK  
Content-Type:application/json  
{  
    "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	NotificationConfigurationNotExist	The specified notification configuration not exist for the scaling group.	The error message returned because the specified notification for scaling activities and resource changes does not exist in the scaling group.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist.

16.5. DescribeNotificationTypes

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

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
Action	String	Yes	DescribeNotificationTypes	The operation that you want to perform. Set the value to DescribeNotificationTypes.

Response parameters

Parameter	Type	Example	Description
NotificationTypes	List	AUTOSCALING:SCALE_OUT_SUCCESS	<p>The types of notifications for scaling activities and resource changes.</p> <ul style="list-style-type: none"> • AUTOSCALING:SCALE_OUT_SUCCESS: The scale-out event is successful. • AUTOSCALING:SCALE_IN_SUCCESS: The scale-in event is successful. • AUTOSCALING:SCALE_OUT_ERROR: The scale-out event fails. • AUTOSCALING:SCALE_IN_ERROR: The scale-in event fails. • AUTOSCALING:SCALE_REJECT: The scaling activity is rejected. • AUTOSCALING:SCALE_OUT_START: The scale-out event is started. • AUTOSCALING:SCALE_IN_START: The scale-in event is started. • AUTOSCALING:SCHEDULE_TASK_EXPIRING: Auto Scaling sends a notification when a scheduled task is about to expire.
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DCODE3C83E	The ID of the request.

Examples

Sample requests

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

Sample success responses

XML format

```
<DescribeNotificationTypesResponse>
  <RequestId>66046F31-AF8C-44B4-9992-453F33CF179E</RequestId>
  <NotificationTypes>
    <NotificationType>AUTOSCALING:SCALE_OUT_SUCCESS</NotificationType>
    <NotificationType>AUTOSCALING:SCALE_IN_SUCCESS</NotificationType>
    <NotificationType>AUTOSCALING:SCALE_OUT_ERROR</NotificationType>
    <NotificationType>AUTOSCALING:SCALE_IN_ERROR</NotificationType>
    <NotificationType>AUTOSCALING:SCALE_REJECT</NotificationType>
    <NotificationType>AUTOSCALING:SCALE_OUT_START</NotificationType>
    <NotificationType>AUTOSCALING:SCALE_IN_START</NotificationType>
    <NotificationType>AUTOSCALING:SCHEDULE_TASK_EXPIRING</NotificationType>
  </NotificationTypes>
</DescribeNotificationTypesResponse>
```

JSON format

```
{
  "RequestId": "66046F31-AF8C-44B4-9992-453F33CF179E",
  "NotificationTypes": {
    "NotificationType": [
      "AUTOSCALING:SCALE_OUT_SUCCESS",
      "AUTOSCALING:SCALE_IN_SUCCESS",
      "AUTOSCALING:SCALE_OUT_ERROR",
      "AUTOSCALING:SCALE_IN_ERROR",
      "AUTOSCALING:SCALE_REJECT",
      "AUTOSCALING:SCALE_OUT_START",
      "AUTOSCALING:SCALE_IN_START",
      "AUTOSCALING:SCHEDULE_TASK_EXPIRING"
    ]
  }
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

17.Instance

17.1. SetInstanceHealth

You can call this operation to set the health status of the ECS instances in a scaling group.

Description

Auto Scaling performs health checks on ECS instances in a scaling group and removes unhealthy ECS instances. If you want to reserve an ECS instance, you can switch the ECS instance to the Standby or Protected state. For more information, see [EnterStandby](#) and [SetInstancesProtection](#).

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
Action	String	Yes	SetInstanceHealth	The operation that you want to perform. Set the value to SetInstanceHealth.
HealthStatus	String	Yes	Healthy	The health status of an ECS instance. Valid values: <ul style="list-style-type: none">• Healthy• Unhealthy
InstanceId	String	Yes	i-bp1ap6bro51a7fsa***	The ID of an ECS instance in a scaling group.

Response parameters

Parameter	Type	Example	Description
RequestId	String	B755AE57-6093-43E4-938E-DEA422A9B10F	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=SetInstanceHealth  
&HealthStatus=Unhealthy  
&InstanceId=i-bplap6bro51a7fsa****  
&<Common request parameters>
```

Sample success responses

XML format

```
<SetInstanceHealthResponse>  
    <RequestId>B755AE57-6093-43E4-938E-DEA422A9B10F</RequestId>  
</SetInstanceHealthResponse>
```

