

弹性伸缩服务

# 用户指南

文档版本 27  
发布日期 2021-11-20



版权所有 © 华为技术有限公司 2022。保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

## 商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

---

# 目录

<b>1 伸缩组</b>	<b>1</b>
1.1 创建伸缩组	1
1.2 (可选) 添加负载均衡器到伸缩组	5
1.3 更换伸缩组的伸缩配置	6
1.4 启用伸缩组	7
1.5 停用伸缩组	7
1.6 修改伸缩组	8
1.7 删除伸缩组	8
<b>2 伸缩配置</b>	<b>10</b>
2.1 创建伸缩配置	10
2.2 使用已有云服务器创建伸缩配置	10
2.3 使用新模板创建伸缩配置	12
2.4 复制伸缩配置	16
2.5 删除伸缩配置	16
<b>3 伸缩策略</b>	<b>17</b>
3.1 伸缩策略介绍	17
3.2 创建伸缩策略	18
3.3 管理伸缩策略	23
<b>4 伸缩活动</b>	<b>26</b>
4.1 动态扩展资源	26
4.2 按计划扩展资源	28
4.3 手动扩展资源	28
4.4 实例移除策略	30
4.5 查询伸缩活动	30
4.6 生命周期挂钩	31
4.7 实例保护	36
4.8 实例备用	37
<b>5 伸缩带宽</b>	<b>39</b>
5.1 创建伸缩带宽策略	39
5.2 查看伸缩带宽策略详情	44
5.3 管理伸缩带宽策略	45

<b>6 伸缩组和实例的监控</b> .....	<b>47</b>
6.1 弹性伸缩健康检查.....	47
6.2 为伸缩组配置通知.....	47
6.3 记录弹性伸缩.....	48
6.4 标记伸缩组和实例.....	51
6.5 监控指标说明.....	52
6.6 查看监控指标数据.....	56
6.7 设置监控告警规则.....	57
<b>7 权限管理</b> .....	<b>58</b>
7.1 创建用户并授权使用 AS.....	58
7.2 创建 AS 自定义策略.....	59
<b>A 修订记录</b> .....	<b>61</b>

# 1 伸缩组

## 1.1 创建伸缩组

### 操作场景

伸缩组是具有相同属性和应用场景的云服务器和伸缩策略的集合，是启停伸缩策略和进行伸缩活动的基本单位。您可以使用伸缩策略设定的条件自动增加、减少伸缩组中的实例数量，或维持伸缩组中固定的实例数量。

创建伸缩组，需要配置最大实例数、最小实例数、期望实例数和负载均衡器等参数。

### 创建须知

不同可用区支持的云服务器类型可能不同。

- 如果伸缩组中所有可用区均不支持伸缩配置中的云服务器类型，此时：
  - 如果伸缩组当前为停用状态，则无法启用伸缩组。
  - 如果伸缩组当前为启用状态，则在扩容操作时，伸缩组状态变为异常。
- 如果伸缩组中仅有部分可用区支持伸缩配置中的云服务器类型，则在弹性伸缩活动中自动添加的云服务器只分布在支持该类型云服务器的可用区中，不能均匀的分布在所有可用区中。
- 创建伸缩组前请先创建好您所需的伸缩配置。

### 操作步骤


1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 选择“计算 > 弹性伸缩 > 伸缩实例”。
4. 单击“创建弹性伸缩组”。
5. 配置名称、最大实例数、最小实例数、期望实例数等。重点配置数据说明如表1-1所示。

表 1-1 伸缩组参数说明

参数	解释	取值样例
区域	区域也称为Region。创建的伸缩组所在的区域。 不同区域的资源之间内网不互通。请选择靠近您客户的区域，可以降低网络时延、提高访问速度。	-
可用区	可用区也称为AZ（Availability Zone）。可用区指在同一区域下，电力、网络隔离的物理区域，可用区之间内网互通，不同可用区之间物理隔离。 <ul style="list-style-type: none"><li>如果您需要提高应用的高可用性，建议您将云服务器创建在不同的可用区内。</li><li>如果您需要较低的网络时延，建议您将云服务器创建在相同的可用区内。</li></ul>	-
多可用区扩展策略	可选择“均衡分布”或“选择优先”。 <ul style="list-style-type: none"><li>均衡分布：云服务器扩容时优先保证选择的可用区列表中各可用区下云服务器数量均衡，当无法在目标可用区下完成云服务器扩容时，按照选择优先原则选择其他可用区。</li><li>选择优先：云服务器扩容时目标可用区的选择按照选择的可用区列表的顺序进行优先级排序。</li></ul> <b>说明</b> 当选择两个及以上可用区时，才需要配置该选项。	均衡分布
名称	创建的伸缩组的名称。 伸缩组名称（1~64个字符）只能由中文、英文字母、数字、下划线、和中划线组成。	-
最大实例数	最大实例数是指伸缩组中云服务器个数的最大值。	1/0台
期望实例数	期望实例数是指伸缩组中期望的云服务器数量。 创建后可以手工修改该值，修改该值就会触发一次弹性伸缩活动。	0台
最小实例数	最小实例数是指伸缩组中云服务器个数的最小值。	0台
伸缩配置	为伸缩组选择所需的伸缩配置。伸缩配置用于定义伸缩组资源扩展时的云服务器的规格。包括云服务器的操作系统镜像、系统盘大小等。您需要在创建伸缩组之前创建好所需的伸缩配置。	-
虚拟私有云	弹性云服务器使用的网络是虚拟私有云（VPC）提供的。 同一伸缩组内的弹性云服务器均属于该VPC。	-

参数	解释	取值样例
子网	<p>您最多可以选择五个子网，伸缩组会自动为创建的实例绑定所有网卡。您选择的第一个子网默认作为云服务器的主网卡，其它子网作为云服务器的扩展网卡。</p> <p>自动分配IPv6地址：当且仅当选择支持IPv6的AZ、且VPC子网开启了IPv6功能时，该参数可见。子网如何开启IPv6功能，请参见“<a href="#">IPv4/IPv6双栈网络</a>”。</p> <p>系统默认分配IPv4地址，勾选后，网卡的IP地址为IPv6类型。在同一VPC内，云服务器通过IPv6地址在双栈ECS之间进行内网访问。如需访问外网，您需要开启“IPv6带宽”并选择共享带宽，此时云服务器可以使用IPv6地址访问互联网。</p>	-
源/目的检查	<p>用于控制系统是否检查弹性云服务器发送的报文中源IP地址的正确性。</p> <ul style="list-style-type: none"> <li>关闭“源/目的检查”，弹性云服务器可以处理非特定目的的网络流量。对于作为SNAT服务器或绑定了虚拟IP地址的弹性云服务器，应关闭源/目的检查。</li> <li>启用“源/目的检查”，系统会检查弹性云服务器发送的报文中源IP地址是否正确，否则不允许弹性云服务器发送该报文。这有助于防止伪装报文攻击，提升安全性。</li> </ul> <p>默认情况下，该功能处于启用状态。</p>	-
负载均衡	<p>可选参数。选择使用负载均衡器后，访问流量将自动分发到伸缩组内的所有弹性云服务器，扩展应用系统对外的服务能力，实现更高水平的应用程序容错性能。</p> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>一个伸缩组可最多添加6个负载均衡器。</li> <li>添加多个负载均衡器后，可同时监听多个业务，从而提高业务的可扩展性。同时，如果您选用“负载均衡健康检查”，弹性云服务器在任何一个监听器下的状态变为异常时，伸缩组会将该弹性云服务器替换掉。</li> </ul> <p>当使用弹性负载均衡时，需要配置以下四个参数：</p> <ul style="list-style-type: none"> <li>负载均衡器</li> <li>后端云服务器组</li> <li>后端端口：是指后端云服务器监听的端口。</li> <li>权重：是指后端云服务器经分发得到的请求数量的比例。</li> </ul> <p>更多关于弹性负载均衡的信息，请参见《<a href="#">弹性负载均衡用户指南</a>》。</p>	-

参数	解释	取值样例
实例移除策略	<p>实例优先被移除的策略。当满足条件时，会触发实例移除活动，包括如下四种方式：</p> <ul style="list-style-type: none"> <li>● 根据较早创建的配置较早创建的实例：根据“较早创建的配置”较早创建的“实例”优先被移除伸缩组。</li> <li>● 根据较早创建的配置较晚创建的实例：根据“较早创建的配置”较晚创建的“实例”优先被移除伸缩组。</li> <li>● 较早创建的实例：创建时间较早的实例优先被移除伸缩组。</li> <li>● 较晚创建的实例：创建时间较晚的实例优先被移除伸缩组。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>● 当可用区不均衡时，移出实例时会优先保证可用区均衡。</li> <li>● 手动移入伸缩组的云服务器不会遵循“实例移除策略”的要求，实例移除优先级最低，且移除时，系统不会删除该云服务器。当有多个手工加入伸缩组的云服务器时，移除规则是：先进先出。</li> </ul>	根据较早创建的配置较早创建的实例
弹性公网IP	若伸缩组的伸缩配置使用了弹性公网IP，在进行扩容活动时，会给创建出来的云服务器绑定一个弹性公网IP。若选择“释放”，当进行缩容活动时，会将云服务器上的弹性公网IP释放，否则仅做解绑定操作，保留弹性公网IP资源。	-
数据盘	若伸缩组的伸缩配置使用了数据盘，在进行扩容活动时，会给创建出来的云服务器挂载一个数据盘。若选择“释放”，则当云服务器从伸缩组中移除时，云服务器挂载的数据盘将被删除。否则，数据盘只会从云服务器上卸载。	-
健康检查方式	<p>健康检查会将异常的云服务器从伸缩组中移除，并重新创建新的云服务器，伸缩组的健康检查方式包括以下几种。</p> <ul style="list-style-type: none"> <li>● 云服务器健康检查：是指对云服务器的运行状态进行检查，如关机、删除都是云服务器异常状态。默认为此选项，伸缩组会定期使用云服务器健康检查结果来确定每个云服务器的运行状况。如果未通过云服务器健康检查，则伸缩组会将该云服务器移出伸缩组。</li> <li>● 负载均衡健康检查：是指根据负载均衡对云服务器的健康检查结果进行的检查。当伸缩组使用负载均衡器时，您还可以选择“负载均衡健康检查”方式来做健康检查，所有监听器下检测到的云服务器状态必须均为正常，否则伸缩组会将该弹性云服务器移出伸缩组。</li> </ul>	-



参数	解释	取值样例
健康检查间隔	伸缩组执行健康检查的周期。您可以根据实际情况设置合理的健康检查间隔（10秒、1分钟、5分钟、15分钟、1小时、3小时）。	5分钟
企业项目	选择伸缩组归属的企业项目。当伸缩组配置了企业项目时，由该伸缩组创建的弹性云服务器将归属于该企业项目。当没有指定企业项目时，将默认使用项目名称为default的企业项目。 <b>说明</b> <ul style="list-style-type: none"> <li>“default”为默认企业项目，帐号下原有资源和未选择企业项目的资源均在默认企业项目内。</li> <li>企业项目是统一身份认证项目的升级版，可针对不同项目间的资源进行分组和管理。</li> </ul>	-
标签	当存在相同类型的许多资源时，标签可以提供灵活的资源管理能力，用户可以根据分配给资源的标签快速识别特定资源。 每个标签均包含一个“键”和一个“值”，您可为每个标签指定键和值。 <ul style="list-style-type: none"> <li>键： <ul style="list-style-type: none"> <li>不能为空。</li> <li>对于同一伸缩组，“键”唯一。</li> <li>长度不超过36个字符。不能包含非打印字符ASCII(0-31), “=”, “*”, “&lt;”, “&gt;”, “\”, “,”, “ ”, “/”。</li> </ul> </li> <li>值： <ul style="list-style-type: none"> <li>可以为空字符串。</li> <li>一个“键”只能添加一个“值”。</li> <li>长度不超过43个字符。不能包含非打印字符ASCII(0-31), “=”, “*”, “&lt;”, “&gt;”, “\”, “,”, “ ”, “/”。</li> </ul> </li> </ul>	-

6. 参数配置完成后，单击“立即创建”。
7. （可选）添加伸缩策略。
  - a. 在伸缩组所在行的“操作”列下，单击“查看伸缩策略”。
  - b. （可选）您可以单击“添加伸缩策略”为新创建的伸缩组创建伸缩策略，更多关于伸缩策略的信息请参考[伸缩策略介绍](#)。

## 1.2 （可选）添加负载均衡器到伸缩组

弹性负载均衡（Elastic Load Balance，简称ELB）是将访问流量根据转发策略分发到后端多台云服务器流量分发控制服务。弹性负载均衡可以通过流量分发扩展应用系统对外的服务能力，通过消除单点故障提升应用系统的可用性。

若您需要使用弹性负载均衡提供的功能，请参考此章节为您的伸缩组添加负载均衡器。将负载均衡器添加到伸缩组后，可确保在伸缩组内添加和删除实例时，所有实例均可分配到应用程序的流量。

弹性伸缩只能添加已创建的负载均衡器，且弹性伸缩组和负载均衡器必须处于同一VPC内。如何创建负载均衡器请参见《[弹性负载均衡用户指南](#)》。为伸缩组添加负载均衡器方法如下：

- 在创建伸缩组时，可通过配置“负载均衡”参数添加负载均衡器。操作可参考[创建伸缩组](#)。
- 伸缩组没有正在进行的伸缩活动时，可以通过修改伸缩组的负载均衡配置，添加负载均衡器。操作可参考[修改伸缩组](#)。

## 1.3 更换伸缩组的伸缩配置

### 操作场景

当伸缩组中所需的弹性云服务器规格变更，需要为伸缩组更换伸缩配置时，可以参考此章节进行更换伸缩配置。

### 更换伸缩配置后的生效时间

若伸缩组正在进行伸缩活动，则当前伸缩活动中的实例配置以更换之前的伸缩配置为准；待下一次伸缩活动开始后，伸缩活动中的实例配置就会更改为更换后的伸缩配置。

例如：伸缩组当前的伸缩配置为as-config-A，更换后的伸缩配置为as-config-B，当伸缩组正在进行伸缩活动时，则当前伸缩活动中的实例配置仍然为as-config-A；

待下一次伸缩活动开始后，伸缩活动中的实例配置就会更改为as-config-B。

图 1-1 更换伸缩配置举例



名称	状态	规格	镜像	系统盘	数据盘(个)	登录方式	创建时间	计费模式	操作
as-config-B	未绑定	kc1.large.2   2vCPUs   4GB	makeImageEulerOS28	普通IO   50GB	0	密码	2020/09/24 10:00:46 ...	按量计费	复制 删除
as-config-A	未绑定	s2.small.1   1vCPUs   1GB	Public_CentOS_7.6_64bit_Uniagent	普通IO   40GB	0	密码	2020/09/24 09:59:18 ...	按量计费	复制 删除

