Auto Scaling

Developer Guide

Developer Guide

API usage instructions

API usage instructions

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

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

Calling method

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.

Request structure

Request structure

Service address

The access address for the Auto Scaling API service is **ess.aliyuncs.com**.

Communication protocols

The system supports request communication through the HTTP or HTTPS channel. We recommend that you send requests over HTTPS for enhanced security.

Request methods

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

Request parameters

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

Character encoding

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

Common parameters

This section describes the common request parameters and the common request parameters

Common request parameters

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

Name	Type	Required	Description
Action	String	Yes	The target API.
AccessKeyId	String	Yes	Equivalent to a logon password. However, an AccessKey is used to call APIs, while logon password is used to log on to the console. For more information, see Create an AccessKey.
Signature	String	Yes	Your signature. For more information, see Signature.
SignatureMethod	String	Yes	Signature method. Value: HMAC-SHA1.
Timestamp	String	Yes	Request time stamp. Use the offset from Coordinated Universal Time (UTC), the time display is based on ISO8601. Format: YYYY-MM-DDThh:mm:ssZ. For example, 2018-01-01T12:00:00Z indicates 20:00:00, Jan 01, 2018, Beijing time (UTC+8).
SignatureVersion	String	Yes	Signature algorithm version. Value: 1.0.
SignatureNonce	String	Yes	Unique random number, which is used to prevent network replay attacks. Different random numbers must be used for different requests.
Version	String	Yes	The API version to use. Value: 2014-08-28.

Format	String	No	Type of the response parameters. Optional values: json | xml. Default value: json.

Request example

```
https://ess.aliyuncs.com/?Action=XXXXXX
&Format=xml
&Version=2014-08-28
&Signature=Pc5WB8gokVn0xfeu%2FZV%2BiNM1dgI%3D
&SignatureMethod=HMAC-SHA1
&SignatureNonce=15215528852396
&SignatureVersion=1.0
&AccessKeyId=key-test
&Timestamp=2012-06-01T12:00:00Z
...
```

Common response parameters

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

Signature mechanism

Signature mechanism

Auto Scaling performs authentication on each access request. Therefore, each request, whether being sent via HTTP or HTTPS, must contain signature information. By using Access Key ID and Access Key Secret, Auto Scaling performs symmetric encryption to authenticate the request sender. The Access Key ID and Access Key Secret are officially issued to visitors by Alibaba Cloud (visitors can apply for and manage them at Alibaba Cloud’ s official website). The Access Key ID indicates the identity of the visitor. The Access Key Secret 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:

The Canonicalized Query String is constructed using the request parameters.

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 "&").

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:

i. English letters A–Z and a–z, digits 0–9, and characters "-", "_", "." , and "~" are not encoded.
ii. 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".
iii. Extended UTF-8 characters are encoded in the "%XY%ZA..." format.
iv. 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.

Connect the encoded parameter names and values with the equal sign "=".

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.

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, namely, "%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.

According to RFC2104 definitions, use the above signature sting to calculate the signature' s HMAC value.

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

According to Base64 encoding rules, encode the above HMAC value into a string. This gives you the signature value.

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&SignatureMe
thod=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%3Dc
n-qingdao&SignatureMethod%3DHMAC-SHA1&SignatureNonce%3D1324fd0e-e2bb-4bb1-917c-
bd6e437f1710&SignatureVersion%3D1.0&TimeStamp%3D2014-08-
15T11%253A10%253A07Z&Version%3D2014-08-28
```

Assume that the Access Key ID is "testid", the Access Key Secret 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 details about request signing and submission, refer to the appendix How to Call Interfaces.

Response results

We return the results in either XML or JSON format, but JSON is the default choice. You can switch to XML schema by specifying the request parameter Format. For more information, see Common parameters.

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

Success response example

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

XML format

```
<?xml version="1.0" encoding="UTF-8"?> <!--Result root node-->
<ActionResponse> <!--Response request tag-->
<RequestId>4C467B38-3910-447D-87BC-AC049166F223</RequestId> <!--Response result data-->
</ActionResponse>
```

JSON format

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

Error response example

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

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

XML format

```
<?xml version="1.0" encoding="UTF-8"?> <!--Result root node-->
<Error>
<RequestId>540CFF28-407A-40B5-B6A5-74Bxxxxxxxxx</RequestId> <!--Request ID-->
<HostId>ess.aliyuncs.com</HostId> <!--Endpoint-->
<Code>ServiceUnavailable </Code> <!--Error code-->
<Message>The request has failed due to a temporary failure of the server.</Message> <!--Error message-->
</Error>
```

JSON format

```
{
"RequestId": "540CFF28-407A-40B5-B6A5-74Bxxxxxxxxx", /* Request ID */
"HostId": "ess.aliyuncs.com", /* Endpoint */
"Code": "ServiceUnavailable", /* Error code */
"Message": "The request has failed due to a temporary failure of the server." /* Error message */
}
```

Common error codes

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

NoSuchVersion	The specified version does not exist.	400	The specified API version does not exist.
ResourceNotAvailable	Resource you requested is not available in this region or zone.	400	ESS service is unavailable in the specified region.
Throttling	Request was denied due to request throttling.	400	You have made too many frequent requests in short time. Please try again later.
UnsupportedOperation	The specified action is not supported.	400	Unable to call the specified API.
Forbidden	Users are not authorized to operate on the specified resource.	403	You cannot perform the specified action.
Forbidden.RiskControl	This operation is forbidden by Aliyun Risk Control system.	403	The specified action is under risk control.
Forbidden.Unsubscribed	Do not have permission to access this API.	403	You must enable the ESS service before calling this API.
SignatureDoesNotMatch	The signature we calculated does not match the one you provided.	403	The signature calculated by us is different from the one you provide.
InternalError	The request processing has failed due to some unknown error, exception or failure.	500	Internal error.
ServiceUnavailable	The request has failed due to a temporary failure of the server.	503	The server cannot respond to your request. Please try again later.

Scaling group

Create a scaling group

A scaling group is a collection of ECS instances with the same application scenarios.

It defines the maximum and minimum numbers of ECS instances in the group, and their associated Server Load Balancer instances, RDS instances, and other attributes.

Description

This operation creates a scaling group according to input parameters.

MaxSize and MinSize respectively define the maximum and minimum number of ECS instances in a scaling group.

- When the current number (total capacity) of ECS instances in the scaling group is smaller than MinSize, the Auto Scaling service automatically attaches ECS instances to the scaling group to make odds even.
- When the current number (total capacity) of ECS instances in the scaling group is greater than MaxSize, the Auto Scaling service automatically removes ECS instances from the scaling group to make odds even.

Default Cooldown indicates the default cool-down time of the scaling group.

- During the cool-down time after a scaling activity (adding or removing ECS instances) is run, the scaling group cannot perform any other scaling activity.
- Currently, this only applies to scaling activities triggered by CloudMonitor’ s alarm tasks.

Removal Policy is used to select the ECS instances you want to remove from the scaling group when multiple candidates for removal exist.

- If a Server Load Balancer instance is specified in the scaling group, the scaling group automatically attaches its ECS instances to the Server Load Balancer instance.
- The Server Load Balancer instance must be enabled.
- Health check must be enabled for all listener ports configured for the Server Load Balancer instance; otherwise, creation fails.
- The default weight of an ECS instance attached to the Server Load Balancer instance is 50.

If an RDS instance is specified in the scaling group, the scaling group automatically attaches the intranet IP addresses of its ECS instances to the RDS access whitelist.

- The specified RDS instance must be in running status.

- The specified RDS instance’ s whitelist must have room for more IP addresses.

The scaling group does not take effect immediately after being created. It must be enabled to support scaling rule trigger and perform scaling activities.

For more information about how to remove ECS instances from the scaling group, see Removal policies.

The scaling group, Server Load Balancer instance, and RDS instance must be in the same region.

You can create up to 20 scaling groups.

Request parameters

Name	Type	Required	Description
Action	String	Yes	Operation interface name, required parameter. Value: CreateScalingGroup
RegionId	String	Yes	ID of the region where a scaling group is located.
MaxSize	Integer	Yes	Maximum number of ECS instances in the scaling group. Value range: [0, 100].
MinSize	Integer	Yes	Minimum number of ECS instances in the scaling group. Value range: [0, 100].
ScalingGroupName	String	No	Name shown for the scaling group, which must contain 2-40 characters (English or Chinese). The name must begin with a number, an upper/lower-case letter or a Chinese character and may contain numbers, “_”, “-” or “.”. The account name is unique in the same region.

			If this parameter is not specified, the default value is ScalingGroupId.
DefaultCooldown	Integer	No	Default cool-down time (in seconds) of the scaling group. Value range: [0, 86400]. The default value is 300s.
RemovalPolicy.N	String	No	Policy for removing ECS instances from the scaling group. Optional values: - OldestInstance: removes the first ECS instance attached to the scaling group. - NewestInstance: removes the first ECS instance attached to the scaling group. - OldestScalingConfiguration: removes the ECS instance with the oldest scaling configuration. Default values: OldestScalingConfiguration and OldestInstance. You

			can enter up to two removal policies.
LoadBalancerIds	String	No	ID list of a Server Load Balancer instance. A Json Array with format: [“lb-id0” , “lb-id1” , ... “lb-idz”], support up to 5 Load Balancer instance.
DBInstanceIds	String	No	ID list of an RDS instance. A Json Array with format: [“rm-id0” , “rm-id1” , ... “rm-idz”], support up to 8 RDS instance.
VSwitchId	String	No	If you create a VPC scaling group, you must specify the ID of a VSwitch.
VSwitchIds.N	String	No	Parameter VSwitchIds.N is used to create instance in multiple zones. Parameter VSwitchIds.N has a priority over parameter VSwitchId. - The valid range of N is [1, 5], and you can specify at most 5 VSwitches in a VPC. - The priority of VSwitches descends from 1 to 5, and 1 indicates the highest priority. - When you

			fail to create an instance in the zone to which a specified VSwitch belongs, another VSwitch with less priority replaces the specified one automatically.

Return parameters

Name	Type	Description
ScalingGroupId	String	ID of a scaling group, generated by the system and globally unique.

Error codes

For errors common to all interfaces, see Client Error Table or Server Error Table.

Error message	Error code	Description	HTTP status code
The scaling group name already exists.	InvalidScalingGroupName.Duplicate	The specified value of parameter <parameter name> is duplicated.	400
The scaling group quota is exceeded.	QuotaExceeded.ScalingGroup	Scaling group quota exceeded.	400
The specified region does not exist.	InvalidRegionId.NotFound	The specified region does not exist.	404
The specified MinSize is greater than MaxSize.	InvalidParameter.Conflict	The value of parameter <parameter name> and parameter	400

		<parameter name> are conflict.	
The specified Server Load Balancer instance does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404
The specified Server Load Balancer instance and scaling group are not in the same region.	InvalidLoadBalancerId.RegionMismatch	The specified Load Balancer and the specified scaling group are not in the same Region.	400
The specified Server Load Balancer instance is not active.	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	400
Health check is not enabled for the specified Server Load Balancer instance.	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	400
The specified Server Load Balancer instance becomes private after the VswitchID is specified.	InvalidLoadBalancerId.IncorrectAddressType	The current address type of specified load balancer does not support this action.	400
The network type of the ECS instance contained in the specified Server Load Balancer is different from the network type of the scaling group.	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified Load Balancer does not support this action.	400
The ECS instance contained in the specified Server Load Balancer and VSwitchId are not in the same VPC.	InvalidLoadBalancerId.VPCMismatch	The specified VSwitch and the instance in specified Load Balancer are not in the same VPC.	400
The specified RDS instance does not exist in this account.	InvalidDBInstanceId.NotFound	DB instance “XXX” does not exist.	404
The specified RDS instance and scaling group are not in the same region.	InvalidDBInstanceId.RegionMismatch	DB instance “XXX” and the specified scaling group are not in the same Region.	400
The specified RDS instance is not running.	IncorrectDBInstanceStatus	The current status of DB instance “XXX” does not support	400

		this action.	
The number of IP addresses in the access whitelist of the specified RDS instance exceeds the upper limit.	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance “XXX”.	400
The specified VSwitch does not exist.	InvalidVSwitchId.NotFound	The specified VSwitch does not exist.	404
The private IP address quota of the VSwitch is exceeded.	QuotaExceeded.PrivateIpAddress	Private IP address quota exceeded in the specified VSwitch.	400
The instance quota of the VPC is exceeded.	QuotaExceeded.VPCInstance	Instance quota exceeded in the specified VPC.	400
The VSwitch is unavailable and instances cannot be created.	IncorrectVSwitchStatus	The current status of VSwitch does not support this operation.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=CreateScalingGroup
&RegionId=cn-qingdao
&MaxSize=20
&MinSize=2
&LoadBalancerId=147b46d767c-cn-qingdao-cm5-a01
&DBInstanceId.1=rdszzzyyunybaeu
&DBInstanceId.2=rdsia3u3yia3u3y
&<Public Request Parameters>
```

