

ALIBABA CLOUD

# Alibaba Cloud

弹性伸缩 ESS  
快速入门

文档版本：

 阿里云

## 法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

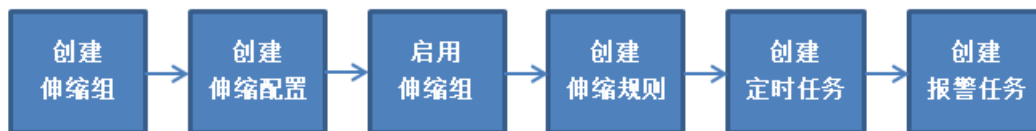
1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

# 目录

1.使用流程	04
2.开通和授权服务	05
3.创建伸缩组和伸缩配置	08
4.实现自动扩张	12
5.实现自动收缩	14

# 1.使用流程

本文主要介绍创建完整弹性伸缩方案的步骤。



1. 创建伸缩组（CreateScalingGroup），配置伸缩资源的最小值（MinSize）、最大值（MaxSize）及需要关联的负载均衡实例和RDS实例。
2. 创建伸缩配置（CreateScalingConfiguration），指定需要弹性伸缩的 ECS 实例的相关属性，如 ImageID、InstanceType 等。
3. 以第二步创建的伸缩配置启用伸缩组（EnableScalingGroup）。
4. 创建伸缩规则（CreateScalingRule），如加 N 台 ECS 实例的伸缩规则。
5. 创建定时任务（CreateScheduledTask），如创建 12: 00 触发第四步伸缩规则的定时任务。
6. 创建报警任务（云监控 API PutAlarmRule），如创建 CPU 平均值（也可以是最大值或最小值）大于等于 80% 则增加一台 ECS 实例的报警任务。

## 2. 开通和授权服务

本文介绍如何开通弹性伸缩服务并授予弹性伸缩服务访问云资源的权限。

### 背景信息

首次使用弹性伸缩服务时才需要开通和授权服务，如果您不是首次使用弹性伸缩，请跳过本文所述步骤前往创建伸缩组和伸缩配置，具体操作请参见[创建伸缩组和伸缩配置](#)。


弹性伸缩服务关联角色（AliyunServiceRoleForAutoScaling）是为弹性伸缩服务获取其他云服务的访问权限而提供的RAM角色。通过弹性伸缩服务关联角色，弹性伸缩服务可以获得云服务器ECS、专有网络VPC、云数据库RDS、负载均衡SLB、运维编排OOS、消息服务MNS、云监控CMS服务的访问权限。更多服务关联角色的说明，请参见[服务关联角色](#)。

开通弹性伸缩服务时自动创建服务关联角色，弹性伸缩服务关联角色的RAM角色名称为AliyunServiceRoleForAutoScaling，AliyunServiceRoleForAutoScaling包含系统权限策略AliyunServiceRolePolicyForAutoScaling。

 **说明** 不支持修改或删除服务关联角色的系统权限策略。

您可以在RAM管理控制台查看AliyunServiceRolePolicyForAutoScaling的最新策略内容，更多信息请参见[查看权限策略基本信息](#)。