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 单击需要更换伸缩配置的伸缩组名称，在“概览”页面的“配置名称”右方，单击“更换配置”。  
或在需要更换伸缩配置的伸缩组所在行的“操作”列下，选择“更多 > 更换伸缩配置”。
4. 在弹出的“更换伸缩配置”对话框中，重新为伸缩组选择伸缩配置。
5. 单击“确定”。

## 1.4 启用伸缩组

### 操作场景

当需要伸缩组实现自动创建或收缩实例时，您可以启用伸缩组。

启用伸缩组后，伸缩组的状态会变为“已启用”。只有状态为“已启用”的伸缩组，系统才会监控该伸缩组的伸缩策略，才可能触发伸缩活动。启用伸缩组后，当伸缩组内的当前云服务器数量小于或大于“期望实例数”时，系统自动添加或减少相应数量的云服务器，便会触发一次伸缩活动。

- 仅当伸缩组状态为“已停用”时，才可以启用伸缩组；
- 仅当伸缩组状态为“异常”时，可以选择“更多 > 强制启用”进行启用伸缩组（强制启用伸缩组，不会产生不良后果）；
- 在完成伸缩组和伸缩配置的创建后，伸缩组会自动启用。

### 启用伸缩组

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 在伸缩组列表中，伸缩组所在行的“操作”列下，单击“启用”。您也可以单击伸缩组名称，在页面右上角单击“启用”。
4. 在弹出“启用伸缩组”的对话框中，单击“是”。

### 强制启用伸缩组

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 在伸缩组列表中，伸缩组所在行的“操作”列下，选择“更多 > 强制启用”。您也可以单击伸缩组名称，在页面右上角单击“强制启用”。
4. 在弹出“强制启用伸缩组”的对话框中，单击“是”。

## 1.5 停用伸缩组

### 操作场景

当您需要对伸缩组中的实例进行关机配置或者升级时，为了避免健康检查将该实例删除，您可以先停用伸缩组，然后对实例进行操作，待实例状态恢复为运行中后再启用伸缩组。

当您伸缩组的伸缩活动一直失败重试时（比如创建云服务器失败，失败原因可以在云服务器页面查看），可通过以下两种方式停止失败重试。

- 先停用伸缩组，此时正在进行的伸缩活动失败后系统不会再立即进行重试，待环境恢复后或者更换伸缩配置后再启用伸缩组。
- 先停用伸缩组，修改期望实例数等于当前实例数，当本次伸缩活动失败结束后，不会再进行新的重试。

停用伸缩组后，伸缩组的状态会变为“已停用”，已停用状态的伸缩组，不会自动触发任何弹性伸缩活动。当伸缩组正在进行伸缩活动，即使停用，伸缩活动也不会立即停止。

当伸缩组状态是“已启用”或者“异常”时，可以停用伸缩组。

## 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 在伸缩组列表中，伸缩组所在行的“操作”列下，单击“停用”。您也可以单击伸缩组名称，在页面右上角单击“停用”。
4. 在弹出“停用伸缩组”的对话框中，单击“是”。

## 1.6 修改伸缩组

### 操作场景

在使用伸缩组的过程中，您可以根据需要修改伸缩组。伸缩组可以修改的参数有：名称、最大实例数、最小实例数、期望实例数、健康检查方式、健康检查间隔、实例移除策略。

#### 说明

当修改“期望实例数”时，会触发弹性伸缩活动。系统自动增加或减少实例以达到期望实例数。

伸缩组为非启用状态、实例数为0且没有正在进行的伸缩活动时，可以修改伸缩组的子网。伸缩组没有正在进行的伸缩活动时，可以修改可用区和负载均衡配置。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 在伸缩组列表中，在待修改的伸缩组所在行的“操作”列下，单击“更多 > 修改”。  
或单击待修改的伸缩组的名称，进入该伸缩组概览页面，单击页面右上方“修改”。
4. 在弹出的“修改伸缩组”对话框中，修改相关数据，例如修改期望实例数。
5. 修改完成后单击“确定”。

## 1.7 删除伸缩组

### 操作场景

当您不再需要某个伸缩组时，可以删除该伸缩组。

- 如果您仅在某段时间不需要启用伸缩组，建议您采用停用伸缩组的方式，而不建议删除。
- 当伸缩组存在云服务器实例或者有正在进行的伸缩活动时，如果您确定需要强制删除伸缩组并移出和释放ECS实例。系统首先会将该伸缩组置于“删除中”状态，

拒绝接收新的伸缩活动请求，然后等待已有的伸缩活动完成，最后将伸缩组内所有ECS实例移出伸缩组（用户手动添加的ECS实例会被移出伸缩组，弹性伸缩自动创建的ECS实例会被自动删除）并删除伸缩组。在上述过程中，您无法对“删除中”状态的伸缩组再进行其他操作。

- 删除伸缩组，包括删除相关伸缩策略以及在该伸缩组创建的告警策略产生的告警规则。

## 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 在伸缩组列表中，伸缩组所在行的“操作”列下，单击“更多 > 删除”。
4. 在弹出的“删除伸缩组”对话框中，单击“是”。

# 2 伸缩配置

## 2.1 创建伸缩配置

伸缩配置用于定义伸缩组资源扩展时的云服务器的规格。包括云服务器的操作系统镜像、系统盘大小等。

### 创建伸缩配置的两种方式

- 使用已有弹性云服务器快速创建伸缩配置  
当您已有云服务器时，您可以使用已有的弹性云服务器快速创建伸缩配置，创建配置时，vCPU、内存、镜像、磁盘和云服务器类型参数信息将默认与选择的云服务器规格保持一致，详细内容请参考[使用已有云服务器创建伸缩配置](#)。
- 使用新模板创建一个全新的伸缩配置  
若您对扩展的云服务器的规格有特殊的要求，可通过使用新模板创建伸缩配置，可按照您的需求配置新模板的规格参数，详细内容请参考[使用新模板创建伸缩配置](#)。

## 2.2 使用已有云服务器创建伸缩配置

### 操作场景

您可以使用已有的弹性云服务器快速创建伸缩配置。此时，伸缩配置中的云服务器类型、vCPU、内存、镜像、磁盘参数信息将默认与选择的云服务器规格保持一致。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例”。
3. 单击“创建伸缩配置”。
4. 填写弹性伸缩配置信息，例如，名称、配置模板等。配置数据说明如[表2-1](#)所示。

表 2-1 伸缩配置数据说明

参数	解释	取值样例
区域	区域也称为Region。创建的伸缩配置所在的区域。	-
名称	创建伸缩配置的名称。	-
配置模板	选择“使用已有云服务器规格为模板 > 请选择云服务器”。 创建配置时，云服务器类型、vCPU、内存、镜像、磁盘参数信息将默认与选择的云服务器规格保持一致。若选择的云服务器是创建在专属主机上的，云服务器的类型、vCPU、内存、镜像、磁盘参数信息也默认与选择的云服务器规格保持一致。	使用已有云服务器规格为模板
弹性公网IP	弹性公网IP是指将公网IP地址和路由网络中关联的弹性云服务器绑定，以实现虚拟私有云内的弹性云服务器通过固定的公网IP地址对外提供访问服务。 您可以根据实际情况选择以下两种方式： <ul style="list-style-type: none"> <li>不使用： 弹性云服务器不能与互联网互通，仅可作为私有网络中部署业务或者集群所需弹性云服务器进行使用。</li> <li>自动分配： 自动为每台弹性云服务器分配独享带宽的弹性公网IP，您可以在“带宽”中设置带宽值。</li> </ul> <b>说明</b> 当您选择自动分配时，需要配置线路、计费方式和带宽等参数。	自动分配
登录方式	云平台提供两种弹性云服务器鉴权方式。 <ul style="list-style-type: none"> <li>密钥对 指使用密钥作为弹性云服务器的鉴权方式。如果选择此方式，请在密钥页面先创建或导入密钥对。 <b>说明</b> 如果您直接从下拉列表中选择已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录弹性云服务器。</li> <li>密码 指使用设置root用户（Linux操作系统）和Administrator用户（Windows操作系统）的初始密码方式作为弹性云服务器的鉴权方式，如果选择此方式，您可以通过用户名密码方式登录弹性云服务器。</li> </ul>	Admin@123

参数	解释	取值样例
高级配置	高级配置可对实例自定义数据注入和云服务器组进行配置。 可选择“暂不配置”和“现在配置”。	-
实例自定义数据注入	可选配置，主要用于创建弹性云服务器时向弹性云服务器注入用户数据。配置实例自定义数据注入后，弹性云服务器首次启动时会自行注入数据信息。 详细内容请参见《 <a href="#">弹性云服务器用户指南</a> 》。 主要分为以下两种： <ul style="list-style-type: none"><li>以文本形式：在下方文本框内输入用户数据内容；</li><li>以文件形式：主要用于创建云主机时向弹性云服务器注入脚本文件或其他文件。</li></ul> <b>说明</b> <ul style="list-style-type: none"><li>对于Linux弹性云服务器，如果选择密码方式登录鉴权，此时不能使用实例自定义数据注入功能。</li><li>如果选择的镜像不支持实例自定义数据注入，则不能使用实例自定义数据注入功能。</li></ul>	-
云服务器组	通过云服务器组功能，弹性云服务器在创建时，将尽量分散地创建在不同的主机上，提高业务的可靠性。	-

5. 参数配置完成后，单击“立即创建”。
6. 如果您需要立即使用新创建的伸缩配置，则需要将伸缩配置添加到伸缩组，请参考[更换伸缩组的伸缩配置](#)。

## 2.3 使用新模板创建伸缩配置

### 操作场景

若您对扩展的云服务器的规格有特殊的要求，可通过使用新模板创建伸缩配置，可按照您的需求配置新模板的规格参数，使得伸缩组内云服务器的规格均符合创建新模板的规格。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例”。
3. 单击“创建伸缩配置”。
4. 填写弹性伸缩配置信息，例如，名称、配置模板、云服务器的镜像、云服务器类型等。配置数据说明如[表2-2](#)所示。



表 2-2 伸缩配置数据说明

参数	解释	取值样例
计费模式	伸缩配置所选实例规格的计费模式。 按需计费：后付费模式，按弹性云服务器的实际使用时长计费，可以随时创建/删除弹性云服务器。	按需计费
区域	区域也称为Region。创建的伸缩配置所在的区域。	-
名称	创建伸缩配置的名称。	-
配置模板	选择“使用新模板”。 重新选择云服务器类型、vCPU、内存、镜像、磁盘等参数信息，创建新的弹性伸缩配置。	使用新模板
CPU架构	CPU架构分为以下两种： <ul style="list-style-type: none"> <li>• x86计算：x86采用复杂指令集（CISC）；</li> <li>• 鲲鹏计算：鲲鹏采用精简指令集（RISC）。</li> </ul> <b>说明</b> 当前Region下存在两种CPU架构的弹性云服务器时，才显示该项。	x86计算
专属主机	可选配置，伸缩组自动创建的实例将被创建在专属主机上。 可选择自动放置或者选择指定专属主机。 <ul style="list-style-type: none"> <li>• 选择自动放置时，系统会自动将实例创建在符合规格的专属主机中剩余内存最多的那台上，使各专属主机尽量均衡负载。</li> <li>• 选择指定的专属主机时，系统会自动将实例创建在该专属主机上。</li> </ul> <b>说明</b> 当前用户下存在专属主机时，才显示该项。	自动放置
规格	公有云提供了多种类型的弹性云服务器供您选择，针对不同的应用场景，可以选择不同规格的弹性云服务器。 更多内容，请参见《 <a href="#">弹性云服务器用户指南</a> 》。	内存优化型
镜像	<ul style="list-style-type: none"> <li>• 公共镜像 常见的标准操作系统镜像，所有用户可见，包括操作系统以及预装的公共应用。请根据您的实际情况自助配置应用环境或相关软件。</li> <li>• 私有镜像 用户基于弹性云服务器创建的个人镜像，仅用户自己可见。包含操作系统、预装的公共应用以及用户的私有应用。选择私有镜像创建弹性云服务器，可以节省您重复配置弹性云服务器的时间。</li> <li>• 共享镜像 用户将接受公有云其他用户共享的私有镜像，作为自己的镜像进行使用。</li> </ul>	公共镜像

参数	解释	取值样例
磁盘	<p>包括系统盘和数据盘。</p> <ul style="list-style-type: none"> <li>系统盘           <p>普通IO：是指由SATA存储提供资源的磁盘类型。</p> <p>高IO：是指由SAS存储提供资源的磁盘类型。</p> <p>通用型SSD：是指由GPSSD存储提供资源的磁盘类型。</p> <p>超高IO：是指由SSD存储提供资源的磁盘类型。</p> <p>如果镜像为整机镜像，系统盘会通过磁盘备份进行恢复，界面上只能修改卷类型及卷大小，并且卷大小不得小于磁盘备份大小。</p> </li> <li>数据盘           <p>您可以为云服务器添加多块数据盘，或者从指定的数据盘镜像导出数据到某个数据盘。</p> <p>当您选择的云服务器镜像为整机镜像时，通过整机镜像中磁盘备份恢复的数据盘，可以修改卷类型、卷大小以及加密属性，其中卷大小不得小于磁盘备份大小，加密属性只有当整机镜像中磁盘备份为region内可用时才可以修改。</p> </li> </ul>	“系统盘”选为“普通IO”
安全组	<p>安全组是一个逻辑上的分组，用来实现安全组内和组间弹性云服务器的访问控制，加强弹性云服务器的安全保护。用户可以在安全组中定义各种访问规则，当弹性云服务器加入该安全组后，即受到这些访问规则的保护。</p>	-
弹性公网IP	<p>弹性公网IP是指将公网IP地址和路由网络中关联的弹性云服务器绑定，以实现虚拟私有云内的弹性云服务器通过固定的公网IP地址对外提供访问服务。</p> <p>您可以根据实际情况选择以下两种方式：</p> <ul style="list-style-type: none"> <li>不使用：弹性云服务器不能与互联网互通，仅可作为私有网络中部署业务或者集群所需弹性云服务器进行使用。</li> <li>自动分配：自动为每台弹性云服务器分配独享带宽的弹性公网IP，带宽值可以由您设定。</li> </ul> <p><b>说明</b> 当您选择自动分配时，需要配置规格、带宽类型和带宽等参数。</p>	自动分配
带宽	<p>带宽大小，单位Mbit/s。</p> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>该参数仅在“弹性公网IP”选择“自动分配”时显示。</li> <li>独享带宽支持选择计费方式，您可以选择按带宽计费或按流量计费。</li> <li>共享带宽仅支持按带宽计费，您可以选择想要加入的共享带宽名称。</li> </ul>	100