Return example

XML format:

```
<CreateScalingGroupResponse>
<ScalingGroupId>dP8VqSd9ENXPc0ciVmbcrBT1</ScalingGroupId>
<RequestId>536E9CAD-DB30-4647-AC87-AA5CC38C5382</RequestId>
</CreateScalingGroupResponse>
```

JSON format:

```
{
"RequestId": "536E9CAD-DB30-4647-AC87-AA5CC38C5382",
"ScalingGroupId": "dP8VqSd9ENXPc0ciVmbcrBT1"
}
```

Modify a scaling group

Modify a scaling group

Description

Modifies the attributes of a scaling group. However, the following attributes cannot be modified:

- RegionId
- LoadBalancerId
- DBInstanceId

The interface can be called only when the scaling group is active or inactive.

When the scaling configuration specified for the scaling group needs to be modified, the instance type attribute of the modified scaling configuration must be consistent with that of the active scaling configuration.

- After a new scaling configuration is added to the scaling group, the running ECS instances which are created based on the previous scaling configuration remain unchanged.

When the number (total capacity) of ECS instances in the scaling group does not meet the modified MaxSize or MinSize specification, the Auto Scaling service automatically attaches or removes ECS instances to/from the group to make odds even.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value:

			ModifyScalingGroup
ScalingGroupId	String	Yes	Scaling group ID
ScalingGroupName	String	No	Name shown for the scaling group, which must contain 2-40 characters (English or Chinese). The name must begin with a number, an upper/lower-case letter or a Chinese character and may contain numbers, "_", "-" or ".". The account name is unique in the same region.
ActiveScalingConfigurationId	String	No	ID of the active scaling configuration in the scaling group.
MinSize	Integer	No	Minimum number of ECS instances in the scaling group. Value range: [0, 100].
MaxSize	Integer	No	Maximum number of ECS instances in the scaling group. Value range: [0, 100].
DefaultCooldown	Integer	No	Default cool-down time (in seconds) of the scaling group. Value range: [0, 86400].
RemovalPolicy.N	String	No	Policy for removing ECS instances from the scaling group. Options: - OldestInstance: removes the first ECS instance attached to the scaling group. - NewestInstance: removes the first ECS instance attached to the scaling group. - OldestScalingConfiguration: removes the ECS instance with

			the oldest scaling configuration. You can enter up to two removal policies.

Return parameters

Public parameters.

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404
The scaling group name already exists.	InvalidScalingGroupName.Duplicate	The specified value of parameter <parameter name> is duplicated.	400
The specified scaling configuration does not exist in the scaling group.	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	404
The instance types of the specified scaling configuration and the active scaling configuration do not match.	InvalidScalingConfigurationId.InstanceTypeMismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	400
The specified MinSize is greater than MaxSize.	InvalidParameter.Conflict	The value of parameter <parameter name> and parameter <parameter name> are conflict.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=ModifyScalingGroup
```

```
&ScalingGroupId=cqS5QbbhmvGLcJbWoDbWLj2V
&ScalingGroupName=ScalingGroup
&<Public Request Parameters>
```

Return example

XML format:

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

JSON format:

```
"RequestId": "6469DCD0-13AC-487E-85A0-CE4922908FDE"
```

Enable a scaling group

Enable a scaling group

Description

Enables the specified scaling group.

- After the scaling group is successfully enabled (the group is active), the ECS instances specified by the interface are attached to the group.
- If the current number of ECS instances in the scaling group is still smaller than MinSize after the ECS instances specified by the interface are attached, the Auto Scaling service automatically creates ECS instances in Pay-As-You-Go mode to make odds even. For example, a scaling group is created with MinSize = 5. Two existing ECS instances are specified by the InstanceId.N parameter when the scaling group is enabled. Three additional ECS instances are automatically created after the two ECS instances are attached by the Auto Scaling service to the scaling group.

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

If the scaling group has no active scaling configurations, you need to input scaling configurations when enabling the scaling group.

- A single scaling group can have only one active scaling configuration at a time.
- If an active scaling configuration has been created before the scaling group is enabled, input of a new active scaling configuration through the interface makes the previous scaling configuration inactive.

Restrictions on attaching ECS instances:

- The attached ECS instance and the scaling group must be in the same region.
- The attached ECS instance and the instance with active scaling configurations must be of the same type.
- The attached ECS instance must in the running state.
- The attached ECS instance has not been attached to other scaling groups.
- The attached ECS instance supports Subscription and Pay-As-You-Go payment methods.
- If the VswitchID is specified for a scaling group, you cannot attach Classic ECS instances or ECS instances on other VPCs to the scaling group.
- If the VswitchID is not specified for the scaling group, ECS instances of the VPC type cannot be attached to the scaling group

The call fails if the number (total capacity) of instances specified by the interface plus instances in the scaling group is greater than MaxSize.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter; value: EnableScalingGroup
ScalingGroupId	String	Yes	Scaling group ID.
ActiveScalingConfigurationId	String	No	ID of the scaling configuration to be activated in a scaling group.
InstanceId.N	String	No	ID of the ECS instance to be attached to the scaling group after it is enabled. You can input up to 20 IDs.

Return parameters

Public parameters

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404
Open API is not fully authorized to the Auto Scaling service.	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	403
The specified scaling group is in the deleting state.	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	400
The specified scaling configuration does not exist in the scaling group.	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	404
The instance types of the specified scaling configuration and the active scaling configuration do not match.	InvalidScalingConfigurationId.InstanceTypeMismatch	The specified scaling configuration and the existing active scaling configuration have different instance types.	400
No active scaling configuration is specified for the scaling group.	MissingActiveScalingConfiguration	An active scaling configuration for the specified scaling group is not supplied.	400
The specified ECS instance does not exist in this account.	InvalidInstanceId.NotFound	Instance “XXX” does not exist.	404
The specified ECS instance and the scaling group are not in the same region.	InvalidInstanceId.RegionMismatch	Instance “XXX” and the specified scaling group are not in the same Region.	400
The instance types of the specified ECS instance and the scaling	InvalidInstanceId.InstanceTypeMismatch	Instance “XXX” and the existing active scaling configuration have	400

configuration do not match.		different instance types.	
The specified ECS instance is not in the running status.	IncorrectInstanceStatus	The current status of instance "XXX" does not support this action.	400
The network types of the specified ECS instance and the scaling configuration do not match.	InvalidInstanceId.NetworkTypeMismatch	The network type of instance "XXX" does not support this action.	400
The specified scaling group and the attached ECS instance are not in the same VPC.	InvalidInstanceId.VPCMismatch	Instance "XXX" and the specified scaling group are not in the same VPC.	400
The specified ECS instance has been attached to another scaling group.	InvalidInstanceId.InUse	Instance "XXX" is already attached to another scaling group.	400
The specified Server Load Balancer instance is not active.	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	400
Health check is not enabled for the specified Server Load Balancer instance.	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	400
The network type of the ECS instance contained in the specified Server Load Balancer is different from the network type of the scaling group.	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified load balancer does not support this action.	400
The ECS instance contained in the specified Server Load Balancer and VSwitchId are not in the same VPC.	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in specified load balancer are not in the same VPC.	400
The specified RDS instance is not running.	IncorrectDBInstanceStatus	The current status of DB instance "XXX" does not support this action.	400
Total Capacity after the ECS instance is attached is greater	IncorrectCapacity.MaxSize	To attach the instances, the total capacity will be	400

than MaxSize.		greater than the max size.	

Example

Request example

```
http://ess.aliyuncs.com/?Action=EnableScalingGroup
&ScalingGroupId=dmIDKNcyWfzncX9MWX1bwFV
&InstanceId.1=i-283vvyytn
&<Public Request Parameters>
```

Return example

XML format:

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

JSON format:

```
{
"RequestId": "6469DCD0-13AC-487E-85A0-CE4922908FDE"
}
```

Disable a scaling group

Disable a scaling group

Description

This operation disables a specified scaling group.

- The scaling activities in progress before the scaling group is disabled are continued until completion, whereas scaling activities triggered after the scaling group is disabled are rejected.
- The interface can be called only when the scaling group is active.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: DisableScalingGroup
ScalingGroupId	String	Yes	Scaling group ID

Return parameters

Public parameters

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP Status Code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404

Example

Request example

```
http://ess.aliyuncs.com/?Action=DisableScalingGroup
&ScalingGroupId=dmIDKNcyWfzncX9MWX1bwFV
&<Public Request Parameters>
```

Return example

XML example:

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

JSON format:

```
```

```
{
"RequestId": "6469DCD0-13AC-487E-85A0-CE4922908FDE"
}
```

Delete a scaling group

Delete a scaling group

Description

This operation deletes a specified scaling group.

ForceDelete indicates whether to forcibly delete a scaling group and remove and release ECS instances if the scaling group has ECS instances or scaling activities are in progress.

If ForceDelete is set to false, the scaling group can be deleted only when the following conditions are met:

- Condition 1: No scaling activities are in progress in the scaling group.
- Condition 2: The current number (total capacity) of ECS instances in the scaling group is 0.
- When the two conditions are met, the scaling group is disabled and then deleted.

When ForceDelete is set to true

- The scaling group is disabled to reject new scaling activity requests. When the existing scaling activity is completed, all ECS instances are removed from the scaling group and the group is then deleted (manually attached ECS instances are removed from the scaling group, whereas ECS instances automatically created by the Auto Scaling service are deleted).

Deleting a scaling group also deletes scaling configurations, rules, activities, and requests.

The following tasks or instances are not deleted: scheduled tasks, Cloud Monitor alarm tasks, Server Load Balancer instances, and RDS instances.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: DeleteScalingGroup
ScalingGroupId	String	Yes	Scaling group ID
ForceDelete	Bool	No	Indicates whether to forcibly delete a scaling group and remove and release ECS instances if the scaling group has ECS instances or scaling activities are in progress. The default value is false, indicating that the scaling group is not forcibly deleted.

Return parameters

Public parameters

Error code

Error	Error code	Description	HTTP status code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404
Open API is not fully authorized to the Auto Scaling service.	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	403
The specified scaling group still has ECS instances	InstanceInUse	You cannot delete a scaling configuration or scaling group while there is an instance associated with it.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=DeleteScalingGroup
&ScalingGroupId=dmIDKNcyWfzncX9MWX1bwFV
&<Public Request Parameters>
```

Return example

XML format:

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

JSON format:

```
{
"RequestId": "6469DCD0-13AC-487E-85A0-CE4922908FDE"
}
```

Query a scaling group

Query a scaling group

Description

This operation queries the information of a scaling group. Scaling groups have the following life cycle states:

- Active: In this state, the scaling group can receive scaling rule execution requests and trigger scaling activities.
- Inactive: In this state, the scaling group does not receive scaling rule execution requests.
- Deleting: The scaling group is being deleted and does not receive scaling rule execution requests.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface

			name, required parameter. Value: DescribeScalingGroups
RegionId	String	Yes	ID of the region where the scaling group is located.
ScalingGroupId.N	String	No	Scaling group ID. You can enter up to 20 IDs. Invalid scaling group IDs are not displayed in query results, and no error is reported.
ScalingGroupName.N	String	No	Scaling group name. You can enter up to 20 names. Invalid scaling group names are not displayed in query results, and no error is reported.
PageNumber	Integer	No	Page number of the scaling group list, staring from 1. Default value: 1.
PageSize	Integer	No	When querying by page, this parameter indicates the number of lines per page. Maximum value: 50; default value: 10.

