弹性伸缩

用户指南

文档版本	01
发布日期	2020-11-05

华为技术有限公司

商标声明


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默示的声明或保证。

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

1 产品介绍

1.1 什么是弹性伸缩?

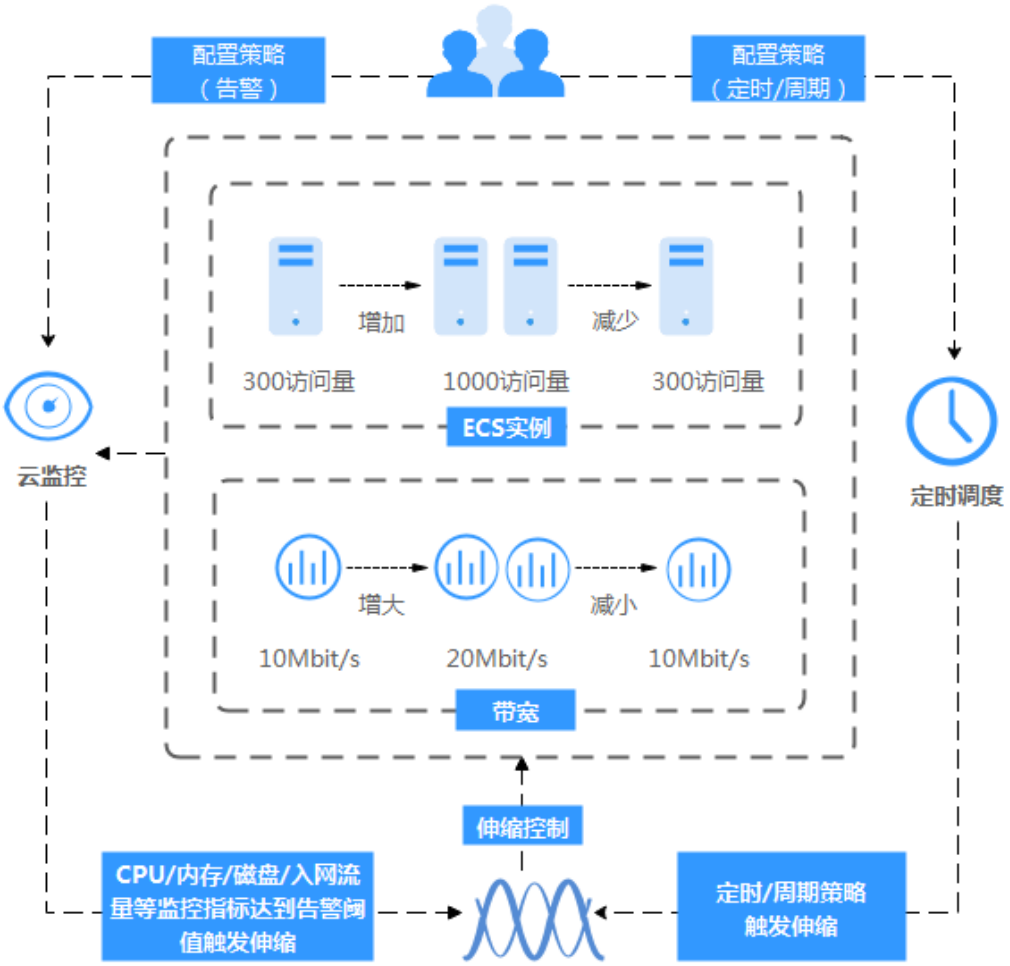
弹性伸缩（Auto Scaling）是根据用户的业务需求，通过策略自动调整其业务资源的服务。您可以根据业务需求自行定义伸缩策略，从而降低人为反复调整资源，以应对业务变化和负载高峰的工作量，帮您节约资源和人力运维成本。弹性伸缩支持自动调整弹性云服务器和带宽资源。

产品架构

通过伸缩控制可以实现弹性云服务器（ECS）实例伸缩和带宽伸缩：

- 伸缩控制：配置策略设置指标阈值/伸缩活动执行的时间，通过云监控监控指标是否达到阈值，通过定时调度，实现伸缩控制。
- 配置策略：可以根据业务需求，配置告警策略/定时策略/周期策略。
- 配置告警策略：可配置CPU、内存、磁盘、入网流量等监控指标。
- 配置定时策略：通过配置触发时间可以配置定时策略。
- 配置周期策略：通过配置重复周期、触发时间、生效时间可以配置周期策略。
- 云监控监控到所配置的告警策略中的某些指标达到告警阈值，从而触发伸缩活动，实现ECS实例的增加/减少或带宽的增大/减小。
- 到达所配置的触发时间时，触发伸缩活动，实现ECS实例的增加/减少或带宽的增大/减小。

图 1-1 弹性伸缩产品架构图



访问方式

公有云提供了Web化的服务管理平台，即管理控制台和基于HTTPS请求的API（Application programming interface）管理方式。

- API方式

  如果用户需要将公有云平台上的弹性伸缩服务集成到第三方系统，用于二次开发，请使用API方式访问弹性伸缩服务，具体操作请参见**《弹性伸缩API参考》**。
- 控制台方式

  其他相关操作，请使用管理控制台方式访问弹性伸缩服务。

  如果用户已注册公有云，可直接登录管理控制台，从主页选择“弹性伸缩”。

1.2 弹性伸缩的优势

弹性伸缩服务可根据用户的业务需求，通过策略自动调整其业务的资源。具有自动调整资源、节约成本开支、提高可用性和容错能力的优势。适用以下场景：

- 访问流量较大的论坛网站，业务负载变化难以预测，需要根据实时监控到的云服务器CPU使用率、内存使用率等指标对云服务器数量进行动态调整。

- 电商网站，在进行大型促销活动时，需要定时增加云服务器数量和带宽大小，以保证促销活动顺利进行。
- 视频直播网站，每天14:00~16:00播出热门节目，每天都需要在该时段增加云服务器数量，增大带宽大小，保证业务的平稳运行。

自动调整资源

弹性伸缩能够实现应用系统自动按需调整资源，即在业务增长时能够实现自动增加实例数量和带宽大小，以满足业务需求，业务下降时能够实现应用系统自动缩容，保障业务平稳运行。

- 按需调整云服务器资源

  向应用系统中添加弹性伸缩，能够实现按需调整资源，即能够实现在业务增长时增加实例，业务下降时减少实例，这样加强了应用系统的成本管理。调整资源主要包括以下几种方式：

  - 动态调整资源

    动态调整资源是通过告警策略的触发来调整资源。详细内容请参阅**动态扩展资源**。
  - 计划调整资源

    计划调整资源是通过定时策略或周期策略的触发来调整资源。详细内容请参阅**按计划扩展资源**。
  - 手工调整资源

    通过修改期望实例数或手动移入、移出实例来调整资源。详细内容请参阅**手动扩展资源**。

  例如，运行在公有云上的基本Web应用程序。此应用程序允许乘客购买火车票。在每年中期时段，人员流动性较低，此应用程序的使用率较低。每年年底和年初，人员流动性较高，因此对此应用程序的需求会显著提高。一般系统会采用添加足够多的服务器，如**图1-2**所示，或添加处理应用程序平均需求所需的容量，如**图1-3**所示，来满足业务需求。但这两种方案会造成资源浪费或无法满足高峰期的需求。当您给应用程序中添加弹性伸缩后，弹性伸缩会自动根据需求调整服务器的数量，如**图1-4**所示，为您节约成本并且满足高峰期的需求。

  图 1-2 服务器资源冗余

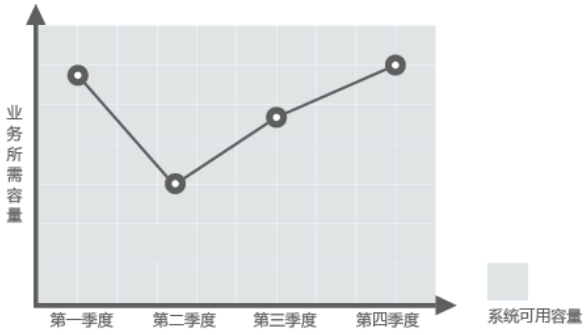


图 1-3 服务器资源不足

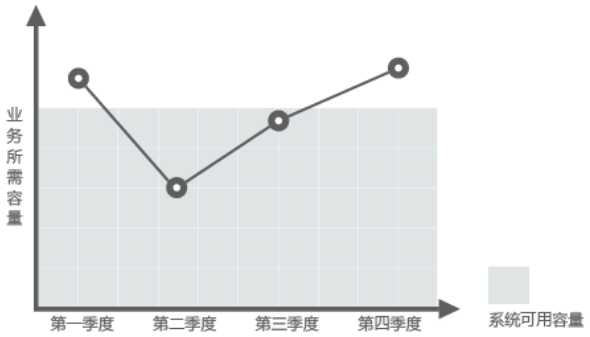
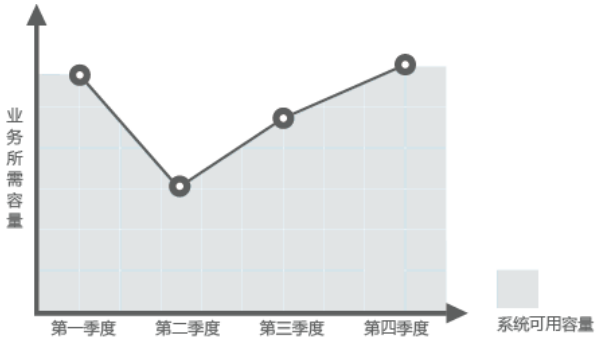


图 1-4 向应用程序中添加弹性伸缩



- 按需调整带宽资源

  弹性伸缩能够实现按需调整带宽，即能够实现在业务增长时扩大带宽，业务下降时减小带宽，加强了应用系统的成本管理。

  您可以根据实际情况选择如下伸缩带宽策略来实现按需调整IP带宽：

  - 告警策略

    可设置出网流量、出网带宽等告警触发条件，系统检测到触发条件满足时，会自动调整带宽的大小。

  - 定时策略

    系统可根据定时策略在固定的时间自动将带宽增大、减小或者调整到固定的值。

  - 周期策略

    系统可根据周期策略周期性的调整带宽大小，减少了人工重复设置带宽的工作量。

  以告警策略的使用为例说明如下：

  某视频直播网站，在不同时间段业务负载变化难以预测，需要根据出网流量、入网流量等指标在10Mbit/s到30Mbit/s之间动态调整带宽资源。弹性伸缩可以实现自动按需调整带宽，很好的解决这个问题。您只需选择需要调整的弹性公网IP，同时创建两个告警策略，一个策略设置在出网流量大于XXXbyte时，增加2Mbit/s，限制值为30Mbit/s；另一个策略在出网流量小于XXXbyte时，减少2Mbit/s，限制值为10Mbit/s。

- 按可用区均匀分配实例

  按可用区均匀分配实例是指尽可能地将实例均匀的分布在不同的可用区中，来降低电力、网络等可能出现的故障对整个系统稳定性的影响。

区域指弹性云服务器云主机所在的物理位置。每个区域包含许多不同的称为“可用区”的位置，即在同一区域下，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。每个可用区都被设计成不受其他可用区故障影响的模式，并提供低价、低延迟的网络连接，以连接到同一地区其他可用区。

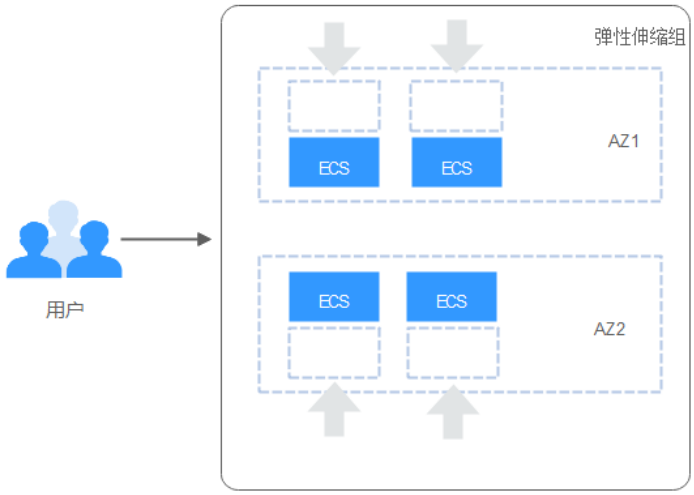
伸缩组可以包含来自同一区域的一个或多个可用区的实例。在资源调整时，弹性伸缩会通过实例分配和再均衡两种方法尽可能的将实例均匀分配到可用区中。