参数	解释	取值样例
登录方式	<p>云平台提供两种弹性云服务器鉴权方式。</p> <ul style="list-style-type: none"> <li>● 密钥对 指使用密钥作为弹性云服务器的鉴权方式。如果选择此方式，请在密钥对页面先创建或导入密钥对。</li> </ul> <p><b>说明</b> 如果您直接从下拉列表中选择已有的密钥，请确保您已在本地获取该文件，否则，将影响您正常登录弹性云服务器。</p> <ul style="list-style-type: none"> <li>● 密码 指使用设置root用户（Linux操作系统）和Administrator用户（Windows操作系统）的初始密码方式作为弹性云服务器的鉴权方式，如果选择此方式，您可以通过用户名密码方式登录弹性云服务器。</li> </ul>	Admin@123
高级配置	<p>高级配置可对实例自定义数据注入和云服务器组进行配置。</p> <p>可选择“暂不配置”和“现在配置”。</p>	-
实例自定义数据注入	<p>可选配置，主要用于创建弹性云服务器时向弹性云服务器注入用户数据。配置实例自定义数据注入后，弹性云服务器首次启动时会自行注入数据信息。</p> <p>详细内容请参见《<a href="#">弹性云服务器用户指南</a>》。</p> <p>主要分为以下几种：</p> <ul style="list-style-type: none"> <li>● 以文本形式：在下方文本框内输入用户数据内容；</li> <li>● 以文件形式：主要用于创建云主机时向云主机注入脚本文件或其他文件。</li> </ul> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>● 对于Linux弹性云服务器，如果选择密码方式登录鉴权，此时不能使用实例自定义数据注入功能。</li> <li>● 如果选择的镜像不支持实例自定义数据注入，则不能使用实例自定义数据注入功能。</li> </ul>	-
云服务器组	<p>通过云服务器组功能，弹性云服务器在创建时，将尽量分散地创建在不同的主机上，提高业务的可靠性。</p>	-

5. 单击“立即创建”。系统会提示伸缩配置创建成功，并根据界面提示返回伸缩配置页面。
6. 如果您需要立即使用新创建的伸缩配置，则需要将伸缩配置添加到伸缩组，请参考[创建伸缩组](#)或[更换伸缩组的伸缩配置](#)。

## 多规格配置说明

伸缩配置支持多规格可最大程度减少因当前规格售罄或不可用导致的扩容失败概率，尽可能确保在用户业务高峰期能够扩容成功。

适用场景：

- 对于伸缩组创建的实例规格无特殊要求，不限定于某一种规格。
- 对伸缩组成功创建实例有较高的成功率和较低的时延要求。
- 业务需要使用较高规格的实例。
- 业务是无状态、可横向扩展的。

伸缩组对多个规格的选择排序有以下两种方式：

- 选择优先：伸缩组扩容时按照选择规格的顺序进行。当您选择的第一个规格库存不足或者因为其他原因创建失败时，系统会尝试您选择的第二个规格创建实例，其余规格的使用以此类推。
- 成本优先：伸缩组扩容时按照价格最优原则进行优先级排序。伸缩组创建实例时将选择成本最优的规格，当成本最优的规格无法成功创建实例时，系统将在剩余规格中选择成本最优的规格创建实例，其余规格的使用以此类推。

## 2.4 复制伸缩配置

### 操作场景

在原有伸缩配置基础上，复制一条伸缩配置。

复制伸缩配置时，用户可在原有伸缩配置基础上，通过更改伸缩配置名称、云服务器规格和镜像等参数，快速创建一个新的伸缩配置。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例”。
3. 选择“伸缩配置”页签，在需要复制的伸缩配置所在行的“操作”列下，单击“复制”。
4. 在弹出的“复制伸缩配置”界面，修改伸缩配置名称、云服务器规格和镜像等参数，并设置弹性云服务器的登录方式等。
5. 单击“确定”，完成复制伸缩配置操作。

## 2.5 删除伸缩配置

### 操作场景

当您不再使用某个伸缩配置时，可以删除该伸缩配置。只有当伸缩配置不被任何弹性伸缩组使用时，才允许被删除。删除伸缩配置时，可单个删除也可批量删除。

### 操作步骤

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例”。
3. 选择“伸缩配置”页签，选择需要删除的伸缩配置所在行“操作”列下，单击“删除”，或勾选多个伸缩配置，并单击列表上方的“删除”，进行批量删除。

# 3 伸缩策略

## 3.1 伸缩策略介绍

伸缩策略可以触发伸缩活动，是对伸缩组中实例数量或带宽进行调整的一种方式。伸缩策略规定了伸缩活动触发需要满足的条件及需要执行的操作，当满足伸缩条件时，系统会自动触发一次伸缩活动。

### 说明

当多个伸缩策略应用于同一个伸缩组时，在伸缩策略不冲突的前提下，只要满足相应的伸缩策略条件，均会触发伸缩活动。

### 目前系统支持的 3 种伸缩策略

- 告警策略：基于云监控系统告警数据（例如CPU使用率），自动增加、减少或设置指定数量的云服务器。
- 定时策略：基于配置的某个时间点，自动增加、减少或设置指定数量的云服务器。
- 周期策略：按照配置周期（按天、按周、按月），周期性地增加、减少或设置指定数量的云服务器。

### 目前系统支持的三种资源调整模式

- 动态模式  
动态模式使用告警策略调整实例数量或带宽大小。  
当业务负载难以预测时，选择告警策略，系统会根据实时的监控数据（如CPU使用率）触发伸缩活动，动态调整伸缩组内的实例数量或带宽大小。
- 按计划模式  
按计划模式使用定时或周期策略调整实例数量或带宽大小。  
当业务负载的变化有规律时，可以使用定时策略或周期策略调整伸缩组内的实例数量或带宽大小。
- 手动模式  
通过手动将实例移入到伸缩组、手动将实例移出伸缩组或手动修改期望实例数，扩展资源。

## 3.2 创建伸缩策略

### 操作场景

用户可以通过伸缩策略对伸缩组中的实例进行管理。本章节介绍如何创建伸缩策略。

### 创建告警策略

1. 登录管理控制台。
1. 选择“计算 > 弹性伸缩 > 伸缩实例”。
2. 在伸缩组所在行的“操作”列下，单击“查看伸缩策略”。
3. 在“伸缩策略”页签，单击“添加伸缩策略”。
4. 根据界面进行参数配置，可参考[表3-1](#)。

表 3-1 告警策略参数配置

参数名称	参数说明	取值样例
策略名称	创建伸缩策略的名称。	as-policy-p6g5
策略类型	选择“告警策略”。	告警策略
告警规则	可选择“现在创建”或“使用已有”。 已有告警规则的设置请参见 <a href="#">设置监控告警规则</a> 。 选择新建告警时，支持系统监控和自定义监控。 <ul style="list-style-type: none"><li>● 系统监控需配置<a href="#">表3-2</a>所示参数。</li><li>● 自定义监控需配置<a href="#">表3-3</a>所示参数。</li></ul>	-

参数名称	参数说明	取值样例
执行动作	<p>设置伸缩活动执行动作及实例的个数或实例百分比。</p> <p>执行动作包括：</p> <ul style="list-style-type: none"> <li>● 增加 当执行伸缩活动时，向伸缩组增加实例。</li> <li>● 减少 当执行伸缩活动时，从伸缩组中减少实例。</li> <li>● 设置为 将伸缩组中的期望实例数设置为固定值。</li> </ul>	<ul style="list-style-type: none"> <li>● 增加1个实例</li> <li>● 增加10%的实例 增加10%的实例，即增加的实例个数是该伸缩组当前实例个数的10%。若伸缩组当前实例个数与实例百分比的乘积是非整数，则系统会自动按照如下规则进行舍入： <ul style="list-style-type: none"> <li>- 大于1的值向下取整。例如，12.7取整为12。</li> <li>- 大于0且小于1的值取整为1。例如，0.67取整为1。</li> </ul> </li> </ul> <p>例如，某伸缩组当前有10个实例，有一个执行动作为“增加15%的实例”的伸缩策略。当该策略执行时，系统会按照规则将1.5向下取整为1。因此，此次伸缩活动结束后伸缩组的当前实例个数为11。</p>

参数名称	参数说明	取值样例
冷却时间	<p>为了避免告警策略频繁触发，必须设置冷却时间。</p> <p>冷却时间是指冷却伸缩活动的时间，在每次伸缩活动完成之后，系统开始计算冷却时间。</p> <p>伸缩组在冷却时间内，会拒绝由告警策略触发的伸缩活动，其他类型的伸缩策略（如定时策略和周期策略）触发的伸缩活动不受限制，但会重新开始计算冷却时间，单位为秒。</p> <p>例如：冷却时间设置为300秒，定时策略设置了10:32进行伸缩活动，10:30告警触发的伸缩活动结束，则在10:30-10:35时间内，伸缩组会拒绝新告警触发的伸缩活动，但不会拒绝在10:32时定时策略触发的伸缩活动；若10:36定时策略触发的伸缩活动结束，则冷却时间为10:36-10:41。</p> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>如果伸缩活动是伸缩策略触发的，以伸缩策略的冷却时间为准。</li> <li>如果是手工修改期望实例数量或者其他方式引起的伸缩活动，则以伸缩组的冷却时间为准，默认为300秒。</li> </ul>	300秒

表 3-2 系统监控参数

参数名称	参数说明	取值样例
告警规则名称	新建告警规则的名称。	as-alarm-7o1u
监控类型	定义监控指标的类型，是系统支持的或是自定义的。选择“系统监控”。	系统监控
触发条件	选择弹性伸缩支持的监控指标并对监控指标设定告警条件。	CPU使用率最大值 >70%
监控周期	告警规则刷新告警状态的周期。	5分钟
连续出现次数	触发告警时的采样点数目。例如：连续出现次数配置为n，则告警规则的采样点是连续n个监控周期的采样点，当这些采样点全部满足触发条件后，告警规则的状态变为告警状态，从而触发伸缩活动。	3次



表 3-3 自定义监控参数

参数名称	参数说明	取值样例
告警规则名称	新建告警规则的名称。	as-alarm-7o1u
监控类型	选择自定义监控。自定义监控可以自行设置，可以满足您多种场景下的对监控指标的需求。	自定义监控
资源类型	配置告警规则监控的服务名称。	AGT.ECS
维度	用于指定告警规则对应指标的维度名称。	instance_id
监控对象	用来配置该告警规则针对的具体资源。	-
触发条件	选择弹性伸缩支持的监控指标并对监控指标设定告警条件。	CPU使用率最大值 >70%
监控周期	告警规则刷新告警状态的周期。	5分钟
连续出现次数	触发告警时的采样点数目。例如：连续出现次数配置为n，则告警规则的采样点是连续n个监控周期的采样点，当这些采样点全部满足触发条件后，告警规则的状态变为告警状态，从而触发伸缩活动。	3次

5. 单击“确定”。

在“伸缩策略”页签中可查看新创建的伸缩策略，新创建的伸缩策略默认的状态为“已启用”。

## 创建定时/周期策略

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例”。
3. 在伸缩组所在行的“操作”列下，单击“查看伸缩策略”。
4. 在“伸缩策略”页签，单击“添加伸缩策略”。
5. 根据界面进行参数配置，定时策略或周期策略可参见表3-4进行参数配置。

表 3-4 参数配置

参数名称	参数说明	取值样例
策略名称	创建伸缩策略的名称。	as-policy-p6g5

参数名称	参数说明	取值样例
策略类型	<p>计划扩展资源的策略类型可选择定时策略和周期策略，在指定的时间段进行扩展资源。</p> <p>若选择周期策略除了配置表格中的参数外，还需配置以下两个参数：</p> <ul style="list-style-type: none"> <li>● 重复周期 <ul style="list-style-type: none"> <li>- 按天</li> <li>- 按周</li> <li>- 按月</li> </ul> </li> <li>● 生效时间 伸缩策略触发的时间段。</li> </ul>	-
时区	<p>为默认值：GMT+08:00 代表格林尼治标准时间加8小时，即北京时间。</p>	GMT+08:00
触发时间	设定伸缩策略触发时间。	-
执行动作	<p>设置伸缩活动执行动作及实例的个数。</p> <p>执行动作包括：</p> <ul style="list-style-type: none"> <li>● 增加 当执行伸缩活动时，向伸缩组增加实例。</li> <li>● 减少 当执行伸缩活动时，从伸缩组中减少实例。</li> <li>● 设置为 将伸缩组中的期望实例数设置为固定值。</li> </ul>	<ul style="list-style-type: none"> <li>● 增加1个实例</li> <li>● 增加10%的实例 增加10%的实例，即增加的实例个数是该伸缩组当前实例个数的10%。若伸缩组当前实例个数与实例百分比的乘积是非整数，则系统会自动按照如下规则进行舍入： <ul style="list-style-type: none"> <li>- 大于1的值向下取整。例如，12.7取整为12。</li> <li>- 大于0且小于1的值取整为1。例如，0.67取整为1。</li> </ul> </li> </ul> <p>例如，某伸缩组当前有10个实例，有一个执行动作为“增加15%的实例”的伸缩策略。当该策略执行时，系统会按照规则将1.5向下取整为1。因此，此次伸缩活动结束后伸缩组的当前实例个数为11。</p>

参数名称	参数说明	取值样例
冷却时间	<p>为了避免告警策略频繁触发，必须设置冷却时间。</p> <p>冷却时间是指冷却伸缩活动的时间，在每次伸缩活动完成之后，系统开始计算冷却时间。</p> <p>伸缩组在冷却时间内，会拒绝由告警策略触发的伸缩活动，其他类型的伸缩策略（如定时策略和周期策略）触发的伸缩活动不受限制，但会重新开始计算冷却时间，单位为秒。</p> <p>例如：冷却时间设置为300秒，定时策略设置了10:32进行伸缩活动，10:30告警触发的伸缩活动结束，则在10:30-10:35时间内，伸缩组会拒绝新告警触发的伸缩活动，但不会拒绝在10:32时定时策略触发的伸缩活动；若10:36定时策略触发的伸缩活动结束，则冷却时间为10:36-10:41。</p> <p><b>说明</b></p> <ul style="list-style-type: none"><li>• 如果伸缩活动是伸缩策略触发的，以伸缩策略的冷却时间为准。</li><li>• 如果是手工修改期望实例数量或者其他方式引起的伸缩活动，则以伸缩组的冷却时间为准，默认为300秒。</li></ul>	300秒