Return parameters

Name	Type	Description
TotalCount	Integer	Total number of scaling groups
PageNumber	Integer	Current page number
PageSize	Integer	Number of lines per page
ScalingGroups	ScalingGroupSetType	Scaling group information set

ScalingGroupSetType is a set of ScalingGroupItemTypes:

Name	Type	Description
ScalingGroup	ScalingGroupItemType	Scaling group information

The attributes of ScalingGroupItemType are listed below:

Name	Type	Description
ScalingGroupId	String	Scaling group ID.
ScalingGroupName	String	Name shown for the scaling group.
ActiveScalingConfigurationId	String	ID of the active scaling configuration in the scaling group.
RegionId	String	ID of the region where the scaling group is located.
MinSize	Integer	Minimum number of ECS instances in the scaling group.
MaxSize	Integer	Maximum number of ECS instances in the scaling group.
DefaultCooldown	Integer	Default cool-down time of the scaling group.
RemovalPolicies	RemovalPolicySetType	A set of policies for removing ECS instances from the scaling group.
LoadBalancerId	String	ID of the Server Load Balancer instance.
DBInstanceIds	DBInstanceIdSetType	ID of the RDS instance.
VSwitchId	String	ID of the virtual switch corresponding to the scaling group.
LifecycleState	String	Status of the scaling group.
TotalCapacity	Integer	Total number of ECS instances in the scaling group.
ActiveCapacity	Integer	Number of ECS instances which have been attached to the scaling group and are running properly.
PendingCapacity	Integer	Number of ECS instances which are being attached to the scaling group with relevant configurations not completed.
RemovingCapacity	Integer	Number of ECS instances which are being removed from the scaling group.
CreationTime	String	Time when the scaling group

		is created.

RemovalPolicySetType is a set of String types:

Name	Type	Description
RemovalPolicy	String	Policy for removing ECS instances from the scaling group.

DBInstanceIdSetType is a set of String types:

Name	Type	Description
DBInstanceId	String	ID of the RDS instance

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Example

Request example

```
http://ess.aliyuncs.com/?Action=DescribeScalingGroups
&RegionId=cn-qingdao
&PageSize=50
&<Public Request Parameters>
```

Return example

XML format:

```
<DescribeScalingGroupsResponse>
<RequestId>6393C3A8-B611-42F2-AFA6-F080FC45D5D0</RequestId>
<TotalCount>1</TotalCount>
<PageNumber>1</PageNumber>
<PageSize>10</PageSize>
<ScalingGroups>
<ScalingGroup>
<ActiveCapacity>1</ActiveCapacity> <ActiveScalingConfigurationId>
dyo713cNYIB4ddEVlKbcpOef
</ActiveScalingConfigurationId>
<DBInstanceIds>
<DBInstanceId>rdszzzyyunybaeu</DBInstanceId>
</DBInstanceIds>
<VSwitchId>vpc-25j4god4l</VSwitchId>
```

```
<DefaultCooldown>20</DefaultCooldown>
<LifecycleState>Active</LifecycleState>
<LoadBalancerId>147b46d767c-cn-qingdao-cm5-a01</LoadBalancerId>
<MaxSize>1</MaxSize>
<MinSize>0</MinSize>
<PendingCapacity>0</PendingCapacity>
<RegionId>cn-qingdao</RegionId>
<RemovingCapacity>0</RemovingCapacity>
<ScalingGroupId>dyrSuvBOtO1dEdIlIbplQb8</ScalingGroupId>
<ScalingGroupName>dyrSuvBOtO1dEdIlIbplQb8</ScalingGroupName>
<RemovalPolicies>
<RemovalPolicy>OldestScalingConfiguration</RemovalPolicy>
<RemovalPolicy>OldestInstance</RemovalPolicy>
</RemovalPolicies>
<TotalCapacity>1</TotalCapacity>
<CreationTime>2014-08-14T10:58Z</CreationTime>
</ScalingGroup>
</ScalingGroups>
</DescribeScalingGroupsResponse>
```

JSON format:

```
{
"RequestId": "68386699-8B9E-4D5B-BC4C-75A28F6C2A00",
"TotalCount": 1,
"PageSize": 10,
"PageNumber": 1,
"ScalingGroups": {
"ScalingGroup": [
{
"ScalingGroupId": "b8pYCVbIV5k9cz4PWpbe0k19",
"ScalingGroupName": "b8pYCVbIV5k9cz4PWpbe0k19",
"RegionId": "cn-qingdao",
"RemovingCapacity": 0,
"DefaultCooldown": 300,
"MinSize": 1,
"MaxSize": 2,
"LifecycleState": "Inactive",
"ActiveScalingConfigurationId": " dyo713cNYIB4ddEVlKbcpOef",
"LoadBalancerId": "147b46d767c-cn-qingdao-cm5-a01",
"PendingCapacity": 0,
"TotalCapacity": 0,
"ActiveCapacity": 0,
"CreationTime": "2014-08-14T10:58Z",
"DBInstanceIds": {
"DBInstanceId": [
"rdsia3u3yia3u3y",
"rdszzzyyunybaeu"
]
},
"VSwitchId":"vpc-25j4god4l",
"RemovalPolicies": {
"RemovalPolicy": [
"OldestScalingConfiguration",
```

```
"OldestInstance"
]
}
}
]
}
}
```

Query the list of ECS instances in a scaling group

Query the list of ECS instances in a scaling group

Description

Queries the list of ECS instances in a scaling group. You can query by scaling group ID, scaling configuration ID, health status, lifecycle status, and creation type.

The ECS instances in a scaling group can be created automatically or attached manually.

- ECS instances can be automatically created by the Auto Scaling service based on your scaling configurations and rules.
- ECS instances can also be manually attached to a scaling group.

An ECS instance in a scaling group may be in the following states during its lifecycle:

- Pending: The ECS instance is being attached to the scaling group, with operations such as instance creation, attaching to Server Load Balancer, and adding to the RDS access whitelist.
- InService: The ECS instance is successfully added to the scaling group and provides services properly.
- Removing: The ECS instance is being removed from the scaling group.

An ECS instance in a scaling group may be in the following health states:

- Healthy: The ECS instance is healthy.
- Unhealthy: The ECS instance is unhealthy if it is not running.

The Auto Scaling service automatically removes unhealthy ECS instances from scaling groups.

- If the unhealthy ECS instances are automatically created, the Auto Scaling service disables and releases them.
- If the unhealthy ECS instances are manually attached, the Auto Scaling service does not disable or release them.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: DescribeScalingInstances.
RegionId	String	Yes	ID of the region where the scaling group is located.
ScalingGroupId	String	No	Scaling group ID.
ScalingConfigurationId	String	No	ID of the associated scaling configuration.
InstanceId.N	String	No	ECS instance ID. You can input up to 20 IDs. Invalid instance IDs are not displayed in query results, and no error is reported.
HealthStatus	String	No	Health status of an ECS instance in the scaling group. Options: Healthy and Unhealthy.
LifecycleState	String	No	Lifecycle status of an ECS instance in the scaling group. Options: - InService: the ECS instance has been added to the scaling group and runs properly. - Pending: the ECS instance is being attached to the scaling group with relevant configurations not

			completed. - Removing: the ECS instance is being removed from the scaling group.
CreationType	String	No	ECS instance creation type. Options: - AutoCreated: the ECS instance is automatically created by the Auto Scaling service in the scaling group. - Attached: the ECS instance is created outside the Auto Scaling service and manually attached to the scaling group.
PageNumber	Integer	No	Page number of the ECS instance list, staring from 1. Default value: 1
PageSize	Integer	No	When querying by page, this parameter indicates the number of lines per page. Maximum value: 50; default value: 10

Return parameters

Name	Type	Description
TotalCount	Integer	Total number of ECS instances
PageNumber	Integer	Current page number
PageSize	Integer	Number of lines per page
ScalingInstances	ScalingInstanceSetType	A set of ECS instance information

ScalingInstanceSetType is a set of ScalingInstanceItemTypes:

Name	Type	Description
ScalingInstance	ScalingInstanceItemType	ECS instance information

The attributes of ScalingInstanceItemType are listed below:

Name	Type	Description
InstanceId	String	ECS instance ID.
ScalingGroupId	String	ID of the scaling group to which the ECS instance belongs.
ScalingConfigurationId	String	ID of the associated scaling configuration.
HealthStatus	String	Health status of the ECS instance in the scaling group.
LifecycleState	String	Lifecycle status of the ECS instance in the scaling group.
CreationTime	String	Time when the ECS instance is attached to the scaling group.
CreationType	String	ECS instance creation type.

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Example

Request example

```
http://ess.aliyuncs.com/?Action=DescribeScalingInstances
&RegionId=cn-qingdao
&ScalingGroupId=dBCYxE26IHKcGq1xPcTNBwV
&<Public Request Parameters>
```

Return example

XML format:

```
<DescribeScalingInstancesResponse>
<RequestId>DFF8797F-5B73-4BD7-A7D0-03479C458F7A</RequestId>
<TotalCount>1</TotalCount>
<PageNumber>1</PageNumber>
<PageSize>10</PageSize>
<ScalingInstances>
<ScalingInstance>
<CreationTime>2014-08-15T17:37Z</CreationTime>
<CreationType>AutoCreated</CreationType>
```

```
<HealthStatus>Healthy</HealthStatus>
<InstanceId>i-283vvyytn</InstanceId>
<LifecycleState>InService</LifecycleState> <ScalingConfigurationId>
cGsGHrdMBa3DcDrrBVcc4k2H
</ScalingConfigurationId>
<ScalingGroupId>dBCYxE26IHKcGq1xPcTNBwV</ScalingGroupId>
</ScalingInstance>
</ScalingInstances>
</DescribeScalingInstancesResponse>
```

JSON format:

```
{
"RequestId": "13305F2D-A4C2-4E6B-B7C7-0F2150842EA3",
"TotalCount": 1,
"PageNumber": 1,
"PageSize": 50,
"ScalingInstances": {
"ScalingInstance": [
{
"ScalingConfigurationId": "bU5uZHcAgtzwcL4IeDeavqTS",
"CreationType": "AutoCreated",
"InstanceId": "i-28sov3exk",
"CreationTime": "2014-08-14T10:59Z",
"HealthStatus": "Healthy",
"LifecycleState": "InService",
"ScalingGroupId": "dE9YbOdCHqaFdFZHXVdDjQCB"
}
]
}
}
```

Query a scaling activity

Query a scaling activity

Description

Queries a scaling activity.

- You can specify a scaling group ID to query all scaling activities in this scaling group.
- You can filter the query results based on the scaling activity status.