实例分配

弹性伸缩尝试在为伸缩组使用的可用区之间均匀分配实例。弹性伸缩通过尝试向实例最少的可用区中移入新实例来实现此目标。

例如，伸缩组目前有四个实例均匀分布在两个可用区内，若该伸缩组下一个伸缩活动增加四个实例时，会在两个可用区内分别增加两个实例，以实现可用区之间均匀分配实例。

图 1-5 均匀实例分配

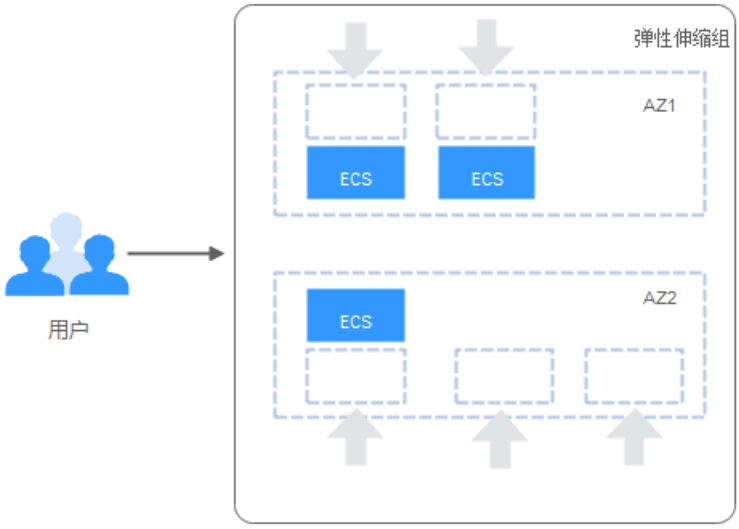


再均衡

手工加入或移出实例后，伸缩组中的实例没有均匀分配在可用区时，后续进行的伸缩活动会优先在可用区内均匀分配实例。

例如，伸缩组中目前有三个实例分布在两个可用区内，若该伸缩组下一个伸缩活动增加五个实例时，会在有两个实例的可用区内增加两个实例，在有一个实例的可用区增加三个实例，以实现可用区之间均匀分配实例。

图 1-6 再均衡



加强成本管理

弹性伸缩能够实现按需使用实例和带宽，并自动调整系统中的资源，节省了资源和人为调整资源带来的损耗，为您最大程度节约了成本。

提高可用性

弹性伸缩可确保应用系统始终拥有合适的容量以满足当前流量需求。当弹性伸缩和负载均衡器结合后，伸缩组会自动地为新加入的实例绑定负载均衡监听器。访问流量将通过负载均衡监听器自动分发到伸缩组内的所有实例。

弹性伸缩和负载均衡结合使用

当您在使用弹性伸缩时，业务增长时应用系统自动扩容，业务下降时应用系统自动缩容，在伸缩组添加和删除实例时，须确保所有实例均可分配到应用程序的流量。弹性伸缩和负载均衡结合使用可以解决这个问题。

使用负载均衡后，伸缩组会自动地将加入伸缩组的实例绑定负载均衡监听器。访问流量将通过负载均衡监听器自动分发到伸缩组内的所有实例，提高了应用系统的可用性。若伸缩组中的实例上部署了多个业务，还可以添加多个负载均衡监听器到伸缩组，同时监听多个业务，从而提高业务的可扩展性。

提高容错能力

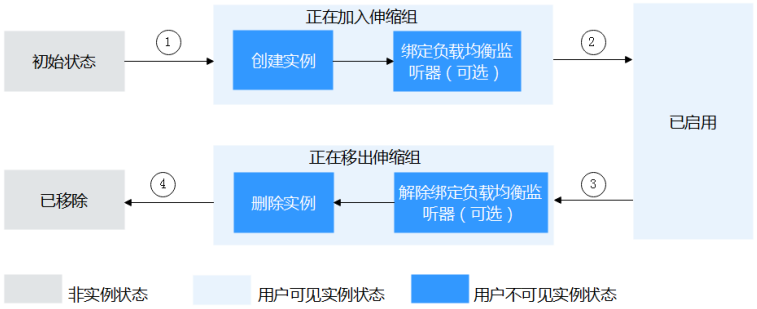
弹性伸缩可以检测到应用系统中实例的运行状况，并启用新实例以替换运行状况不佳的实例。

1.3 生命周期

伸缩组中的实例生命周期，从创建实例开始，到该实例从伸缩组中移除结束。

伸缩组中未添加生命周期挂钩时，实例生命周期内状态之间的过渡如图1-7所示。

图 1-7 实例生命周期内状态之间的过渡



触发条件②和④表示系统自发触发实例状态的改变。

表 1-1 实例的状态

实例所处状态	子状态	实例状态含义	触发条件
初始状态	-	即实例还没状态。	触发条件①包括有两种情况，只要其中一种情况就能够触发实例进入“正在加入伸缩组”状态。 ● 手动修改期望实例数或满足伸缩策略的条件时，触发伸缩活动，进行扩容。 ● 手动添加已有实例至伸缩组。
正在加入伸缩组	创建实例	在触发条件①的作用下，伸缩组开始扩容，创建实例。	
	绑定负载均衡监听器（可选）	在触发条件①的作用下，创建实例完成后，实例绑定负载均衡监听器。	
已启用	-	实例进入伸缩组，开始接受处理业务流量。	触发条件③包括有三种情况，只要其中一种情况就能够触发实例从“已启用”状态到“正在移出伸缩组”状态： ● 手动修改期望实例数或满足伸缩策略的条件时，触发伸缩活动，进行缩容。 ● 实例进入启用状态后，开始健康检查，健康检查失败后，移出实例。 ● 手动将实例移出伸缩组。
正在移出伸缩组	解除绑定负载均衡监听器（可选）	在触发条件③的作用下，伸缩组开始缩容，实例解除绑定负载均衡监听器。	
	删除实例	实例解除绑定负载均衡监听器后，从伸缩组中移出。	
已移除	-	实例在伸缩组中的生命周期已结束，即实例没有状态。	-

通过手动添加实例和伸缩活动向伸缩组添加实例，实例经过正在加入伸缩组、已启用和正在移出伸缩组状态后，实例将移出伸缩组。

伸缩组中已添加生命周期挂钩后，实例生命周期内状态之间的过渡如图1-8所示。当伸缩组进行伸缩活动时，实例将被生命周期挂钩挂起并置于等待状态，实例将保持此状

态直至超时时间结束或者用户手动回调。用户能够在实例保持等待状态的时间段内执行自定义操作，例如，用户可以在新移入的实例上安装或配置软件，也可以在实例终止前从实例中下载日志文件。

图 1-8 实例生命周期内状态之间的过渡

触发条件②、④、⑥、⑧表示系统自发触发实例状态的改变。

表 1-2 实例状态

实例所处状态	子状态	实例状态含义	触发条件含义
初始状态	-	即实例还没状态。	触发条件①包括有两种情况，只要其中一种情况就能够触发实例进入“正在加入伸缩组”状态。 ● 手动修改期望实例数或满足伸缩策略的条件时，触发伸缩活动，进行扩容。 ● 手动添加已有实例至伸缩组。
正在加入伸缩组	创建实例	在触发条件①的作用下，伸缩组开始扩容，创建实例。	
等待（正在加入伸缩组）	-	正在加入伸缩组的实例被生命周期挂钩挂起，将实例至于等待的状态。	触发条件③包括有两种情况，只要其中一种情况就能够触发实例从“等待（正在加入伸缩组）”到“正在加入伸缩组”状态。 ● 默认回调操作 ● 手动回调操作
正在加入伸缩组	绑定负载均衡监听器（可选）	在触发条件③的作用下，实例将继续正在加入伸缩组，绑定负载均衡监听器。	
已启用	-	实例进入伸缩组，开始接受处理业务流量。	触发条件⑤包括有三种情况，只要其中一种情况就能够触发实例从“已启用”状态到“正在移出伸缩组”状态：

实例所处状态	子状态	实例状态含义	触发条件含义
正在移出伸缩组	解除绑定负载均衡监听器（可选）	在触发条件⑤的作用下，伸缩组开始缩容，实例解除绑定负载均衡监听器。	• 手动修改期望实例数或满足伸缩策略的条件时，触发伸缩活动，进行缩容。 • 实例进入启用状态后，开始健康检查，健康检查失败后，移出实例。 • 手动将实例移出伸缩组。
等待（正在移出伸缩组）	-	正在移出伸缩组的实例被生命周期挂钩挂起，将实例至于等待的状态。	触发条件⑦包括有两种情况，只要其中一种情况就能够触发实例从“等待（正在移出伸缩组）”到“正在移出伸缩组”状态。 • 默认回调操作 • 手动回调操作
正在移出伸缩组	删除实例	在触发条件⑦的作用下，实例将继续正在移出伸缩组，删除实例。	
已移除	-	实例在伸缩组中的生命周期已结束，即实例没有状态。	-

1.4 使用限制

在应用系统中添加弹性伸缩后，使用时有一定的限制，使用限制如下所示：

- 弹性伸缩的云服务器中运行的应用需要是无状态、可横向扩展的。

  说明

  - 无状态：关于应用的既往事务，没有任何记录和参考，每项事务处理均是从头开始。
    无状态应用运行的实例不会在本地存储需要持久化的数据。
    例如：可以将无状态事务看作一台自动售货机：一个请求对应一个响应。
  - 有状态：是可以周而复始、反复发生的应用和流程，操作是在之前的事务背景下执行的，当前事务可能会受到之前事务的影响。
    有状态应用运行的实例会在本地存储需要持久化的数据。
    例如：可以将有状态事务看作网上银行或电子邮件，有上下文记录。
- 弹性伸缩会自动释放云服务器，所以弹性伸缩组内的云服务器不可以保存应用的状态信息（例如session）和相关数据（如数据库、日志等）。如果应用中需要云服务器保存状态或日志信息，可以考虑把相关信息保存到独立的服务器中。
- 弹性伸缩无法纵向扩展，即弹性伸缩无法自动升降ECS实例的vCPU和内存等配置。
- 弹性伸缩对用户的资源数量或容量做的配额限制如**表1-3**所示。

表 1-3 配额一览表

类别	描述	默认值
弹性伸缩组	用户可以创建的最多伸缩组个数。	10
弹性伸缩配置	用户可以创建的最多伸缩配置个数。	100
弹性伸缩策略	某个弹性伸缩组下可以创建的最多伸缩策略个数。	10
弹性伸缩实例	某个弹性伸缩组下可以创建的最多实例个数。	300
伸缩带宽策略	用户最多可以创建的伸缩带宽策略个数。	10

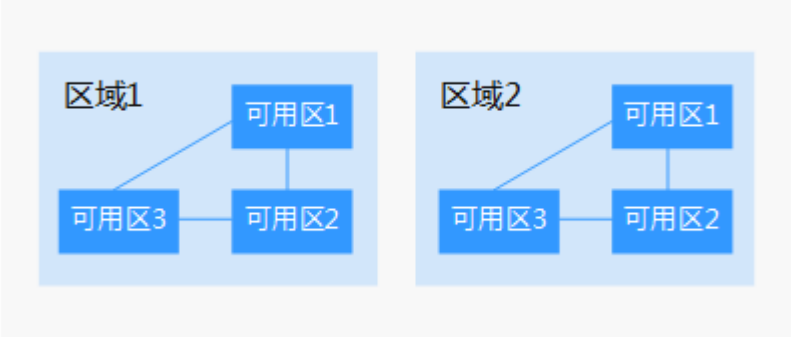
1.5 区域和可用区

我们用区域和可用区来描述数据中心的位置，您可以在特定的区域、可用区创建资源。

- 区域（Region）：从地理位置和网络时延维度划分，同一个Region内共享弹性计算、块存储、对象存储、VPC网络、弹性公网IP、镜像等公共服务。Region分为通用Region和专属Region，通用Region指面向公共租户提供通用云服务的Region；专属Region指只承载同一类业务或只面向特定租户提供业务服务的专用Region。
- 可用区（AZ，Availability Zone）：一个AZ是一个或多个物理数据中心的集合，有独立的风火水电，AZ内逻辑上再将计算、网络、存储等资源划分成多个集群。一个Region中的多个AZ间通过高速光纤相连，以满足用户跨AZ构建高可用性系统的需求。

图1-9阐明了区域和可用区之间的关系。

图 1-9 区域和可用区



如何选择区域?