JSON format

```
{  
    "RequestId": "B755AE57-6093-43E4-938E-DEA422A9B10F"  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidInstanceId.NotFound	Instance "%s" does not exist.	The error message returned because the specified instance does not exist in the scaling group.
400	InvalidParameter	The specified value of parameter "%s" is not valid.	The error message returned because a specified parameter is invalid.

17.2. EnterStandby

Puts an Elastic Compute Service (ECS) instance into the Standby state.

Description

- If you call the operation to put an ECS instance in a scaling group that is associated with a Server Load Balancer (SLB) instance into the Standby state, the weight of the ECS instance as the backend server of the SLB instance is set to 0.
- You can remove an instance that is in the Standby state from a scaling group, and then release the instance.
- ECS instances that are in the Standby state are not removed from the scaling group during scale-in

- activities triggered by event-triggered tasks.
- If Auto Scaling considers an ECS instance that is in the Standby state unhealthy, for example, the ECS instance is being stopped or being restarted, Auto Scaling does not update the health status of the ECS instance or trigger scale-in activities to remove the ECS instance from the scaling group. Auto Scaling updates the health status of the ECS instance only when the ECS instance is no longer in the Standby state.

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
Action	String	Yes	EnterStandby	The operation that you want to perform. Set the value to EnterStandby .
ScalingGroupId	String	Yes	asg-bp1fo0dbtsbmq a9h***	The ID of the scaling group.
ClientToken	String	No	123e4567-e89b- 12d3-a456- 42665544***	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests.</p> <p>The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence.</p>
Async	Boolean	No	test	<p>Specifies whether to put the ECS instance into the Standby state in an asynchronous manner. Valid values:</p> <ul style="list-style-type: none">• true• false <p>Default value: false.</p>
InstanceId.N	String	Yes	i-28wt4***	The ID of ECS instance N.

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F-4DC5-B3DB-A3DC0DE3****	The ID of the request.
ScalingActivityId	String	asa-2zeb04oym05qaceq****	The ID of the scaling activity.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=EnterStandby
&ScalingGroupId=asg-bp1fo0dbtsbmqa9h****
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&InstanceId=["i-28wt4****"]
&Async=false
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<EnterStandbyResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>
  <ScalingActivityId>asa-2zeb04oym05qaceq****</ScalingActivityId>
</EnterStandbyResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3****",
  "ScalingActivityId" : "asa-2zeb04oym05qaceq****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because the RAM user is not authorized to call the operation. Contact the owner of the Alibaba Cloud account for the authorization and try again.
404	InvalidInstanceId.NotFound	Instance "XXX" does not exist.	The error message returned because the ECS instance does not exist.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the scaling group does not exist.

17.3. ExitStandby

Removes an Elastic Compute Service (ECS) instance from the Standby state.

Description

If a scaling group is associated with Server Load Balancer (SLB) instances, the weight of the ECS instance in the scaling group is set to the value that is specified in the scaling configuration when the ECS instance is removed from the Standby state.

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
Action	String	Yes	ExitStandby	The operation that you want to perform. Set the value to ExitStandby .

Parameter	Type	Required	Example	Description
ScalingGroupId	String	Yes	asg-bp1fo0dbtsbmq a9h***	The ID of the scaling group.
ClientToken	String	No	123e4567-e89b- 12d3-a456- 42665544***	<p>The client token that is used to ensure the idempotence of the request. You can use the client to generate the value, but you must ensure that it is unique among different requests.</p> <p>The token can only contain ASCII characters and cannot exceed 64 characters in length. For more information, see How to ensure idempotence.</p>
Async	Boolean	No	false	<p>Specifies whether to asynchronously remove the ECS instance from the Standby state. Valid values:</p> <ul style="list-style-type: none"> • true • false <p>Default value: false.</p>
RegionId	String	No	cn-hangzhou	The ID of the region where the scaling group resides.
InstanceId.N	String	Yes	i-28wt4***	The ID of ECS instance N. The value of this parameter can be a JSON array that consists of up to 20 instance IDs. Separate multiple instance IDs with commas (,).

Response parameters

Parameter	Type	Example	Description
RequestId	String	473469C7-AA6F- 4DC5-B3DB- A3DCODE3***	The ID of the request.
ScalingActivityId	String	asa- 2zeb04oym05qaceq ***	The ID of the scaling activity.

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=ExitStandby
&ScalingGroupId=asg-bp1fo0dbtsbmqa9h****
&ClientToken=123e4567-e89b-12d3-a456-42665544****
&InstanceId=["i-28wt4****"]
&Async=false
&RegionId=cn-hangzhou
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<ExitStandbyResponse>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3****</RequestId>
  <ScalingActivityId>asa-2zeb04oym05qaceq****</ScalingActivityId>
</ExitStandbyResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3****",
  "ScalingActivityId" : "asa-2zeb04oym05qaceq****"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because the RAM user is not authorized to call this operation. Contact the owner of the Alibaba Cloud account to obtain the required permissions and try again.

HTTP status code	Error code	Error message	Description
404	InvalidInstanceId.NotFound	A required authorization for the specified action is not supplied.	The error message returned because the specified ECS instance does not exist.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist within the Alibaba Cloud account.

17.4. RebalanceInstances

Rebalances Elastic Compute Service (ECS) instances in a multi-zone scaling group.

Description

To rebalance ECS instances in a multiple-zone scaling group, Auto Scaling creates ECS instances to replace the existing instances. Auto Scaling starts the created instances before it terminates the existing instances. This ensures that the performance and availability of your applications are not affected.

- You can call this operation only if you set the MultiAZPolicy parameter of the multi-zone scaling group to BALANCE.
- You can rebalance ECS instances in a scaling group only when the distribution of ECS instances in the scaling group is significantly imbalanced.
- You can replace up to 20 ECS instances during a rebalancing activity.
- If the number of ECS instances in the scaling group approaches or reaches the maximum number of instances that is specified by the MaxSize parameter during a rebalancing activity and the rebalancing activity must continue, Auto Scaling allows the number of ECS instances in the scaling group to temporarily exceed the maximum number. The excess number of instances ranges from one to 10% of the value of the MaxSize parameter. In most cases, Auto Scaling requires one to six minutes to rebalance the instances in the scaling group when the number of instances in the scaling group exceeds the maximum number.

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
-----------	------	----------	---------	-------------

Parameter	Type	Required	Example	Description
Action	String	Yes	RebalanceInstances	The operation that you want to perform. Set the value to RebalanceInstances .
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dloq****	The ID of the scaling group.
RegionId	String	No	cn-hangzhou	The ID of the region where the scaling group resides.

Response parameters

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

Examples

Sample requests