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

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: DescribeScalingActivities
RegionId	String	Yes	Region of the scaling activity
ScalingGroupId	String	No	Scaling group ID
ScalingActivityId.N	String	No	Scaling group ID. You can enter up to 10 IDs. Invalid scaling group IDs are neglected in the query results, and no error is reported.
StatusCode	String	No	Scaling activity status. Optional values: - Successful: Successful scaling activities. - Warning: Partially successful scaling activities. - Failed: Failed scaling activities. - InProgress. Scaling activities in progress. - Rejected: The scaling activity request is rejected.
PageNumber	Integer	No	Page number of the scaling activity list starting from 1. The default value is 1.
PageSize	Integer	No	When querying by page, this parameter indicates the number of lines per page. Maximum value: 50; default value: 10

Return parameters

Name	Type	Description
TotalCount	Integer	Total number of the scaling activities
PageNumber	Integer	Current page number
PageSize	Integer	Number of lines per page
ScalingActivities	ScalingActivitySetType	Set of the scaling activity information

ScalingActivitySetType is a set consisting of the ScalingActivityItemType:

Name	Type	Description
ScalingActivity	ScalingActivityItemType	Scaling activity information

Attributes of the ScalingActivityItemType are introduced as follows:

Name	Type	Description
ScalingActivityId	String	Scaling activity ID
ScalingGroupId	String	Scaling group ID
Description	String	Description on the scaling activity
Cause	String	Cause that triggers the scaling activity
StartTime	String	Start time of the scaling activity
EndTime	String	End time of the scaling activity
Progress	Integer	Running speed of the scaling activity
StatusCode	String	Current status of the scaling activity
StatusMessage	String	Message about the scaling activity status

Error code

For errors common to all interfaces, see Client Error Table or Server Error Table.

Example

Request example

```
http://ess.aliyuncs.com/?Action=DescribeScalingActivities
&RegionId=cn-qingdao
&PageSize=50
&<Public Request Parameters>
```

Return example

XML format:

```
<DescribeScalingActivitiesResponse>
<RequestId>8FAAE99E-EB43-4838-85AD-93F62454904C</RequestId>
<TotalCount>1</TotalCount>
<PageNumber>1</PageNumber>
<PageSize>10</PageSize>
<ScalingActivities>
<ScalingActivity>
<Cause>A scheduled task excuete scaling rule "srtest", changing the Total Capacity from "0" to "1".</Cause>
<Description>Add "1" ECS instance</Description>
<EndTime>2014-08-17T12:39Z</EndTime>
<Progress>100</Progress>
<ScalingActivityId>ebta5WbUzC8gcwUWvfchyT4U</ScalingActivityId>
<ScalingGroupId>AG6CQdPU8OKdwLjgZcJ2eaQ</ScalingGroupId>
<StartTime>2014-08-17T12:39Z</StartTime>
<StatusCode>Successful</StatusCode>
<StatusMessage>"1" ECS instances is added. </StatusMessage>
</ScalingActivity>
</ScalingActivities>
</DescribeScalingActivitiesResponse>
```

JSON format:

```
{
"RequestId":"0A016AD6-A91F-4210-9DEE-D5BBD4270F45",
"TotalCount":2582,
"PageNumber":1,
"PageSize":1,
"ScalingActivities":{
"ScalingActivity":[
{
"ScalingActivityId":"bBxOR5dhKFoccGkbTrcyHE2g",
"StartTime":"2014-08-18T20:49Z",
"EndTime":"2014-08-18T20:49Z",
"Cause":" A scheduled task excuete scaling rule \"srtest\", changing the Total Capacity from \"0\" to \"1\".",
"Description":" Add \"1\" ECS instance ",
"Progress":100,
```

```
"ScalingGroupId":"c7acXJbAJmpPcGE7G3bwwbS9",
"StatusCode":"Successful",
"StatusMessage":"\"1\" ECS instances is added. "
}
]
},
}
```

Scaling configuration

Create a scaling configuration

Description

This operation creates a scaling configuration according to input parameters.

- Scaling configurations cannot be changed. To use a new scaling configuration, you must create it and activate it in a scaling group (by changing the ActiveScalingConfigurationId attribute).
- When creating a scaling configuration, you must select an image to determine the system disk configuration of the new instance. The image contains OS and the application software configuration. After an instance is created based on an image, the system disk of the instance is a clone of the image.
- When creating an instance, you must add it to a security group. Therefore, you need to specify a security group in a scaling configuration. The security group must be created in advance. Mutual access is allowed between instances in the same security group in the intranet. By default, firewalls are deployed between security groups, and mutual access between security groups is disabled. However, firewall permissions of a security group can be set through security group authorization (by using the interface for granting security group permissions). Instances in one security group cannot exceed 1,000. If the instances in a security group exceed this limit, and the security group is specified when creating an instance, a failure prompt is displayed.
- When you create an instance through a scaling configuration, if InternetChargeType is set to PayByBandwidth, the value of InternetMaxBandwidthOut is the selected fixed bandwidth value; if InternetChargeType is set to PayByTraffic, InternetMaxBandwidthOut specifies the upper limit of the bandwidth value, and billing is based on the actual network traffic. When setting InternetChargeType and InternetMaxBandwidthOut, calculate the possible bandwidth

expenses carefully.
- The value of InternetMaxBandwidthIn is unrelated to billing in all circumstances. Incoming data traffic of the instance is free of charge.
- During instance creation, the system allocates a system disk with an appropriate size for the instance based on the image specified by the user. Moreover, the system will also specify the type of the system disk as a cloud disk or an ephemeral disk if you have the permission to activate an ephemeral disk. (By default, a new user does not have this permission.)
- The capacity of a cloud disk cannot exceed 2,000 GB and the capacity of an ephemeral disk cannot exceed 1 TB (1,024 GB).
- The Portable attribute of a cloud disk created with an instance is False, that is, the cloud disk cannot be detached or attached. The DeleteWithInstance attribute of all created disks is True.
- A maximum of four data disks can be added to an instance. The total capacity of ephemeral disks for an instance cannot exceed 2 TB (2,048 GB). (This value does not include the capacity of the system disk.)
- When the system disk is a cloud disk, data disks cannot be ephemeral disks; when the system disk is an ephemeral disk, data disks can be ephemeral disks or cloud disks.
- During instance creation, the ECS system allocates a system disk with an appropriate size for the instance based on the image specified by the user.
- If the instance memory is 512 MB or less, Windows OS is not supported. If the instance memory is 4 GB or more, a 32-bit OS is not supported.
- When the VSwitchId value is not set in a scaling group, the security group with the VPCId value set cannot be used to create a scaling configuration.
- When the VSwitchId value is set in a scaling group, the security group of the classic network type or with a different VPCId value cannot be used to create a scaling configuration.
- A maximum of 10 scaling configurations can be created in a scaling group.

Request parameters

Name	Type	Required	Description
Action	String	Yes	Operation interface, required. The parameter value is CreateScalingConfiguration.
ScalingGroupId	String	Yes	ID of the scaling group of a scaling configuration.
ImageId	String	Yes	ID of an image file, indicating the image resource selected when an instance is enabled.
InstanceType	String	Yes	Resource type of an ECS instance.

SecurityGroupId	String	Yes	ID of the security group to which a newly created instance belongs. Mutual access is allowed between instances in one security group.
ScalingConfigurationName	String	No	Name shown for the scheduled task. The name must contain 2-40 English or Chinese characters, and start with a number, a letter in upper or lower case or a Chinese character. The name can contain numbers, “_”, “-” or “.”. The account name in the same scaling group is unique in the same region. If this parameter value is not specified, the default value is ScalingConfigurationId.
InternetChargeType	String	No	Network billing type, Values: PayByBandwidth or PayByTraffic. If this parameter value is not specified, the default value is PayByBandwidth.
InternetMaxBandwidthIn	Integer	No	Maximum incoming bandwidth from the public network, measured in Mbps (Mega bit per second). The value range is [1,200]. If this parameter value is not specified, AliyunAPI automatically sets the value to 200 Mbps.
InternetMaxBandwidthOut	Integer	No	Maximum outgoing bandwidth from the public network, measured in Mbps

			(Mega bit per second). The value range for PayByBandwidth is [1,100]. If this parameter value is not specified, AliyunAPI automatically sets the value to 0 Mbps. The value range for PayByTraffic is [1,100]. If this parameter value is not specified, an error is reported.
IoOptimized	String	No	Whether the instance is I/O optimized or not. Optional value: - optimized: The specified instance is I/O optimized. Instances of generation I instance types are not I/O optimized by default. For non I/O optimized instance, IoOptimized is an optional parameter.
SystemDisk.Category	String	No	Category of the system disk. The parameter value options are cloud and ephemeral. The default value is cloud.
DataDisk.N.Size	Integer	No	Size of data disk N (ranging from 1 to 4), in GB. The value ranges from 5 to 2,000 for a cloud disk and from 5 to 1,024 for an ephemeral disk. A maximum of four values can be entered.

DataDisk.N.Category	String	No	Category of data disk N (ranging from 1 to 4). The parameter value options are cloud and ephemeral. The default value is cloud. A maximum of four values can be entered.
DataDisk.N.SnapshotId	String	No	Snapshot used for creating the data disk N (ranging from 1 to 4). If this parameter is specified, the DataDisk.N.Size parameter is neglected, and the size of the created disk is the size of the snapshot. If this snapshot is created before July 15, 2013 (included), the snapshot cannot be called, and InvalidSnapshot.TooOld is returned in Response. A maximum of four values can be entered.
DataDisk.N.Device	String	No	Attaching point of the data disk N (ranging from 1 to 4). If this parameter is empty, the ECS automatically assigns the attaching point when an ECS is created. The parameter value ranges from **/dev/xvdb** to **/dev/xvdz**. The default value is empty. A maximum of four values can be entered.
DataDisk.n.DeleteWithInstance	String	No	Whether the data disk will be released along with the instance. Optional values:

			- true: Release the data disk along with the instance. - false: Retain the data disk when you release the instance. Default value: true DataDisk.n.DeleteWithInstance is valid only for independent cloud disks, for which the parameter DataDisk.n.Category is set to cloud, cloud_efficiency or cloud_ssd. Otherwise, an error returns.
LoadBalancerWeight	Integer	No	The weight of the backend server, the value range is [0, 100] and the default value is 50.
UserData	String	No	The user-defined data of the instance. The UserData of an instance must be encoded in Base64 format. The maximum size of the raw data is 16 KB.
KeyPairName	String	No	Key pair name. - If a Windows ECS instance is being created, ignore this parameter.

			By default, no value is set. If a value is set for the KeyPairName parameter, the password content is still bound to the instance - the user password authentication method will be disabled during the initialization of a Linux instance.
RamRoleName	String	No	The name of the instance RAM role. You can query the name of an instance RAM role by using the RAM API ListRoles . You can also see the API CreateRole and ListRoles for more information.
Tags	String	No	The tags of an instance. You should input the information of the tag with the format of the Key-Value, such as { “key1” :” value1” ,” key2” :” value2” , … “key5” :” value5” } .

			At most 5 tags can be specified. Key - It can be up to 64 characters in length. - Cannot begin with aliyun. - Cannot begin with http:// or https://. - Cannot be a null string. Value - It can be up to 128 characters in length. - Cannot begin with aliyun. - Cannot begin with http:// or https://. - Can be a null string.
InstanceTypes.N	String	No	- InstanceType will be ignored if InstanceTypes.N is applied. - N ranges from 1 to 10, which means one fleet can contain up

			to 10 instance types. - InstanceTypes.N N is the priority of each instance type in your fleet, 1 means the highest, the greater the number, the lower the instance type’ s priority is. - If you cannot launch an instance from the instance type with the relatively higher priority, the system will automatically use the instance type a level lower.
InstanceName	String	No	The name of the instance launched from the current fleet.

Return parameters

Name	Type	Description
ScalingConfigurationId	String	ID of a scaling configuration. It is generated by the system and is globally unique.

Error code

For errors common to all interfaces, see the Client Error Table or Server Error Table.