建议就近选择靠近您或者您的目标用户的区域，这样可以减少网络时延，提高访问速度。

如何选择可用区?

是否将资源放在同一可用区内，主要取决于您对容灾能力和网络时延的要求。

- 如果您的应用需要较高的容灾能力，建议您将资源部署在同一区域的不同可用区内。
- 如果您的应用要求实例之间的网络延时较低，则建议您将资源创建在同一可用区内。

区域和终端节点

当您通过API使用资源时，您必须指定其区域终端节点。有关公有云的区域和终端节点的更多信息，请参阅**地区和终端节点**。

1.6 与其他服务的关系

除直接使用弹性伸缩提供的对资源进行调整的功能外，若您同时购买了云服务中的其他产品，可以结合其他产品一起使用，能满足您多种场景下对云产品的需求。

弹性伸缩服务与周边服务的依赖关系如**图1-10**所示。

图 1-10 弹性伸缩服务与其他服务的关系示意图

表 1-4 弹性伸缩与其他服务的关系

服务名称	说明	交互功能	相关内容
弹性负载均衡（Elastic Load Balance）	当配置了负载均衡服务后，弹性伸缩组在添加或移除云服务器时，自动会为云服务器绑定或解绑负载均衡监听器。	使伸缩组中每一个实例均可分配到应用程序流量	**（可选）添加负载均衡器到伸缩组**
云监控服务（Cloud Eye）	弹性伸缩配置了告警触发策略时，会根据云监控的告警条件触发弹性伸缩活动。	通过监控伸缩组内实例的状态指标调节资源。	**弹性伸缩支持的监控指标**
弹性云服务器（Elastic Cloud Server）	弹性伸缩活动中添加的云服务器可以通过弹性云服务器进行管理和维护。	自动调整弹性云服务器数量	**动态扩展资源**和**按计划扩展资源**
虚拟私有云（Virtual Private Cloud）	弹性伸缩支持自动调整虚拟私有云中创建的弹性公网IP带宽或共享带宽大小。	自动调整带宽大小	**创建伸缩带宽策略**
消息通知服务（Simple Message Notification）	用户使用消息通知功能后，系统会将伸缩组的多种情况及时推送给用户，便于用户及时了解伸缩组的状态。	消息通知	**为伸缩组配置通知**
云审计服务（Cloud Trace Service	开通云审计服务后，可以记录弹性伸缩相关的操作事件，便于日后的查询、审计和回溯。	日志审计	**记录弹性伸缩**
标签管理服务（Tag Management Service）	当您具有许多相同类型的弹性伸缩资源时，标签可以为您提供灵活的资源管理能力。	标签	**标记伸缩组和实例**

服务名称	说明	交互功能	相关内容
云数据库服务（Relational Database Service）	伸缩出来的实例，可以直接访问RDS实例的前提条件是： • 该实例与目标RDS实例必须处于同一VPC内； • 该实例必须处于目标RDS实例所属安全组允许访问的范围内；	伸缩出来的实例可以访问云数据库实例	通过内网连接MySQL实例

1.7 基本概念

伸缩组

伸缩组是具有相同应用场景的实例的集合，是启停伸缩策略和进行伸缩活动的基本单位。

伸缩配置

伸缩配置是伸缩组内实例（弹性云服务器）的模板，定义了伸缩组内待添加的实例的规格数据。包括云服务器类型、vCPU、内存、镜像、磁盘、登录方式等。

伸缩策略

伸缩策略可以触发伸缩活动，是对伸缩组中实例数量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。

伸缩活动

伸缩组中增加或减少实例的过程称为伸缩活动。伸缩活动的目的是使应用系统中当前实例数和期望实例数保持一致，或达到已设置的伸缩策略触发条件时，执行增加或减少实例数量的操作，保证业务正常运行。

冷却时间

为了避免告警策略频繁触发，必须设置冷却时间。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。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。

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。

伸缩带宽

伸缩带宽可以根据用户配置的伸缩带宽策略自动调整带宽资源。弹性伸缩仅支持对按需购买的弹性公网IP带宽和共享带宽进行调整，不支持对包年包月的带宽进行调整。

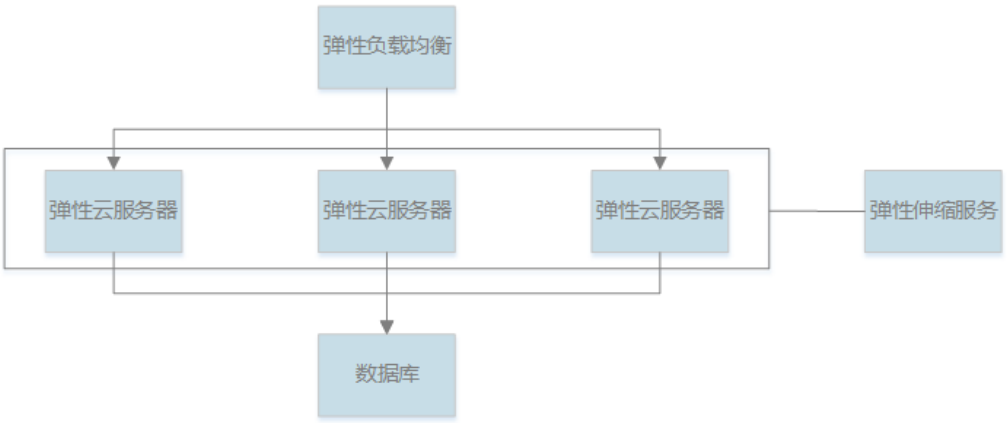
2 最佳实践

2.1 搭建可自动伸缩的 Discuz!论坛网站

场景介绍

通过使用弹性伸缩，您可以在需要时向应用程序添加新实例，并在不需要时将其移出。对于预期内的营销活动或者未知的业务高峰，无需提前准备大量云服务器，从而提高了系统运行的稳定性，同时降低了成本。

本例以搭建Discuz!论坛为例，介绍了如何使用弹性伸缩、弹性云服务器、虚拟私有云、弹性负载均衡等服务搭建一个可自动横向伸缩的web服务。



场景要求

1. 在虚拟私有云页面已完成虚拟私有云和弹性公网IP的申请。
2. 在弹性负载均衡页面已完成负载均衡和监听器的创建，创建负载均衡时选择的VPC为1中申请的VPC。

操作步骤

创建一台弹性云服务器用于安装数据库

数据库可以使用云平台的关系型数据库，也可以自行创建弹性云服务器安装所需的数据库。这里是在创建的弹性云服务器上安装MySQL数据库。

1. 创建弹性云服务器时，根据界面提示配置参数，选择相应规格的云服务器，网络相关参数选择刚创建的虚拟私有云、安全组以及弹性公网IP，完成弹性云服务器的创建。具体创建弹性云服务器的操作请参考**《弹性云服务器用户指南》**。
2. 待弹性云服务器页面上刚创建的弹性云服务器的状态为“运行中”时，即表示该弹性云服务器创建完成，就可以对这台弹性云服务器进行操作了，使用XFtp、Xshell等工具连接弹性云服务器的弹性公网IP，完成MySQL数据库的安装配置。

创建一台弹性云服务器用于安装Discuz!论坛

1. 创建弹性云服务器，此时可以不添加弹性公网IP，具体创建弹性云服务器的操作请参考**《弹性云服务器用户指南》**。
2. 解绑和绑定弹性公网IP。

   由于可通过私网访问数据库，因此可以将之前用于绑定数据库节点的弹性公网IP解绑以节省资源。将弹性公网IP和数据库节点解绑定，后将新创建的弹性云服务器和该弹性公网IP绑定，具体操作请参考**《虚拟私有云用户指南》**。此时即可通过公网访问该云服务器，安装PHP、Apache等环境。
3. 安装论坛。

   环境安装完成后，即可进行Discuz!论坛的安装，安装方法可参考Discuz官方文档。在进行配置参数时，数据库服务器填写之前安装MySQL云服务器的私网IP，而数据库用户名和密码为安装MySQL时所授权远程访问的用户名和密码，完成全部的安装操作后，可选择将弹性公网IP解绑释放以节省资源。

制作自定义镜像

将已安装了Discuz!论坛的云服务器制作成私有镜像，用于弹性伸缩组的自动扩容模板。

1. 只有关机状态的云服务器才可以制作私有镜像，进入弹性云服务器页面，将刚配置好的discuz云服务器关机，具体操作请参考**《弹性云服务器用户指南》**。
2. 进入镜像的页面，选择已安装有discuz云服务器制作私有镜像，具体操作请参考**《镜像服务用户指南》**。

创建伸缩组

伸缩组是具有相同属性和应用场景的云服务器、伸缩配置和伸缩策略的集合，也是弹性伸缩中启停伸缩策略和进行伸缩活动的基本单位。需要创建一个弹性伸缩组来实现Discuz!论坛服务的自动扩容和缩容。

创建伸缩组的具体操作请参考**创建伸缩组**，在进行界面参数配置时，选择已创建的虚拟私有云、子网、弹性负载均衡、监听器等参数。

创建伸缩配置

为伸缩组添加伸缩配置，伸缩配置指定了伸缩活动中自动添加的云服务器的基本规格。

镜像选择上述步骤中制作的私有镜像，其他参数根据您的需求完成填写。

手动加入实例到伸缩组

完成伸缩组和伸缩配置的创建后，进入伸缩组详情页面，切换到伸缩实例的选项卡，单击“移入伸缩组”，将已经创建好的运行中的discuz云服务器手动加入到伸缩组中。为保证手动加入的discuz云服务器不被自动移除伸缩组，可以对该实例设置实例保护。

创建伸缩策略

伸缩策略则是伸缩活动的触发条件，当满足条件时，会触发一次伸缩活动。在之前创建的伸缩组中，并未添加伸缩策略，为了使伸缩组能够根据业务需求自动扩容或缩容，这里我们需要添加相关弹性伸缩策略。

弹性伸缩支持动态扩展资源，可通过配置告警策略实现动态扩展资源的目的。告警策略即基于告警模式触发实例的伸缩，基于云监控系统告警数据，自动增加或减少云服务器。弹性伸缩同时支持按照可预测的负载变化来扩展资源，可通过配置定时策略和周期策略实现扩展资源。

当伸缩策略创建完成并处于启用状态，且触发条件满足时，可实现弹性伸缩组的扩容、缩容。

3 快速入门

3.1 弹性伸缩向导式使用流程

使用弹性伸缩的流程如图3-1所示。

图 3-1 弹性伸缩向导式使用流程

3.2 快速创建弹性伸缩

若您首次使用弹性伸缩，建议选择向导式创建弹性伸缩，具体操作请参考本章节。

前提条件

- 已经创建所需的VPC、子网、安全组、负载均衡器等。
- 如果使用密钥方式鉴权，还需要准备好密钥对。鉴权方式是指弹性伸缩活动中添加的云服务器的鉴权方式。

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“创建弹性伸缩组”。
4. 填写弹性伸缩组的基本信息，例如，名称、最大实例数、最小实例数、期望实例数。参数配置说明如表3-1所示。

表 3-1 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。	-
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。	-
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/最小实例数	最大/最小实例数是指伸缩组中云服务器个数的最大值/最小值。	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
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。	-
子网	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。	-
负载均衡	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。 **说明** • 一个伸缩组可最多添加6个负载均衡器。 • 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“弹性负载均衡健康检查”，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。	-

参数	解释	取值样例
实例移除策略	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： ● 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。 ● 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。 ● 较早创建的实例：创建时间较早的实例优先被移除伸缩组。 ● 较晚创建的实例：创建时间较晚的实例优先被移除伸缩组。 **说明** ● 当可用区不均衡时，移出实例时会优先保证可用区均衡。 ● 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。	-
健康检查方式	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： ● 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。	-
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。 **说明** ● “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。 ● 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。	-
高级配置	高级配置可对通知进行配置。 可选择“暂不配置”或“现在配置”。	-

参数	解释	取值样例
通知	通过消息通知服务提供的功能，将伸缩组的伸缩活动的结果及时推送给用户。 ● 每当伸缩组：每当伸缩组出现以下一种或几种场景时，向用户发送通知。 – 扩容成功 – 减容成功 – 异常 – 扩容失败 – 减容失败 ● 发送通知到：选择已经创建成功主题。请参见《消息通知服务用户指南》创建主题。	-
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。 ● 键： – 不能为空。 – 对于同一伸缩组，“键”唯一。 – 长度不超过36个字符。不能包含非打印字符ASCII(0-31)，“=”，“*”，“<”，“>”，“\”，“,”，“|”，“/”。 ● 值 – 可以为空字符串。 – 一个“键”只能添加一个“值”。 – 长度不超过43个字符。不能包含非打印字符ASCII(0-31)，“=”，“*”，“<”，“>”，“\”，“,”，“|”，“/”。	-

