

# Alibaba Cloud Auto Scaling

Quick Start

Issue: 20181115

# Legal disclaimer

---

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

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.
2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company, or individual in any form or by any means without the prior written consent of Alibaba Cloud.
3. The content of this document may be changed due to product version upgrades, adjustments, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and the updated versions of this document will be occasionally released through Alibaba Cloud-authorized channels. You shall pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.
4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides the document in the context that Alibaba Cloud products and services are provided on an "as is", "with all faults" and "as available" basis. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not bear any liability for any errors or financial losses incurred by any organizations, companies, or individuals arising from their download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, bear responsibility for any indirect, consequential, exemplary, incidental, special, or punitive damages, including lost profits arising from the use or trust in this document, even if Alibaba Cloud has been notified of the possibility of such a loss.
5. By law, all the content of the Alibaba Cloud website, including but not limited to works, products, images, archives, information, materials, website architecture, website graphic layout, and webpage design, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade

secrets. No part of the Alibaba Cloud website, product programs, or content shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates).

6. Please contact Alibaba Cloud directly if you discover any errors in this document.



# Generic conventions

Table -1: Style conventions

Style	Description	Example
	This warning information indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	 <b>Danger:</b> Resetting will result in the loss of user configuration data.
	This warning information indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	 <b>Warning:</b> Restarting will cause business interruption. About 10 minutes are required to restore business.
	This indicates warning information, supplementary instructions, and other content that the user must understand.	 <b>Note:</b> Take the necessary precautions to save exported data containing sensitive information.
	This indicates supplemental instructions, best practices, tips, and other content that is good to know for the user.	 <b>Note:</b> You can use <b>Ctrl + A</b> to select all files.
>	Multi-level menu cascade.	<b>Settings &gt; Network &gt; Set network type</b>
<b>Bold</b>	It is used for buttons, menus, page names, and other UI elements.	Click <b>OK</b> .
Courier font	It is used for commands.	Run the <code>cd /d C:/windows</code> command to enter the Windows system folder.
<i>Italics</i>	It is used for parameters and variables.	<code>bae log list --instanceid Instance_ID</code>
[ ] or [a b]	It indicates that it is a optional value, and only one item can be selected.	<code>ipconfig [-all -t]</code>
{ } or {a b}	It indicates that it is a required value, and only one item can be selected.	<code>swich {stand   slave}</code>

# Contents

---

<b>Legal disclaimer.....</b>	<b>I</b>
<b>Generic conventions.....</b>	<b>I</b>
<b>1 Step 1. Activate and authorize the Auto Scaling service.....</b>	<b>1</b>
<b>2 Step 2. Create a scaling solution.....</b>	<b>5</b>

# 1 Step 1. Activate and authorize the Auto Scaling service

---

This topic introduces the API information for activating and authorizing the Auto Scaling service.

## Authorization procedure

1. Log on to the [Auto Scaling console](#).
2. Click **Authorize** to go to the RAM console and grant authorization.
3. Select **AliyunESSDefaultRole** and click **Agree to Authorize**.
4. Return to the Auto Scaling console and refresh the page.

## Next step

After completing authorization, you can use the Auto Scaling service. Next, you can create your first scaling group in the desired region (such as China (Shanghai)). For more information, see [create a scaling group](#).

## Permission list

By default, the role `AliyunESSDefaultRole` allows Auto Scaling to call the following Alibaba Cloud resources for you:

- ECS permissions

Permission name	Description
ecs:RunInstances	Create one or more ECS instances as needed.
ecs:CreateInstance	Create ECS instances.
ecs:StartInstance	Start ECS instances.
ecs:AllocatePublicIpAddress	Allocate public IP addresses to ECS instances.
ecs:StopInstance	Stop ECS instances.
ecs>DeleteInstance	Delete ECS instances.
ecs:DescribeInstances	Query ECS instance lists.
ecs:DescribeInstanceAttribute	Query ECS instance attributes.
ecs:ModifyInstanceAttribute	Modify ECS instance attributes.
ecs:DescribeSecurityGroupAttribute	Query security group attributes.

Permission name	Description
ecs:DescribeSnapshots	Query snapshot lists.
ecs:DescribeKeyPairs	Query key pair lists.

- SLB permissions

Permission name	Description
slb:DescribeLoadBalancerAttribute	Query SLB instance information.
slb:RemoveBackendServers	Delete backend servers from a SLB instance.
slb:DescribeHealthStatus	Check the health of the backend servers of a SLB instance.
slb:AddBackendServers	Add backend servers to a SLB instance.
slb:SetBackendServers	Configure backend server weights.

- RDS permissions

Permission name	Description
rds:ModifySecurityIps	Modify the IP address whitelist of an RDS instance.
rds:DescribeDBInstanceAttribute	View RDS instance details.
rds:DescribeTaskInfo	Query RDS task information.
rds:DescribeDBInstanceIPArrayList	View the IP address whitelist of an RDS instance.

- VPC permissions

Permission name	Description
vpc:DescribeVpcs	Query VPC lists.
vpc:DescribeVSwitches	Query VSwitch lists.

- MNS permissions

Permission name	Description
mns:ListTopic	List topic lists under an account.
mns:ListQueue	List queue lists under an account.
mns:SendMessage	Send messages.
mns:PublishMessage	Publish messages.