```
http(s)://ess.aliyuncs.com/?Action=RebalanceInstances
&ScalingGroupId=asg-bp18p2yfxow2dloq****
&RegionId=cn-hangzhou
&<Common request parameters>
```

Sample success responses

XML format

```
HTTP/1.1 200 OK
Content-Type:application/xml
<RebalanceInstancesResponse>
  <ScalingActivityId>asa-kjgffgdfadah****</ScalingActivityId>
  <RequestId>473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E</RequestId>
</RebalanceInstancesResponse>
```

JSON format

```
HTTP/1.1 200 OK
Content-Type:application/json
{
  "ScalingActivityId" : "asa-kjgffgdafadah****",
  "RequestId" : "473469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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 scaling group is disabled.
400	OperationDenied	This operation is denied because the specified scaling group does not support this action.	The error message returned because the MultiAZPolicy parameter of the scaling group is not set to BALANCE, or because the distribution of ECS instances in the scaling group is slightly imbalanced.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because you are not authorized to call the RebalanceInstances operation.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the scaling group does not exist within the Alibaba Cloud account.

17.5. SetInstancesProtection

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

Description

After an ECS instance is put into the Protected state, the following limits apply to the instance:

- The instance remains in the Protected state until you remove the instance from the Protected state.
- Even if a scale-in event is triggered for a scaling group due to a change in the number of ECS instances or due to an event-triggered task, Auto Scaling does not remove the ECS instances in the Protected state. You must remove the ECS instance from the Protected state before Auto Scaling can remove the ECS instance from the scaling group and release the instance. For information about how to remove an ECS instance from a scaling group, see [Remove an ECS instance from a scaling group](#).

- When the ECS instance is stopped or restarted, the health status of the instance remains unchanged.

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
Action	String	Yes	SetInstancesProtection	The operation that you want to perform. Set the value to SetInstancesProtection.
InstanceId.N	RepeatList	Yes	i-28wt4***	The ID of ECS instance N. Valid values of N: 1 to 20.
ProtectedFromScaleIn	Boolean	Yes	true	Specifies whether to enable protection for an ECS instance to prevent the ECS instance from being stopped or removed from the scaling group during scale-in events. Valid values: <ul style="list-style-type: none">• true: enables protection for the ECS instance.• false: disables protection for the ECS instance.
ScalingGroupId	String	Yes	asg-bp18p2yfxow2dl0q***	The ID of the scaling group.

Response parameters

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

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=SetInstancesProtection  
&ScalingGroupId=asg-bp18p2yfxow2dlog****  
&InstanceId.1=i-28wt4****  
&InstanceId.2=i-bp1j1****  
&ProtectedFromScaleIn=true  
&<Common request parameters>
```

Sample success responses

XML format

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

JSON format

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

Error codes

For a list of error codes, visit the [API Error Center](#).

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 scaling group is disabled.
403	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	The error message returned because you are not authorized to call the SetInstancesProtection operation.

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.
404	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	The error message returned because the specified scaling group does not exist.

18.Region

18.1. DescribeRegions

You can call this operation to query the regions where Auto Scaling is available.

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
Action	String	Yes	DescribeRegions	The operation that you want to perform. Set the value to DescribeRegions.
AcceptLanguage	String	No	zh-CN	<p>The language that is used as a filter condition to filter returned results. For more information, see RFC 7231. Valid values:</p> <ul style="list-style-type: none">• zh-CN: Chinese• en-US: English• ja: Japanese <p>Default value: zh-CN.</p>

Response parameters

Parameter	Type	Example	Description
Regions	Array of Region		Details about the regions.
Region			
ClassicUnavailable	Boolean	false	<p>Indicates whether the region supports scaling groups of the classic network type. Valid values:</p> <ul style="list-style-type: none">• true: The region does not support scaling groups of the classic network type.• false: The region supports scaling groups of the classic network type.

Parameter	Type	Example	Description
LocalName	String	China (Beijing)	The name of the region.
RegionEndpoint	String	ess.aliyuncs.com	The endpoint of the region.
RegionId	String	cn-beijing	The region ID.
VpcUnavailable	Boolean	false	Indicates whether the region supports scaling groups of the VPC type. Valid values: <ul style="list-style-type: none">• true: The region does not support scaling groups of the VPC type.• false: The region supports scaling groups of the VPC type.
RequestId	String	73469C7-AA6F-4DC5-B3DB-A3DC0DE3C83E	The ID of the request.

Examples

Sample requests

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

Sample success responses

XML format

```
<DescribeRegionsResponse>
  <RequestId>38EC7366-F5A9-46B1-BDB1-0FDC2E296397</RequestId>
  <Regions>
    <Region>
      <ClassicUnavailable>false</ClassicUnavailable>
      <RegionId>cn-beijing</RegionId>
      <VpcUnavailable>false</VpcUnavailable>
      <RegionEndpoint>ess.aliyuncs.com</RegionEndpoint>
      <LocalName>China (Beijing)</LocalName>
    </Region>
    <Region>
      <ClassicUnavailable>false</ClassicUnavailable>
      <RegionId>cn-shanghai</RegionId>
      <VpcUnavailable>false</VpcUnavailable>
      <RegionEndpoint>ess.aliyuncs.com</RegionEndpoint>
      <LocalName>China (Shanghai)</LocalName>
    </Region>
  </Regions>
</DescribeRegionsResponse>
```

JSON format

```
{
  "RequestId": "38EC7366-F5A9-46B1-BDB1-0FDC2E296397",
  "Regions": {
    "Region": [
      {
        "ClassicUnavailable": false,
        "RegionId": "cn-beijing",
        "VpcUnavailable": false,
        "RegionEndpoint": "ess.aliyuncs.com",
        "LocalName": "China (Beijing)"
      },
      {
        "ClassicUnavailable": false,
        "RegionId": "cn-shanghai",
        "VpcUnavailable": false,
        "RegionEndpoint": "ess.aliyuncs.com",
        "LocalName": "China (Shanghai)"
      }
    ]
  }
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

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.