5. 单击“下一页”。
6. 在“伸缩配置”页面，您可以选择使用已有的伸缩配置或者即时创建新的伸缩配置。
7. 单击“下一页”。
8. （可选）为伸缩组添加伸缩策略。

   在“伸缩策略”页面，单击“添加伸缩策略”。

   配置伸缩活动触发的策略类型、执行动作、冷却时间等，根据界面提示进行参数配置。

   说明

   - 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。
   - 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。

9. 单击“立即创建”。
10. 请核对弹性伸缩组、伸缩配置和伸缩策略的信息。单击 “提交”。
11. 请仔细核对创建结果，并根据界面提示返回弹性伸缩组列表。
    创建伸缩组成功后，伸缩组状态变为“已启用”。

4 伸缩管理

4.1 伸缩组

4.1.1 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。

您在创建伸缩组时，需为伸缩组指定伸缩配置，同时也可以为伸缩组添加一条或多条伸缩策略。

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。因此，用户在创建弹性伸缩组时，需要根据可用区支持的云服务器类型，选择合适的伸缩配置。

- 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，则在进行扩容操作时，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。

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“创建弹性伸缩组”。
4. 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4-1**所示。

表 4-1 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。	-
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。	-
多可用区扩展策略	可选择“均衡分布”或“选择优先”。 ● 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。 ● 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。 **说明** 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/最小实例数	最大/最小实例数是指伸缩组中云服务器个数的最大值/最小值。	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
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。	-
子网	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。	-
负载均衡	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。 **说明** ● 一个伸缩组可最多添加6个负载均衡器。 ● 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“弹性负载均衡健康检查”，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。	-

参数	解释	取值样例
实例移除策略	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： ● 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。 ● 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。 ● 较早创建的实例：创建时间较早的实例优先被移除伸缩组。 ● 较晚创建的实例：创建时间较晚的实例优先被移除伸缩组。 **说明** ● 当可用区不均衡时，移出实例时会优先保证可用区均衡。 ● 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。	根据较早创建的配置较早创建的实例
健康检查方式	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。 ● 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。 ● 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。只有当伸缩组使用弹性负载均衡器时，您才可以选择“弹性负载均衡健康检查”，所有监听器下检测到的云服务器状态必须均为正常，否则伸缩组会将该云服务器移出伸缩组。	-

5. 单击“下一页”，跳转至创建伸缩配置页面，您可以选择使用已有伸缩配置或者创建新伸缩配置，更多信息请参见**使用已有云服务器创建伸缩配置**和**使用新模板创建伸缩配置**。
6. 单击“立即创建” 。
7. 请核对弹性伸缩组和伸缩配置的信息，单击 “提交”。
8. 您可以为伸缩组添加伸缩策略，请参见**创建伸缩策略**章节。

4.1.2 （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。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**。

4.1.3 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 4-1 更换伸缩配置举例

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“基本信息”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”。

4.1.4 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，才可以选择“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，在伸缩组的“基本信息”页面中，单击“状态”右侧的“启用”。
4. 在弹出“启用伸缩组”的对话框中，单击“是”。

4.1.5 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（比如创建失败或者创建云硬盘失败，失败原因可以在页面查看），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，在伸缩组的“基本信息”页面中，单击“状态”右侧的“停用”。
4. 在弹出“停用伸缩组”的对话框中，单击“是”。

4.1.6 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，单击待修改的伸缩组的名称，进入该伸缩组基本信息页面，单击页面右上方“修改”。

   或在待修改的伸缩组所在行的“操作”列下，单击“更多 > 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”。

4.1.7 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”。

4.2 伸缩配置

4.2.1 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。

创建伸缩配置的入口

- 在创建伸缩组时，创建相应的伸缩配置或使用已有的伸缩配置。
- 在伸缩实例页面创建所需的伸缩配置。
- 在伸缩组详情页更换伸缩配置。

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**。

4.2.2 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4-2**所示。

表 4-2 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。	-

参数	解释	取值样例
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。	使用已有云服务器规格为模板
弹性IP	弹性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： ● 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。 ● 自动分配： 自动为每台弹性云服务器分配独享带宽的弹性IP，您可以在“带宽”中设置带宽值。 **说明** 当您选择自动分配时，需要配置规格、带宽类型、计费方式和带宽等参数。	自动分配
带宽类型	您可选择独享带宽或共享带宽。 ● 独享带宽：一个带宽只能被一个弹性公网IP地址使用。 ● 共享带宽：一个带宽中可以加入多个弹性公网IP地址，多个弹性公网IP地址共用一个带宽。 **说明** ● 该参数仅在“弹性公网IP”选择“自动分配”时显示。 ● 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。 ● 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。	共享带宽

参数	解释	取值样例
登录方式	云平台提供两种弹性云服务器鉴权方式。 ● 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。 **说明** 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。 ● 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。	Admin@123
高级配置	高级配置可对用户数据注入进行配置。 可选择“暂不配置”和“现在配置”。	-
用户数据注入	可选配置，主要用于创建云主机时向云主机注入用户数据。配置用户数据注入后，云主机首次启动时会自行注入数据信息。 详细内容请参见**《弹性云服务器用户指南》**。 主要分为以下几种： ● 以文本形式：在下方文本框内输入用户数据内容； ● 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。配置文件注入后，系统在创建云主机时自动将文件注入到指定目录下。 – Linux系统请输入注入文件保存路径，例如“/etc/foo.txt”。 – Windows系统注入文件自动保存在C盘根目录，只需要输入保存文件名，例如“foo”，文件名只能包含字母和数字。 **说明** ● 对于Linux弹性云服务器，如果选择密码方式登录鉴权，此时不能使用用户数据注入功能。 ● 如果选择的镜像不支持用户数据注入，则不能使用用户数据注入功能。	-

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**。
7. （可选）启动伸缩组。

   如果伸缩组状态是“未启用”状态，启用伸缩组，请参考**启用伸缩组**。

4.2.3 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4-3**所示。

表 4-3 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： ● x86计算：x86采用复杂指令集（CISC）； ● 鲲鹏计算：鲲鹏采用精简指令集（RISC）。	x86计算
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**《弹性云服务器用户指南》**。 根据您选择的“云服务器类型”的类型，配置相应的规格参数，包括vCPU、内存、镜像类型和磁盘。	内存优化型
镜像	● 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。 ● 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。 ● 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。	公共镜像

参数	解释	取值样例
磁盘	包括系统盘和数据盘。 ● 系统盘 普通IO：是指由SATA存储提供资源的磁盘类型。 高IO：是指由SAS存储提供资源的磁盘类型。 超高IO：是指由SSD存储提供资源的磁盘类型。 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。 ● 数据盘 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。 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。	“系统盘”选为“普通IO”
安全组	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。	-
弹性IP	弹性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： ● 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。 ● 自动分配：自动为每台弹性云服务器分配独享带宽的弹性IP，带宽值可以由您设定。 **说明** 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。	自动分配
带宽	您可选择独享带宽或共享带宽。 ● 独享带宽：一个带宽只能被一个弹性公网IP地址使用。 ● 共享带宽：一个带宽中可以加入多个弹性公网IP地址，多个弹性公网IP地址共用一个带宽。 **说明** ● 该参数仅在“弹性公网IP”选择“自动分配”时显示。 ● 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。 ● 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。	共享带宽

参数	解释	取值样例
登录方式	云平台提供两种弹性云服务器鉴权方式。 ● 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。 **说明** 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。 ● 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。	Admin@123
高级配置	高级配置可对用户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”。	-
用户数据注入	可选配置，主要用于创建云主机时向云主机注入用户数据。配置用户数据注入后，云主机首次启动时会自行注入数据信息。 详细内容请参见**《弹性云服务器用户指南》**。 主要分为以下几种： ● 以文本形式：在下方文本框内输入用户数据内容； ● 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。配置文件注入后，系统在创建云主机时自动将文件注入到指定目录下。 – Linux系统请输入注入文件保存路径，例如“/etc/foo.txt”。 – Windows系统注入文件自动保存在C盘根目录，只需要输入保存文件名，例如“foo”，文件名只能包含字母和数字。 **说明** ● 对于Linux弹性云服务器，如果选择密码方式登录鉴权，此时不能使用用户数据注入功能。 ● 如果选择的镜像不支持用户数据注入，则不能使用用户数据注入功能。	-

5. 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**。

4.2.4 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。

4.2.5 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。

4.3 伸缩策略

4.3.1 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。

4.3.2 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4-4**。

表 4-4 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

参数名称	参数说明	取值样例
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**设置监控告警规则**。 选择新建告警时，支持系统监控和自定义监控。 ● 系统监控需配置**表4-5**所示参数。 ● 自定义监控需配置**表4-6**所示参数。	-
执行动作	设置伸缩活动执行动作及实例的个数或实例百分比。 执行动作包括： ● 增加 当执行伸缩活动时，向伸缩组增加实例。 ● 减少 当执行伸缩活动时，从伸缩组中减少实例。 ● 设置为 将伸缩组中的期望实例数设置为固定值。	● 增加1个实例 ● 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： – 大于1的值向下取整。例如，12.7取整为12。 – 大于0且小于1的值取整为1。例如，0.67取整为1。 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。

参数名称	参数说明	取值样例
冷却时间	为了避免告警策略频繁触发，必须设置冷却时间。 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。 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。 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。 **说明** • 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。 • 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。	300秒

表 4-5 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 4-6 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4-7进行参数配置。

表 4-7 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	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。 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： ● 重复周期 – 按天 – 按周 – 按月 ● 生效时间 伸缩策略触发的时间段。	-
时区	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。	GMT+08:00
触发时间	设定伸缩策略触发时间。	-
执行动作	设置伸缩活动执行动作及实例的个数。 执行动作包括： ● 增加 当执行伸缩活动时，向伸缩组增加实例。 ● 减少 当执行伸缩活动时，从伸缩组中减少实例。 ● 设置为 将伸缩组中的期望实例数设置为固定值。	● 增加1个实例 ● 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： ● 大于1的值向下取整。例如，12.7取整为12。 ● 大于0且小于1的值取整为1。例如，0.67取整为1。 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。

参数名称	参数说明	取值样例
冷却时间	为了避免告警策略频繁触发，必须设置冷却时间。 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。 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。 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。 **说明** • 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。 • 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。	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。

4.3.3 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”。

4.4 伸缩活动

4.4.1 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4-2**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4-3**进行配置。

图 4-2 告警策略 01

图 4-3 告警策略 02

4.4.2 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。

4.4.3 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，您可以手动将移入指定伸缩组。

将成功移入指定伸缩组必须满足如下条件：

- 不能存在于其它伸缩组中。
- 待移入的所在的VPC必须和伸缩组所在的VPC相同。
- 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。

将移入伸缩组的步骤如下。

1. 单击具体的伸缩组名称。
2. 在“伸缩实例”页签，单击“移入伸缩组”。
3. 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。

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。
- 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。

将移出伸缩组的步骤如下。

1. 单击具体的伸缩组名称。
2. 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”。
3. 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。

4.4.4 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。当有多个手工加入伸缩组的时，移除规则是：先进先出。