Error message	Error code	Description	HTTP status code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404
The scaling configuration name already exists.	InvalidScalingConfigurationName.Duplicate	The specified value of parameter <parameter name> is duplicated.	400
Scaling configuration quantity exceeds the upper limit for a user to use.	QuotaExceeded.ScalingConfiguration	Scaling configuration quota exceeded in the specified scaling group.	400
Specified image is not in this account.	InvalidImageId.NotFound	The specified image does not exist.	404
Specified security group is not in this account.	InvalidSecurityGroupId.NotFound	The specified security group does not exist.	404
Specified image does not support the specified instance type.	InvalidImageId.InstanceTypeMismatch	The specified image does not support the specified instance type.	400
Specified network type is inconsistent for the specified security group and the scaling group.	InvalidSecurityGroupId.IncorrectNetworkType	The network type of specified Security Group does not support this action.	400
The specified security group and the virtual switch are not in the same VPC.	InvalidSecurityGroupId.VPCMismatch	The specified security group and the specified virtual switch are not in the same VPC.	400
The number of ECS instances attached to the specified security group	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	400

exceeds the upper limit.			
The specified scaling configuration and the existing scaling configuration have different instance types.	InstanceType.Mismatch	The specified scaling configuration and existing active scaling configuration have different instance type.	400
Ephemeral disk capacity exceeds 2 TB (2,048 GB).	QuotaExceeded.EphemeralDiskSize	Ephemeral disk size quota exceeded.	403
Type of the specified system disk conflicts with that of the data disk.	InvalidParameter.Conflict	The value of parameter <parameter name> and parameter <parameter name> are conflict.	400
You are unauthorized to create an ephemeral system disk.	InvalidSystemDiskCategory.ValueUnauthorized	The system disk category is not authorized.	403
Specified snapshot does not exist.	InvalidDataDiskSnapshotId.NotFound	Snapshot “XXX” does not exist.	404
Capacity of the specified snapshot exceeds the upper limit of the disk size.	InvalidDataDiskSnapshotId.SizeNotSupported	The capacity of snapshot “XXX” exceeds the size limit of the specified disk category.	400
The snapshot is created before July 15, 2013 (included), and thus cannot be called.	InvalidSnapshot.TooOld	This operation is denied because the specified snapshot is created before 2013-07-15.	403
Data disk attaching point has been occupied.	InvalidDevice.InUse	Device “XXX” has been occupied.	403
The public network bandwidth and the billing type cannot be specified for a scaling group for which VSwitch is specified.	InvalidParameter.Conflict	The value of parameter InternetChargeType and parameter VSwitchId are conflict.	400
The value of parameter InternetChargeType and parameter VSwitchId are conflict.	InvalidParameter.Conflict	The value of parameter InternetChargeType and parameter VSwitchId are conflict.	400

The specified KeyPairName does not exist.	InvalidKeyPairName.NotFound	The specified KeyPairName does not exist in our records.	400
The parameter KeyPairName is only valid for a Windows instance.	InvalidParameter	The specified value of parameter KeyPairName is not valid.	400
The RamRoleName is only applicable to VPC instances.	InvalidNetworkType.ForRAMRole	RAMRole can’t be used For classic instance.	400
The specified RamRoleName does not exist.	InvalidRamRole.NotFound	The specified RamRoleName does not exist.	400
The specified UserData should be encoded in Base64 format.	InvalidUserData.Base64FormatInvalid	The specified parameter UserData must be base64 encoded	400
The size of the specified UserData exceeds 16 KB	InvalidUserData.SizeExceeded	The specified parameter UserData exceeds the size.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=CreateScalingConfiguration
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ2eaQ
&SecurityGroupId=sg-280ih3w4b
&ImageId=centos6u5_64_20G_aliaegis_20140703.vhd
&InstanceType=ecs.t1.xsmall
&<Public Request Parameters>
```

Return example

XML format:

```
<CreateScalingConfigurationResponse>
<ScalingConfigurationId>eOs27Kb0oXvQcUYjEGelJqUy</ScalingConfigurationId>
<RequestId>5CC0AD41-08ED-4559-A683-6F56355FE068</RequestId>
</CreateScalingConfigurationResponse>
```

JSON format:

```
{
"RequestId": "5CC0AD41-08ED-4559-A683-6F56355FE068",
"ScalingConfigurationId": "eOs27Kb0oXvQcUYjEGelJqUy",
}
```

Query a scaling configuration

Query a scaling configuration

Description

This operation queries scaling configuration information. You can query all scaling configurations in a scaling group by specifying the scaling group ID.

Scaling configuration states (LifecycleState) can be set to either of the following:

- Active: Scaling groups use scaling configurations in the active status to automatically create ECS instances.
- Inactive: The scaling configuration is inactive in a scaling group. The scaling group will not use scaling configurations in inactive state to automatically create ECS instances.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface, required. The parameter value is DescribeScalingConfigurations.
RegionId	String	Yes	Region ID of the scaling group of a scaling configuration.
ScalingGroupId	String	No	Scaling group ID.
ScalingConfigurationId.N	String	No	ID of a scaling configuration. A maximum of 10 values can be entered. Invalid

			scaling configuration IDs are neglected in the query result and no error is reported.
ScalingConfigurationName.N	String	No	Name of a scaling configuration. A maximum of 10 values can be entered. Invalid scaling configuration names are neglected in the query result and no error is reported.
PageNumber	Integer	No	Page number of the scaling configuration list. The initial value and default value are both 1.
PageSize	Integer	No	When querying by page, this parameter indicates the number of lines per page. Maximum value: 50; default value: 10.

Return parameters

Name	Type	Description
TotalCount	Integer	Total number of scaling configurations
PageNumber	Integer	Current page number
PageSize	Integer	Number of lines per page
ScalingConfigurations	ScalingConfigurationSetType	Scaling configuration information set

ScalingConfigurationSetType is a set of ScalingConfigurationItemTypes:

Name	Type	Description
ScalingConfiguration	ScalingConfigurationItemType	Scaling configuration information

Attributes of the ScalingConfigurationItemType are listed below:

Name	Type	Description

ScalingConfigurationId	String	ID of a scaling configuration
ScalingConfigurationName	String	Name shown for a scaling configuration
ScalingGroupId	String	ID of the scaling group of a scaling configuration
ImageId	String	ID of an image file
InstanceType	String	Resource rule of an instance
SecurityGroupId	String	ID of a security group
InternetChargeType	String	Network billing type
InternetMaxBandwidthIn	Integer	Maximum incoming bandwidth from the public network, measured in Mbps (Mega bit per second)
InternetMaxBandwidthOut	Integer	Maximum outgoing bandwidth from the public network, measured in Mbps (Mega bit per second)
SystemDisk.Category	String	Category of the system disk
DataDisks	DataDiskSetType	Data disk information set
LifecycleState	String	Scaling configuration status in a scaling group
CreationTime	String	Time when a scaling configuration is created

DataDiskSetType is a set of DataDiskItemTypes:

Name	Type	Description
DataDisk	DataDiskItemType	Data disk information

Attributes of the DataDiskItemType are listed below:

Name	Type	Description
Size	Integer	Disk capacity
Category	String	Disk category
SnapshotId	String	ID of the snapshot used for creating the data disk
Device	String	Data disk attaching point

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Example

Request example

```
http://ess.aliyuncs.com/?Action=CreateScalingConfiguration
&RegionId=cn-qingdao
&PageSize=50
&<Public Request Parameters>
```

Return example

XML format:

```
<DescribeScalingConfigurationsResponse>
<RequestId>804F240A-8D3E-40A1-BD68-6B333DEA2CA8</RequestId>
<TotalCount>1</TotalCount>
<PageNumber>1</PageNumber>
<PageSize>50</PageSize>
<ScalingConfigurations>
<ScalingConfiguration>
<CreationTime>2014-08-14T10:58Z</CreationTime>
<ImageId>centos6u5_64_20G_aliaegis_20140703.vhd</ImageId>
<InstanceType>ecs.t1.xsmall</InstanceType>
<InternetChargeType>PayByBandwidth</InternetChargeType>
<InternetMaxBandwidthIn>200</InternetMaxBandwidthIn>
<InternetMaxBandwidthOut>0</InternetMaxBandwidthOut>
<LifecycleState>Active</LifecycleState> <ScalingConfigurationId>
bU5uZHcAgtzwcL4IeDeavqTS
</ScalingConfigurationId>
<ScalingConfigurationName>
c1908dd1-690f-4c9b-ab73-350f1f06e84f
</ScalingConfigurationName>
<ScalingGroupId>dE9YbOdCHqaFdFZHXVdDjQCB</ScalingGroupId>
<SecurityGroupId>sg-280ih3w4b</SecurityGroupId>
<SystemDiskCategory>cloud</SystemDiskCategory>
<DataDisks>
<DataDisk>
<Size>200</Size>
<Category>cloud</Category>
<SnapshotId>s-280s7ngih</SnapshotId>
<Device>/dev/xvdb</Device>
</DataDisk>
</DataDisks>
</ScalingConfiguration>
</ScalingConfigurations>
</DescribeScalingConfigurationsResponse>
```

JSON format:

```
{
"RequestId": "67E4324F-CE14-4C2C-9D60-5422641DB76F",
"TotalCount": 1,
"PageNumber": 1,
"PageSize": 1,
"ScalingConfigurations": {
"ScalingConfiguration": [
{
"ScalingConfigurationId": "eqkz17cfW3clcPExOtLNVlD",
"SecurityGroupId": "sg-28oewzxvg",
"CreationTime": "2014-08-18T21:07Z",
"SystemDiskCategory": "cloud",
"InternetMaxBandwidthIn": 200,
"InternetMaxBandwidthOut": 0,
"LifecycleState": "Inactive",
"InternetChargeType": "PayByBandwidth",
"ImageId": "rhel5u7_64_20G_aliaegis_20131231.vhd",
"InstanceType": "ecs.s2.small",
"ScalingConfigurationName": "LxVdcOqPBV",
"ScalingGroupId": "dRsEAGdvbjR5c4SVc2bqLubj",
 “DataDisks” :{
 “DataDisk” :[{
 “Size” :200,
 “Category” : cloud,
 “SnapshotId” : s-280s7ngih,
 “Device” :” /dev/xvdb”
}]
}
}
]
}
}
```


Delete a scaling configuration

Delete a scaling configuration

Description

Deletes a specified scaling configuration.

- An active scaling configuration in a scaling group cannot be deleted.
- If any ECS instance created according to a scaling configuration is still in the scaling group, the scaling configuration cannot be deleted.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface, required. The parameter value is DeleteScalingConfiguration.
ScalingConfigurationId	String	Yes	ID of a scaling configuration.

Return parameters

Public parameters

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling configuration does not exist in this account.	InvalidScalingConfigurationId.NotFound	The specified scaling configuration does not exist.	404
The specified scaling configuration is not in inactive state.	IncorrectScalingConfigurationLifecycleState	The current lifecycle state of specified scaling configuration does not support this action.	400
The scaling configuration has an associated ECS instance not deleted yet.	InstanceInUse	You cannot delete a scaling configuration or scaling group while there is an instance associated with it.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=DeleteScalingConfiguration
&ScalingConfigurationId=eOs27Kb0oXvQcUYjEGelJqUy
&<Public Request Parameters>
```

Return example

XML format:

```
<DeleteScalingConfigurationResponse>
<RequestId>61D30272-7111-44D9-BB45-FCB55E4A1410</RequestId>
</DeleteScalingConfigurationResponse>
```

JSON format:

```
"RequestId": "61D30272-7111-44D9-BB45-FCB55E4A1410"
```

Scaling rule

Create a scaling rule

Create a scaling rule

A scaling rule defines specific scaling actions, for example, adding or removing n ECS instances.

If the execution of a scaling rule results in a number of ECS instances in the scaling group that is less than the MinSize or greater than the MaxSize, Auto Scaling automatically adjusts the number of ECS instances to be added or removed by executing the “Adjust scaling group instance quantity to MinSize” or “Adjust scaling group instance quantity to MaxSize” rule.

- Example 1: If a scaling group has a MaxSize of 3, the current number of instances (Total Capacity) is 2, and a scaling rule instructs the system to “add 3 ECS instances”, this operation actually only adds one ECS instance. (The values in the scaling rule are not changed.)
- Example 2: If a scaling group has a MinSize of 2, the current number of instances (Total Capacity) is 3, and a scaling rule instructs the system to “remove 5 ECS instances”, this operation actually only removes one ECS instance. (The values in the scaling rule are not changed.)

Description

This operation creates a scaling rule according to input parameters.

When AdjustmentType is TotalCapaciy, the quantity of ECS instances in the current scaling group is adjusted to the specified value, and the corresponding AdjustmentValue must be no less than 0.