19.Tag management

19.1. TagResources

You can call this operation to create and bind tags to specified Auto Scaling resources.

Description

- A maximum of 20 tags can be bound for a scaling group.
- Before you bind a tag to a resource, Alibaba Cloud checks the number of existing tags that are bound to the resource. An error message is returned if the maximum number is reached.

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
Action	String	Yes	TagResources	The operation that you want to perform. Set the value to TagResources.
RegionId	String	Yes	cn-hangzhou	The region ID of the resource. You can call the DescribeRegions operation to query the most recent region list.
ResourceId.N	RepeatList	Yes	asg-bp17xb4x1vr29lg****	The ID of resource N. Valid values of N: 1 to 50.
ResourceType	String	Yes	scalinggroup	The type of the resource. Only scaling groups are supported. Set the value to scalinggroup.
Tag.N.Key	String	No	TestKey	The key of tag N of the resource. Valid values of N: 1 to 20. The tag key cannot be an empty string. It can be up to 128 characters in length and cannot start with acs: or aliyun. It cannot contain http:// or https://.

Parameter	Type	Required	Example	Description
Tag.N.Value	String	No	TestValue	The value of tag N of the resource. Valid values of N: 1 to 20. The tag value can be an empty string. It can be up to 128 characters in length and cannot start with acs: or aliyun. It cannot contain http:// or https://.

Response parameters

Parameter	Type	Example	Description
RequestId	String	74C4E313-8570-479F-8791-DC25360D5627	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=TagResources
&RegionId=cn-hangzhou
&ResourceId.1=asg-bp17xb4x1vr291gt****
&ResourceId.2=asg-bp12qqpsd8bzra3t****
&ResourceType=scalinggroup
&<Common request parameters>
```

Sample success responses

XML format

```
<TagResourcesResponse>
  <RequestId>74C4E313-8570-479F-8791-DC25360D5627</RequestId>
</TagResourcesResponse>
```

JSON format

```
{
  "RequestId": "74C4E313-8570-479F-8791-DC25360D5627"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

19.2. ListTagResources

You can call this operation to query the tags that are bound to one or more Auto Scaling resources.

Description

- Set ResourceId.N or set Tag.N that consists of Tag.N.Key and Tag.N.Value in the request to specify query objects.
- If you set Tag.N and ResourceId.N at the same time, the Auto Scaling resources that match both the parameters are 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
Action	String	Yes	ListTagResources	The operation that you want to perform. Set the value to ListTagResources.
RegionId	String	Yes	cn-hangzhou	The region ID of the resource. You can call the DescribeRegions operation to query the most recent region list.
ResourceType	String	Yes	scalinggroup	The type of the resource. Only scaling groups are supported. Set the value to scalinggroup.
ResourceId.N	RepeatList	No	asg-bp17xb4x1vr29lg t****	The ID of resource N. Valid values of N: 1 to 50.

Parameter	Type	Required	Example	Description
Tag.N.Key	String	No	TestKey	<p>The key of tag N. It is used for exact match of Auto Scaling resources. The key must be 1 to 128 characters in length. Valid values of N: 1 to 20.</p> <p>Tag.N is used for exact match of Auto Scaling resources to which the specified tags are bound and consists of one key-value pair.</p> <ul style="list-style-type: none"> If you specify only Tag.N.Key, all Auto Scaling resources whose tags contain the specified tag key are returned. If you specify only Tag.N.Value, the MissingParameter.TagKey error code is returned. If you specify multiple tag key-value pairs at the same time, only Auto Scaling resources that match all tag key-value pairs are returned.
Tag.N.Value	String	No	TestValue	The value of tag N. It is used for exact match of Auto Scaling resources. The value must be 1 to 128 characters in length. Valid values of N: 1 to 20.
NextToken	String	No	caeba0bbb2be03f84eb48b699f0a4883	The token that is used to start the next query.

Response parameters

Parameter	Type	Example	Description
NextToken	String	caeba0bbb2be03f84eb48b699f0a4883	The token that is used to start the next query.
RequestId	String	DE65F6B7-7566-4802-9007-96F2494AC5XX	The ID of the request.
TagResources	Array of TagResource		A collection of resources and tags, including the resource ID, resource type, and tag key-value pair.
TagResource			

Parameter	Type	Example	Description
ResourceId	String	asg-bp17xb4x1vr29lgt**	The ID of the resource.
ResourceType	String	ALIYUN::ESS::SCALINGGROUP	The type of the resource.
TagKey	String	TestKey	The key of the tag.
TagValue	String	TestValue	The value of the tag.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=ListTagResources
&RegionId=cn-hangzhou
&ResourceType=scalinggroup
&ResourceId.1=asg-bp17xb4x1vr29lgt****
&<Common request parameters>
```

Sample success responses

XML format

```
<ListTagResourcesResponse>
  <RequestId>6E79F465-6101-4071-A852-19B76E238717</RequestId>
  <NextToken></NextToken>
  <TagResources>
    <TagResource>
      <ResourceId>asg-bp17xb4x1vr29lgt****</ResourceId>
      <TagKey>TestKey</TagKey>
      <ResourceType>ALIYUN::ESS::SCALINGGROUP</ResourceType>
      <TagValue>TestValue</TagValue>
    </TagResource>
  </TagResources>
</ListTagResourcesResponse>
```

JSON format

```
{  
    "RequestId": "6E79F465-6101-4071-A852-19B76E238717",  
    "NextToken": "",  
    "TagResources": {  
        "TagResource": [  
            {  
                "ResourceId": "asg-bp17xb4x1vr29lg****",  
                "TagKey": "TestKey",  
                "ResourceType": "ALIYUN::ESS::SCALINGGROUP",  
                "TagValue": "TestValue"  
            }  
        ]  
    }  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

19.3. UntagResources

You can call this operation to unbind tags from the specified Auto Scaling resources. After a tag is unbound, the tag is automatically deleted if it is not bound to any other resources.