4.4.5 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

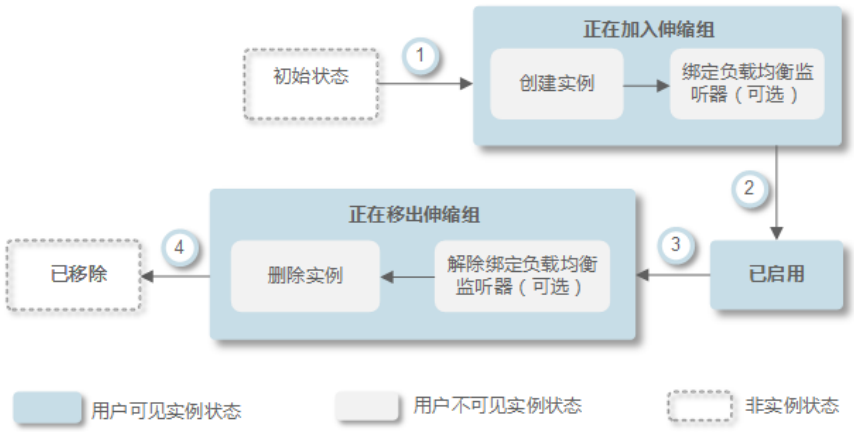
下面介绍如何查看伸缩组的伸缩活动。

1. 登录管理控制台。
2. 单击具体的伸缩组名称。

4.4.6 生命周期挂钩

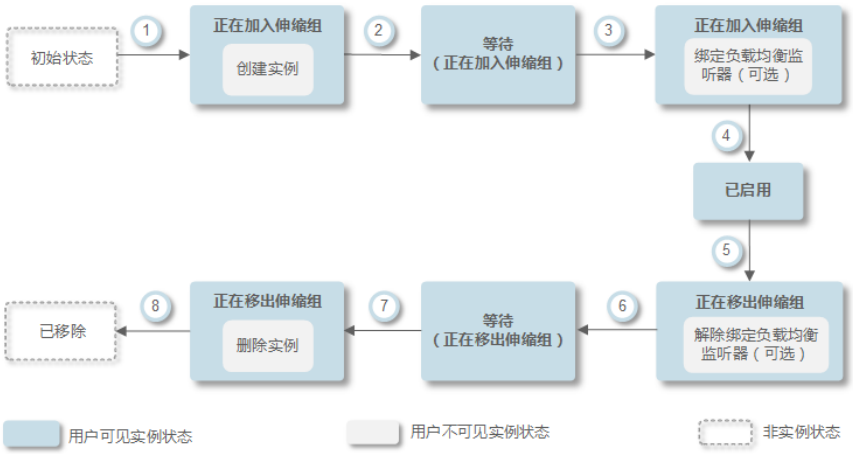
生命周期挂钩功能提供灵活控制伸缩组内ECS实例创建和移出过程的能力，以便用户灵活管理ECS实例的生命周期。伸缩组中未添加生命周期挂钩时，实例生命周期状态如图4-4所示。

图 4-4 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4-5所示。

图 4-5 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4-5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，需要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4-8**所示。

表 4-8 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
默认回调操作	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。 在当前实例正在加入伸缩组时，默认回调操作的含义为： ● 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。 ● 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。 在当前实例正在移出伸缩组时，默认回调操作的含义为： ● 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。 ● 终止：无论有几个个挂钩，只要有一个挂钩状态为“终止”，会直接将实例移出伸缩组。	继续
超时时间	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。 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**进行回调操作**。	3600秒

参数名称	参数说明	取值样例

通知主题	为生命周期挂钩定义一个通知目标（请参见《消息通知服务用户指南》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：
	{
	"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"
	}
	}
		-

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4-6所示。

图 4-6 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4-7所示。

图 4-7 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4-8。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4-8**，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”。

4.4.7 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。

   也可在单个实例所在行的操作列下，单击“设置实例保护”，在弹出的“设置实例保护”页面，单击“是”，为该实例设置实例保护。

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。

也可在单个实例所在行的操作列下，单击“取消实例保护”，在弹出的“设置实例保护”页面，单击“是”，对该实例取消实例保护。

4.5 伸缩带宽

4.5.1 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 4-9 “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。 策略名称只能由英文字母、数字、下划线、和中划线组成。	-
弹性IP	需要进行伸缩管理的公网IP。 **说明** 当前仅支持对按需付费模式的弹性IP进行伸缩，包周期的弹性IP不能创建弹性伸缩。	-
策略类型	选择“告警策略”。	告警策略

参数	解释	取值样例
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**新建告警规则**。 若选择新建告警，需配置如下参数： ● 告警规则名称 新建告警的名称，例如as-alarm-7o1u。 ● 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**表4-10**所示。例如上行流量平均值>100bit/s。 ● 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。 ● 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。	-
执行动作	设置伸缩策略执行动作。 执行动作包括： ● 增加 当执行伸缩活动时，增加带宽大小。 ● 减少 当执行伸缩活动时，减少带宽大小。 ● 调整到 将带宽大小设置为固定值。 **说明** 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。 ● 小于等于300Mbit/s：默认步长为1Mbit/s。 ● 300Mbit/s~1000Mbit/s：默认步长为50Mbit/s。 ● 大于1000Mbit/s：默认步长为500Mbit/s。	-
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 4-10 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。

指标名称	解释
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。

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4-8**和**表4-11**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 4-8 创建告警规则

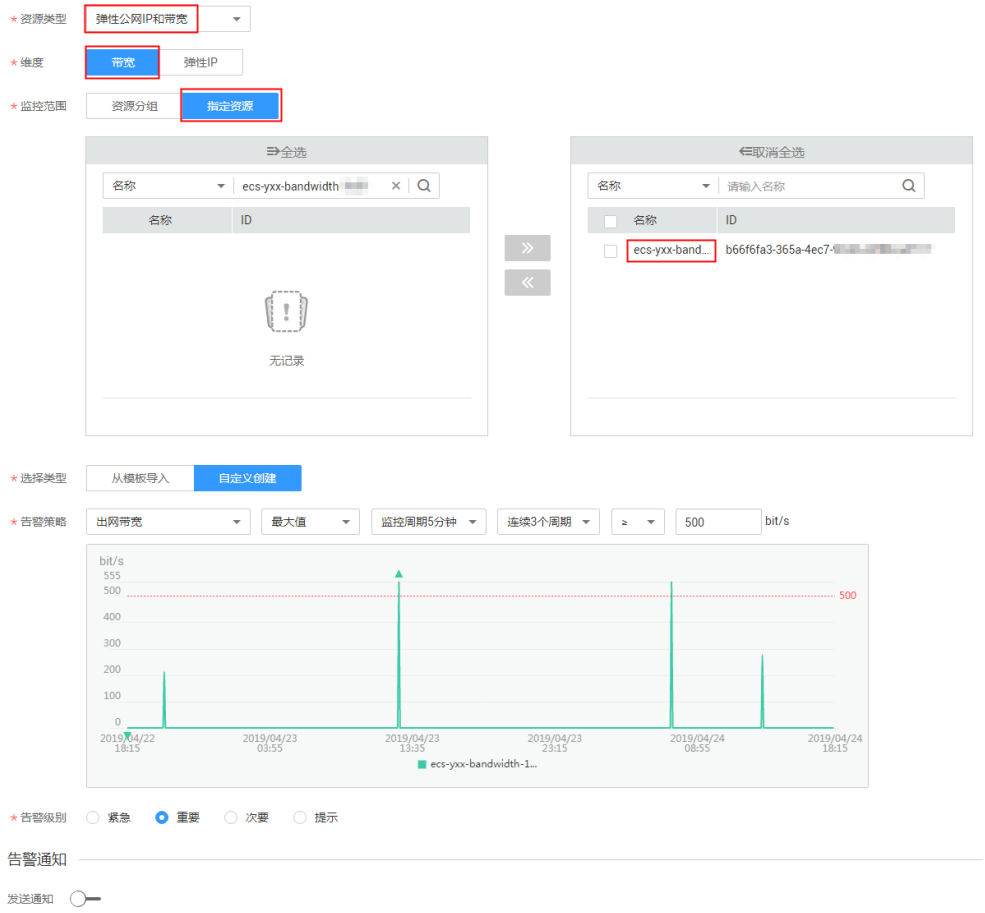


表 4-11 创建告警规则关键参数

参数	解释	配置示例
名称	告警规则的名称。	alarm-bandwidth
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
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4-10**。	-

4. 参数配置完成后，单击“立即创建”。
5. 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

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**表4-12**进行配置。

表 4-12 “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。 策略名称只能由英文字母、数字、下划线、和中划线组成。	as-policy-p6g5
弹性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。 **说明** 当前仅支持对按需付费模式的弹性IP进行伸缩，包周期的弹性IP不能创建弹性伸缩。	-

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： ● 生效时间 选择伸缩策略触发的时间段。 ● 重复周期 – 按天 – 按周 – 按月	-
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： ● 增加 当执行伸缩活动时，增加带宽大小。 ● 减少 当执行伸缩活动时，减少带宽大小。 ● 调整到 将带宽大小设置为固定值。 **说明** 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。 ● 小于等于300Mbit/s：默认步长为1Mbit/s。 ● 300Mbit/s~1000Mbit/s：默认步长为50Mbit/s。 ● 大于1000Mbit/s：默认步长为500Mbit/s。	-
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”。

4.5.2 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。参考**查看伸缩带宽策略详情**操作可进入该策略的基本信息页面。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。

4.5.3 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。

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，建议您采用停用的方式，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。

4.6 伸缩组和实例的监控

4.6.1 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“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，系统将其移出伸缩组的同时也会将其删除。对于手动移入伸缩组的实例，系统仅将其移出伸缩组。

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。

4.6.2 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，以便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4-13所示。

表 4-13 参数配置

参数名称	参数说明	取值样例
发送通知到	选择已经创建成功主题。请参见《消息通知服务用户指南》创建主题。	f123

参数名称	参数说明	取值样例
每当伸缩组	每当伸缩组出现以下一种或几种场景时，向用户发送通知。 ● 扩容成功 ● 扩容失败 ● 减容成功 ● 减容失败 ● 异常	-

4. 单击“保存”。

4.6.3 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 4-14 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

操作名称	资源类型	事件名称
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

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“管理与部署 > 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。
- 起始时间、结束时间：可通过选择时间段查询操作事件。