#### 6. 单击“确定”。

在“伸缩策略”页签中可查看新创建的伸缩策略，新创建的伸缩策略默认的状态为“已启用”。

#### 说明

如果创建了同一时间触发的定时或周期策略，当达到触发时间时，系统会选取创建时间最晚的策略执行。告警策略不受该限制约束。

## 3.3 管理伸缩策略

### 操作场景

伸缩策略规定了触发伸缩活动的条件和执行的动作，当满足条件时，会触发一次伸缩活动。

本章节介绍对伸缩策略的管理，包括修改伸缩策略、启用伸缩策略、停用伸缩策略、立即执行伸缩策略和删除伸缩策略。

## 修改伸缩策略

当现有伸缩策略已不能满足现有业务需求时，可通过修改伸缩策略的基本参数，满足业务需求。

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 在伸缩组所在行的“操作”列下，单击“查看伸缩策略”，在伸缩策略所在行的“操作”列下，单击“更多 > 修改”。
4. 在弹出的“修改伸缩策略”对话框中，修改相应参数，完成后单击“确定”。

## 启用伸缩策略

只有当伸缩策略和伸缩组均处于启用状态时，伸缩策略才能触发伸缩活动。伸缩组可以启用一个，也可以启用多个伸缩策略。

- 启用多个策略时，需要您保证多个伸缩策略的条件不冲突。
- 仅当伸缩策略状态为“已停用”时，才可以启用伸缩策略。

在伸缩组所在行的“操作”列下，单击“查看伸缩策略”，在伸缩策略所在行的“操作”列下，单击“启用”。若需同时启用多个伸缩策略，勾选需要启用的伸缩策略后，单击伸缩策略列表上方的“启用”。

## 停用伸缩策略

如果不希望某个伸缩策略在某个时间段触发伸缩活动，可选择停用指定伸缩策略。

- 如果停用所有的伸缩策略，将不会触发任何由伸缩策略触发的伸缩活动。但手工修改期望实例数时，同样会触发伸缩活动。
- 仅当伸缩策略状态为“已启用”时，才可以停用伸缩策略。

在伸缩组所在行的“操作”列下，单击“查看伸缩策略”，在伸缩策略所在行的“操作”列下，单击“停用”。若需同时停用多个伸缩策略，勾选需要停用的伸缩策略后，单击伸缩策略列表上方的“停用”。

## 立即执行伸缩策略

为了使伸缩组当前实例数立即达到期望实例数。

- 立即执行伸缩策略与伸缩条件是否满足没有关系。
- 仅当伸缩组状态和伸缩策略状态都为“已启用”时，才可以立即执行伸缩策略：  
在伸缩组所在行的“操作”列下，单击“查看伸缩策略”，在伸缩策略所在行的“操作”列下，单击“立即执行”。

## 删除伸缩策略

某个伸缩策略不再用于触发伸缩活动时，可选择删除该伸缩策略。

如果被删除的策略已经被触发，伸缩活动正在进行，也可以删除该策略，伸缩活动不受影响。

在伸缩组所在行的“操作”列下，单击“查看伸缩策略”，在伸缩策略所在行的“操作”列下，单击“更多 > 删除”。

若需同时删除多个伸缩策略，勾选需要删除的伸缩策略后，单击伸缩策略列表上方的“删除”。

# 4 伸缩活动

## 4.1 动态扩展资源

弹性伸缩进行伸缩活动时，需定义如何按照不断变化的需求进行伸缩活动，即动态扩展资源。

当业务需求变化频繁且无固定规律时，可通过配置告警策略实现动态扩缩资源的目的。当满足伸缩策略的条件时，系统自动修改期望实例数，从而触发伸缩活动进行资源的扩张或收缩。如何创建告警策略请参考[创建伸缩策略](#)进行操作。

例如，一个支持用户进行购买火车票的Web应用程序，当运行该应用程序的实例的CPU使用率上升到90%时，需要增加一个实例，以确保业务正常运行，在CPU使用率下降到30%时需删除一个实例，以减少资源的浪费。根据以上情况，可以配置两条告警策略，第一条告警策略设置触发条件为：CPU使用率最大值大于90%，执行动作为：增加一个实例。可参考[图 告警策略01](#)进行配置。第二条告警策略设置触发条件为：CPU使用率最小值小于30%，执行动作为：减少一个实例。可参考[图 告警策略02](#)进行配置。

图 4-1 告警策略 01

### 添加伸缩策略

策略名称	<input type="text" value="as-policy-001"/>
策略类型	<input checked="" type="radio"/> 告警策略 <input type="radio"/> 定时策略 <input type="radio"/> 周期策略
告警规则	<input checked="" type="radio"/> 现在创建 <input type="radio"/> 使用已有
告警规则名称	<input type="text" value="as-alarm-cpu01"/>
监控类型	<input checked="" type="radio"/> 系统监控 <input type="radio"/> 自定义监控
触发条件	<input type="text" value="CPU使用率"/> <input type="text" value="最大值"/> <input type="text" value="&gt;"/> <input type="text" value="90"/> %
	<small>不同操作系统的监控指标有所不同。<a href="#">了解更多</a> 如要使用Agent监控指标，请确认伸缩组中实例均已安装了Agent插件。<a href="#">如何安装插件</a></small>
监控周期	<input type="text" value="5分钟"/>
连续出现次数	<input type="text" value="3"/> <input data-bbox="900 913 922 945" type="button" value="?"/>
执行动作	<input type="text" value="增加"/> <input type="text" value="1"/> <input type="text" value="个实例"/>
冷却时间(秒)	<input type="text" value="900"/> <input data-bbox="900 1041 922 1072" type="button" value="?"/>
	<input type="button" value="确定"/> <input type="button" value="取消"/>

图 4-2 告警策略 02

**添加伸缩策略**

策略名称	<input type="text" value="as-policy-002"/>
策略类型	<input checked="" type="radio"/> 告警策略 <input type="radio"/> 定时策略 <input type="radio"/> 周期策略
告警规则	<input checked="" type="radio"/> 现在创建 <input type="radio"/> 使用已有
告警规则名称	<input type="text" value="as-alarm-cpu02"/>
监控类型	<input checked="" type="radio"/> 系统监控 <input type="radio"/> 自定义监控
触发条件	<input type="text" value="CPU使用率"/> <input type="text" value="最小值"/> <input type="text" value="&lt;"/> <input type="text" value="30"/> %
<small>不同操作系统的监控指标有所不同。<a href="#">了解更多</a> 如要使用Agent监控指标，请确认伸缩组中实例均已安装了Agent插件。<a href="#">如何安装插件</a></small>	
监控周期	<input type="text" value="5分钟"/>
连续出现次数	<input type="text" value="3"/> <input type="button" value="?"/>
执行动作	<input type="text" value="减少"/> <input type="text" value="1"/> <input type="text" value="个实例"/>
冷却时间(秒)	<input type="text" value="900"/> <input type="button" value="?"/>

## 4.2 按计划扩展资源

弹性伸缩进行伸缩活动时，应对需求有规律变化的场景，需按照规律定期或者周期性的进行伸缩活动，可通过配置定时策略和周期策略来调整资源。如何创建定时或周期策略可参考[创建伸缩策略](#)。

例如，假如有一个Web应用程序，该应用程序支持学生选择选修课程，在每学期开始时，该应用程序的使用率显著提高，但在每学期其余时间该应用程序使用率较低。则可以在每学期的开始时分别配置两条定时策略，第一条定时策略的执行动作设置为：增加一个实例，第二条定时策略的执行动作设置为：减少一个实例。触发时间分别选择选课开始时段和选课结束时间。弹性伸缩便会在设定的时间（即选课开始时间）增加一个实例，在选课结束时减少一个实例，满足学生的使用需求，同时节约了成本。

## 4.3 手动扩展资源

### 操作场景

通过手动将实例移入到伸缩组、手动将实例移出伸缩组或手动修改期望实例数，扩展资源。

### 操作步骤

#### 将实例移入伸缩组



伸缩组没有正在进行的伸缩活动、处于“已启用”状态，且当前实例数小于最大实例数时，您可以手动将云服务器移入指定伸缩组。

将云服务器成功移入指定伸缩组必须满足如下条件：

- 云服务器不能存在于其它伸缩组中。
- 待移入的云服务器所在的VPC必须和伸缩组所在的VPC相同。
- 云服务器所在的可用区必须包含于伸缩组所配置的可用区内。
- 批量添加后实例数后的总实例数不能大于伸缩组的最大实例数。
- 单次最多批量操作实例个数为10个。

将云服务器移入伸缩组的步骤如下。

1. 选择“计算 > 弹性伸缩 > 伸缩实例”。
2. 单击具体的伸缩组名称。
3. 在“伸缩实例”页签，单击“移入伸缩组”。
4. 选择待移入的实例名称，单击“确定”。

### 将实例移出伸缩组

您可以将实例移出伸缩组，更新实例或排查实例的问题，然后将实例重新移入伸缩组。移出伸缩组后实例不再处理应用程序流量。

例如，您可以随时更改伸缩组使用的伸缩配置，后续实例将使用此配置。不过，伸缩组不会更新当前正在运行的实例。您可以终止这些实例并让伸缩组替换这些实例，也可以先将实例移出伸缩组，更新软件，然后将该实例移入伸缩组。

云服务器移出伸缩组的相关约束。

- 伸缩组没有正在进行的伸缩活动，伸缩实例状态为“已启用”，且批量移出后实例数不能小于伸缩组的最小实例数时，您可以手动将云服务器移出伸缩组。
- 伸缩组没有进行伸缩活动，选择的实例是自动伸缩，生命周期状态是已启用并且不在存储容灾服务中使用，才可以对实例进行移出伸缩组并删除操作。
- 自动添加的云服务器默认计费方式是按需计费，当您选择“移出伸缩组”时，系统仅将其移出伸缩组，云服务器实例可以作为其他用途，当您选择“移出伸缩组并删除”时，系统将云服务器移出伸缩组时，也会将其删除。
- 如果您手动将伸缩组中云服务器的计费方式修改为包年包月，那么当您选择“移出伸缩组并删除”时，系统仅将其移出伸缩组，不进行删除操作。
- 对于手动移入伸缩组的实例，只能进行移出伸缩组操作，不能进行移出伸缩组并删除操作。
- 单次最多批量操作实例个数为10个。

将云服务器移出伸缩组的步骤如下。

1. 选择“计算 > 弹性伸缩 > 伸缩实例”。
2. 单击具体的伸缩组名称。
3. 在“伸缩实例”页签中的实例所在行的“操作”列下，单击“移出伸缩组”或“移出伸缩组并删除”。
4. 如果您要删除多个实例，可以勾选多个实例，单击“移出伸缩组”或“移出伸缩组并删除”。

如果您要删除所有实例，可以勾选参数“实例名称”左侧的方框，单击“移出伸缩组”或“移出伸缩组并删除”。

### 修改期望实例数

手工修改期望实例数可以增加或减少伸缩组中的实例，实现资源的扩展。

请参见[修改伸缩组](#)章节，进行修改期望实例数。

## 4.4 实例移除策略

当您的伸缩组自动移除实例时，如果伸缩组内存在不属于当前配置的可用区的实例，移除实例时，会优先移除这些实例。其次，会评估伸缩组当前配置的可用区是否存在不平衡。如果某个可用区的实例数多于其他可用区，移除实例时会优先保证可用区均衡。如果该组使用的可用区是平衡的，则实例会按照您配置的实例移除策略被移除。

弹性伸缩目前支持的实例移除策略，包括如下四种方式：

- 较早创建的实例：创建时间较早的实例优先被移除伸缩组。当您将伸缩组中的实例升级为新的实例类型，可以逐渐将较旧类型的实例替换为较新类型的实例时，此策略非常有用。
- 较晚创建的实例：创建时间较晚的实例优先被移除伸缩组。如果要测试新的伸缩配置但不想在生产中保留它时，此策略非常有用。
- 较早创建的配置中较早创建的实例：较早创建的配置中较早创建的“实例”优先被移除伸缩组。如果要更新某个组并且逐步淘汰先前配置中的实例时，此策略非常有用。
- 较早创建的配置中较晚创建的实例：较早创建的配置中较晚创建的“实例”优先被移除伸缩组。

### 📖 说明

手动移入伸缩组的云服务器不会遵循“实例移除策略”的要求，实例移除优先级最低，且移除时，系统不会删除该云服务器。当有多个手工加入伸缩组的云服务器时，移除规则是：先进先出。

## 4.5 查询伸缩活动

### 操作场景

若您需要查看伸缩活动是否成功及查看伸缩活动的详情，请参考本章节[查询伸缩活动](#)。

### 查询伸缩活动

下面介绍如何查看伸缩组的伸缩活动。


1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例”。
3. 单击具体的伸缩组名称。
4. 在“监控”页签中，展示伸缩活动详情。可以查看实例数量、CPU使用率等监控指标的变化情况。

### 查询活动历史

下面介绍如何查看伸缩组内伸缩活动的历史记录。

1. 选择“计算 > 弹性伸缩 > 伸缩实例”。
2. 单击具体的伸缩组名称。
3. 切换到“活动历史”页签。该页签展示进行伸缩活动的历史记录，支持查看实例伸缩和ELB迁移这两种伸缩活动的记录。

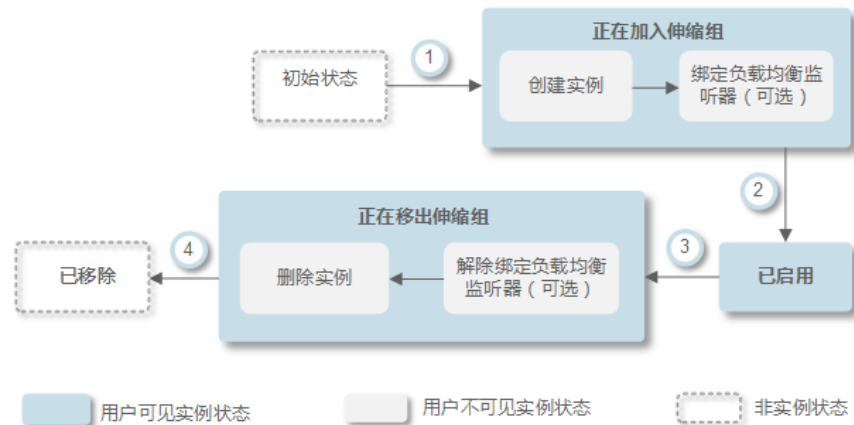
该页签记录了伸缩活动的ID、状态、活动类型、描述、开始时间和结束时间信

息。单击具体的伸缩活动ID前的  按钮，可以查看资源名称、状态和失败原因信息。您也可以使用页签右上角的过滤功能，查看指定时间段内的伸缩活动。

## 4.6 生命周期挂钩

生命周期挂钩功能提供灵活控制伸缩组内ECS实例创建和移出过程的能力，以使用户灵活管理ECS实例的生命周期。伸缩组中未添加生命周期挂钩时，实例生命周期状态如图4-3所示。

