



天翼云·弹性伸缩

用户指南

天翼云科技有限公司

目录

| | |
|---------------------|-----------|
| 1 产品动态 | 8 |
| 1.1 产品动态..... | 8 |
| 2 产品介绍 | 10 |
| 2.1 什么是弹性伸缩服务..... | 10 |
| 2.2 弹性伸缩相关术语解释..... | 10 |
| 2.3 弹性伸缩服务的优势..... | 12 |
| 2.4 弹性伸缩服务的功能..... | 13 |
| 2.5 使用限制..... | 14 |
| 2.6 与其他服务的关联关系..... | 15 |
| 2.7 弹性伸缩使用流程..... | 17 |
| 3 快速入门 | 18 |
| 3.1 产品价格..... | 18 |
| 3.2 快速创建弹性伸缩..... | 18 |
| 4 操作指南 | 26 |
| 4.1 伸缩组管理..... | 26 |
| 3.1.1. 创建伸缩组..... | 26 |

| | |
|-------------------------------|----|
| 3.1.2. (可选) 添加负载均衡器到伸缩组 | 32 |
| 3.1.3. 修改伸缩组信息 | 33 |
| 3.1.4. 更换伸缩组的伸缩配置 | 35 |
| 3.1.5. 启用伸缩组 | 36 |
| 3.1.6. 停用伸缩组 | 37 |
| 3.1.7. 删除伸缩组 | 38 |
| 4.2 伸缩配置管理 | 38 |
| 4.2.1 创建伸缩配置 | 38 |
| 4.2.2 复制伸缩配置 | 47 |
| 4.2.3 删除伸缩配置 | 48 |
| 4.3 伸缩策略管理 | 49 |
| 4.3.1 创建伸缩策略 | 49 |
| 4.3.2 修改伸缩策略 | 53 |
| 4.3.3 启用伸缩策略 | 55 |
| 4.3.4 停用伸缩策略 | 55 |
| 4.3.5 立即执行伸缩策略 | 56 |
| 4.3.6 删除伸缩策略 | 56 |
| 4.4 伸缩活动管理 | 57 |
| 4.4.1 查询伸缩活动 | 57 |

| | |
|--------------------------|-----------|
| 4.4.2 将实例移入伸缩组..... | 58 |
| 4.4.3 将实例移出伸缩组..... | 59 |
| 4.4.4 实例移除策略..... | 60 |
| 4.4.5 实例保护..... | 61 |
| 4.4.6 动态资源扩展..... | 62 |
| 4.4.7 按计划扩展资源..... | 63 |
| 4.5 伸缩带宽..... | 63 |
| 4.5.1 创建伸缩带宽策略..... | 63 |
| 4.5.2 查看伸缩带宽策略详情..... | 68 |
| 4.5.3 管理伸缩带宽策略..... | 69 |
| 4.6 伸缩组和实例的监控..... | 73 |
| 4.6.1 弹性伸缩健康检查..... | 73 |
| 4.6.2 记录弹性伸缩..... | 74 |
| 4.6.3 监控指标说明..... | 76 |
| 4.6.4 查看监控指标数据..... | 79 |
| 4.6.5 设置监报告警规则..... | 80 |
| 5 最佳实践..... | 81 |
| 5.1 搭建可自动伸缩的 WEB 服务..... | 81 |
| 5.1.1 场景介绍..... | 81 |