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
Action	String	Yes	UntagResources	The operation that you want to perform. Set the value to UntagResources.
RegionId	String	Yes	cn-hangzhou	The region ID of the resource. You can call the DescribeRegions operation to query the most recent region list.
ResourceId.N	RepeatList	Yes	asg-bp17xb4x1vr29lg****	The ID of resource N. Valid values of N: 1 to 50.
ResourceType	String	Yes	scalinggroup	The type of the resource. Only scaling groups are supported. Set the value to scalinggroup.

Parameter	Type	Required	Example	Description
TagKey.N	RepeatList	No	TestKey	The key of tag N of the resource. Valid values of N: 1 to 20.
All	Boolean	No	false	<p>Specifies whether to unbind all tags from the resource. This parameter takes effect only when the TagKey.N parameter is not specified. Valid values:</p> <ul style="list-style-type: none"> • true: All tags are unbound from the resource. • false: No tags are unbound. <p>Default value: false</p>

Response parameters

Parameter	Type	Example	Description
RequestId	String	3AEBB1B9-5B13-4311-951F-C3C7FA2B73ED	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=UntagResources
&RegionId=cn-hangzhou
&ResourceType=scalinggroup
&ResourceId.1=asg-bp17xb4x1vr29lgt****
&<Common request parameters>
```

Sample success responses

XML format

```
<UntagResourcesResponse>
  <RequestId>3AEBB1B9-5B13-4311-951F-C3C7FA2B73ED</RequestId>
</UntagResourcesResponse>
```

JSON format

```
{
  "RequestId": "3AEBB1B9-5B13-4311-951F-C3C7FA2B73ED"
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

19.4. ListTagKeys

You can call this operation to query the tag keys of Auto Scaling resources.

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
Action	String	Yes	ListTagKeys	The operation that you want to perform. Set the value to ListTagKeys.
RegionId	String	Yes	cn-hangzhou	The region ID of the Auto Scaling resource.
ResourceType	String	Yes	scalinggroup	The type of the Auto Scaling resource. Set the value to scalinggroup, which indicates that the tag is bound to a scaling group.
NextToken	String	No	caeba0bbb2be03f84eb48b699f0a***	The token that is used to start the next query. If the NextToken parameter is empty, no subsequent requests are sent.
PageSize	Integer	No	10	The number of entries to return on each page. Valid values: 1 to 50. Default value: 10

Response parameters

Parameter	Type	Example	Description
Keys	List	ESS	The information of the tag keys.
NextToken	String	caeba0bbb2be03f84eb48b699f0a****	The token that is used to start the next query. If the NextToken parameter is empty, no subsequent requests are sent.

Parameter	Type	Example	Description
PageSize	Integer	10	The number of entries returned per page.
RequestId	String	DC09A6AA-2713-4E10-A2E9-E6C5C43A8842	The ID of the request.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=ListTagKeys
&RegionId=cn-hangzhou
&ResourceType=scalinggroup
&<Common request parameters>
```

Sample success responses

XML format

```
<ListTagKeysResponse>
  <RequestId>CA84E8A6-A2BB-41B6-896F-C62987A8BAFE</RequestId>
  <NextToken></NextToken>
  <PageSize>10</PageSize>
  <Keys>
    <Key>ESS</Key>
    <Key>Tag002</Key>
    <Key>Tag001</Key>
  </Keys>
</ListTagKeysResponse>
```

JSON format

```
{
  "RequestId": "CA84E8A6-A2BB-41B6-896F-C62987A8BAFE",
  "NextToken": "",
  "PageSize": 10,
  "Keys": {
    "Key": [
      "ESS",
      "Tag002",
      "Tag001"
    ]
  }
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidResourceType.No tFound	The ResourceType provided does not exist in our records.	The error message returned because the specified the ResourceType parameter is invalid.

19.5. ListTagValues

You can call this operation to query the tag values of Auto Scaling tag keys.

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
Action	String	Yes	ListTagValues	The operation that you want to perform. Set the value to ListTagValues.
Key	String	Yes	ESS	The key of the tag.
RegionId	String	Yes	cn-hangzhou	The region ID of the Auto Scaling resource.
ResourceType	String	Yes	scalinggroup	The type of the Auto Scaling resource. Set the value to scalinggroup, which indicates that the tag is bound to a scaling group.
NextToken	String	No	caeba0bbb2be03 f84eb48b699f0a* ***	The token that is used to start the next query. If the NextToken parameter is empty, no subsequent requests are sent.
PageSize	Integer	No	10	The number of entries to return on each page. Valid values: 1 to 50. Default value: 10

Response parameters

Parameter	Type	Example	Description
NextToken	String	caeba0bbb2be03f8 4eb48b699f0a****	The token that is used to start the next query. If the NextToken parameter is empty, no subsequent requests are sent.
PageSize	Integer	10	The number of entries returned per page.
RequestId	String	AB444F46-1CFF- 4B06-B8F0- B45D3158F95B	The ID of the request.
Values	List	Doc	The information of the tag values.