图 4-3 实例生命周期状态



添加生命周期挂钩后，实例生命周期状态如图4-4所示。

图 4-4 添加生命周期挂钩后实例生命周期状态



当伸缩组中发生伸缩活动，触发生命周期挂钩时，伸缩活动将被挂起，正在进行伸缩活动的实例会被置为等待状态，如图4-4中②和⑥。您可在实例保持等待状态时进行一些自定义操作。例如，在新实例移入伸缩组时，您可以在其上安装或配置软件。有以下两种方式可以结束被挂起的伸缩活动：

- 实例保持等待的时间大于超时时间。
- 手动执行回调操作，主动结束实例等待状态。

## 使用场景

- 伸缩组中新移入的实例，需要先进行初始化（安装或配置软件等）并检测服务正常运行后，再绑定到负载均衡监听器对外提供服务。
- 伸缩组中的实例被释放之前，需先从负载均衡监听器上解绑以确保不再接收新的请求，待检测已经接收到的请求处理完毕后进行释放。
- 伸缩组中的实例被释放之前，需要执行数据备份操作或者下载日志文件。
- 其它需要执行自定义操作的场景。

## 工作原理

将生命周期挂钩添加到伸缩组后，生命周期挂钩将按照如下方式工作：

- 实例移入伸缩组  
实例移入伸缩组并且初始化完成之后，自动触发挂钩类型为“实例启动”的生命周期挂钩，实例进入“等待（正在加入伸缩组）”状态，即实例被挂钩挂起。若您配置了一个通知目标，则系统会向该目标发送消息。收到消息后，您可以执行自定义操作，例如在实例上安装软件。自定义操作执行完成后，您可以手动执行回调操作，结束实例等待状态。或等待超时时间结束，系统自动结束实例等待状态。实例等待状态结束之后的默认回调操作有两种执行方案，“继续”或“终止”。这两种执行方案解释如下：
  - 继续：处于等待状态的实例将加入伸缩组。
  - 终止：处于等待状态的实例将被直接删除并重新创建新实例。

若配置了多个“实例启动”类型的生命周期挂钩，实例移入伸缩组会触发多个生命周期挂钩，若有一个挂钩默认回调操作为“终止”时，将会直接删除实例并重

新创建新实例。若所有挂钩默认回调操作都为“继续”时，则会等待最后一个挂钩挂起结束后，将实例加入伸缩组。

- 实例移出伸缩组

实例移出伸缩组时，先进入正在移出伸缩组状态，触发生命周期挂钩后，实例进入“等待（正在移出伸缩组）”状态。系统会向您配置的通知目标发送消息，收到消息后，您可以执行自定义操作，如卸载实例上的软件、备份数据等。自定义操作执行完成之后，您可以选择手动执行默认回调操作或等待超时时间超时来结束实例等待状态。等待状态结束后实例有两种执行方案，继续或终止，这两种执行方案解释如下：

- 继续：将实例移出伸缩组
- 终止：将实例移出伸缩组

当有多个挂钩时，“继续”表示继续等待其他挂钩挂起超时，只有所有挂钩状态都为“继续”时，才会将实例移出伸缩组。只要有一个挂钩默认回调操作为“终止”，会直接将实例移出伸缩组。

## 使用限制

- 添加挂钩、删除挂钩、修改挂钩等操作都是在该伸缩组未进行伸缩活动时可操作。
- 一个伸缩组内最多可添加5个生命周期挂钩。

## 添加挂钩

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 单击需要添加生命周期挂钩的弹性伸缩组名称，进入伸缩组的“基本信息”界面，选择“生命周期挂钩”页签，单击“添加生命周期挂钩”。
4. 在“添加生命周期挂钩”界面，根据界面提示进行参数配置，如表4-1所示。

表 4-1 参数配置

参数名称	参数说明	取值样例
挂钩名称	生命周期挂钩名称(1~32个字符)，只能由字母、数字、下划线、中划线组成。	we12_w
挂钩类型	挂钩类型包括“实例启动”和“实例终止”，它们分别在实例加入伸缩组和实例移出伸缩组时将实例置于“等待（正在加入伸缩组）”或者“等待（正在移出伸缩组）”状态。	实例启动

参数名称	参数说明	取值样例
默认回调操作	<p>默认回调操作是指当实例为等待状态且等待状态的时间已经达到超时时间后的系统默认操作。</p> <p>在当前实例正在加入伸缩组时，默认回调操作的含义为：</p> <ul style="list-style-type: none"> <li>继续：当有一个挂钩时，表示继续将实例加入伸缩组；当有多个挂钩时，表示继续等待其他挂钩的状态，只有所有挂钩状态都为“继续”时，才会继续将实例加入伸缩组。</li> <li>终止：无论有几个挂钩，只要有一个挂钩状态为“终止”，将会直接删除实例并重新创建新实例。</li> </ul> <p>在当前实例正在移出伸缩组时，默认回调操作的含义为：</p> <ul style="list-style-type: none"> <li>继续：当只有一个挂钩时，“继续”表示直接将实例移出伸缩组；当有多个挂钩时，表示继续等待其他挂钩的状态，只有所有挂钩状态都为“继续”时，才会继续将实例移出伸缩组。</li> <li>终止：无论有几个挂钩，只要有一个挂钩状态为“终止”，会直接将实例移出伸缩组。</li> </ul>	继续
超时时间	<p>默认情况下，实例保持等待状态的时间。取值范围为：300秒~86400秒。</p> <p>您可以延长超时时间，也可以在超时时间结束前进行“继续”或“终止”操作。关于回调操作更多信息请参见<a href="#">进行回调操作</a>。</p>	3600秒
通知主题	<p>为生命周期挂钩定义一个通知目标（请参见<a href="#">《消息通知服务用户指南》</a>创建主题），当实例被挂钩挂起时向该通知目标发送消息。该消息包含实例的基本信息、用户自定义通知消息，以及可用于控制生命周期操作的令牌信息。消息样例如下：</p> <pre>{   "service": "AutoScaling",   "tenant_id": "93075aa73f6a4fc0a3209490cc57181a",   "lifecycle_hook_type": "INSTANCE_LAUNCHING",   "lifecycle_hook_name": "test02",   "lifecycle_action_key": "4c76c562-9688-45c6-b685-7fd732df310a",   "notification_metadata": "xxxxxxxxxxxx",   "scaling_instance": {     "instance_id": "89b421e4-5fa6-4733-bf40-6b07a8657256",     "instance_name": "as-config-kxeg_RM6OCREY",     "instance_ip": "192.168.0.202"   },   "scaling_group": {     "scaling_group_id": "fe376277-50a6-4e36-bdb0-685da85f1a82",     "scaling_group_name": "as-group-wyz01",     "scaling_config_id": "16ca8027-b6cc-45fc-af2d-5a79996f685d",     "scaling_config_name": "as-config-kxeg"   } }</pre>	-

参数名称	参数说明	取值样例
自定义通知消息	当配置了通知目标时，可向其发送用户自定义的通知内容。	-

5. 单击“确定”。  
在生命周期挂钩页签可查看到新添加的生命周期挂钩。

## 进行回调操作

1. 在伸缩实例页面，单击需要进行回调操作的伸缩组名称，进入弹性伸缩组的“基本信息”界面。
2. 选择“伸缩实例”页签。
3. 单击被挂钩挂起实例“生命周期状态”列下的“等待（正在加入伸缩组）”或“等待（正在移出伸缩组）”，如图4-5所示。

图 4-5 回调操作入口



### 📖 说明

- 只有被生命周期挂钩挂起的实例，可执行回调操作。
4. 在弹出的“伸缩实例挂起信息”界面，可查看某个实例的挂起信息及当前伸缩组中所有的挂钩，并且可对每个挂钩执行回调操作。如图4-6所示。

图 4-6 伸缩实例挂起信息页面



回调操作包括：

- 继续
- 终止
- 延长超时时间

如果您在超时时间结束前已完成自定义操作，选择“继续”或“终止”完成生命周期操作。“继续”或“终止”操作含义请参见表4-1。如果您需要更多时间完成自定义操作，选择“延长超时间”延长超时时间，实例保持等待状态的时间将增加3600秒。

## 修改挂钩

在“生命周期挂钩”页签，选择要修改的挂钩“操作”列下的“修改”，修改时界面的参数配置请参考表4-1，您可以修改除挂钩名称外的其余参数，例如挂钩类型、默认回调操作、超时时间等。

## 删除挂钩

在“生命周期挂钩”页签，选择要删除的挂钩“操作”列下的“删除”。

# 4.7 实例保护

## 操作场景

如果您希望伸缩组中特定的实例不被自动移出伸缩组，请使用实例保护。您可以对伸缩组中一个或多个正常状态的实例启用实例保护设置，当伸缩组发生缩容活动时，设置了实例保护的实例将不会被移出伸缩组。

## 前提条件

以下场景实例保护无法确保实例不被移出伸缩组：



- 实例未通过健康检查而被移出。
- 手动移出实例。

#### 📖 说明

- 健康状态异常实例无法提供服务，伸缩组优先保证组内实例均正常，因此实例保护无法保护异常实例。
- 伸缩组新创建的实例或者被新加入到伸缩组中的实例默认不启用实例保护。
- 实例一旦被移出伸缩组，为其设置的实例保护属性就会失效。

## 设置实例保护

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 单击需设置实例保护的弹性伸缩组名称，进入伸缩组基本信息页面。
4. 选择“伸缩实例”页签，勾选一个或多个实例，再选择“更多 > 设置实例保护”，在弹出的“设置实例保护”页面，单击“是”，同时为一个或多个实例设置实例保护。

## 取消实例保护

1. 登录管理控制台。
1. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
2. 单击需取消实例保护的弹性伸缩组名称，进入伸缩组基本信息页面。
3. 选择“伸缩实例”页签，勾选一个或多个实例，再选择“更多 > 取消实例保护”，在弹出的“取消实例保护”页面，单击“是”，可同时取消多个或单个实例保护。

## 4.8 实例备用

若您需要伸缩组中的部分实例暂时停止承担业务流量且不被移出伸缩组，您可以使用弹性伸缩提供的实例备用功能。您可以对伸缩组中的一个或多个实例设置实例备用，实例备用能在保证实例不被移出伸缩组的同时对实例进行关机、重启等操作。

### 应用场景

对于伸缩组内管理的ECS实例，用户无法控制其生命周期。而伸缩组对实例的非健康状态的释放操作，也阻碍了用户对伸缩组中的ECS实例进行停机、重启等操作，使得用户无法充分利用ECS服务提供一些功能，如重置密码、重装系统、切换操作系统等。

通过将伸缩组中的实例转入备用状态，ECS实例的生命周期控制权将移至用户手中，方便用户对ECS实例进行停机等相关操作，极大的增强了用户对伸缩组中实例的管理能力，适用于多种业务场景下的需求。

- 需要对弹性伸缩弹出的ECS实例进行切换操作系统、重启等变更操作的场景，用户通过将目标ECS实例转入备用状态，用户可进行ECS服务所支持的全部操作，操作完成后，再将实例移出备用状态，实例将进入伸缩组中的正常运行状态。

例如，您可以随时为伸缩组更换伸缩配置，伸缩组后续启动的任何实例都将使用此配置。不过，伸缩组不会更新已经正在运行的实例。您可以终止这些实例并让伸缩组替换这些实例，也可以将其转入备用状态，更新实例上的软件，然后将实例重新置于运行状态（即移出备用状态）。

- 用户通过伸缩组配置负载均衡的方式来管理业务机器，当伸缩组中的某台实例出现业务问题，用户可以通过转入备用操作对这台实例的承载的流量进行分流，进行一系列离线排查验证（登录，排查，重启等）操作后，在确认该实例已经正常后，再移出备用状态，重新处理业务流量。

## 工作原理

- 将实例转入备用

当您选择将实例转入备用，系统会自动将该实例从伸缩组相关联的负载均衡解绑。备用状态的实例仍在伸缩组中，但是健康检查会停止对该实例的检测。由于没有流量通过负载均衡进入到备用实例，因此会增加伸缩组中其他实例的负载，您可以根据需要在转入备用操作时勾选“向伸缩组中加入与备用主机等量的新实例”，可以避免引起其他实例的额外负载，保证业务的正常运行。

### 说明

- 伸缩组没有进行伸缩活动，选择的实例生命周期状态是已启用状态，才可以将该实例转入备用状态。
- 伸缩组中由伸缩策略触发的自动扩容的伸缩活动不会移除处于备用状态的实例。
- 实例处于备用状态时，可以手动执行移出伸缩组操作。
- 将实例移出备用

您可以通过将实例移出备用状态使实例恢复正常运行状态，实例会重新承载伸缩组的业务。若伸缩组设置了负载均衡，系统会为该实例绑定伸缩组的负载均衡。实例恢复正常运行状态后，健康检查会恢复对该实例健康状态的检测。

### 说明

伸缩组没有进行伸缩活动，选择的实例生命周期状态是已备用状态，才可以将该实例移出备用状态。

## 将实例转入备用

1. 登录管理控制台。
2. 单击服务列表。
3. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
4. 单击需要进行实例转入备用操作的弹性伸缩组名称，进入伸缩组基本信息页面。
5. 选择“伸缩实例”页签，勾选一个或多个实例，再选择“更多 > 转入备用”，在弹出的“转入备用”页面，可根据实际情况选择是否勾选“向伸缩组中加入与备用主机等量的新实例”，完成后单击“是”，将被勾选的实例转入备用状态。

## 将实例移出备用

1. 登录管理控制台。
1. 单击服务列表。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
3. 单击需要进行实例移出备用的弹性伸缩组名称，进入伸缩组基本信息页面。
4. 选择“伸缩实例”页签，勾选一个或多个实例，再选择“更多 > 移出备用”，在弹出的“移出备用”页面，单击“是”，将被勾选的实例移出备用状态。

# 5 伸缩带宽

## 5.1 创建伸缩带宽策略

### 操作场景

用户可以通过伸缩带宽策略对购买的弹性公网IP带宽和共享带宽进行调整。本章节介绍如何创建伸缩带宽策略。

您可以通过创建伸缩带宽策略来实现自动调整带宽。创建伸缩带宽策略时，需要配置对应的基本信息，系统支持告警策略、定时策略、周期策略三种伸缩带宽策略。

创建伸缩带宽策略的基本信息，包括配置策略名称、资源类型、策略类型、触发条件等。

### 创建伸缩带宽告警策略

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩带宽”。
3. 单击“创建伸缩带宽策略”。
4. 配置策略名称、策略类型、触发条件等参数，请参考[创建伸缩带宽策略](#)进行配置。

表 5-1 “告警策略”参数说明

参数	解释	取值样例
区域	创建的伸缩带宽策略所在的区域。	-
策略名称	创建的伸缩带宽策略的名称。 策略名称只能由中文、英文字母、数字、下划线、和中划线组成。	-
资源类型	需要进行管理的带宽的类型，可选择“弹性公网IP”或“共享带宽”。	弹性公网IP