The complete policy list for the role AliyunESSDefaultRole is as follows:

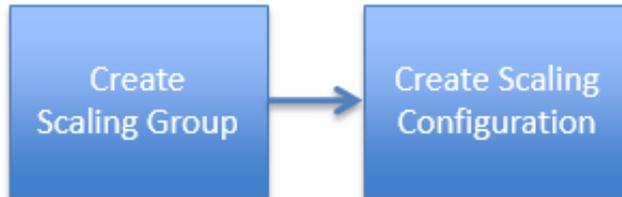
```
{
  "Version": "1",
  "Statement": [
    {
      "Action": [
        "ecs:CreateInstance",
        "ecs:RunInstances",
        "ecs:StartInstance",
        "ecs:AllocatePublicIpAddress",
        "ecs:StopInstance",
        "ecs>DeleteInstance",
        "ecs:DescribeInstances",
        "ecs:DescribeInstanceAttribute",
        "ecs:ModifyInstanceAttribute",
        "ecs:DescribeSecurityGroupAttribute",
        "ecs:DescribeImages",
        "ecs:DescribeSnapshots",
        "ecs:DescribeKeyPairs",
        "slb:DescribeLoadBalancerAttribute",
        "slb:RemoveBackendServers",
        "slb:DescribeHealthStatus",
        "slb:AddBackendServers",
        "slb:SetBackendServers",
        "rds:ModifySecurityIps",
        "rds:DescribeDBInstanceAttribute",
        "rds:DescribeTaskInfo",
        "rds:DescribeDBInstanceIPArrayList"
      ],
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "vpc:DescribeVpcs",
        "vpc:DescribeVSwitches"
      ],
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": [
        "mns:ListTopic",
        "mns:ListQueue",
        "mns:SendMessage",
        "mns:PublishMessage"
      ],
      "Resource": "*",
      "Effect": "Allow"
    },
    {
      "Action": "ram:PassRole",
      "Resource": "*",
      "Effect": "Allow",
      "Condition": {
        "StringEquals": {
          "acs:Service": "ecs.aliyuncs.com"
        }
      }
    }
  ]
}
```

```
}
```

## 2 Step 2. Create a scaling solution

---

It takes two main steps as follows to create a simple scaling solution.



Creating a scaling group is the first step to use Auto Scaling. A scaling group is a set of ECS instances with the same scenarios. You can define the maximum and minimum numbers of ECS instances in the group, its associated Server Load Balancer and RDS instances, and other attributes. After you create a scaling group, its scope attributes cannot be changed. For more information, see [regions and zones](#). Here is an example of how to create a scaling group and complete the scaling configurations in a region such as China East 2 (Shanghai).

### Create a scaling group

1. Log on to the [Auto Scaling console](#).
2. Select a **Region**, such as China East 2.
3. Click **Create Scaling Group**.
4. On the Create Scaling Group page:
  - a. Enter the **Scaling Group Name**, such as ScalingGroupTest.
  - b. Set the **Maximum Number of Instances Allowed for Scaling**, such as 4.
  - c. Set the **Minimum Number of Instances Allowed for Scaling**, such as 1.
  - d. Set the **Default Cool-down Time (Seconds)**, such as 600.
  - e. Set the **Removal Policy**. For example, filter the **Earliest Instances** first, and filter **Instances Corresponded to the Earliest Scaling Configurations** from the result.
  - f. Set **Network Type**, such as **VPC**. For more information, see .
  - g. Configure a Server Load Balancer instance Health check must be enabled for all listening ports of the specified Server Load Balancer instance, otherwise, the instance cannot be added to the scaling group.
  - h. Configure an RDS instance
  - i. Click **Submit** to complete the creation.

5. After the scaling group is created successfully, you can [Create Scaling Configuration](#) directly, or click **Later**.

## Create scaling configurations

### Prerequisite

You have created an ECS security group. For more information, see [create a security group](#).

### Procedure

1. Log on to the [Auto Scaling console](#).
2. In the left-side navigation pane, click **Scaling Groups**.
3. Select a **region**, such as China East 2.
4. Find the target scaling group, and click **Add Scaling Configuration**.
5. On the Basic Configurations page:
  - a. Enter the **Scaling Configuration Name**, such as ScalingGroupTest. 1.
  - b. Select a **Billing Method**, such as **Pay-As-You-Go**. For more information, see [Pay-As-You-Go](#) and [preemptible instance](#).
  - c. Select an **instance** type, such as ecs.xn4.small and ecs.g5.2xlarge. For more information, see [instance type families](#).



#### Note:

We recommend that you configure multiple instance types to avoid ECS creation failure due to insufficient resource of certain instance type.

- d. Select an **image**, such as CentOS 7.4 64. If you want to realize functions like Web server automatic startup, automatic code and script downloading, and so on, select **Custom Image**.
  - e. Select the **storage** type, such as 40 GB Ultra Disk.
  - f. Select the **public network bandwidth**, such as 1 Mbit/s by traffic.
  - g. Select a security group.
  - h. Click **Next**.
6. On the System Configurations page:
    - a. (Optional) Select a tag. For more information, see [add a tag to resources](#).
    - b. Set logon credentials, such as SSH key pair. For more information, see [create an SSH key pair](#).
    - c. Set the **Instance Name**, such as ScalingGroupTest.

- d. Click **Next**.
7. Confirm the scaling configuration information, including the scaling group set in the previous two steps and the scaling configuration information, and the estimated cost of the scaling configuration. Check the protocol, and click **Confirm**.

**Note:**

The cost only includes the cost of the ECS service. Auto Scaling is for free. For more information, see [billing](#).

8. In the dialog box, click **Back** or **Enable Configuration**, such as **Enable Configuration**.

**Note:**

After you enable the scaling configuration, the minimum number of ECS instances you set in the scaling configuration are automatically created and added to the specified Server Load Balancer instance, and the private IPs of the instances are added to the specified RDS instance whitelist.

**What to do next**

You can enable your first scaling group in the corresponding region, such as China East 2.