您可以在RAM管理控制台删除弹性伸缩服务关联角色，更多信息请参见[删除RAM角色](#)。

 **说明** 您必须提前释放依赖弹性伸缩服务关联角色的所有云资源，包括伸缩组、定时任务和报警任务。

### 步骤一：开通弹性伸缩服务

1. 登录弹性伸缩控制台。
2. 单击开通弹性伸缩。



3. 选中我已阅读并同意《弹性伸缩服务ESS（按量付费）服务协议》复选框，然后单击立即开通。



4. 在开通完成页面，单击管理控制台返回控制台页面。

## 步骤二：授予弹性伸缩服务权限

在上一步开通服务时，服务关联角色AliyunServiceRoleForAutoScaling会自动创建。

如果自动创建成功，请跳过本小节的步骤，您已经开通并授权服务。

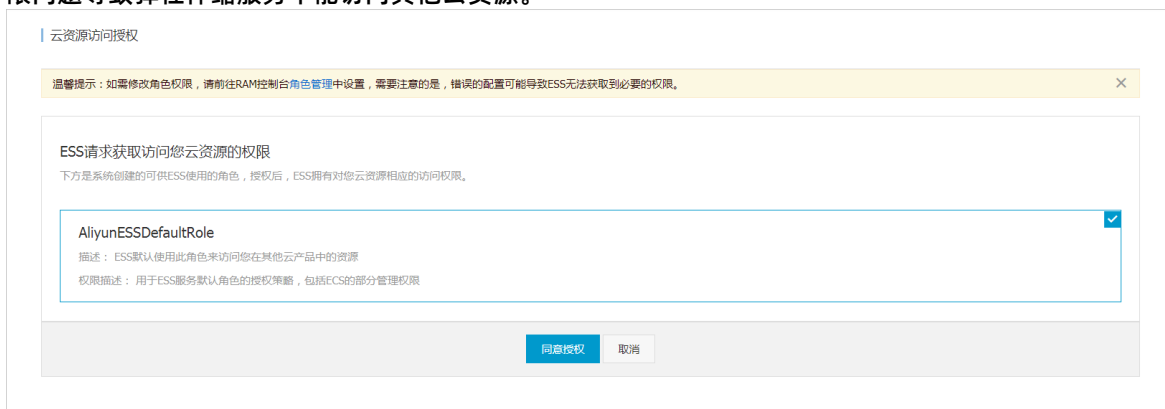
如果自动创建失败，返回控制台页面后会提示您手动授权弹性伸缩服务访问其他云资源，请按照本小节步骤完成。

### 1. 单击前往授权。



### 2. 选中AliyunESSDefaultRole，然后单击同意授权。

AliyunESSDefaultRole是普通服务角色，默认包含系统权限策略AliyunESSRolePolicy，也用于授权弹性伸缩服务访问其他云资源。普通服务角色支持修改和删除权限策略，但建议您保持默认，避免因权限问题导致弹性伸缩服务不能访问其他云资源。



## 后续步骤

创建伸缩组和伸缩配置，具体操作请参见[创建伸缩组和伸缩配置](#)。

## 3. 创建伸缩组和伸缩配置

创建伸缩组和伸缩配置是实现自动伸缩的第一步，伸缩组是具有相同应用场景的ECS实例的集合，伸缩配置是伸缩组自动创建ECS实例时的模板。本章节演示创建一个最小实例数为1的伸缩组，并为伸缩组创建伸缩配置以启用伸缩组。

### 前提条件

已完成以下操作：

- [开通和授权服务](#)
- [创建安全组](#)

 说明 安全组所在地域和伸缩组所在地域必须一致。

### 背景信息

创建伸缩组后，地域属性无法变更，更多信息请参见[地域和可用区](#)。

### 步骤一：创建伸缩组

本示例步骤使用简洁的配置，更全面的伸缩组信息请参见[创建伸缩组](#)。



1. 登录[弹性伸缩控制台](#)。
2. 在顶部菜单栏处，选择地域。
3. 在左侧导航栏中，单击[伸缩组管理](#)。
4. 在页面右上角，单击[创建伸缩组](#)。
5. 完成伸缩组配置，然后单击[创建伸缩组](#)。  
本示例中采用以下配置，未提及的配置保持默认即可。

配置类型	配置项	示例	描述
伸缩组内实例模板来源	来源类型	从零开始创建	如果从零开始创建伸缩组，伸缩组创建完成后没有可用的伸缩配置来源，您必须继续创建伸缩配置。
伸缩组基本信息	伸缩组名称	MyFirstScalingGroup	无
	组内最大实例数	3	伸缩组内最多有3台ECS实例，多出的ECS实例会被自动移出。
	组内最小实例数	1	伸缩组内最少有1台ECS实例，不足1台时会自动创建ECS实例。
	标签	ESS:Documentation	标签用于归类伸缩组，方便管理。
组内实例扩缩容配置	网络类型	经典网络	创建伸缩配置时，只能选择支持经典网络的实例规格。