Examples

Sample requests

```
https://ess.aliyuncs.com/?Action=ListTagValues  
&RegionId=cn-hangzhou  
&ResourceType=scalinggroup  
&Key=ESS  
&<Common request parameters>
```

Sample success responses

XML format

```
<ListTagValuesResponse>  
  <RequestId>178E635A-11AF-4214-A65E-50B67B3CA548</RequestId>  
  <NextToken></NextToken>  
  <PageSize>1</PageSize>  
  <Values>  
    <Value>Doc</Value>  
  </Values>  
</ListTagValuesResponse>
```

JSON format

```
{  
    "RequestId": "178E635A-11AF-4214-A65E-50B67B3CA548",  
    "NextToken": "",  
    "PageSize": 1,  
    "Values": {  
        "Value": [  
            "Doc"  
        ]  
    }  
}
```

Error codes

For a list of error codes, visit the [API Error Center](#).

HTTP status code	Error code	Error message	Description
400	InvalidResourceType.NotFound	The ResourceType provided does not exist in our records.	The error message returned because the specified the ResourceType parameter is invalid.
400	InvalidTagKey.Malformed	The specified tag key \"%s\" is not valid.	The error message returned because the specified Tag.N.Key parameter does not exist.

20.Error codes

20.1. Client errors

This topic describes client errors.

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

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

20.2. Server errors

This topic describes server errors.

HTTP status code	Error code	Error message	Description
500	InternalError	The request processing has failed due to some unknown error, exception or failure.	The error message returned because the server fails to process this request.
503	ServiceUnavailable	The request has failed due to a temporary failure of the server.	The error message returned because the server cannot process this request.

21. 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**.

22. Call API operations over the internal network

This topic describes how to configure Alibaba Cloud DNS PrivateZone so that VPC-type ECS instances in a scaling group can initiate API requests over the Alibaba Cloud internal network.

Context

Auto Scaling provides public network endpoints. However, if your ECS instance does not have a public bandwidth package or a public IP address, the instance cannot initiate an API request by using tools such as Alibaba Cloud CLI or corresponding SDKs. Alibaba Cloud provides Alibaba Cloud DNS PrivateZone to ensure that your instance can send API requests over the Alibaba Cloud internal network. You can use PrivateZone to associate the VPC with the region to which your ECS instance belongs.

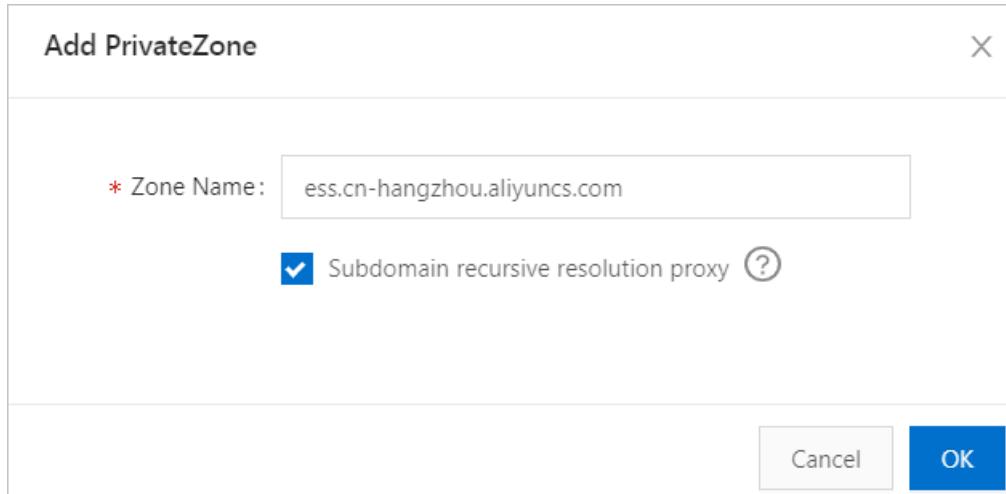
- You can only configure PrivateZone for regions that contain VPC-type ECS instances. You cannot configure PrivateZone across multiple regions.
- We recommend that you create ECS instances in a scaling group by using custom images that have Alibaba Cloud CLI or the SDK deployed. This way, your instances can add relevant dependencies when they have no access to the public network.
- The following table describes the endpoints that support PrivateZone. Make sure that you use the endpoint listed in the table.

Alibaba Cloud region	Region ID	CNAME record value	Public network endpoint
China (Beijing)	cn-beijing	popunify-vpc.cn-beijing.aliyuncs.com	ess.cn-beijing.aliyuncs.com
China (Hangzhou)	cn-hangzhou	popunify-vpc.cn-hangzhou.aliyuncs.com	ess.cn-hangzhou.aliyuncs.com
China (Shanghai)	cn-shanghai	popunify-vpc.cn-shanghai.aliyuncs.com	ess.cn-shanghai.aliyuncs.com
China (Shenzhen)	cn-shenzhen	popunify-vpc.cn-shenzhen.aliyuncs.com	ess.cn-shenzhen.aliyuncs.com
China (Hong Kong)	cn-hongkong	popunify-vpc.cn-hongkong.aliyuncs.com	ess.cn-hongkong.aliyuncs.com
Singapore	ap-southeast-1	popunify-vpc.ap-southeast-1.aliyuncs.com	ess.ap-southeast-1.aliyuncs.com

Procedure

1. Log on to the [Alibaba Could DNS console](#).

2. In the left-side navigation pane, click **PrivateZone**. On the page that appears, click **Add Zone**.
3. Configure the following parameters and click **OK**.
 - o **Zone Name**: Enter an ECS endpoint that supports PrivateZone. In this example, enter *ess.cn-hangzhou.aliyuncs.com*.
 - o **Subdomain recursive resolution proxy**: If you select this check box, the name resolved on the public network is used for DNS domain name checks for the specified Zone Name but is not included in the Zone file.