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"
}
```


4.6.4 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4-15所示。

表 4-15 标签命名规则

参数	规则	样例
键	• 不能为空。 • 对于同一伸缩组，“键”唯一。 • 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。	Organization

参数	规则	样例
值	• 可以为空字符串。 • 一个“键”只能添加一个“值”。 • 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。	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“基本信息”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4-15所示。
   单击“删除”后，将伸缩组已添加的该标签删除。

4.6.5 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4-16所示。

表 4-16 弹性伸缩支持的监控指标

指标	指标名称	指标含义	取值范围	测量对象&维度	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%	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	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 **说明** 如果用户使用的镜像未安装vmtools，则无法获取该监控指标。	≥0%	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	5分钟
network_outgoing_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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	5分钟
instance_num	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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	5分钟
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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	5分钟

指标	指标名称	指标含义	取值范围	测量对象&维度	监控周期（原始指标）
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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	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 %	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	1分钟
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 %	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	1分钟

指标	指标名称	指标含义	取值范围	测量对象&维度	监控周期（原始指标）
load_average1	（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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	1分钟
load_average5	（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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	1分钟
load_average15	（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	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	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 %	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	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 %	测量对象：弹性伸缩组 测量维度： AutoScalingGroup	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）。

维度

Key	Value
AutoScalingGroup	弹性伸缩组的ID

4.6.6 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

1. 登录管理控制台。
2. 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。
3. 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”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

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“管理与部署 > 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”和“近12小时”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。

4.6.7 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. 选择“管理与部署 > 云监控服务”。
3. 在左侧导航树栏，选择“告警 > 告警规则”。
4. 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。
5. 规则参数设置完成后，单击“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。

5 常见问题

5.1 通用类

5.1.1 弹性伸缩有什么限制?

弹性伸缩的云服务器中运行的应用需要是无状态、可横向扩展的。因为AS会自动释放云服务器，所以弹性伸缩组内的云服务器不可以保存应用的状态信息（例如session会话）和相关数据（如数据库、日志等）。

如果应用中需要保存状态或日志信息，可以考虑把相关信息保存到独立的服务器中。

弹性伸缩对用户的资源数量或容量做的配额限制如表5-1所示。

表 5-1 配额一览表

类别	描述	默认值
弹性伸缩组	用户可以创建的最多伸缩组个数。	10
弹性伸缩配置	用户可以创建的最多伸缩配置个数。	100
弹性伸缩策略	某个弹性伸缩组下可以创建的最多伸缩策略个数。	10
弹性伸缩实例	某个弹性伸缩组下可以创建的最多实例个数。	300
伸缩带宽策略	用户最多可以创建的伸缩带宽策略个数。	10

5.1.2 弹性伸缩一定要搭配弹性负载均衡、云监控才能使用吗?

弹性伸缩可以单独使用，也可以同弹性负载均衡（ELB），云监控（CES）一起使用。

其中，CES服务为免费服务，系统默认开通；ELB服务在有需求时可以部署，例如，有分布式集群需求的场景下，可以使用ELB。

5.1.3 弹性伸缩是否收取费用?

弹性伸缩服务本身不收取费用，但伸缩组自动创建的按需付费实例需要支付相应的费用。实例使用的弹性公网IP也需支付相应的费用。伸缩组进行缩容时，自动创建的实例会被移出伸缩组并删除，删除后将不再收取费用。而手动移入的实例只会被移出伸缩组，系统仍会收取该实例的使用费用。若您不再需要使用该实例，请自行在ECS页面进行退订。

例如，弹性伸缩进行扩容活动创建了两台实例，使用一个小时后，进行了缩容活动，这两台实例被移出伸缩组并删除了，则系统只收取这两台实例使用一小时产生的费用。

5.1.4 弹性伸缩是否会因监控指标突变导致误伸缩?

不会。弹性伸缩的监控数据基于云监控来获取的，监控周期可配置多个档位，如：五分钟，二十分钟，一小时等。不会因为一次指标的高峰而导致错误伸缩。

同时，弹性伸缩还支持配置冷却时间，防止由于监控的变化造成伸缩组的反复无效变化。该时间可由用户进行自定义。

5.1.5 我能创建和使用多少个伸缩策略和配置?

您默认可以创建10个弹性伸缩组，100个弹性伸缩配置。每个弹性伸缩组同一时刻支持使用1个伸缩配置，10个伸缩策略。

如果系统提供的默认配额不能满足您的需求，请联系管理员进行处理。

5.1.6 弹性伸缩是否能够自动升降云服务器的 CPU、内存和带宽?

弹性伸缩目前仅支持纵向扩展带宽资源。

5.1.7 弹性伸缩的配额是什么?

什么是配额?

为防止资源滥用，平台限定了各服务资源的配额，对用户的资源数量和容量做了限制。如您最多可以创建多少个伸缩组。如果有需要，您可以申请扩大配额。

本节指导您如何查询指定区域下，弹性伸缩服务各资源的使用情况，以及总配额。

怎样查看我的配额?

1. 登录管理控制台。
2. 单击管理控制台左上角的 ，选择区域和项目。
3. 单击页面右上角的“My Quota”图标 。
   系统进入“服务配额”页面。
4. 您可以在“服务配额”页面，查看各项资源的总配额及使用情况。
   如果当前配额不能满足业务要求，请参考后续操作，申请扩大配额。

5.1.8 同账户下不同用户操作弹性伸缩资源时，为什么提示密钥对不存在而拦截操作?

密钥对作为用户级资源，无法跨用户使用。如果伸缩配置中配置的是同账号下其他用户的密钥对，那么在弹性伸缩下，无法使用该伸缩配置手动触发资源下发。

如果需要用户之前互相操作伸缩配置关联资源且不受密钥对权限限制，请使用密码方式进行虚拟机鉴权。

5.2 伸缩组类

5.2.1 伸缩组启用失败如何处理?

请参考“伸缩组异常情况下如何处理？”章节中描述的可能原因和处理方法。

5.2.2 伸缩组异常情况下如何处理?

伸缩组异常情况及其处理方法：

- 情况描述：云服务器配额不足、云硬盘配额不足、弹性公网IP配额不足等。

  可能原因：配额不足。

  处理方法：申请扩大配额或者删除不需要的资源，之后重新启用伸缩组。
- 情况描述：伸缩组使用的VPC不存在、子网不存在等。

  可能原因：虚拟私有云服务异常或者相关资源被删除。

  处理方法：等待虚拟私有云服务恢复或者修改伸缩组中VPC、子网相关参数，之后重新启用伸缩组。
- 情况描述：负载均衡监听器不存在、后端云服务器组不存在、负载均衡器不可用等。

  可能原因：负载均衡服务异常或者相关资源被删除。

  处理方法：等待负载均衡服务恢复或者修改伸缩组中负载均衡相关参数，之后重新启用伸缩组。
- 情况描述：您添加到负载均衡监听器的后端云服务器超过最大限制。

  可能原因：伸缩组使用经典型负载均衡，会自动地将加入伸缩组的实例添加到负载均衡监听器。一个监听器最多可添加300个后端云服务器。

  处理方法：您可以从监听器上移除不需要的且不在该伸缩组中的后端云服务器，之后重新启用伸缩组。
- 情况描述：伸缩配置使用的镜像不存在、规格不存在、密钥对不存在等。

  可能原因：相关资源被删除。

  处理方法：为伸缩组更换新的伸缩配置，之后重新启用伸缩组。
- 情况描述：生命周期挂钩使用的通知主题不存在。

  可能原因：伸缩组添加了生命周期挂钩，且伸缩活动开始前生命周期挂钩使用的通知主题被删除。如果伸缩活动开始后通知主题被删除，则下次伸缩活动时伸缩组会异常。

  处理方法：修改生命周期挂钩使用的通知主题或者删除生命周期挂钩，之后重新启用伸缩组。

- 情况描述：您选择的子网下私有IP不足。

  可能原因：伸缩组使用的子网下私有IP地址被用尽。

  处理方法：修改伸缩组中的子网信息，之后重新启用伸缩组。

- 情况描述：您选择的可用区下的该类型云服务器资源已售罄。

  可能原因：您为伸缩组选择的可用区下该类型云服务器资源售罄，或者该可用区不支持该类型云服务器资源。该类型云服务器指的是伸缩配置里选择的云服务器规格。

  处理方法：您可以为伸缩组更换新的伸缩配置，之后重新启用伸缩组。当您的伸缩组中没有实例时，还可以修改伸缩组的可用区信息，之后重新启用伸缩组。

- 情况描述：您选择的规格和磁盘不匹配。

  可能原因：伸缩配置中的云服务器类型和磁盘类型不匹配导致创建云服务器失败。

  处理方法：为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：您选择的规格和镜像不匹配。

  可能原因：伸缩配置中的云服务器类型和镜像不匹配导致创建云服务器失败。

  处理方法：为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：您选择的可用区下的该类型存储资源已售罄。

  可能原因：您为伸缩组选择的可用区下该类型存储资源售罄，或者该可用区不支持该类型存储资源。该类型存储资源指的是伸缩配置里选择的磁盘（包括系统盘和数据盘）类型。

  处理方法：您可以为伸缩组更换新的伸缩配置，之后重新启用伸缩组。当您的伸缩组中没有实例时，还可以修改伸缩组的可用区信息，之后重新启用伸缩组。

- 情况描述：您选择的伸缩配置中定义的共享带宽不存在。

  可能原因：相关资源被删除。

  处理方法：您可以使用新购买的共享带宽或者其他存在的共享带宽资源重新创建伸缩配置，为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：伸缩配置中定义的共享带宽绑定的EIP个数超过最大限制。

  可能原因：单个共享带宽最多可以添加20个公网IP。

  处理方法：您可以申请扩大配额或者从共享带宽上移出不需要的公网IP，或者为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：您的伸缩配置中选择的专属主机不存在，请更换伸缩配置。

  可能原因：相关资源被删除。

  处理方法：您可以使用新购买的专属主机或者其他存在的专属主机资源重新创建伸缩配置，为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：无可用的专属主机资源，请确保当前有可用的专属主机资源。

  处理方法：您可以排除专属主机的故障使其恢复可用状态或者开启专属主机自动放置属性，之后重新启用伸缩组。还可以使用新购买的专属主机资源重新创建伸缩配置，为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：您的伸缩配置中选择的专属主机可用容量不足。

  处理方法：您可以删除专属主机上不需要的云服务器资源之后重新启用伸缩组，或者使用新购买的专属主机资源重新创建伸缩配置，为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：您的伸缩组中选择的可用区下无可用的专属主机资源。

处理方法：您可以在该可用区下购买新的专属主机资源，重新创建伸缩配置，为伸缩组更换新的伸缩配置，之后重新启用伸缩组。当您的伸缩组中没有实例时，还可以修改伸缩组的可用区信息，之后重新启用伸缩组。

- 情况描述：您的伸缩配置中选择的专属主机不支持该类型云服务器，请更换伸缩配置。

  处理方法：您可以选择专属主机支持的规格重新创建伸缩配置，为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：系统异常。

  可能原因：弹性伸缩或者周边服务异常、网络异常等。

  处理方法：请稍后重试，或联系技术支持。

- 情况描述：您选择的伸缩配置中定义的规格不可用。

  处理方法：您可以通过错误提示信息通过重新创建伸缩配置更换新的规格，并为伸缩组更换新的伸缩配置，之后重新启用伸缩组

- 情况描述：你选择的伸缩配置无法被当前伸缩组使用。

  处理方法：您可以通过错误提示信息重新创建伸缩配置，并为伸缩组更换新的伸缩配置，之后重新启用伸缩组。

- 情况描述：您的账户已欠费或余额不足。

  可能原因：账户欠费或余额不足时，伸缩组无法进行扩容。

  处理方法：您可以在充值之后重新启用伸缩组。

5.2.3 停用伸缩组后，什么操作会暂停?

停用伸缩组之后，伸缩组将不再自动触发伸缩活动，但是已开始的伸缩活动会继续执行。伸缩策略不会自动触发伸缩活动。手动调整期望实例数后，尽管当前实例数与期望实例数不相等，但是不会触发伸缩活动。

健康检查会继续检查实例的健康状态，但不会执行移除操作。

5.3 伸缩策略类

5.3.1 我能启用多少个伸缩策略?

伸缩策略可以启用一个，也可以启用多个。

5.3.2 告警策略支持的告警触发条件有哪些?

可以针对CPU使用率、内存使用率、带内网络流入速率、带内网络流出速率、磁盘读速率、磁盘写速率、磁盘读操作速率、磁盘写操作速率等指标进行监测告警，自动增加或减少ECS实例。

5.3.3 什么是冷却时间，为什么需要冷却时间?

冷却时间是指冷却伸缩活动的时间。每次伸缩活动完成之后，系统开始计算冷却时间。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。

实例加入伸缩组投入使用之前，需要使用配置脚本安装和配置软件，大约需要两到三分钟（实际时间取决于诸多因素，如实例规格和是否有启动脚本等）。因此实例从启

动到投入使用如果没有冷却时间，系统会在负载降下来前不断扩容，新加入的实例接管业务后，发现负载过低，然后又缩容。冷却时间避免了伸缩组重复进行不必要的伸缩活动。

冷却时间工作原理举例：

业务出现流量高峰，触发告警策略，按照配置AS会自动新增一个实例到伸缩组来帮助处理增加的需求。但是存在一个问题：该实例需要几分钟的时间才能启动，并且启动后到可以从ELB接收请求也需要一段时间。在此期间，告警可能会持续触发，从而导致告警每次触发时都会新增一个实例。若您设置了冷却时间，AS在启动一个实例后，将暂停告警策略引起的扩展活动，直至经过了该指定时间段（默认值为300秒）。这样，新启动的实例有时间开始处理应用程序流量。冷却时间过后，如果告警再次触发，AS才会启动另一个实例，而冷却时间也会再次生效。

5.3.4 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?

弹性伸缩支持根据自定义监控进行动态伸缩。

5.3.5 未安装 VM Tools 对弹性伸缩组监控指标有什么影响?

未安装VM Tools，云监控无法监控弹性云服务器的内存使用率、带内网络流入速率和带内网络流出速率三个指标。但可以监控带外网络流入速率和带外网络流出速率指标，这样导致CPU使用率指标的精确性可能会降低。

如果弹性云服务器类型是IO优化实例，无论是否安装了vmtools，监控弹性云服务器的监控指标不包含磁盘使用率、带内网络流入速率和带内网络流出速率指标。

因此，当上述情况发生时，弹性伸缩支持的内存使用率、带内网络流入速率和带内网络流出速率三个指标会受到影响，无法获取监控数据。

5.3.6 伸缩策略启用失败如何处理?

- 情况描述：告警规则不存在。
  可能原因：告警策略中使用的告警规则被删除。
  处理方法：修改告警策略中使用的告警规则，之后重新启用策略。
- 情况描述：周期策略的触发时间不包含在策略的生效时间内。
  可能原因：周期策略的生效时间已过期。
  处理方法：修改周期策略生效的起始时间和结束时间，之后重新启用策略。
- 情况描述：定时策略触发时间必须晚于当前时间。
  可能原因：定时策略触发时间已过期。
  处理方法：修改定时策略的触发时间，之后重新启用策略。
- 情况描述：系统异常。
  处理方法：请稍后重试，或联系技术支持。

5.3.7 如需使用 Agent 监控指标，如何为伸缩组中的实例安装 Agent 插件?

问题背景

伸缩策略的类型为告警策略时，支持使用Agent监控指标触发伸缩活动。Agent监控即操作系统监控，相比基础监控，操作系统监控可以为用户提供服务器的系统级、主动

式、细颗粒度监控服务。如需使用Agent监控指标，伸缩组中的实例必须均已安装Agent插件，本文提供了详细的操作指导。

操作步骤

1. 登录管理控制台，选择“计算 > 弹性云服务器”。

   进入云服务器控制台。

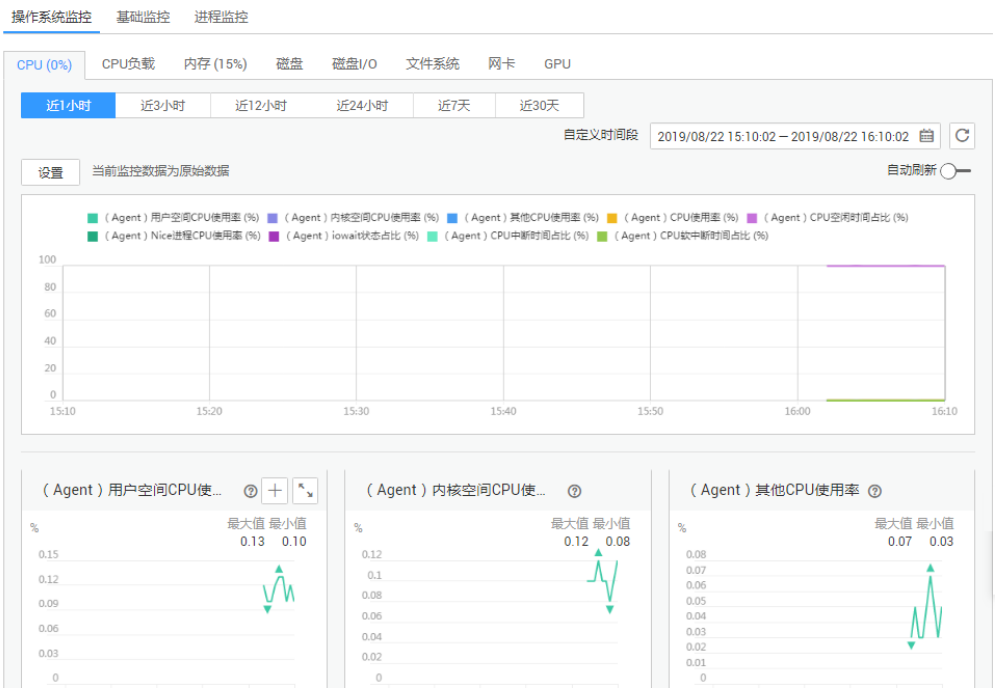
2. 创建一台弹性云服务器，并安装Agent插件。

   请参考**“Agent安装配置方式说明”**选择一种方式进行插件安装。

3. 待Agent插件安装成功后，进入云监控控制台，选择“主机监控 > 弹性云服务器”，确保插件状态为“运行中”且能够采集Agent监控指标数据。

图 5-1 查看插件状态

图 5-2 查看 Agent 监控指标



4. 向弹性云服务器的conf.json配置文件中添加AccessKey/SecretKey（AK/SK）信息。

   a. 单击用户名，选择“我的凭证 > 访问密钥”，获取AK/SK。

      - 如已有访问密钥，查看创建时下载保存的credentials.csv文件，获取文件中记录的Key值。
      - 如未创建，则通过“新增访问密钥”创建新的密钥，妥善保存credentials.csv文件，并获取文件中记录的Key值。

b. 登录弹性云服务器，执行**cd /usr/local/telescope/bin**进入Agent安装路径。

c. 执行**vi conf.json**打开配置文件，输入已获取的AK/SK。