参数	解释	取值样例
弹性公网IP	需要进行伸缩管理的公网IP。 <b>说明</b> 当前仅支持对按需付费模式的弹性公网IP进行伸缩，包周期的弹性公网IP不能创建弹性伸缩。	-
策略类型	选择“告警策略”。	告警策略
告警规则	可选择已有的告警和新建告警。也可单击右侧“新建告警规则”在云监控服务页面创建告警规则，详细操作请参考 <a href="#">新建告警规则</a> 。 若选择新建告警，需配置如下参数： <ul style="list-style-type: none"> <li>告警规则名称 新建告警的名称，例如as-alarm-7o1u。</li> <li>触发条件 选择弹性带宽支持的监控指标及设定该监控指标的触发条件，支持的监控指标如<a href="#">表5-2</a>所示。例如出网带宽平均值&gt;100bit/s。</li> <li>监控周期 设定对弹性带宽支持的监控指标监控的周期，例如5分钟。</li> <li>连续出现次数 在监控周期内，连续达到触发条件几次后，开始执行伸缩活动，例如1次。</li> </ul>	-
执行动作	设置伸缩策略执行动作。 执行动作包括： <ul style="list-style-type: none"> <li>增加 当执行伸缩活动时，增加带宽大小。</li> <li>减少 当执行伸缩活动时，减少带宽大小。</li> <li>设置为 将带宽大小设置为固定值。</li> </ul> <b>说明</b> 由于带宽在不同的取值范围内步长（即可调整的最小单位）不同，最终调整后的带宽会根据实际步长自动调整为就近值。 <ul style="list-style-type: none"> <li>小于等于300Mbit/s：默认步长为1Mbit/s。</li> <li>300Mbit/s~1000Mbit/s：默认步长为50Mbit/s。</li> <li>大于1000Mbit/s：默认步长为500Mbit/s。</li> </ul>	-
冷却时间	冷却时间是指冷却伸缩活动的时间，在每次伸缩活动完成之后，系统开始计算冷却时间。在冷却时间内，会拒绝由告警策略触发的伸缩活动，其他类型的伸缩策略（如定时策略和周期策略）触发的伸缩活动不受限制，但会重新开始计算冷却时间，单位为秒。	300秒

表 5-2 告警策略监控指标说明

指标名称	解释
入网带宽	该指标用于统计测量对象入云平台的网络速度。
入网流量	该指标用于统计测量对象入云平台的网络流量。
出网带宽	该指标用于统计测量对象出云平台的网络速度。
出网流量	该指标用于统计测量对象出云平台的网络流量。
出网带宽使用率	该指标用于统计测量对象出云平台的带宽使用率，以百分比为单位。

5. 参数配置完成后，单击“立即创建”。

在“伸缩带宽”页面中可查看新创建的伸缩带宽策略，新创建的策略默认的状态为“已启用”。

## 新建告警规则

创建伸缩带宽告警策略时，您可以通过单击“告警规则名称”右侧的“新建告警规则”创建所需的告警规则。可参考如下步骤进行操作。

1. 单击“告警规则名称”右侧的“新建告警规则”，跳转到云监控的告警规则页面。
2. 单击页面右上角“创建告警规则”。
3. 参考图5-1和表5-3进行参数配置。了解更多关于创建告警规则的信息请参考《云监控用户指南》。

图 5-1 创建告警规则

\* 资源类型 弹性公网IP和带宽

\* 维度 带宽 弹性IP

\* 监控范围 资源分组 指定资源

⇨ 全选

名称 ecs-yxx-bandwidth × Q

名称	ID
无记录	

⇨ 取消全选

名称 请输入名称 Q

名称	ID
<input type="checkbox"/> ecs-yxx-band...	b66f6fa3-365a-4ec7...

>> <<

\* 选择类型 从模板导入 自定义创建

\* 告警策略 出网带宽 最大值 监控周期5分钟 连续3个周期 >= 500 bit/s

\* 告警级别  紧急  重要  次要  提示


告警通知 \_\_\_\_\_

发送通知

表 5-3 创建告警规则关键参数

参数	解释	配置示例
名称	告警规则的名称。	alarm-bandwidth
描述	告警规则描述（此参数非必填项）。	-
企业项目	告警规则所属的企业项目。只有拥有该企业项目权限的用户才可以查看和管理该告警规则。	default
资源类型	告警规则监控的服务名称，需要选择“弹性公网IP和带宽”。	弹性公网IP和带宽
维度	指定被监控的服务的具体模块。伸缩带宽是调整带宽大小的，需要选择“带宽”。	带宽
监控范围	告警规则适用的资源范围，需要选择“指定资源”。选择资源时可通过带宽名称或ID进行搜索，带宽名称或ID请在需要调整的EIP的详情页面获取。	指定资源

参数	解释	配置示例
选择类型	选择创建告警的方式，需要选择“自定义创建”。	自定义创建
告警策略	触发告警规则的告警策略，请按需要进行配置。监控指标的含义请参考表5-2。	-

- 参数配置完成后，单击“立即创建”。
- 返回伸缩带宽策略创建页面，单击“告警规则名称”右侧的  按钮，然后为“告警规则名称”选择刚才创建的告警规则。

您还可以在创建伸缩带宽策略前，在云监控页面创建好所需的告警规则，创建告警规则时选择的指定资源必须是创建伸缩带宽策略时的选择的EIP资源对应的带宽资源。创建完成后，在创建伸缩带宽策略时可直接选择该告警规则。

## 创建伸缩带宽定时/周期策略

- 登录管理控制台。
- 选择“计算 > 弹性伸缩 > 伸缩带宽”。
- 单击“创建伸缩带宽策略”。
- 配置策略名称、策略类型、触发条件等参数，请参考表5-4进行配置。

表 5-4 “定时策略”或“周期策略”参数说明

参数名称	参数说明	取值样例
区域	创建的伸缩带宽策略所在的区域。	-
策略名称	创建伸缩带宽策略的名称。 策略名称只能由中文字符、英文字母、数字、下划线、和中划线组成。	as-policy-p6g5
资源类型	需要进行管理的带宽的类型，可选择“弹性公网IP”或“共享带宽”。	弹性公网IP
弹性公网IP	需要进行伸缩管理的公网IP。当资源类型选择“弹性公网IP”时需要配置该项。 <b>说明</b> 当前仅支持对按需付费模式的弹性公网IP进行伸缩，包周期的弹性公网IP不能创建弹性伸缩。	-
共享带宽	需要进行伸缩管理的共享带宽。当资源类型选择“共享带宽”时需要配置该项。	-

参数名称	参数说明	取值样例
策略类型	可选择定时策略和周期策略，在指定的时间段进行调整带宽。 若选择周期策略除了配置表格中的参数还需配置以下两个参数： <ul style="list-style-type: none"><li>生效时间 选择伸缩策略触发的时间段。</li><li>重复周期<ul style="list-style-type: none"><li>按天</li><li>按周</li><li>按月</li></ul></li></ul>	-
触发时间	设定伸缩策略触发时间。	-
执行动作	设置伸缩活动的执行动作。 执行动作包括： <ul style="list-style-type: none"><li>增加 当执行伸缩活动时，增加带宽大小。</li><li>减少 当执行伸缩活动时，减少带宽大小。</li><li>设置为 将带宽大小设置为固定值。</li></ul> <b>说明</b> 由于带宽在不同的取值范围内步长（即可调整的最小单位）不同，最终调整后的带宽会根据实际步长自动调整为就近值。 <ul style="list-style-type: none"><li>小于等于300Mbit/s：默认步长为1Mbit/s。</li><li>300Mbit/s~1000Mbit/s：默认步长为50Mbit/s。</li><li>大于1000Mbit/s：默认步长为500Mbit/s。</li></ul>	-
冷却时间	冷却时间是指冷却伸缩活动的时间，在每次伸缩活动完成之后，系统开始计算冷却时间。在冷却时间内，会拒绝由告警策略触发的伸缩活动，其他类型的伸缩策略（如定时策略和周期策略等）触发的伸缩活动不受限制，但会重新开始计算冷却时间，单位为秒。	300秒

5. 参数配置完成后，单击“立即创建”。

## 5.2 查看伸缩带宽策略详情

### 操作场景

用户可以通过查看伸缩带宽策略详情，了解该伸缩带宽策略的基本信息及执行日志。策略执行日志记录了策略执行的详细情况。本章节介绍如何查看伸缩带宽策略详情。



## 查看伸缩带宽策略详情

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩带宽”。
3. 在“伸缩带宽”页面，单击需要查看的伸缩带宽策略名称，可跳转至该策略的“基本信息”页面，查看详情。可以看到伸缩带宽策略的策略类型、触发条件、执行动作等基本信息。

## 查看伸缩带宽策略执行日志

在需要查看的伸缩带宽策略的基本信息页面，您可以看到策略执行日志。策略执行日志记录了伸缩带宽策略的执行状态、策略执行时间、伸缩原始值和伸缩目标值等信息。

# 5.3 管理伸缩带宽策略

## 操作场景

用户可以通过伸缩带宽策略来调整带宽的大小。

本章节介绍对伸缩带宽策略的管理，包括启用、停用、修改、删除、立即执行伸缩带宽策略。

### 说明

已释放EIP对应的伸缩带宽策略仍然会占用配额，且只能使用华为云帐号和有全局权限的IAM用户进行管理，授予部分企业项目的IAM用户不能对该伸缩带宽策略进行管理。

## 启用伸缩带宽策略

只有当伸缩带宽策略状态为“已停用”时，可以启用策略。

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩带宽”。
3. 在伸缩带宽策略列表中，待启用的策略所在行的“操作”列下，单击“启用”。
4. 在弹出“启用伸缩带宽策略”的对话框中，单击“是”。

## 停用伸缩带宽策略

当伸缩带宽策略状态是“已启用”时，可以停用策略。

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩带宽”。
3. 在伸缩带宽策略列表中，待停用的策略所在行的“操作”列下，单击“停用”。
4. 在弹出“停用伸缩带宽策略”的对话框中，单击“是”。

### 说明

停用伸缩带宽策略后，策略的状态会变为“已停用”，已停用状态的策略，不会自动触发任何弹性伸缩活动。

## 修改伸缩带宽策略

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩带宽”。
3. 在伸缩带宽策略列表中，单击待修改的策略的名称，进入该策略基本信息页面。  
在“基本信息”页面右上方，单击“修改”。  
或在待修改的策略所在行的“操作”列下，单击“更多 > 修改”。
4. 修改相关数据。伸缩带宽策略可以修改的参数有：策略名称、弹性公网IP、策略类型、执行动作、冷却时间等。
5. 单击“确定”。

### 说明

如果伸缩带宽策略状态是“执行中”，则无法修改。

## 删除伸缩带宽策略

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩带宽”。
3. 在伸缩带宽策略列表中，在待删除的策略所在行的“操作”列下，单击“更多 > 删除”。
4. 在弹出“删除伸缩带宽策略”的对话框中，单击“是”。  
您也可以同时勾选一个或多个伸缩带宽策略，单击列表上方的“删除”，来删除一个或多个伸缩带宽策略。

### 说明

- 当您不再需要某个伸缩带宽策略时，可以删除该策略。如果您仅在某段时间不需要该策略，建议您采用停用方式，而不建议删除。
- 伸缩带宽策略状态为非执行中，才可以被删除。

## 立即执行伸缩带宽策略

通过立即执行已创建的伸缩带宽策略，您可以将带宽值立即调整到伸缩策略中设置的执行动作，不用等待伸缩带宽策略的触发条件被满足。

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩带宽”。
3. 在伸缩带宽策略列表中，单击需要立即执行的伸缩带宽策略所在行的“立即执行”。
4. 在弹出“执行伸缩带宽策略”的对话框中，单击“是”。

也可在需要立即执行的伸缩带宽策略的基本信息页面，单击右上角的“立即执行”按钮。

### 说明

- 只有当伸缩带宽策略是启用状态，并且当前没有正在执行的伸缩带宽策略时，才可以进行立即执行操作。
- 对伸缩带宽策略进行立即执行操作后，达到该策略的触发条件时，系统仍会按照配置的执行动作调整带宽。

# 6 伸缩组和实例的监控

## 6.1 弹性伸缩健康检查

健康检查会将异常的实例从伸缩组中移除，伸缩组会重新创建新的实例以维持伸缩组的期望实例数和当前实例数保持一致，伸缩组的健康检查方式主要包括以下两种。

- 云服务器健康检查：是指对云服务器的运行状态进行检查，如关机、删除都是云服务器异常状态。伸缩组的健康检查方式默认是“云服务器健康检查”方式，指伸缩组会定期使用云服务器健康检查结果来确定每个云服务器的运行状况。如果未通过云服务器健康检查，则伸缩组会将该云服务器移出伸缩组。
- 弹性负载均衡健康检查：是指根据ELB对云服务器的健康检查结果进行的检查。仅当伸缩组使用弹性负载均衡器时，可以选择“弹性负载均衡健康检查”方式来做健康检查。如果您将多个负载均衡器添加到伸缩组，则只有在所有负载均衡器均检测到云服务器状态为正常的情况下，才会认为该弹性云服务器正常。否则只要有一个负载均衡器检测到云服务器状态异常，伸缩组会将该弹性云服务器移出伸缩组。

以上两种健康检查方式，检查的结果均是将异常的云服务器从伸缩组中移除，移出伸缩组的实例，是否会将云服务器删除，如下所述：

- 弹性伸缩活动中自动添加的云服务器默认计费方式是按需计费，这种情况下，系统将云服务器移出伸缩组时，也会将其删除。对于手动移入伸缩组的实例，系统仅将其移出伸缩组。
- 如果您手动将计费方式修改为包年包月，那么系统将云服务器移出伸缩组时，仅将其移出伸缩组，不进行删除操作。

值得注意的是，当伸缩组为停用状态时，对实例的健康状态会继续进行检查，但不会执行移除操作。

## 6.2 为伸缩组配置通知

### 操作场景

当用户申请开通消息通知服务后，可通过消息通知服务提供的功能，将伸缩组的扩容成功、扩容失败、减容成功、减容失败和异常等情况及时推送给用户，以使用户能够及时了解伸缩组的各种状态。

给弹性伸缩组配置通知，需配置一个通知事件和通知主题。每个伸缩组最多可以配置5个通知，通知主题由用户先在消息通知服务界面创建，当通知主题对应的通知场景出现时，伸缩组会向用户发送通知。

## 为伸缩组配置通知

1. 登录管理控制台。
1. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
2. 选择需要添加“通知”的伸缩组，在“基本信息”界面，选择“通知 > 添加通知”。
3. 根据界面提示进行参数配置，如表6-1所示。