6. 在创建伸缩组状态向导对话框，单击[查看伸缩组详情](#)。

## 步骤二：创建伸缩配置并启用伸缩组

本示例步骤使用简洁的配置，更全面的伸缩配置信息请参见[创建伸缩配置](#)。

1. 在左侧导航栏中，单击[组内实例配置信息来源](#)。
2. 在页面左上角，单击[伸缩配置](#)。
3. 单击[创建伸缩配置](#)。

4. 完成伸缩配置，然后单击确认创建。  
本示例中采用以下配置，未提及的配置保持默认即可。

配置页面	配置项	示例	描述
基础配置	付费模式	按量付费	弹性伸缩服务免费，但是加入伸缩组的ECS实例需要按云服务器ECS的定价支付费用，更多信息请参见 <a href="#">计费概述</a> 。
	实例	共享通用型mn4的实例规格ecs.mn4.small	专有网络伸缩组支持更多更新的实例规格，专有网络伸缩组创建说明请参见 <a href="#">创建伸缩组</a> 。
	镜像	公共镜像CentOS 7.6 64位	实例启动后，系统盘将完整复制镜像的操作系统和应用数据。
	公网IP	按使用流量，峰值1 Mbit/s	按公网出方向实际发生的网络流量计费，带宽不会超过1 Mbit/s。
	安全组	sg-bp18kz60mefsicg****	选择您提前创建好的安全组。
系配配置	登录凭证	创建后设置	ECS实例创建完成后，再手动为ECS实例设置密码。
确认配置	伸缩配置名称	MyFirstScalingConfiguration	无

- 在伸缩配置创建成功对话框，单击启用配置。
- 在选用伸缩配置对话框，单击确定。
- 在左侧导航栏中，单击基本信息。
- 在页面右上角，单击启用。

## 执行结果

由于伸缩组的最小实例数为1，伸缩组启用后会自动创建1台ECS实例。您可以前往ECS实例列表查看实例，自动创建出的ECS实例配置和伸缩配置保持一致，具体操作请参见[查询ECS实例列表](#)。



## 后续步骤

伸缩组启用成功后，您可以为伸缩组创建伸缩规则，然后搭配定时任务和报警任务实现自动伸缩，具体操作请参见：

- [实现自动扩张](#)
- [实现自动收缩](#)

## 4. 实现自动扩张

弹性伸缩能够通过定时任务自动执行伸缩规则，添加或移出ECS实例。本章节介绍如何自动扩张，向伸缩组添加2台ECS实例。

### 前提条件

已完成快速入门中的以下操作：

- [开通和授权服务](#)
- [创建伸缩组和伸缩配置](#)


### 背景信息

本章节中使用已创建的伸缩组MyFirstScalingGroup和伸缩配置MyFirstScalingConfiguration演示自动扩张的效果，伸缩组中已经有1台ECS实例。

### 步骤一：创建伸缩规则

1. 登录[弹性伸缩控制台](#)。
2. 在左侧导航栏中，单击[伸缩组管理](#)。
3. 找到待操作的伸缩组，选择一种方式打开伸缩组详情页面。
  - 在[伸缩组名称/ID](#)区域，单击伸缩组ID。
  - 在[操作区域](#)，单击[管理](#)。
4. 在左侧导航栏中，单击[伸缩规则](#)。
5. 在页面右上角，单击[创建伸缩规则](#)。
6. 设置伸缩规则的属性，然后单击[创建伸缩规则](#)。  
本示例中采用以下配置，未提及的配置保持默认即可。

配置项	示例
伸缩规则名称	Add2
伸缩规则类型	简单规则
执行的操作	增加2台

7. 在页面左上角，单击  图标，返回[伸缩组管理列表](#)页面。

## 步骤二：创建定时任务

1. 在左侧导航栏中，单击自动触发任务管理 > 定时任务。
2. 在页面右上角，单击创建定时任务。
3. 设置定时任务的属性，然后单击提交。  
本示例中采用以下配置，未提及的配置保持默认即可。

配置项	示例	说明
任务名称	ScheduledScalingOut	无
描述	定时添加2台实例。	无
执行时间	2019-11-11 16:35	当前时间的5分钟后。
伸缩组	MyFirstScalingGroup	为伸缩组MyFirstScalingGroup执行定时任务。
伸缩方式	选择已有伸缩规则	无
伸缩规则	Add2	执行伸缩规则Add2，向伸缩组添加2台ECS实例。

## 执行结果

到执行时间2019-11-11 16:35后，定时任务ScheduledScalingOut自动执行伸缩规则Add2，向伸缩组MyFirstScalingGroup添加2台ECS实例。您可以前往伸缩活动列表查看详情。

伸缩活动 列表 图表

伸缩活动	变化后总实例数	开始时间	停止时间	描述	状态(全部)	操作
asa-bp-...	3	2019年11月11日 16:35	2019年11月11日 16:36	Add "2" ECS ins...	成功	<a href="#">查看详情</a>
asa-bp-...	1	2019年11月11日 16:11	2019年11月11日 16:11	Add "1" ECS ins...	成功	<a href="#">查看详情</a>

共有2条, 每页显示: 10 条

伸缩活动ID: asa-bp-... 状态: 成功

开始时间: 2019年11月11日 16:35 停止时间: 2019年11月11日 16:36

活动起因: A scheduled task executes scaling rule "asr-bp1-...", changing the Total Capacity from "1" to "3"(Max Size).