When AdjustmentType is QuantityChangeInCapacity or PercentChangeInCapacity, if the corresponding AdjustmentValue is a positive number, additional ECS instances are increased; if the corresponding AdjustmentValue is a negative number, ECS instances are decreased.

When AdjustmentType is PercentChangeInCapacity, the Auto Scaling service uses the formula: **current number of instances (Total Capacity) * AdjustmentValue/100** and follow standard rounding principles to confirm the number of ECS instances to be added or removed.

When a cool-down time (Cooldown) is specified in a scaling rule, the specified Cooldown is applied to the scaling group after the scaling activity to perform this rule is completed. Otherwise, the DefaultCooldown is applied to the scaling group.

A maximum of 10 scaling rules can be created in a scaling group.

The only identifier of a scaling rule returned (ScalingRuleAri) can be primarily used by the following interfaces:

- By specifying the ExecuteScalingRule’ s ScalingRuleAri parameter, you can manually perform a scaling rule.
- By specifying the CreateScheduledTask’ s ScheduledAction parameter, you can schedule the performing of a scaling rule.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface, required. The parameter value is CreateScalingRule.
ScalingGroupId	String	Yes	ID of the scaling group of a scaling rule.
AdjustmentType	String	Yes	Adjustment mode of

			a scaling rule. Optional values: - QuantityChangeInCapacity: It is used to increase or decrease a specified number of ECS instances. - PercentChangeInCapacity: It is used to increase or decrease a specified proportion of ECS instances. - TotalCapacity: It is used to adjust the quantity of ECS instances in the current scaling group to a specified value.
AdjustmentValue	Integer	Yes	Adjusted value of a scaling rule. Value range: - QuantityChangeInCapacity: (0, 100] U (-100, 0] - PercentChangeInCapacity: [0, 10000] U [-10000, 0] - TotalCapacity: [0, 100]
ScalingRuleName	String	No	Name shown for the scaling group, which is a string containing 2 to 40 English or Chinese characters. It must begin with a number, a letter (case-insensitive) or a Chinese character and can contain numbers, “_”, “-” or “.”. The account name in the same scaling group is unique in the same region. If this parameter value is not specified, the default value is ScalingRuleId.
Cooldown	Integer	No	Cool-down time of a

			scaling rule. Value range: [0, 86,400], in seconds. The default value is empty.

Return parameters

Name	Type	Description
ScalingRuleId	String	ID of a scaling rule, generated by the system and globally unique.
ScalingRuleAri	String	Unique identifier of a scaling rule.

Error code

For errors common to all interfaces, see Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404
The scaling rule name already exists.	InvalidScalingRuleName.Duplicate	The specified value of parameter <parameter name> is duplicated.	400
Scaling rule quantity exceeds the upper limit for a user to use.	QuotaExceeded.ScalingRule	Scaling rule quota exceeded in the specified scaling group.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=CreateScalingRule
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ2eaQ
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=-10
&<Public Request Parameters>
```

Return example

XML format:

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

JSON format:

```
{
"RequestId": "04F0F334-1335-436C-A1D7-6C044FE73368",
"ScalingRuleId": "eMKWG8SRNb9dBLAjweNI1Ik",
"ScalingRuleAri":"ari:acs:ess:cn-qingdao:1344371:scalingrule/eMKWG8SRNb9dBLAjweNI1Ik"
}
```

ModifyScalingRule

Description

Modifies the attributes of a scaling rule.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface, required. The parameter value is ModifyScalingRule.
ScalingRuleId	String	Yes	ID of a scaling rule.
AdjustmentType	String	No	Adjustment mode of a scaling rule. Optional values: - QuantityChangeInCapacity: It is used to increase or decrease a specified number of ECS instances. -

			PercentChangeInCapacity: It is used to increase or decrease a specified proportion of ECS instances. - TotalCapacity: It is used to adjust the quantity of ECS instances in the current scaling group to a specified value.
AdjustmentValue	Integer	No	Adjusted value of a scaling rule. Value range: - QuantityChangeInCapacity: (0, 100] U [-100, 0) - PercentChangeInCapacity: [0, 10,000] U [-10,000, 0] - TotalCapacity: [0, 100]
ScalingRuleName	String	No	Name shown for the scaling group, which is a string containing 2 to 40 English or Chinese characters. It must begin with a number, a letter (case-insensitive) or a Chinese character and can contain numbers, "_", "-" or ".". The account name in the same scaling group is unique in the same region. If this parameter value is not specified, the default value is ScalingRuleId.
Cooldown	Integer	No	Cool-down time of a scaling rule. Value range: [0, 86,400], in seconds. The default value is empty.

Response parameters

Name	Type	Description

ScalingRuleId	String	ID of a scaling rule, generated by the system and globally unique.
ScalingRuleAri	String	Unique identifier of a scaling rule.

Example

Request example

```
http://ess.aliyuncs.com/?Action=CreateScalingRule
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ2eaQ
&AdjustmentType=QuantityChangeInCapacity
&AdjustmentValue=-10
&<Common Request Parameters>
```

Response example

XML format

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

JSON format

```
{
"RequestId": "04F0F334-1335-436C-A1D7-6C044FE73368",
"ScalingRuleId": "eMKWG8SRNb9dBLAjweNI1Ik",
"ScalingRuleAri":"ari:acs:ess:cn-qingdao:1344371:scalingrule/eMKWG8SRNb9dBLAjweNI1Ik"
}
```

Error codes

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error code	Error message	HTTP status code	Description
InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404	The specified scaling group does not exist in this account.
InvalidScalingRuleName.Duplicate	The specified value of parameter <parameter name>	400	The scaling rule name already exists

	is duplicated.		
Scaling rule quota exceeded in the specified scaling group.	400	Scaling rule quantity exceeds the upper limit for a user to use	QuotaExceeded.ScalingRule

Query a scaling rule

Query a scaling rule

Description

Queries information of a scaling rule. You can query all scaling rules in a scaling group by specifying the scaling group ID.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface, required. The parameter value is DescribeScalingRules.
RegionId	String	Yes	Region ID of the scaling group of a scaling rule.
ScalingGroupId	String	No	Scaling group ID.
ScalingRuleId.N	String	No	ID of a scaling rule. A maximum of 10 values can be entered. Invalid scaling rule IDs are neglected in the query result and no error is reported.
ScalingRuleName.N	String	No	Name of a scaling rule. A maximum of 10 values can be entered. Invalid scaling rule names

			are neglected in the query result and no error is reported.
ScalingRuleAri.N	String	No	Unique identifier of a scaling rule. A maximum of 10 values can be entered. Invalid unique identifiers of scaling rules are neglected in the query result and no error is reported.
PageNumber	Integer	No	Page number of the scaling rule list. The initial value and default value are both 1.
PageSize	Integer	No	When querying by page, this parameter indicates the number of lines per page. Maximum value: 50; default value: 10.

Return parameters

Name	Type	Description
TotalCount	Integer	Total number of scaling rules
PageNumber	Integer	Current page number
PageSize	Integer	Number of lines per page
ScalingRules	ScalingRuleSetType	Scaling rule information set

ScalingRuleSetType is a set of ScalingRuleItemTypes:

Name	Type	Description
ScalingRule	ScalingRuleItemType	Scaling rule information

The attributes of ScalingRuleItemType are listed below:

Name	Type	Description
ScalingRuleId	String	ID of a scaling rule
ScalingGroupId	String	Scaling group ID
ScalingRuleName	String	Name of a scaling rule

Cooldown	Integer	Cool-down time
AdjustmentType	String	Adjustment mode
AdjustmentValue	Integer	Adjustment value
ScalingRuleAri	String	Unique identifier of a scaling rule

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Example

Request example

```
http://ess.aliyuncs.com/?Action=DescribeScalingRules
&RegionId=cn-qingdao
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ2eaQ
&PageSize=50
&<Public Request Parameters>
```

Return example

XML format:

```
<DescribeScalingRulesResponse>
<PageNumber>1</PageNumber>
<PageSize>50</PageSize>
<ScalingRules>
<ScalingRule>
<AdjustmentType>QuantityChangeInCapacity</AdjustmentType>
<AdjustmentValue>1</AdjustmentValue>
<Cooldown>20</Cooldown>
<ScalingGroupId>AG6CQdPU8OKdwLjgZcJ2eaQ</ScalingGroupId> <ScalingRuleAri>
ari:acs:ess:cn-qingdao:1344371:scalingRule/eMKWG8SRNb9dBLAjweNI1Ik
</ScalingRuleAri>
<ScalingRuleId>eMKWG8SRNb9dBLAjweNI1Ik</ScalingRuleId>
<ScalingRuleName>eMKWG8SRNb9dBLAjweNI1Ik</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": "efcqrZdjlookc0UkE3dA5I0a",
"ScalingRuleAri": "ari:acs:ess:cn-qingdao:1344371:scalingRule/efcqrZdjlookc0UkE3dA5I0a",
"Cooldown": 500,
"ScalingGroupId": "ccMvs9dcZlE5c9CtrwbXzizr",
"AdjustmentType": "TotalCapacity",
"ScalingRuleName": "KFJoxGKXXt",
"AdjustmentValue": 5
}
]
}
}
```


Delete a scaling rule

Delete a scaling rule

Description

Deletes a specified scaling rule.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface, required. The parameter value is DeleteScalingRule.
ScalingRuleId	String	Yes	ID of a scaling rule.

Return parameters

Public parameters.

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling rule does not exist in this account	InvalidScalingRuleId.NotFound	The specified scaling rule does not exist.	404

Example

Request example

```
http://ess.aliyuncs.com/?Action=DeleteScalingRule
&ScalingRuleId=eMKWG8SRNb9dBLAjweNI1Ik
&<Public Request Parameters>
```

Return example

XML format:

```
<DeleteScalingRuleResponse>
<RequestId>F0595173-CA3A-4597-B0D3-97A07A042B9C</RequestId>
</DeleteScalingRuleResponse>
```

JSON format:

```
"RequestId": "F0595173-CA3A-4597-B0D3-97A07A042B9C"
```

Trigger task

Execute a scaling rule

Execute a scaling rule

Description

Executes a specified scaling rule.

- The interface can be called only when the scaling group is active.
- The interface can be called only when no scaling activity in the scaling group is in progress.
- When no scaling activity in the scaling group is in progress, the interface can be directly executed without cooldown.
- Successfully calling this interface only means that the Auto Scaling service has accepted the call request, and the scaling activity can be executed, but does not necessarily mean that the scaling activity can be successfully executed. You can use the returned ScalingActivityId to check the status of the scaling activity.
- When the total capacity of to-be-attached ECS instances specified by this scaling rule plus the instances in the current scaling group is greater than MaxSize, the rule **Total Capacity = MaxSize** is applied.
- When the total capacity of instances in the current scaling group minus the number of ECS instances to be reduced according to the scaling rule is smaller than MinSize, the rule **Total Capacity = MinSize** is applied.
- You can only perform auto scaling for at most 100 ECS instances of all scaling groups in all regions. (The number only includes the number of ECS instances automatically created, but does not include the number of manually attached ECS instances.)

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: ExecuteScalingRule.
ScalingRuleAri	String	Yes	Unique identifier of the scaling rule.
ClientToken	String	No	Used to ensure request idempotence. The value is generated by a client. It must be unique among all requests and can contain a maximum of 64 ASCII characters. For details, refer to the appendix How to

			ensure idempotence.

Return parameters

Name	Type	Description
ScalingActivityId	String	Scaling activity ID.

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling rule does not exist in this account.	InvalidScalingRuleAri.NotFound	The specified scaling rule Ari does not exist.	404
Open API is not fully authorized to the Auto Scaling service.	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	403
The specified scaling rule is not active.	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	400
The scaling group to which the scaling rule belongs has a scaling activity in progress.	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.	400
Your account balance is not enough.	InsufficientBalance	Your account does not have enough balance.	400
Your ECS instance quota is exceeded.	QuotaExceed.Instance	Living instance quota exceeded.	400
The Server Load Balancer instance in the scaling group to which the scaling rule belongs is not active.	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	400
Health check is not enabled for the Server Load Balancer	IncorrectLoadBalancerHealthCheck	The current health check type of specified load	400

in the scaling group to which the specified scaling rule belongs.		balancer does not support this action.	
The network type of the ECS instance contained in the specified Server Load Balancer is different from the network type of the scaling group.	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified load balancer does not support this action.	400
The ECS instance contained in the specified Server Load Balancer and VSwitchId are not in the same VPC.	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in specified load balancer are not in the same VPC.	400
The RDS instance in the scaling group to which the specified scaling rule belongs is not running.	IncorrectDBInstanceStatus	The current status of DB instance “XXX” does not support this action.	400
The number of IP addresses in the white list that can access the RDS instance in the scaling group to which the specified scaling rule belongs has reached the upper limit.	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance “XXX”.	400
The number of ECS instances attached to the specified security group exceeds the upper limit.	QuotaExceeded.SecurityGroupInstance	Instance quota exceeded in the specified security group.	400
The number of instances in the scaling group does not change after the scaling rule is executed.	IncorrectCapacity.NoChange	To execute the specified scaling rule, the total capacity will not change.	400
The Auto Scaling ECS instance quota is exceeded.	QuotaExceeded.ScalingInstance	Scaling instance quota exceeded.	400
The Pay-As-You-Go ECS instance quota is exceeded.	QuotaExceeded.AfterpayInstance	Living afterpay instance quota exceeded.	400

The specified ECS instance type or disk type cannot be created in the specified region.	ResourceNotAvailable.ECS	The specified region or zone does not offer the specified disk or instance category.	

Example

Request example