表 6-1 参数配置

参数名称	参数说明	取值样例
通知事件	每当伸缩组出现以下一种或几种场景时，向用户发送通知。 <ul style="list-style-type: none"><li>• 扩容成功</li><li>• 减容成功</li><li>• 异常</li><li>• 扩容失败</li><li>• 减容失败</li></ul>	-
通知主题	选择已经创建成功主题。请参见 <a href="#">创建主题</a> 。	-

4. 单击“确定”。

## 6.3 记录弹性伸缩

### 操作场景

弹性伸缩支持使用云审计记录服务资源操作。云审计记录的操作类型有三种，通过云平台帐户登录管理控制台执行的操作，通过云服务支持的API执行的操作，以及系统内部触发的操作。

如果用户开通了云审计，AS服务的API被调用时，调用信息将会上报到云审计，云审计会将操作信息定时的转储到用户指定的对象存储桶。通过云审计服务，您可以记录与弹性伸缩相关的操作事件，便于日后的查询、审计和回溯。

### Cloud Trace Service 中的 AS 信息



在您的应用系统中启用云审计服务后，将在日志文件记录对弹性伸缩执行的API调用的操作。您可以在云审计服务管理控制台查询近7天内的操作记录。如果需要保存7天之前的操作记录，您可以通过对象存储服务（Object Storage Service，以下简称OBS），将操作记录实时同步保存至OBS。

云审计服务支持的AS操作列表如表1所示。

表 6-2 云审计服务支持的 AS 操作列表

操作名称	资源类型	事件名称
创建伸缩组	scaling_group	createScalingGroup
修改伸缩组	scaling_group	modifyScalingGroup
删除伸缩组	scaling_group	deleteScalingGroup
启用伸缩组	scaling_group	enableScalingGroup
停用伸缩组	scaling_group	disableScalingGroup
创建伸缩配置	scaling_configuration	createScalingConfiguration
删除伸缩配置	scaling_configuration	deleteScalingConfiguration
批量删除伸缩配置	scaling_configuration	batchDeleteScalingConfiguration
创建伸缩策略	scaling_policy	createScalingPolicy
修改伸缩策略	scaling_policy	modifyScalingPolicy
删除伸缩策略	scaling_policy	deleteScalingPolicy
启用伸缩策略	scaling_policy	enableScalingPolicy
停用伸缩策略	scaling_policy	disableScalingPolicy
执行伸缩策略	scaling_policy	executeScalingPolicy
移除实例	scaling_instance	removeInstance
批量移除实例	scaling_instance	batchRemoveInstances
批量添加实例	scaling_instance	batchAddInstances
批量设置实例保护	scaling_instance	batchProtectInstances
批量取消实例保护	scaling_instance	batchUnprotectInstances
配置通知	scaling_notification	putScalingNotification
删除通知	scaling_notification	deleteScalingNotification
创建生命周期挂钩	scaling_lifecycle_hook	createLifecycleHook
修改生命周期挂钩	scaling_lifecycle_hook	modifyLifecycleHook
删除生命周期挂钩	scaling_lifecycle_hook	deleteLifecycleHook
删除标签	scaling_tag	deleteScalingTag
创建/更新标签	scaling_tag	updateScalingTag

## 查看审计日志

1. 登录管理控制台。
2. 在管理控制台左上角单击  图标，选择区域和项目。
3. 单击“服务列表”，选择“管理与监管 > 云审计服务”，进入云审计服务信息页面。
4. 单击左侧导航树的“事件列表”，进入事件列表信息页面。
5. 事件列表支持通过筛选来查询对应的操作事件。当前事件列表支持四个维度的组合查询，详细信息如下：
  - 事件类型、事件来源、资源类型和筛选类型。  
在下拉框中选择查询条件。  
其中筛选类型选择事件名称时，还需选择某个具体的事件名称。  
选择资源ID时，还需选择或者手动输入某个具体的资源ID。  
选择资源名称时，还需选择或手动输入某个具体的资源名称。
  - 操作用户：在下拉框中选择某一具体的操作用户，此操作用户指用户级别，而非租户级别。
  - 事件级别：可选项为“所有事件级别”、“normal”、“warning”、“incident”，只可选择其中一项。
  - 时间范围：可在页面右上角选择查询最近1小时、最近1天、最近1周及自定义时间段的操作事件。
6. 在需要查看的记录左侧，单击  展开该记录的详细信息。
7. 在需要查看的记录右侧，单击“查看事件”，弹出的窗口显示了该操作事件结构的详细信息。

## 日志记录条目

云审计中每个记录条目由一个JSON格式的事件组成。一个日志条目表示一条SMN接口请求，内容主要包括所请求的操作、操作的时间和日期、所操作的参数以及生成该请求的用户信息，其中用户信息来自统一身份认证服务。

以下示例显示了CreateScalingPolicy操作的CloudTrace记录条目。

```
{
  "time": "2016-12-15 15:27:40 GMT+08:00",
  "user": {
    "name": "xxxx",
    "id": "62ff83d2920e4d3d917e6fa5e31ddebe",
    "domain": {
      "name": "xxx",
      "id": "30274282b09749adbe7d9cabeebcbe8b"
    }
  },
  "request": {
    "scaling_policy_name": "as-policy-oonb",
    "scaling_policy_action": {
      "operation": "ADD",
      "instance_number": 1
    },
    "cool_down_time": "",
    "scheduled_policy": {
```

```
"launch_time": "2016-12-16T07:27Z"  
},  
"scaling_policy_type": "SCHEDULED",  
"scaling_group_id": "ec4051a7-6fbd-42d2-840f-2ad8cdabee34"  
},  
"response": {  
"scaling_policy_id": "6a38d488-664b-437a-ade2-dc45f96f7f4c"  
},  
"code": 200,  
"service_type": "AS",  
"resource_type": "scaling_policy",  
"resource_name": "as-policy-oonb",  
"resource_id": "6a38d488-664b-437a-ade2-dc45f96f7f4c",  
"source_ip": "10.190.205.233",  
"trace_name": "createScalingPolicy",  
"trace_rating": "normal",  
"trace_type": "ConsoleAction",  
"api_version": "1.0",  
"record_time": "2016-12-15 15:27:40 GMT+08:00",  
"trace_id": "f627062b-c297-11e6-a606-eb2c0f48bec5"  
}
```

## 6.4 标记伸缩组和实例

### 操作场景

当您具有相同类型的许多资源时，标签可以为您提供灵活的资源管理能力，您可以根据分配给资源的标签快速识别特定资源。

通过标签的形式将自定义数据分配给每个伸缩组，您可以对伸缩组进行组织和管理，例如可以通过用途、所有者或环境对伸缩组资源进行分类。

每个标签均包含一个“键”和一个“值”，您可为每个标签指定键和值。键可以是具有特定关联值的一般类别，如“usage”、“owner”或“environment”。

例如，要区分测试环境和生产环境，您可以为每个伸缩组分配一个标签，其键为“environment”，如果伸缩组是测试环境则设置对应的值为“test”，如果伸缩组是生产环境则设置对应的值为“production”。建议您根据需要使用一组或多组具有一致性的标签来更轻松地管理您的伸缩组资源。

当您为伸缩组设置标签后，系统会自动将伸缩组的标签添加到该伸缩组自动创建的实例上。对伸缩组添加或者修改标签，新的标签会被添加到伸缩组自动创建的实例。创建、删除或修改伸缩组的标签，不会对已经在伸缩组中运行的实例进行这些更改。

### 使用标签的限制

使用标签的基本限制包括如下几方面：

- 每个伸缩组最多可以添加10个标签。
- 每个标签均包含一个“键”和一个“值”。
- 您可以将标签的值设为空字符串。
- 如果删除伸缩组，则该伸缩组的所有标签也会被删除。

### 为伸缩组添加标签

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。

3. 选择需要添加“标签”的伸缩组，在“基本信息”界面，选择“标签 > 添加标签”。
4. 根据界面提示进行参数配置，如表6-3所示。

表 6-3 标签命名规则

参数	规则	样例
键	<ul style="list-style-type: none"><li>• 不能为空。</li><li>• 对于同一伸缩组，“键”唯一。</li><li>• 键的长度最大36字符，由英文字母、数字、下划线、中划线、中文字符组成。</li></ul>	Organization
值	<ul style="list-style-type: none"><li>• 可以为空字符串。</li><li>• 一个“键”只能添加一个“值”。</li><li>• 值的长度最大43字符，由英文字母、数字、下划线、点、中划线、中文字符组成。</li></ul>	Apache

5. 单击“确定”。

## 修改/删除伸缩组的标签

1. 登录管理控制台。
1. 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
2. 选择需要添加“标签”的伸缩组，在伸缩组详情界面，单击“标签”。
3. 选择需要修改/删除的标签所在行的操作列下的“编辑”或“删除”。  
单击“编辑”后，根据界面提示进行参数配置，如表6-3所示。  
单击“删除”后，将伸缩组已添加的该标签删除。

## 6.5 监控指标说明

### 功能说明

本节定义了弹性伸缩上报云监控的监控指标的命名空间，监控指标列表，各项监控指标的具体含义与使用说明，用户可以通过云监控检索弹性伸缩服务产生的监控指标和告警信息。

### 命名空间

SYS.AS

### 监控指标

弹性伸缩支持的监控指标如表6-4所示。



表 6-4 弹性伸缩支持的监控指标

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
cpu_util	CPU使用率	该指标用于统计弹性伸缩组的CPU使用率。 计算公式：伸缩组中的所有云服务器的CPU使用率之和/伸缩组实例数 单位：百分比	≥0%	弹性伸缩组	5分钟
mem_util	内存使用率	该指标用于统计弹性伸缩组的内存使用率，以百分比为单位。 计算公式：伸缩组中的所有云服务器内存使用率之和/伸缩组实例数 单位：百分比 <b>说明</b> 如果用户使用的镜像未安装vmtools，则无法获取该监控指标。	≥0%	弹性伸缩组	5分钟
instance_num	实例数	该指标用于统计弹性伸缩组中可用的云服务器云主机数量。 计算公式：弹性伸缩组内生命周期状态为“已启用”的云服务器数量之和	≥0	弹性伸缩组	5分钟
network_incoming_bytes_rate_inband	带内网络流入速率	该指标用于统计每秒流入弹性伸缩组的网络流量。 计算公式：伸缩组中所有云服务器的带内网络流入速率之和 / 伸缩组实例数 单位：字节/秒	≥0 Byte/s	弹性伸缩组	5分钟
network_outgoing_bytes_rate_inband	带内网络流出速率	该指标用于统计每秒流出弹性伸缩组的网络流量。 计算公式：伸缩组中所有云服务器的带内网络流出速率之和 / 伸缩组实例数 单位：字节/秒	≥0 Byte/s	弹性伸缩组	5分钟

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
disk_read_bytes_rate	磁盘读速率	该指标用于统计每秒从弹性伸缩组读出的数据量。 计算公式：伸缩组中所有云服务器的磁盘读速率之和 / 伸缩组实例数 单位：字节/秒	≥0 Byte/s	弹性伸缩组	5分钟
disk_write_bytes_rate	磁盘写速率	该指标用于统计每秒写到弹性伸缩组的数据量。 计算公式：伸缩组中所有云服务器的磁盘写速率之和 / 伸缩组实例数 单位：字节/秒	≥0 Byte/s	弹性伸缩组	5分钟
disk_read_requests_rate	磁盘读操作速率	该指标用于统计每秒从弹性伸缩组读取数据的请求次数。 计算公式：伸缩组中所有云服务器的磁盘读操作速率之和 / 伸缩组实例数 单位：请求/秒	≥0 request/s	弹性伸缩组	5分钟
disk_write_requests_rate	磁盘写操作速率	该指标用于统计每秒往弹性伸缩组写数据的请求次数。 计算公式：伸缩组中的所有云服务器的磁盘写操作速率之和 / 伸缩组实例数 单位：请求/秒	≥0 request/s	弹性伸缩组	5分钟
cpu_usage	(Agent) CPU使用率	该指标用于统计弹性伸缩组的 (Agent) CPU使用率。 计算公式：伸缩组中的所有云服务器的 (Agent) CPU使用率之和 / 伸缩组实例数 单位：百分比	0-100 %	弹性伸缩组	1分钟

指标ID	指标名称	指标含义	取值范围	测量对象	监控周期（原始指标）
mem_usedPercent	（Agent）内存使用率	该指标用于统计弹性伸缩组的（Agent）内存使用率，以百分比为单位。 计算公式：伸缩组中的所有云服务器（Agent）内存使用率之和/伸缩组实例数 单位：百分比	0-100%	弹性伸缩组	1分钟
load_ave_rage1	（Agent）1分钟平均负载	该指标用于统计测量对象中所有云服务器过去1分钟的CPU平均负载的均值。该指标无单位。	≥0	弹性伸缩组	1分钟
load_ave_rage5	（Agent）5分钟平均负载	该指标用于统计测量对象中所有云服务器过去5分钟的CPU平均负载的均值。该指标无单位。	≥0	弹性伸缩组	1分钟
load_ave_rage15	（Agent）15分钟平均负载	该指标用于统计测量对象中所有云服务器过去15分钟的CPU平均负载的均值。该指标无单位。	≥0	弹性伸缩组	1分钟
gpu_usage_gpu	（Agent）GPU使用率	该指标用于统计弹性伸缩组的（Agent）GPU使用率，以百分比为单位。 计算公式：伸缩组中的所有云服务器（Agent）GPU使用率之和/伸缩组实例数 单位：百分比	0-100%	弹性伸缩组	1分钟
gpu_usage_mem	（Agent）显存使用率	该指标用于统计弹性伸缩组的（Agent）显存使用率，以百分比为单位。 计算公式：伸缩组中的所有云服务器（Agent）显存使用率之和/伸缩组实例数 单位：百分比	0-100%	弹性伸缩组	1分钟

### 📖 说明

区分带Agent和不带Agent的监控指标：有的操作系统需要安装Agent后才能获取到相应的监控指标，此时，触发条件应选择带有Agent字样的监控指标（如：（Agent）内存使用率）。

### 说明

对于不同的操作系统，监控指标“内存使用率”、“带内网络流出速率”和“带内网络流入速率”是否支持，详细信息请参见《[弹性云服务器用户指南](#)》。

如需使用Agent监控指标，请确认伸缩组中的实例已安装Agent插件，详细信息请参见“[如何为伸缩组中的实例安装Agent插件?](#)”。

## 维度

Key	Value
AutoScalingGroup	弹性伸缩组的ID

## 6.6 查看监控指标数据

### 操作场景

为使用户更好地掌握自己的弹性云服务器运行状态，云平台提供了云监控。通过本节，您可以了解如何查看伸缩组的监控指标详情，更好地了解弹性云服务器的各项性能指标。

### 前提条件

弹性伸缩组中的弹性云服务器正常运行。


### 说明

- 当伸缩组中的实例数为0时，只能查看“实例数”这一项监控指标；CPU使用率、磁盘读速率等指标只有在伸缩组中有实例时才能查看。
- 关机、故障、删除状态的弹性云服务器，无法查看其CPU使用率、磁盘读速率等监控指标。当弹性云服务器再次启动或恢复后，即可正常查看。