详细信息: new ECS instances "i-bp1-...", i-bp1-..." are created.

状态信息: "2" ECS instances are added

## 5. 实现自动收缩

弹性伸缩能够通过报警任务监控伸缩组内ECS实例的使用情况，并在触发报警时自动执行伸缩规则，添加或移出ECS实例。本章节介绍如何使用报警任务实现自动收缩，将2台ECS实例移出伸缩组。

### 前提条件

已完成快速入门中的以下操作：

- [开通和授权服务](#)
- [创建伸缩组和伸缩配置](#)
- [实现自动扩张](#)


### 背景信息

本章节中使用已创建的伸缩组MyFirstScalingGroup演示自动收缩的效果，伸缩组中已经有3台ECS实例。

### 步骤一：创建伸缩规则

1. 登录[弹性伸缩控制台](#)。
2. 在左侧导航栏中，单击[伸缩组管理](#)。
3. 找到待操作的伸缩组，选择一种方式打开伸缩组详情页面。
  - 在[伸缩组名称/ID](#)区域，单击伸缩组ID。
  - 在[操作](#)区域，单击[管理](#)。
4. 在左侧导航栏中，单击[伸缩规则](#)。
5. 在页面右上角，单击[创建伸缩规则](#)。
6. 设置伸缩规则的属性，然后单击[创建伸缩规则](#)。  
本示例中采用以下配置，未提及的配置保持默认即可。

配置项	示例
伸缩规则名称	To1
伸缩规则类型	简单规则
执行的操作	调整至1台

7. 在页面左上角，单击  图标，返回[伸缩组管理列表](#)页面。

## 步骤二：创建报警任务

1. 在左侧导航栏中，单击自动触发任务管理 > 报警任务。
2. 在页面右上角，单击创建报警任务。
3. 设置报警任务的属性，然后单击提交。  
本示例中采用以下配置，未提及的配置保持默认即可。

配置项	示例	说明
任务名称	EventTriggeredScalingIn	无
描述	平均CPU使用率不足10%时移出ECS实例。	无
监控资源	MyFirstScalingGroup	监控伸缩组MyFirstScalingGroup的指标。
监控类型	系统监控	无
监控项	(ECS) CPU使用率	监控伸缩组内ECS实例的CPU使用率。
统计周期	1分钟	1分钟统计1次数据。
统计方法	平均值 <= 10%	如果伸缩组内ECS实例的CPU使用率平均值小于等于10%，则记录1次。
重复几次后报警	5次	ECS实例的CPU使用率统计结果连续5次小于等于10%，则触发报警。
报警触发规则	To1	触发报警后，执行伸缩规则To1，将伸缩组内的ECS实例数量调整为1台。

## 执行结果

报警任务EventTriggeredScalingIn会监控伸缩组MyFirstScalingGroup中3台ECS实例的平均CPU使用率，每分钟统计1次。如果连续5次的统计结果均小于等于10%，则触发报警，报警任务EventTriggeredScalingIn自动执行伸缩规则To1，将伸缩组MyFirstScalingGroup内的ECS实例数量调整为1台。您可以前往伸缩活动列表查看详情。

伸缩活动 列表 图表

伸缩活动	变化后总实例数	开始时间	停止时间	描述	状态(全部)	操作
...	...	...	...	...	失败	查看详情
asa-bp-...	1	2019年11月11日 17:07	2019年11月11日 17:08	Remove "2" ECS ...	成功	查看详情
asa-bp-...	3	2019年11月11日 16:35	2019年11月11日 16:36	Add "2" ECS ins...	成功	查看详情
asa-bp-...	1	2019年11月11日 16:11	2019年11月11日 16:11	Add "1" ECS ins...	成功	查看详情

共有4条, 每页显示: 10 条

伸缩活动ID: asa-bp-... 状态: 成功

开始时间: 2019年11月11日 17:07 停止时间: 2019年11月11日 17:08

活动起因: An alarm task executes scaling rule "asr-bp1-...", changing the Total Capacity from "3" to "1"(Min Size).

详细信息: Instances "i-bp1-...", i-bp1-..." are deleted.

状态信息: "2" ECS instances are removed