```
{
  "InstanceId":"fa07b      4cd0-9aa6-e5c791569e3a",
  "ProjectId": "050b1      572f8cc01f3740bed5",
  "AccessKey": "MK8NR3      7FUMUB",
  "SecretKey": "sPHiTB8                    N4wWv3YCNwcUFqj",
  "RegionId": "cn-north-1"
}
```

如果使用“购买ECS时安装Agent”的安装方式，在注入用户数据时已经添加了AK/SK信息，此处只需检查一下。

d. 按“ESC”，输入**:wq**保存并退出。

5. 进入镜像服务页面，将这台弹性云服务器制作为私有镜像，详细操作请参考“**创建私有镜像**”。

图 5-3 创建私有镜像

6. 进入弹性伸缩页面，使用步骤5中创建的私有镜像创建伸缩配置。

图 5-4 选择私有镜像

在“镜像”区域，单击“私有镜像”，在下拉列表中选择“ces-agent-test”，其他参数按照实际需求配置。

7. 创建伸缩组，并绑定步骤6中的伸缩配置。
8. 为伸缩组添加伸缩策略，策略类型选择“告警策略”，触发条件选择Agent相关监控指标，如：（Agent）内存使用率。

图 5-5 选择触发条件

9. 将步骤2中的云服务器手动移入伸缩组。
10. 等待并验证Agent监控指标是否生效，例如，验证以下项目：
    - 伸缩组详情页“监控”页签下显示有Agent监控指标
    - 达到告警阈值时，在伸缩组详情页“活动历史”页签查看告警策略触发成功，且发生实例的伸缩
    - 伸缩组自动扩容出来的云服务器均有Agent监控数据

5.4 实例类

5.4.1 如何保证手动移入的 ECS 实例不被移出伸缩组?

假设您向伸缩组手动移入了N台ECS实例，并且不希望这些实例被自动移出，那么您可以通过如下两种方法确保这些实例不被移出伸缩组。

方法一

在伸缩组同时进行如下两条配置：

- 将伸缩组的最小实例数设置为N或者大于N的值 。
- 将实例移除策略配置为“根据较早创建的配置较早创建的实例”或“根据较早创建的配置较晚创建的实例”。

根据弹性伸缩的规则，手工添加的实例不会对应任何伸缩配置（因为它们不是通过伸缩配置创建的），所以弹性伸缩会先挑选通过伸缩配置自动创建的实例进行释放，只有当自动创建的实例释放完了，才会挑选手工添加的实例进行释放。由于您将最小实例数设置成N或大于N，所以手工添加的实例是不会被选中。

注意：以上是在您手动移入的实例处于正常的情况下，如果这些实例处于关机或其他异常状态，弹性伸缩会视为他们不健康，并将它们移出伸缩组，因为健康检查需要保证在伸缩组里的实例是健康的。

方法二

为这N台实例设置实例保护，

您可以同时为这N台实例设置实例保护，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。注意，实例若未通过健康检查仍然会被移出伸缩组。

5.4.2 多规格伸缩配置创建实例的选择的规格顺序是什么？

当伸缩配置选择多个规格时，根据伸缩组可用区及多可用区扩展策略的配置不同，创建实例时选择的规格顺序不同，本章节将分单个可用区和多可用区情况说明。

单可用区

当伸缩组只选择了一个可用区时，伸缩组中的实例都会创建在该可用区中。伸缩配置选择多个规格时，可按照如下两种规格使用优先顺序创建实例：

- 选择优先：伸缩组扩容时按照您选择规格的顺序进行。例如，您依次选择了规格2、3、1。系统会按照您选择的顺序优先选择规格2创建实例，当规格2库存不足或者因为其他原因创建失败时，系统会选择规格3创建实例，当规格3也无法创建实例时才会使用规格1。
- 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。例如，您依次选择了规格1、2、3。这三个规格按成本排序为：1>3>2。系统会优先选择规格2（成本最低的规格）创建实例，当规格2无法创建实例时，选择规格3，当规格3也无法创建实例时，才会选择规格1。

多可用区

当伸缩组选择了两个及两个以上的可用区时，需要配置“多可用区扩展策略”（“均衡分布”或“选择优先”）。当您选择不同的多可用区扩展策略时，选择的实例规格的创建顺序也会不同。对不同的创建顺序分情况说明如下：

- 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。创建实例选择AZ和规格的顺序举例如下：

  您依次选择了可用区AZ1、AZ2、AZ3，创建伸缩配置时选择了规格1、2、3，且规格选择的优先级顺序为2、3、1。AZ1、AZ2、AZ3分别有3、2、3台实例，按照均衡分布原则AZ2的实例数较少，优先选择AZ2创建实例。先使用规格2在AZ2中创建实例，若成功则伸缩活动成功，若规格2无法在 AZ2中创建实例，依次尝试使用规格3和规格1创建实例，若均失败，则AZ2中无法创建实例。根据当均衡分布无法创建实例，按照选择优先原则选择其他可用区，将在AZ1中按照规格2、3、1

的顺序依次尝试创建实例。若AZ1仍无法创建云服务器，选择AZ3进行尝试，选择规格的顺序仍是2、3、1。

- 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。创建实例选择AZ和规格的顺序举例如下：

  您依次选择了可用区AZ1、AZ2、AZ3，创建伸缩配置时选择了规格1、2、3，且规格选择的优先级顺序为2、3、1。无论AZ中的实例是否均衡，系统会按照您选择AZ的顺序，即AZ1、AZ2、AZ3依次创建实例。先使用规格2在AZ1中创建实例，若无法成功，使用规格3在AZ1中创建实例，若仍无法成功，再使用规格1在AZ1中创建实例。若使用3种规格在AZ1中均无法创建实例，则尝试在AZ2中创建实例，仍按照规格顺序2、3、1进行创建。若AZ2中仍无法创建实例，选择AZ3创建实例，规格选择顺序仍为2、3、1。

说明

该部分规格的优先级顺序是根据伸缩配置中选择的“规格使用优先策略”确定的，具体的确定方式可参考**单可用区**。

5.4.3 当实例被移出伸缩组并删除后，实例中的数据会保留吗?

不会。弹性伸缩会自动释放ECS实例，您需要确保伸缩组内的实例没有保存应用的状态信息或者重要数据，例如，会话记录（session）、数据库和日志等。如果您的应用需要保存状态信息，可以考虑将状态信息保存到独立的状态服务器（如ECS）、数据库（如RDS）等。

5.4.4 我能添加已经创建的包年包月 ECS 实例吗?

能。目前弹性伸缩默认自动创建按需付费的ECS实例，同时支持手动添加已经创建的包年包月和按需付费ECS实例。

5.4.5 按照伸缩策略增加的云服务器，当我不用时可以自动删除吗?

可以，但是需要您增加一条删除云服务器的伸缩策略。

5.4.6 什么是期望实例数?

期望实例数是指伸缩组中期望运行的弹性云服务器的个数，大小介于最小实例数和最大实例数之间。您可以手动调整期望实例数，也可以通过定时（周期）策略和告警策略触发调整期望实例数。

创建伸缩组时设置期望实例数：当期望实例数不为0时，伸缩组创建完成后会立即开始伸缩活动自动添加相应个数的弹性云服务器。

手动调整期望实例数：当用户通过直接修改的方式更改了期望实例数，系统发现当前实例数与期望实例数不相等，就会触发伸缩活动，使当前实例数等于期望实例数。

伸缩策略触发调整期望实例数：伸缩策略触发后，如果该策略的触发动作是增加2个实例，系统首先会将期望实例数在当前基础上+2。此时，由于伸缩组的当前实例数与期望实例数不再相等，系统就会触发伸缩活动增加2个实例，使当前实例数等于期望实例数。

5.4.7 如何删除通过弹性伸缩创建的云服务器?

处理方法

方法一：

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“弹性伸缩组”名称。
4. 在“弹性伸缩组”详情页面中，选择“伸缩实例”页签。
5. 在“操作”列下选择需要删除的实例，单击“移出伸缩组并删除”。

说明

如果您要删除多个实例，可以依次勾选指定实例名称左侧的方框，单击“移出伸缩组并删除”。

方法二

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“弹性伸缩组”名称。
4. 在“弹性伸缩组”详情页面中，选择“伸缩策略”页签。
5. 在“伸缩策略”页签中增加一条删除云服务器的伸缩策略，可以按照需要减少或者调整至指定数量。

方法三

1. 登录管理控制台。
1. 选择“计算 > 弹性伸缩 > 伸缩实例”。
2. 单击具体的“弹性伸缩组”名称。
3. 在“弹性伸缩组”详情页面中，单击右上方“修改”，进入“修改伸缩组”页面。
4. 手动修改伸缩组的期望实例数。

5.4.8 包年包月的 ECS 实例出现异常后会不会被伸缩组删除?

不会。伸缩组不会删除包周期的ECS实例，包周期的ECS实例出现异常后，只会将ECS实例移出伸缩组，不会进行删除操作。

5.4.9 如何处理伸缩组中状态是“异常”的实例?

正常情况下，您不要处理伸缩组中状态是“异常”的实例，弹性伸缩的健康检查功能会周期性地对伸缩组中实例的健康状态进行检查。当伸缩组为启用状态时，会将异常的实例从伸缩组中移除，然后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。当伸缩组为非启用状态时，对实例的健康状态会继续进行检查，但不会执行移除操作。

值得注意的是，负载均衡健康检查是通过负载均衡系统向后端云服务器发起心跳检查的方式来实现的，而负载均衡系统和云服务器之间是通过内网进行通信的。所以，如果伸缩组使用负载均衡健康检查方式，为确保健康检查工作的正常进行，您需要确保能够通过内网访问您的云服务器，请按照以下方法排查。

1. 在监听器页面，在健康检查异常的监听器所在行，单击“健康检查”列下的“查看”。弹出健康检查配置项提示框。
   - 检查“健康检查方式”：确保后端服务器已配置相应协议并开启端口。

- 检查“检查路径”：如果是使用HTTP协议进行健康检查，还应检查后端主机的健康检查路径是否正确。

2. 检查云服务器中防火墙等软件是否有对来自健康检查源IP的屏蔽。
3. 检查后端云主机所在安全组与网络ACL规则是否配置放行100.125.0.0/16，并配置ELB用于健康检查的协议和端口。健康检查的协议和端口在步骤1中弹出的健康检查配置项提示框中获取。
   - 若采用默认的健康检查方式：需要放行后端云服务器业务端口。
   - 若配置了不同于云服务器业务端口的健康检查端口：需要放行云服务器业务端口与健康检查端口。
4. 如果以上配置检查均正常但问题依然存在，请联系技术支持。

5.4.10 当伸缩组中实例无法通过负载均衡健康检查而频繁地被删除再重新创建时应该怎么办?

实例所在安全组规则必须放通100.125.0.0/16网段，且协议和端口号需要为ELB用于健康检查的协议和端口，否则会导致健康检查失败。

5.4.11 如何阻止伸缩组内的云服务器被自动移除?

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，当伸缩组发生自动缩容活动时，设置了实例保护的实例将不会被移除伸缩组。您还可以设置伸缩组的最小实例数，配合实例移除策略，可以使伸缩组始终至少保持一定数量的ECS实例运行。

值得注意的是，健康检查会将异常的实例从伸缩组中移除，并重新创建新的ECS实例。所以，请不要在ECS页面停止或者删除已经加入到伸缩组中的ECS实例，否则我们认定该ECS实例不健康，并自动移出伸缩组。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。

5.4.12 为什么在伸缩组内移除并删除实例后，ECS 页面还能看到实例?

如果伸缩组内自动伸出的实例被锁定，那么实例移出伸缩组时，我们仅移除实例，不做删除操作，以确保该实例可被其他服务正常使用。

一般情况下，实例正在被其他服务使用时会被锁定，例如：实例正在被镜像服务使用，制作成私有镜像场景；实例被存储容灾服务使用场景。

5.5 其他

5.5.1 如何自动部署应用?

为了使伸缩组自动加入的实例自动部署应用，您需要创建私有镜像，确保该镜像上有应用的程序软件和开机自启动设置。为伸缩组选择镜像类型为您的私有镜像的伸缩配置，新实例加入伸缩组后，就可以实现自动部署应用。详细的操作步骤指导如下：

1. 在创建私有镜像前，您需要在源云服务器中安装所需的应用程序，并设置开机自动启动。
2. 创建私有镜像，你可以根据操作系统选择对应的创建私有镜像的方法，操作指导请参考《镜像服务用户指南》。

3. 创建一个新的伸缩配置，操作可参考《创建伸缩配置》，确保该伸缩配置选择的镜像为2中创建的私有镜像。
4. 切换到“弹性伸缩组”页签，单击所需的伸缩组名称，进入伸缩组详情页面。
5. 单击“配置名称”右侧的“更换配置”，在弹出的“更换伸缩配置”对话框中，选择3中创建的伸缩配置，单击“确定”。

   至此全部设置已经完成，等待下次伸缩活动触发新实例加入伸缩组后，您可以查看新实例是否自动部署了所需的应用，若有问题请联系技术支持处理。

5.5.2 支持 Cloud-Init 特性后，对使用弹性伸缩有哪些影响?

Cloud-init是开源的云初始化程序，能够对新创建弹性云服务器中指定的自定义信息（主机名、密钥和用户数据等）进行初始化配置。在创建伸缩配置时，通过Cloud-Init进行对云服务器的初始化配置。

弹性伸缩组使用的伸缩配置中的私有镜像若没有安装Cloud-Init/Cloudbase-init工具，伸缩活动创建的弹性云服务器会出现如下情况：

- 如使用未安装Cloudbase-init的Windows私有镜像创建的弹性云服务器，在获取弹性云服务器密码时，系统将提示查询不到密码。您只能通过镜像本身的密码登录此。若忘记镜像本身密码，可以通过云服务器页面“重置密码”功能，自助完成云服务器的密码重置。
- 如使用未安装Cloud-Init的Linux私有镜像创建的云服务器，使用创建时输入的密码或秘钥将无法正常登录云服务器。您只能通过镜像本身的密码或秘钥登录此云服务器。若忘记镜像本身密码，或镜像本身的秘钥丢失，可以通过云服务器页面“重置密码”功能，自助完成云服务器的密码重置。
- 如使用未安装Cloud-Init/Cloudbase-init的私有镜像创建云服务器时，用户数据注入会失败。

  鉴于出现上述情况，在使用弹性伸缩时，请检查伸缩配置中的私有镜像是否已经安装并配置了Cloud-Init或者Cloudbase-init工具，请将使用了未安装Cloud-Init/Cloudbase-init的私有镜像的伸缩配置删除，并使用已安装Cloud-Init/Cloudbase-init的私有镜像创建新的伸缩配置。具体操作步骤：

  a. 登录管理控制台。
  b. 选择“计算 > 弹性伸缩 > 伸缩实例”。
  c. 选择“伸缩配置”页签，进入伸缩配置列表页面。
  d. 单击“创建伸缩配置”，选择已安装Cloud-Init或Cloudbase-init工具的私有镜像创建新的伸缩配置。
  e. 在伸缩组中将伸缩配置修改为新创建的伸缩配置。

5.5.3 如何在新扩容的实例上运行已有业务?

5.5.4 为什么使用密钥文件无法正常登录云服务器?

问题描述

用户使用密钥文件登录弹性伸缩组中的云服务器时，登录失败。

可能原因

该弹性伸缩组使用的伸缩配置中的镜像为用户自己制作的私有镜像，且在创建该私有镜像时用户未安装Cloud-init工具。

创建私有镜像时不安装Cloud-init工具，用户将无法对云服务器进行自定义配置，只能使用镜像原有密码或密钥登录云服务器。

处理方法

1. 判断是否需要继续登录该云服务器。
   - 是，请使用镜像原有密码或密钥登录云服务器。

     其中，镜像原有密码或密钥指创建私有镜像时，用户自己设置的操作系统密码或密钥。
   - 否，跳转**2**。
2. 更换弹性伸缩组的伸缩配置。

说明

请确保新伸缩配置中的镜像已安装了Cloud-init/Cloudbase-init工具，Cloud-init/Cloudbase-init工具的安装方法请参见《镜像服务用户指南》。

更换伸缩配置后，弹性伸缩组进行伸缩活动而新增的弹性云服务器可以直接使用密钥文件正常登录，无需再使用镜像原有密码或密钥登录云服务器。

5.5.5 伸缩组中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?

伸缩组中已经添加了负载均衡后，伸缩配置可以不配置弹性公网IP。系统会自动将加入伸缩组的实例添加到负载均衡上，伸缩组中的实例统一通过负载均衡绑定的弹性公网IP对外提供服务。

5.5.6 如何自动初始化弹性伸缩新增的云服务器数据盘?

操作场景

云服务器创建完成后，数据盘需要初始化后才能使用。当使用弹性伸缩为伸缩组增加数量较多的云服务器时，您就需要逐一手动初始化数据盘，将会占用较长时间。

本节为您介绍通过脚本自动化完成初始化磁盘的操作，包括磁盘分区和挂载指定目录。本节介绍的自动初始化脚本示例仅支持初始化一个数据盘。

本节操作以centos 6.5为例。其他操作系统配置方法略有区别，请参考对应操作系统的相关资料进行操作，文档中不对此进行详细说明。

操作步骤

1. 以root用户登录已有云服务器。
2. 执行以下命令，进入脚本存放目录。

   cd /脚本目录

   例如：

   cd /home
3. 执行以下命令，创建脚本。

   vi 脚本名称

   例如：

   vi fdisk_mount.sh

4. 按“i”，进入脚本编辑页面。

   以下脚本为云服务器只有一个数据盘时的自动初始化，仅供参考，请用户根据实际情况修改。