### 在弹性伸缩组页面查看

- 登录管理控制台。
- 选择“计算 > 弹性伸缩 > 伸缩实例 > 弹性伸缩组”。
- 选择需要查看监控数据的伸缩组，单击伸缩组名称进入详情页面。
- 单击“监控”页签，查看伸缩组各项监控指标的数据。


支持查看“近1小时”、“近3小时”、“近12小时”、“近24小时”、“近7天”的数据。如果您想查看更长时间范围的监控曲线，请单击“查看更多指标详

情”跳转至云监控页面，在监控视图中单击  图标，进入大图模式查看。

### 在云监控页面查看

- 登录管理控制台。
- 在管理控制台左上角单击  图标，选择区域和项目。
- 选择“管理与监管 > 云监控服务”。

4. 单击页面左侧的“云服务监控”，选择“弹性伸缩”。
5. 单击“操作”列的“查看监控指标”，查看伸缩组各项监控指标的数据。支持查看“近1小时”、“近3小时”、“近12小时”、“近24小时”、“近7

天”的数据。如果您想查看更长时间范围的监控曲线，请在监控视图中单击图标，进入大图模式查看。

#### 说明


由于监控数据的获取与传输会花费一定时间，因此，请等待一段时间后再查看监控数据。

## 6.7 设置监控告警规则

### 操作场景

通过设置弹性云服务器告警规则，用户可自定义监控目标与通知策略，及时了解弹性云服务器运行状况，从而起到预警作用。

### 操作步骤

1. 登录管理控制台。
2. 在管理控制台左上角单击图标，选择区域和项目。
3. 选择“管理与监管 > 云监控服务”。
4. 在左侧导航树栏，选择“告警 > 告警规则”。
5. 在“告警规则”界面，单击“创建告警规则”创建弹性伸缩的告警规则，或者选择已有的弹性伸缩的告警规则进行修改，设置弹性伸缩的告警规则。
6. 规则参数设置完成后，单击“立即创建”。

#### 说明

- 更多关于设置告警规则的信息，请参见《云监控用户指南》。
- 您可以使用在云监控页面创建的告警规则，实现动态资源扩展。

# 7 权限管理

## 7.1 创建用户并授权使用 AS

### 操作场景

如果您需要对您所拥有的AS进行精细的权限管理，您可以使用[统一身份认证服务](#)（Identity and Access Management，简称IAM），通过IAM，您可以：

- 根据企业的业务组织，在您的华为云帐号中，给企业中不同职能部门的员工创建IAM用户，让员工拥有唯一安全凭证，并使用AS资源。
- 根据企业用户的职能，设置不同的访问权限，以达到用户之间的权限隔离。
- [统一身份认证服务](#)将AS资源委托给更专业、高效的其他华为云帐号或者云服务，这些帐号或者云服务可以根据权限进行代运维。

如果华为云帐号已经能满足您的要求，不需要创建独立的IAM用户，您可以跳过本章节，不影响您使用AS服务的其它功能。

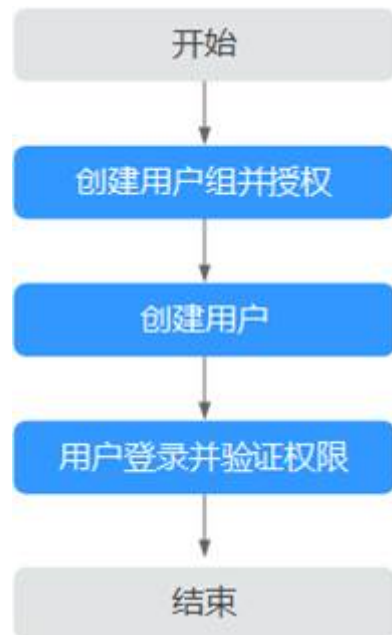
本章节为您介绍对用户授权的方法，操作流程如[图7-1](#)所示。

### 前提条件

给用户组授权之前，请您了解用户组可以添加的AS系统权限，并结合实际需求进行选择，AS支持的系统权限，请参见：[AS系统权限](#)。若您需要对除AS之外的其它服务授权，IAM支持服务的所有策略请参见[系统权限](#)。

## 示例流程

图 7-1 给用户授权 AS 权限流程



### 1. 创建用户组并授权

在IAM控制台创建用户组，并授予弹性伸缩权限“AutoScaling ReadOnlyAccess”。

### 2. 创建用户并加入用户组

在IAM控制台创建用户，并将其加入1中创建的用户组。

### 3. 用户登录并验证权限

新创建的用户登录控制台，切换至授权区域，验证权限：

- 在“服务列表”中选择弹性伸缩服务，进入AS主界面，单击右上角“创建伸缩组”，如果无法创建伸缩组（假设当前权限仅包含AutoScaling ReadOnlyAccess），表示“AutoScaling ReadOnlyAccess”已生效。
- 在“服务列表”中选择除弹性伸缩服务外（假设当前策略仅包含ECS Viewer）的任一服务，若提示权限不足，表示“AutoScaling ReadOnlyAccess”已生效。

## 7.2 创建 AS 自定义策略

### 操作场景

如果系统预置的AS权限，不满足您的授权要求，可以创建自定义策略。自定义策略中可以添加的授权项（Action）请参考[策略及授权项说明](#)。

目前华为云支持以下两种方式创建自定义策略：

- 可视化视图创建自定义策略：无需了解策略语法，按可视化视图导航栏选择云服务、操作、资源、条件等策略内容，可自动生成策略。

- JSON视图创建自定义策略：可以在选择策略模板后，根据具体需求编辑策略内容；也可以直接在编辑框内编写JSON格式的策略内容。

具体创建步骤请参见：[创建自定义策略](#)。本章为您介绍常用的AS自定义策略样例。

## AS 自定义策略样例

- 示例1：授权用户移出伸缩组实例、创建弹性伸缩配置。

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "as:instances:delete",
        "as:configs:create"
      ]
    }
  ]
}
```

- 示例2：拒绝用户删除伸缩组。

拒绝策略需要同时配合其他策略使用，否则没有实际作用。用户被授予的策略中，一个授权项的作用如果同时存在Allow和Deny，则遵循Deny优先。

如果您给用户授予AutoScaling FullAccess的系统策略，但不希望用户拥有AutoScaling FullAccess中定义的删除伸缩组权限，您可以创建一条拒绝删除伸缩组的策略，然后同时将AutoScaling FullAccess和拒绝策略授予用户，根据Deny优先原则，则用户可以对AS执行除了删除伸缩组外的所有操作。拒绝策略示例如下：

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "as:groups:delete"
      ],
      "Effect": "Deny"
    }
  ]
}
```



# A 修订记录

版本日期	变更说明
2021-11-20	第二十七次正式发布。本次变更如下： <a href="#">创建伸缩组</a> ，新增“源/目的检查”参数。
2021-10-29	第二十六次正式发布。 本次变更如下： <a href="#">创建伸缩组</a> 和 <a href="#">使用已有云服务器创建伸缩配置</a> 章节删除“文件注入”功能，修改“用户数据注入”为“实例自定义数据注入”。
2020-10-19	第二十五次正式发布。 本次变更如下： 增加“访问方式”章节。
2020-02-10	第二十四次正式发布。 本次变更如下： AS系统权限名称变更。
2019-08-30	第二十三次正式发布。 本次变更如下： <a href="#">监控指标说明</a> 章节新增Agent监控指标。
2019-05-30	第二十二次正式发布。 本次变更如下： <ul style="list-style-type: none"><li>适配控制台界面变更，刷新伸缩组和伸缩配置的操作步骤和相关参数。</li></ul>
2019-04-30	第二十一次正式发布。 本次变更如下： <ul style="list-style-type: none"><li>优化（可选）<a href="#">添加负载均衡器到伸缩组</a>内容。</li><li>新增<a href="#">伸缩策略介绍</a>章节。</li><li>优化<a href="#">创建伸缩带宽策略</a>章节内容，新增“新建告警规则”相关内容。</li></ul>

版本日期	变更说明
2019-03-31	<p>第二十次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● <a href="#">创建伸缩组</a>章节新增“多可用区扩展策略”参数。</li><li>● <a href="#">使用新模板创建伸缩配置</a>章节新增“规格使用优先策略”参数，支持规格配置功能。</li><li>● <a href="#">使用新模板创建伸缩配置</a>章节新增<a href="#">多规格配置说明</a>。</li><li>● <a href="#">管理伸缩带宽策略</a>章节新增“删除告警策略使用的告警规则”相关内容，删除伸缩带宽策略时支持同时删除告警策略使用的告警规则。</li></ul>
2019-01-30	<p>第十九次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● 优化伸缩策略和伸缩带宽相关内容。</li><li>● 新增<a href="#">实例备用</a>章节。</li></ul>
2018-12-30	<p>第十八次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● <a href="#">使用新模板创建伸缩配置</a>章节新增专属主机参数。</li></ul>
2018-11-30	<p>第十七次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● <a href="#">查询伸缩活动</a>章节新增查询活动历史操作指导。</li><li>● 常见问题添加分类。</li><li>● 新增常见问题“什么是期望实例数？”、“停用伸缩组后，什么操作会暂停？”、“伸缩组中实例为什么无法通过负载均衡健康检查而频繁地被删除后重新创建？”、“告警策略支持的告警触发条件有哪些？”和“伸缩组已经添加了负载均衡，创建伸缩配置时是否还需要配置弹性公网IP？”。</li><li>● 优化常见问题“如何自动部署应用”和“什么是冷却时间，为什么需要冷却时间？”。</li></ul>
2018-10-31	<p>第十六次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● 优化<a href="#">生命周期挂钩</a>内容。</li><li>● 新增<a href="#">创建伸缩配置</a>章节。</li></ul>
2018-09-30	<p>第十五次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● “弹性IP”参数变更为“弹性公网IP”。</li><li>● <a href="#">使用已有云服务器创建伸缩配置</a>和<a href="#">使用新模板创建伸缩配置</a>章节新增“云服务器组”参数。</li><li>● <a href="#">使用已有云服务器创建伸缩配置</a>和<a href="#">使用新模板创建伸缩配置</a>章节新增“带宽类型”参数。</li></ul>

版本日期	变更说明
2018-08-30	<p>第十四次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● 冷却时间计算方式变更为从伸缩活动完成开始，且默认值变更为300秒。</li><li>● 新增常见问题“什么是冷却时间，系统如何计算冷却时间？”。</li><li>● <b>伸缩带宽</b>章节增加伸缩带宽应用场景描述。</li></ul>
2018-07-30	<p>第十三次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● 优化“快速创建弹性伸缩”、“创建伸缩组”、“动态扩展资源”章节。</li><li>● 新增常见问题“当实例被移出伸缩组并删除后，实例中的数据会保留吗？”。</li><li>● 新增常见问题“弹性伸缩是否可以根据云监控中自定义监控进行动态伸缩？”。</li><li>● 新增常见问题“弹性伸缩是否能够自动升降云服务器的CPU、内存和带宽？”。</li><li>● <b>动态扩展资源</b>章节触发告警参数选择新建告警时，支持选择自定义监控。</li><li>● <b>创建伸缩组</b>章节健康检查间隔新增可设置间隔10秒和1分钟。</li></ul>
2018-06-30	<p>第十二次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● 优化“创建伸缩组”、“创建伸缩配置”章节。</li><li>● 新增常见问题“伸缩策略启用失败如何处理？”。</li><li>● 新增常见问题“我能添加已经创建的包年包月ECS实例吗？”。</li><li>● 新增常见问题“如何阻止伸缩组内的ECS实例被自动移除？”。</li></ul>
2018-05-30	<p>第十一次正式发布。</p> <p>本次变更如下：</p> <ul style="list-style-type: none"><li>● 新增弹性带宽特性。</li><li>● <b>使用新模板创建伸缩配置</b>新增“安全组”参数。</li><li>● 优化负载均衡参数说明。</li><li>● 新增常见问题“伸缩组启用失败如何处理？”。</li><li>● 新增常见问题“如何处理伸缩组中状态是“异常”的实例？”。</li></ul>

版本日期	变更说明
2018-04-30	第十次正式发布。 本次变更如下： <ul style="list-style-type: none"><li>● 新增<a href="#">监控指标说明</a>章节。</li><li>● 新增<a href="#">查看监控指标数据</a>章节。</li><li>● 新增<a href="#">设置监控告警规则</a>章节。</li><li>● 新增“如何查看审计日志”内容。</li><li>● 新增常见问题“如何删除通过弹性伸缩创建的云服务器?”。</li></ul>
2018-03-30	第九次正式发布。 本次变更如下： <ul style="list-style-type: none"><li>● 优化关于用户数据注入的说明。</li><li>● 优化对实例的健康状态进行检查的描述。</li></ul>
2018-01-30	第八次正式发布。 本次变更如下： <ul style="list-style-type: none"><li>● 新增强制删除伸缩组的相关描述。</li><li>● 新增弹性负载均衡（经典型）相关描述。</li><li>● 新增整机镜像相关描述。</li></ul>
2017-11-30	第七次正式发布。 本次变更如下： <ul style="list-style-type: none"><li>● 增加了弹性伸缩监控指标。</li><li>● 伸缩策略支持实例百分比方式进行扩容和缩容。</li><li>● 新增标记伸缩组和实例。</li></ul>
2017-09-30	第六次正式发布。 本次变更如下： <ul style="list-style-type: none"><li>● 新增实例保护章节。</li><li>● 新增最佳实践章节。</li></ul>
2017-07-30	第五次正式发布。 <ul style="list-style-type: none"><li>● 优化向导式创建弹性伸缩的流程，即用户创建完伸缩组后必须新建或使用已有伸缩配置。</li><li>● 在伸缩组页面的“操作”列下，新增“查看伸缩策略”，“操作”列下的“更多”新增“更换伸缩配置”。</li></ul>

版本日期	变更说明
2016-12-30	第四次正式发布。 新增： <ul style="list-style-type: none"><li>配置伸缩配置时增加“弹性IP”参数。</li><li>配置伸缩组时增加“可用区”参数。</li><li>生命周期挂钩特性。</li><li>支持使用SMN Topic方式发邮件。</li></ul>
2016-10-19	第三次正式发布。 新增： 支持Cloud-Init功能。
2016-07-15	第二次正式发布。 新增： 镜像类型：共享镜像。
2016-03-14	第一次正式发布。