| | |
|---|-----------|
| 5.1.2 前期准备 | 82 |
| 5.1.3 解决方案 | 82 |
| 5.1.4 操作步骤 | 83 |
| 5.1.5 结果验证 | 89 |
| 6 常见问题 | 90 |
| 6.1 通用类 | 90 |
| 6.1.1 弹性伸缩有什么限制? | 90 |
| 6.1.2 弹性伸缩一定要搭配弹性负载均衡、云监控才能使用吗? | 90 |
| 6.1.3 用户默认能创建和使用多少个伸缩策略和配置? | 91 |
| 6.1.4 弹性伸缩是否能够自动升降云主机的 CPU、内存和带宽? | 91 |
| 6.1.5 伸缩带宽对弹性公网 IP 带宽和共享带宽有什么要求? | 91 |
| 6.2 计费类 | 91 |
| 6.2.1 弹性伸缩服务是否收取费用? | 91 |
| 6.3 开通类 | 92 |
| 6.3.1 天翼云所有节点都支持弹性伸缩服务么? | 92 |
| 6.3.2 退订弹性云主机后弹性伸缩服务是否继续使用? | 92 |
| 6.3.3 弹性云主机关机后弹性伸缩服务会关闭么? | 92 |
| 6.4 操作类 | 92 |
| 6.4.1 弹性伸缩一定要搭配弹性负载均衡、云监控才能使用吗? | 92 |

| | |
|---|-----|
| 6.4.2 弹性伸缩是否会因监控指标突变导致误伸缩? | 93 |
| 6.4.3 能创建和使用多少个伸缩策略和配置? | 93 |
| 6.4.4 按照伸缩策略增加的云主机在我不用时可以自动删除吗? | 93 |
| 6.4.5 伸缩组使用负载均衡健康检查, 为何没有立即生效? | 93 |
| 6.4.6 如何删除通过弹性伸缩创建的云主机? | 94 |
| 6.4.7 如何阻止伸缩组内的云主机被自动移除? | 95 |
| 6.4.8 如何自动初始化弹性伸缩新增的云主机数据盘? | 95 |
| 6.5 伸缩组类..... | 101 |
| 6.5.1 伸缩组启用失败如何处理? | 101 |
| 6.5.2 伸缩组异常情况下如何处理? | 103 |
| 6.5.3 停用伸缩组后, 什么操作会暂停? | 104 |
| 6.5.4 已知弹性云主机实例 ID, 能否查看当前实例所在的伸缩组信息? 以及如何查看伸缩组及其中 实例信息? | 105 |
| 6.6 伸缩策略类..... | 105 |
| 6.6.1 我能启用多少个伸缩策略? | 105 |
| 6.6.2 告警策略支持的告警触发条件有哪些? | 105 |
| 6.6.3 什么是冷却时间, 为什么需要冷却时间? | 106 |
| 6.6.4 未安装 VM Tools 对弹性伸缩组监控指标有什么影响? | 106 |
| 6.6.5 伸缩策略启用失败如何处理? | 107 |

| | |
|--|-----|
| 6.7 实例类: | 107 |
| 6.7.1 如何保证手动移入的 ECS 实例不被移出伸缩组? | 107 |
| 6.7.2 当实例被移出伸缩组并删除后, 实例中的数据会保留吗? | 108 |
| 6.7.3 按照伸缩策略增加的云主机, 当我不用时可以自动删除吗? | 108 |
| 6.7.4 什么是期望实例数? | 109 |
| 6.7.5 如何处理伸缩组中状态是“异常”的实例? | 109 |
| 6.7.6 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? | 110 |
| 6.7.7 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? | 110 |
| 6.8 其他: | 110 |
| 6.8.1 如何自动部署应用? | 110 |
| 6.8.2 支持 Cloud-Init 特性后, 对使用弹性伸缩有哪些影响? | 111 |
| 6.8.3 为什么使用密钥文件无法正常登录云主机? | 112 |
| 6.8.4 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? | 113 |