4. Click **Configure** in the **Actions** column corresponding to the created PrivateZone.
5. On the **Resolution Settings** page that appears, click **Add Record**.
6. In the **Add Record** dialog box that appears, configure the following parameters and click **OK**.
 - o **Record Type**: Select CNAME.
 - o **Resource Records**: Enter "@" to resolve the "@.example.com" domain name.
 - o **Record Value**: Set the value to the CNAME record value of the corresponding region.
 - o **TTL Value**: The time to live value. In this example, select 1 minute(s).

Add Record

X

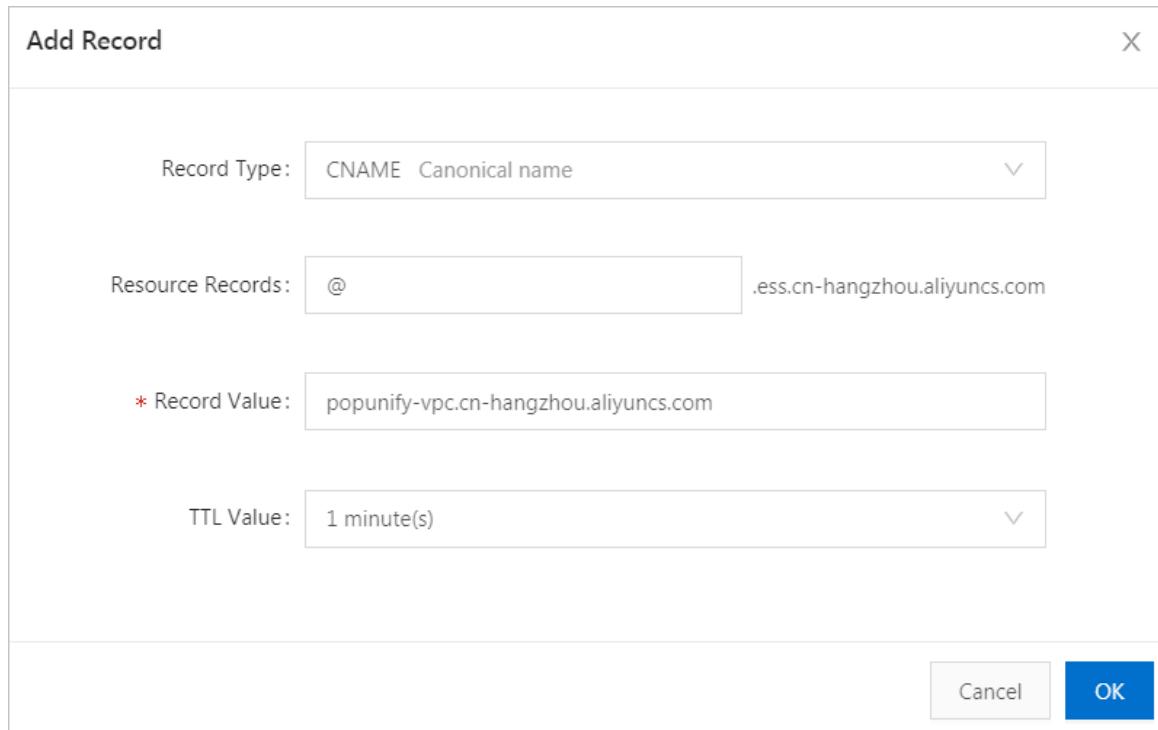
Record Type: CNAME Canonical name

Resource Records: @ .ess.cn-hangzhou.aliyuncs.com

* Record Value: popunify-vpc.cn-hangzhou.aliyuncs.com

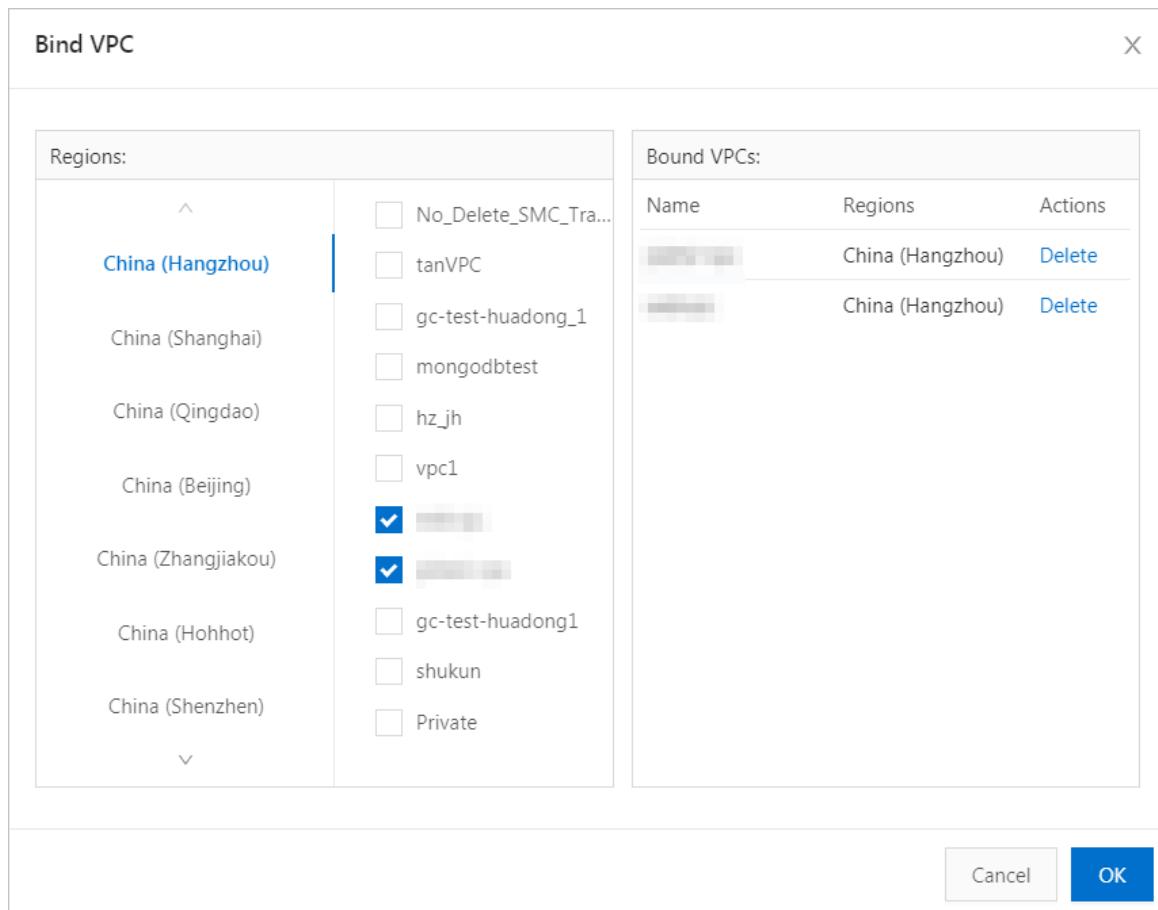
TTL Value: 1 minute(s)

Cancel OK



7. Go back to the PrivateZone page. Click **Bind VPC** in the **Actions** column corresponding to the created PrivateZone.
8. In the dialog box that appears, select the region where the PrivateZone is located. Select one or more VPCs to which your ECS instances belong. Click **OK**.

 Note Select the VPC to which the ECS instance belongs.



Result

After you associate a VPC with an Alibaba Cloud DNS PrivateZone, you can log on to your ECS instance to check whether the instance can access the endpoint of the corresponding region. For more information about how to log on to an ECS instance in a scaling group, see [Connect to a Linux instance by using a password](#).

For example, if the endpoint is `ess.cn-hangzhou.aliyuncs.com`, you can:

- Run a `ping` command to check whether data packets can be properly transmitted and received.

```
ping ess.cn-hangzhou.aliyuncs.com
```

```
[root@... ~]# ping ess.cn-hangzhou.aliyuncs.com
PING popunify-vpc.cn-hangzhou.aliyuncs.com.gds.alibabads.com (...).56(84) bytes of data.