```
#!/bin/bash
bash_scripts_name=fdisk_mount.sh
ini_path=/home/fdisk.ini
disk=
size=
mount=
partition=

function get_disk_from_ini()
{
disk=`cat $ini_path|grep disk| awk -F '=' '{print $2}'`
if [ $disk = "" ]
then
    echo "disk is null in file,exit"
    exit
fi
result=`fdisk -l $disk | grep $disk`
if [ $result = 1 ]
then
    echo "disk path is not exist in linux,exit"
    exit
fi

}

function get_size()
{
size=`cat $ini_path| grep size|awk -F '=' '{print $2}'`
if [ $size = "" ]
then
    echo "size is null,exit"
    exit
fi
}

function make_fs_mount()
{
mkfs.ext4 -T largefile $partition
if [ $? -ne 0 ]
then
    echo "mkfs disk failed,exit"
    exit
fi

dir=`cat $ini_path|grep mount |awk -F '=' '{print $2}'`
if [ $dir = "" ]
then
    echo "mount dir is null in file,exit"
    exit
fi

if [ ! -d $dir ]
then
    mkdir -p $dir
fi

mount $partition $dir
if [ $? -ne 0 ]
then
    echo "mount disk failed,exit"
    exit
fi

echo "$partition $dir ext3 defaults 0 0" >> /etc/fstab
}
```



```
function remove_rc()
{
cat /etc/rc.local | grep $bash_scripts_name
if [ $? ne 0 ]
then
    sed -i '/'$bash_scripts_name'/d' /etc/rc.local
fi
}

################### start ########################
##1、判断配置文件是否存在
if [ ! -f $ini_path ]
then
    echo "ini file not exist,exit"
    exit
fi

##2、获取配置文件中disk指定的设备路径
get_disk_from_ini

##3、获取配置文件中size分区大小
get_size

##4、将磁盘分区
fdisk $disk  <<EOF
n
p
1
1
$size
w
EOF
partition=`fdisk -l $disk 2>/dev/null| grep "^/dev/[xsh].*d" | awk '{print $1}'`

##5、格式化分区，挂载分区到对应目录
make_fs_mount

##6、修改开机启动项，避免重试执行
remove_rc

echo 'SUCESS'
```

5. 按“Esc”键，输入**:wq**，按“Enter”保存并退出编辑。
6. 执行以下命令，创建配置文件。

   vi fdisk.ini
7. 按“i”，进入配置文件编辑页面。

   配置文件中设置了数据盘的盘符、大小和挂载目录信息，用户可参考如下所示信息，根据实际情况修改。

   ```
   disk=/dev/xdev
   size=+100G
   mount=/opt/test
   ```
8. 按“Esc”键，输入**:wq**，按“Enter”保存并退出编辑。
9. 执行以下命令，打开配置文件“rc.local”。

   vi /etc/rc.local
10. 按“i”，在“rc.local”文件中添加如下内容。

    /home/fdisk_mount.sh

    配置“rc.local”后，首次启动时会自动执行初始化数据盘脚本。
11. 按“Esc”键，输入**:wq**，按“Enter”保存并退出编辑。
12. 通过已有创建私有镜像。

13. 创建伸缩配置。

    填写伸缩配置信息时，请选择上述步骤中创建的私有镜像，并选择一个数据盘。

14. 创建伸缩组。

    配置伸缩组时，请选择上述步骤中创建的伸缩配置。

    伸缩组创建成功后，通过该组的伸缩活动产生的，会按照私有镜像中的配置自动初始化数据盘。

A 修订记录

版本日期	变更说明
2020-11-03	第一次正式发布。