```
http://ess.aliyuncs.com/?Action=ExecuteScalingRule
&ScalingRuleAri=ari:acs:ess:cn-qingdao:1344371:scalingRule/cCBpdYdQuBe2cUxOdu6piOk
&<Public Request Parameters>
```

Return example

XML format:

```
<ExecuteScalingRuleResponse>
<ScalingActivityId>ebta5WbUzC8gcwUWvfchyT4U</ScalingActivityId>
<RequestId>262216B9-F9D4-4D16-BE9B-BD1C39A4F42B</RequestId>
</ExecuteScalingRuleResponse>
```

JSON format:

```
"RequestId": "6469DCD0-13AC-487E-85A0-CE4922908FDE",
"ScalingActivityId": "ebta5WbUzC8gcwUWvfchyT4U"
```

Attach an ECS instance

Attach an ECS instance

Description

Attaches an ECS instance to a specified scaling group. Restrictions on the attached ECS instance:

- The attached ECS instance and the scaling group must be in the same region.
- The attached ECS instance must in the running state.
- The attached ECS instance has not been attached to other scaling groups.
- The attached ECS instance supports Subscription and Pay-As-You-Go payment methods.
- If the VswitchID is specified for a scaling group, you cannot attach Classic ECS instances or ECS instances on other VPCs to the scaling group.
- If the VswitchID is not specified for the scaling group, ECS instances of the VPC type cannot be attached to the scaling group.

The interface can be called only when the scaling group is active.

The interface can be called only when the scaling group has no scaling activity in progress.

When the scaling group has no scaling activity in progress, the interface can be directly executed without cooldown.

Successfully calling this interface only means that the Auto Scaling service has accepted the call request, and the scaling activity can be executed, but does not necessarily mean that the scaling activity can be successfully executed. You can use the returned ScalingActivityId to check the status of the scaling activity.

The call attempt may fail when the total capacity of instances specified by this interface plus the instances of the scaling group is greater than MaxSize.

The manually attached ECS instance is not associated with the active scaling configurations of the scaling group.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: AttachInstances.
ScalingGroupId	String	Yes	Scaling group ID.
InstanceId.N	String	Yes	ECS instance ID. You can input up to 20 IDs.

Return parameters

Name	Type	Description
ScalingActivityId	String	Scaling activity ID

Error code

For errors common to all interfaces, see the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404
OpenAPI is not fully authorized to the Auto Scaling service.	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	403
The specified scaling group is not active.	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	400
The specified ECS instance does not exist in this account.	InvalidInstanceId.NotFound	Instance “XXX” does not exist.	404
The specified ECS instance and the scaling group are not in the same region.	InvalidInstanceId.RegionMismatch	Instance “XXX” and the specified scaling group are not in the same Region.	400
The specified ECS instance and the instance with valid scaling configurations do not match.	InvalidInstanceId.InstanceTypeMismatch	Instance “XXX” and existing active scaling configurations have different instance types.	400
The specified ECS instance is not in the running status.	IncorrectInstanceStatus	The current status of instance “XXX” does not support this action.	400
The specified ECS instance and the scaling group are not in the same region.	InvalidInstanceId.NetworkTypeMismatch	The network type of instance “XXX” does not support this action.	400
The specified scaling group and the attached ECS	InvalidInstanceId.VPCMismatch	Instance “XXX” and the specified scaling group are	400

instance are not in the same VPC.		not in the same VPC.	
The specified ECS instance has been attached to another scaling group.	InvalidInstanceId.InUse	Instance “XXX” is already attached to another scaling group.	400
The specified scaling group has an in-progress scaling activity.	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.	400
The Server Load Balancer instance of the specified scaling group is not in the active status.	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	400
Health check is not enabled for the Server Load Balancer of the specified scaling group.	IncorrectLoadBalancerHealthCheck	The current health check type of specified load balancer does not support this action.	400
The network type of the ECS instance contained in the specified Server Load Balancer is different from the network type of the scaling group.	InvalidLoadBalancerId.IncorrectInstanceNetworkType	The network type of the instance in specified load balancer does not support this action.	400
The ECS instance contained in the specified Server Load Balancer and VSwitchId are not in the same VPC.	InvalidLoadBalancerId.VPCMismatch	The specified virtual switch and the instance in specified load balancer are not in the same VPC.	400
The RDS instance in the specified scaling group is not in the running status.	IncorrectDBInstanceStatus	The current status of DB instance “XXX” does not support this action.	400
The number of IP addresses in the whitelist that can access the RDS instance in the scaling group has exceeded the upper limit.	QuotaExceeded.DBInstanceSecurityIP	Security IP quota exceeded in DB instance “XXX”.	400
The number of ECS	QuotaExceeded.Secu	Instance quota	400

instances attached to the specified security group exceeds the upper limit.	rityGroupInstance	exceeded in the specified security group.	
Total Capacity after the ECS instance is attached is greater than MaxSize.	IncorrectCapacity.MaxSize	To attach the instances, the total capacity will be greater than the MaxSize.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=AttachInstances
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ2eaQ
&InstanceId.1=i-28wt48iaa
&<Public Request Parameters>
```

Return example

XML format:

```
<AttachInstancesResponse>
<ScalingActivityId>bybj9OcaOT4ucPMbFhcqHfA3</ScalingActivityId>
<RequestId>DD0309B7-2613-4792-9B86-275906695253</RequestId>
</AttachInstancesResponse>
```

JSON format:

```
"RequestId": "6469DCD0-13AC-487E-85A0-CE4922908FDE",
"ScalingActivityId": "ebta5WbUzC8gcwUWvfchyT4U"
```

Remove an ECS instance

Remove an ECS instance

Description

Removes an ECS instance from a specified scaling group.

- When the ECS instance automatically created by the Auto Scaling service is removed from the scaling group, the ECS instance is disabled and released.
- When the manually attached ECS instance is removed from the scaling group, the ECS instance is neither disabled nor released.
- The interface can be called only when the scaling group is active.
- The interface can be called only when no scaling activity in the scaling group is in progress.
- When no scaling activity in the scaling group is in progress, the interface can be directly executed without cooldown.
- Successfully calling this interface only means that the Auto Scaling service has accepted the call request, and the scaling activity can be executed, but does not necessarily mean that the scaling activity can be successfully executed. You can use the returned ScalingActivityId to check the status of the scaling activity.
- When the total capacity of instances of the scaling group minus instances specified by this interface is smaller than than MinSize, the call fails.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: RemoveInstances.
ScalingGroupId	String	Yes	Scaling group ID.
InstanceId.N	String	Yes	ECS instance ID. You can input up to 20 IDs.

Return parameters

Name	Type	Description
ScalingActivityId	String	Scaling activity ID

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scaling group does not exist in this account.	InvalidScalingGroupId.NotFound	The specified scaling group does not exist.	404

The specified ECS instance does not exist in the scaling group.	InvalidInstanceId.NotFound	Instance “XXX” does not exist.	404
Open API is not fully authorized to the Auto Scaling service.	Forbidden.Unauthorized	A required authorization for the specified action is not supplied.	403
The specified scaling group is not active.	IncorrectScalingGroupStatus	The current status of the specified scaling group does not support this action.	400
The specified scaling group has an in-progress scaling activity.	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.	400
The Server Load Balancer instance in the scaling group to which the scaling rule belongs is not active.	IncorrectLoadBalancerStatus	The current status of the specified load balancer does not support this action.	400
The RDS instance in the scaling group to which the specified scaling rule belongs is not running.	IncorrectDBInstanceStatus	The current status of DB instance “XXX” does not support this action.	400
After instance removal, the total capacity is lower than MinSize.	IncorrectCapacity.MinSize	To remove the instances, the total capacity will be lesser than the MinSize.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=RemoveInstances
&ScalingGroupId=AG6CQdPU8OKdwLjgZcJ2eaQ
&InstanceId.1=i-28wt48iaa
&<Public Request Parameters>
```

Return example

XML format:

```
<RemoveInstancesResponse>
<ScalingActivityId>bybj9OcaOT4ucPMbFhcqHfA3</ScalingActivityId>
<RequestId>DD0309B7-2613-4792-9B86-275906695253</RequestId>
</RemoveInstancesResponse>
```

JSON format:

```
"RequestId": "6469DCD0-13AC-487E-85A0-CE4922908FDE",
"ScalingActivityId": "ebta5WbUzC8gcwUWvfchyT4U"
```

Scheduled task

Create a scheduled task

Create a scheduled task

Description

This operation creates a scheduled task according to input parameters.

- You can create up to 20 scheduled tasks.
- When the trigger of a scheduled task fails because a scaling activity in a scaling group is in progress or the scaling group is disabled, the scheduled task is automatically retried within the LaunchExpirationTime; otherwise, the scheduled trigger task is abandoned.
- If multiple tasks are scheduled at similar times to execute the rule of the same group, the earliest task triggers the scaling activity first, and other tasks make attempts to execute the rule within their Launch Expiration Time because a scaling group executes only one scaling activity at a time. If another scheduled task is still making triggering attempts within its Launch Expiration Time after the scaling activity is finished, the scaling rule is executed and the corresponding scaling activity is triggered.

Request parameters

Name	Type	Required?	Description

Action	String	Yes	Operation interface name, required parameter. Value: CreateScheduledTask.
RegionId	String	Yes	ID of the region in which the scheduled task is located.
ScheduledAction	String	Yes	Operations performed when the scheduled task is triggered. Fill in the unique identifier of the scaling rule.
LaunchTime	String	Yes	Time point at which the scheduled task is triggered. The date format follows the ISO8601 standard and uses UTC time. It is in the format of YYYY-MM-DDThh:mmZ. If RecurrenceType is specified, the time point specified by this attribute is the default time point at which the circle is executed. If RecurrenceType is not specified, the task is executed once on the designated date and time. A time point 90 days after creation or modification cannot be entered.
ScheduledTaskName	String	No	Display name of the scheduled task, which must be 2-40 characters (English or Chinese) long. It must begin with a number, an upper/lower-case letter or a Chinese character and may contain numbers, “_”, “-” or “.”. The account name is unique in the same region.

			If this parameter is not specified, the default value ScheduledScalingTaskId is used.
Description	String	No	Description of the scheduled task, which is 2-200 characters (English or Chinese) long.
LaunchExpirationTime	Integer	No	Time period within which the failed scheduled task is retried. The default value is 600s. Value range: [0, 21600]
RecurrenceType	String	No	Type of the scheduled task to be repeated. Optional values: - Daily: Recurrence interval by day for a scheduled task. - Weekly: Recurrence interval by week for a scheduled task. - Monthly: Recurrence interval by month for a scheduled task. RecurrenceType, RecurrenceValue and RecurrenceEndTime must be specified.
RecurrenceValue	String	No	Value of the scheduled task to be repeated. - Daily: Only one value in the range [1,31] can be filled. - Weekly: Multiple values can be filled. The values of Sunday to Saturday are 0 to 6 in sequence. Multiple values shall be separated by a comma ",". - Monthly: In the format of A-B. The

			value range of A and B is 1 to 31, and the B value must be greater than the A value. RecurrenceType, RecurrenceValue and RecurrenceEndTime must be specified.
RecurrenceEndTime	String	No	End time of the scheduled task to be repeated. The date format follows the ISO8601 standard and uses UTC time. It is in the format of YYYY-MM-DDThh:mmZ. A time point 90 days after creation or modification cannot be entered. RecurrenceType, RecurrenceValue and RecurrenceEndTime must be specified.
TaskEnabled	Bool	No	Whether to enable the scheduled task. - When the parameter is set to true, the task is enabled. - When the parameter is set to false, the task is disabled. The default value is true.

Return parameters

Name	Type	Description
ScheduledTaskId	String	ID of the scheduled task, which is generated by the system and globally unique.

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error Code	Description	HTTP Status Code

The specified region does not exist.	InvalidRegionId.NotFound	The specified region does not exist.	404
The scheduled task name already exists.	InvalidScheduledTaskName.Duplicate	The specified value of parameter <parameter name> is duplicated.	400
The specified ScheduledAction and the specified scheduled task are not in the same region.	ScheduledAction.RegionMismatch	The specified scheduled task and the specified scheduled action are not in the same Region.	400
Your scheduled task quota is exceeded.	QuotaExceeded.ScheduledTask	Scheduled task quota exceeded.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=CreateScheduledTask
&RegionId=cn-qingdao
&LaunchTime=2014-08-17T16:52Z
&RecurrenceType=Daily
&RecurrenceValue=1
&RecurrenceEndTime=2014-08-17T16:55Z
&ScheduledAction=ari:acs:ess:cn-qingdao:1344371:scalingRule/cCBpdYdQuBe2cUxOdu6piOk
&<Public Request Parameters>
```

Return example

XML format:

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

JSON format:

```
"RequestId": "04F0F334-1335-436C-A1D7-6C044FE73368",
"ScheduledTaskId": "edRtShc57WGXdt8TlPbrjsnV"
```

Modify a scheduled task

Modify a scheduled task

Description

Modifies the attributes of a scheduled task.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: ModifyScheduledTask.
ScheduledTaskId	String	Yes	ID of the scheduled task.
ScheduledTaskName	String	No	Display name of the scheduled task, which must be 2-40 characters (English or Chinese) long. It must begin with a number, an upper/lower-case letter or a Chinese character and may contain numbers, “_”, “-” or “.”. The account name is unique in the same region. If this parameter is not specified, the default value ScheduledScalingTaskId is used.
Description	String	No	Description of the scheduled task, which is 2-200 characters (English or Chinese) long.
ScheduledAction	String	No	Operations performed when the scheduled task is triggered. Fill in the unique identifier of

			the scaling rule.
LaunchTime	String	No	Time point at which the scheduled task is triggered. The date format follows the ISO8601 standard and uses UTC time. It is in the format of YYYY-MM-DDThh:mmZ. If RecurrenceType is specified, the time point specified by this attribute is the default time point at which the circle is executed. If RecurrenceType is not specified, the task is executed once on the designated date and time. A time point 90 days after creation or modification cannot be entered.
LaunchExpirationTime	Integer	No	Time period within which the failed scheduled task is retried. The default value is 600s. Value range: [0, 21600]
RecurrenceType	String	No	Type of the scheduled task to be repeated. Optional values: - Daily: Recurrence interval by day for a scheduled task. - Weekly: Recurrence interval by week for a scheduled task. - Monthly: Recurrence interval by month for a scheduled task. After modification, RecurrenceType, RecurrenceValue and RecurrenceEndTime must be set

			simultaneously.
RecurrenceValue	String	No	Value of the scheduled task to be repeated. - Daily: Only one value in the range [1,31] can be filled. - Weekly: Multiple values can be filled. The values of Sunday to Saturday are 0 to 6 in sequence. Multiple values shall be separated by a comma ",". - Monthly: In the format of A-B. The value range of A and B is 1 to 31, and the B value must be greater than the A value. After modification, RecurrenceType, RecurrenceValue and RecurrenceEndTime must be set simultaneously.
RecurrenceEndTime	String	No	End time of the scheduled task to be repeated. The date format follows the ISO8601 standard and uses UTC time. It is in the format of YYYY-MM-DDThh:mmZ. A time point 90 days after creation or modification cannot be entered. After modification, RecurrenceType, RecurrenceValue and RecurrenceEndTime must be set simultaneously.
TaskEnabled	Bool	No	Whether to enable the scheduled task. - When the parameter is set to true, the task is enabled.

			- When the parameter is set to false, the task is disabled. The default value is true.

Return parameters

Public parameters.

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error Code	Description	HTTP Status Code
The specified scheduled task does not exist in your account.	InvalidScheduledTaskId.NotFound	The specified scheduled task does not exist.	404
The scheduled task name already exists.	InvalidScheduledTaskName.Duplicate	The specified value of parameter <parameter name> is duplicated.	400
The specified ScheduledAction and the specified scheduled task are not in the same region.	ScheduledAction.RegionMismatch	The specified scheduled task and the specified scheduled action are not in the same Region.	400

Example

Request example

```
http://ess.aliyuncs.com/?Action=ModifyScheduledTask
&ScheduledTaskId=edRtShc57WGXdt8TlPbrjsnV
&LaunchTime=2014-08-18T10:52Z
&RecurrenceEndTime=2014-08-20T16:55Z
&<Public Request Parameters>
```

Return example

XML format:

```
<ModifyScheduledTaskResponse>
<RequestId>F9372E8D-C163-471F-BEB4-3A02B3CE176E</RequestId>
</ModifyScheduledTaskResponse>
```

JSON format:

```
"RequestId": "04F0F334-1335-436C-A1D7-6C044FE73368"
```

Query a scheduled task

Query a scheduled task

Description

Queries information of a scheduled task.

You can use the ScheduledAction to query the relevant scheduled task.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value: DescribeScheduledTasks.
RegionId	String	Yes	Region of the scheduled task.
ScheduledTaskId.N	String	No	ID of the scheduled task. You can enter at most 20 IDs. Invalid scheduled task IDs are not displayed in the query results, and no error is reported.
ScheduledTaskName.N	String	No	Display name of the schedules task. You can enter at most 20 display names.

			Names of invalid scheduled tasks will be neglected in the query results, and no error is reported.
ScheduledAction.N	String	No	Operations performed when the scheduled task is triggered. You can enter at most 20 operations. Invalid operations are not displayed in the query results, and no error is reported.
PageNumber	Integer	No	Page number of the scheduled task list, starting from 1. The default value is 1.
PageSize	Integer	No	When querying by page, this parameter indicates the number of lines per page. Maximum value: 50; default value: 10.

Return parameters

Name	Type	Description
TotalCount	Integer	Total number of scheduled tasks
PageNumber	Integer	Current page number
PageSize	Integer	Number of lines per page
ScheduledTasks	ScheduledTaskSetType	A set of scheduled task information

ScheduledTaskSetType is a set of ScheduledTaskItemTypes:

Name	Type	Description
ScheduledTask	ScheduledTaskItemType	Information of the scheduled task

The attributes of the ScheduledTaskItemType are listed below:

Name	Type	Description
ScheduledTaskId	String	ID of the scheduled task

ScheduledTaskName	String	Name of the scheduled task
Description	String	Description of the scheduled task
ScheduledAction	String	Operations performed when the scheduled task is triggered
LaunchTime	String	Time point at which the scheduled task is triggered
LaunchExpirationTime	Integer	Retry interval for the failed scheduled task
RecurrenceType	String	Type of the scheduled task to be repeated
RecurrenceValue	String	Value of the scheduled task to be repeated
RecurrenceEndTime	String	End time of the scheduled task to be repeated
TaskEnabled	Bool	Whether to enable the scheduled task

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Example

Request example

```
http://ess.aliyuncs.com/?Action=DescribeScheduledTasks
&RegionId=cn-qingdao
&ScheduledTaskId.1=edRtShc57WGXdt8TlPbrjsnV
&PageSize=50
&<Public Request Parameters>
```

Return example

XML format:

```
<DescribeScheduledTasksResponse>
<RequestId>B9B498DA-E836-45FF-83C7-1930492FDD5A</RequestId>
<TotalCount>1</TotalCount>
<PageNumber>1</PageNumber>
<PageSize>50</PageSize>
<ScheduledTasks>
<ScheduledTask>
```

```
<Description/>
<LaunchExpirationTime>600</LaunchExpirationTime>
<LaunchTime>2014-08-18T10:52Z</LaunchTime>
<RecurrenceEndTime>2014-08-20T16:55Z</RecurrenceEndTime>
<RecurrenceType>Daily</RecurrenceType>
<RecurrenceValue>1</RecurrenceValue>
<ScheduledAction>
ari:acs:ess:cn-qingdao:1344371:scalingrule/cCBpdYdQuBe2cUxOdu6piOk
</ScheduledAction>
<ScheduledTaskId>edRtShc57WGXdt8TlPbrjsnV</ScheduledTaskId>
<ScheduledTaskName>edRtShc57WGXdt8TlPbrjsnV</ScheduledTaskName>
<TaskEnabled>true</TaskEnabled>
</ScheduledTask>
</ScheduledTasks>
</DescribeScheduledTasksResponse>
```

JSON format:

```
{
"RequestId": "43434132-91C4-4264-8343-681130760A5C",
"TotalCount": 1,
"PageSize": 1,
"PageNumber": 1,
"ScheduledTasks": {
"ScheduledTask": [
{
"TaskEnabled": true,
"ScheduledTaskId": "b27CLSc8T478c2iqHr6fqbF",
"Description": "ditingshigechunqingchunan",
"ScheduledTaskName": "9906a33f-14eb-42b8-8bdb-ee8cdf912706",
"LaunchExpirationTime": 120,
"RecurrenceType": "Daily",
"RecurrenceEndTime": "2014-08-13T19:19Z",
"LaunchTime": "2014-08-12T17:55Z",
"RecurrenceValue": "1",
"ScheduledAction": "ari:acs:ess:cn-qingdao:1344371:scalingrule/qGx9feK1giadmp3XKer94cD"
}
]
}
}
```

Delete a scheduled task

Delete a scheduled task

Description

Deletes a specified scaling rule.

Request parameters

Name	Type	Required?	Description
Action	String	Yes	Operation interface name, required parameter. Value:DeleteScheduledTask.
ScheduledTaskId	String	Yes	ID of the scheduled task.

Return parameters

Public parameters.

Error code

For errors common to all interfaces, refer to the Client Error Table or Server Error Table.

Error	Error code	Description	HTTP status code
The specified scheduled task does not exist in your account.	InvalidScheduledTaskId.NotFound	The specified scheduled task does not exist.	404

Example

Request example

```
http://ess.aliyuncs.com/?Action=DeleteScheduledTask
&ScheduledTaskId=edRtShc57WGXdt8TlPbrjsnV
&<Public Request Parameters>
```

Return example

XML format:

```
<DeleteScheduledTaskResponse>
<RequestId>7683D637-CF8A-41DC-85A8-E128061E65FC</RequestId>
</DeleteScheduledTaskResponse>
```

JSON format:

```
{
"RequestId": "04F0F334-1335-436C-A1D7-6C044FE73368"
}
```

How to ensure idempotence

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”.