64 bytes from ...: icmp_seq=1 ttl=102 time=1.30 ms
64 bytes from ...: icmp_seq=2 ttl=102 time=1.34 ms
64 bytes from ...: icmp_seq=3 ttl=102 time=1.31 ms
64 bytes from ...: icmp_seq=4 ttl=102 time=1.33 ms
64 bytes from ...: icmp_seq=5 ttl=102 time=2.29 ms
64 bytes from ...: icmp_seq=6 ttl=102 time=1.30 ms
64 bytes from ...: icmp_seq=7 ttl=102 time=2.99 ms
^C
--- popunify-vpc.cn-hangzhou.aliyuncs.com.gds.alibabads.com ping statistics ---
7 packets transmitted, 7 received, 0% packet loss, time 6009ms
rtt min/avg/max/mdev = 1.308/1.698/2.990/0.626 ms
```

- Use Alibaba Cloud CLI to call [DescribeRegions](#), and specify the value of the `--endpoint` field to the example endpoint.

```
aliyun ecs DescribeRegions --endpoint ess.cn-hangzhou.aliyuncs.com
```

```
[root@... aliyun ess DescribeRegions --endpoint ess.cn-hangzhou.aliyuncs.com]
{"RequestId":"1339AEF0-B414-4E06-B9B9-684B9E551BC2","Regions":[{"Region": {"ClassicUnavailable":false,"RegionId":"cn-hangzhou-test-1","RegionEndpoint":"ess.aliyuncs.com","LocalName":"阿里云杭州测试1","VpcUnavailable":false}, {"ClassicUnavailable":false,"RegionId":"cn-hangzhou-test-2","RegionEndpoint":"ess.aliyuncs.com","LocalName":"阿里云杭州测试2","VpcUnavailable":false}, {"ClassicUnavailable":false,"RegionId":"cn-qingdao","RegionEndpoint":"ess.aliyuncs.com","LocalName":"阿里云青岛","VpcUnavailable":false}, {"ClassicUnavailable":false,"RegionId":"cn-beijing","RegionEndpoint":"ess.aliyuncs.com","LocalName":"阿里云北京","VpcUnavailable":false}, {"ClassicUnavailable":false,"RegionId":"cn-zhangjiakou","RegionEndpoint":"ess.cn-zhangjiakou.aliyuncs.com","LocalName":"阿里云张家口","VpcUnavailable":false}, {"ClassicUnavailable":false,"RegionId":"cn-huhehaote","RegionEndpoint":"ess.cn-huhehaote.aliyuncs.com","LocalName":"阿里云呼伦贝尔","VpcUnavailable":false}, {"ClassicUnavailable":false,"RegionId":"cn-hangzhou","RegionEndpoint":"ess.aliyuncs.com"}]}
```