1 产品动态

1.1 产品动态

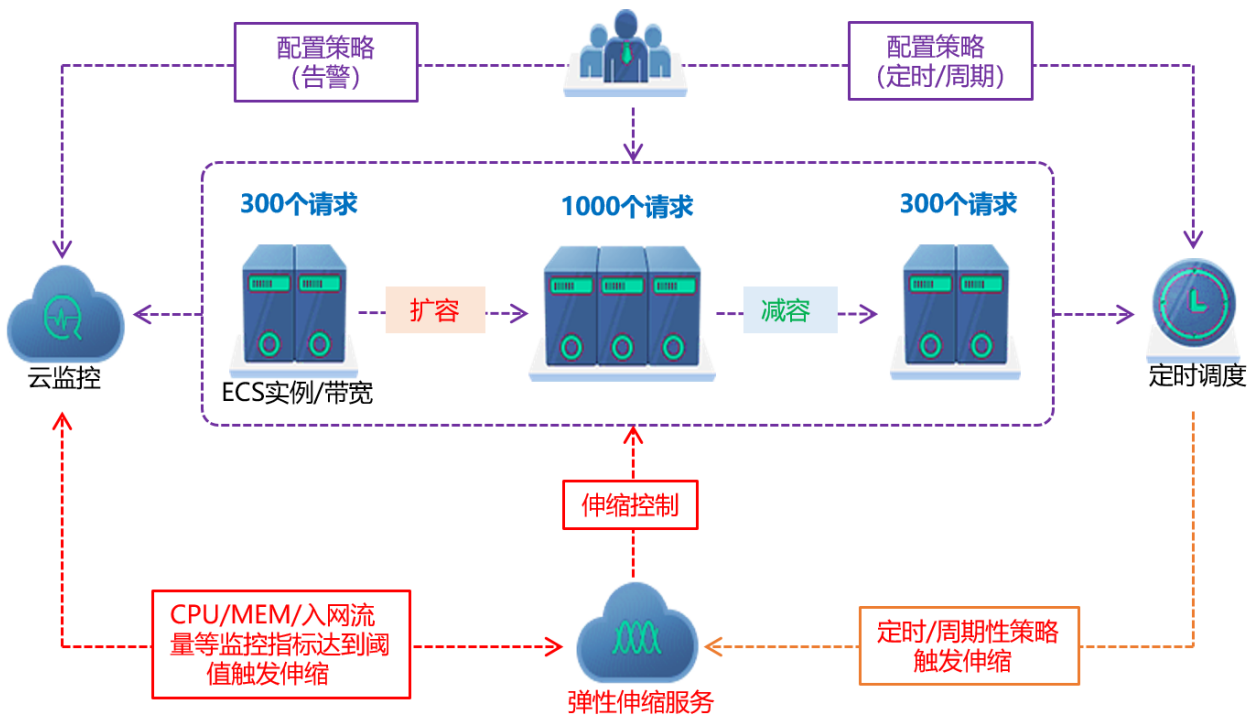
| 序号 | 时间节点 | 功能名称 | 功能描述 | 相关文档 |
|----|---------|------------------|--|-------------------------|
| 8 | 2023.09 | 支持创建多实例规格的伸缩配置 | 创建伸缩配置时支持选择多种实例规格，可最大程度降低因当前规格售罄或不可用导致的扩容失败概率，确保在用户业务高峰期伸缩组能扩容成功。目前苏州、广州已支持。 | 多规格配置说明 |
| 7 | 2023.07 | 支持多可用区扩展策略 | 多可用区扩展策略为您提供多样化的可用区配置策略。目前苏州、广州已支持。 | 创建伸缩组 |
| 6 | 2022.09 | 优化帮助中心 | 优化目录结构，内容更详细，图文并茂，逻辑更清晰，更加方便客户快速熟悉天翼云弹性伸缩服务。 | 弹性伸缩服务 |
| 5 | 2021.10 | 控制台伸缩组详情页面新增概览页签 | 采用图形化的展示方式，期望实例数、正在进行的伸缩活动等伸缩组信息一目了然，让您轻松掌握伸缩组的动态。 | 伸缩组 |
| 4 | 2019.12 | 支持创建伸缩带宽策略 | 可以通过创建伸缩带宽策略实现自动调整带宽资源 | 创建伸缩带宽 |
| 3 | 2019.03 | 支持实例保护功能 | 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。 | 实例保护 |
| 2 | 2018.09 | 支持消息通知功能 | 伸缩组使用消息通知功能，伸缩组的扩容成功、扩容失败、缩容成功、缩容失败和异常等情况将及时推送给您，使您能够及时的了解伸缩组的各种动态。 | 产品优势 |

| | | | | |
|---|---------|---------------|--------------------------------------|----------------------------|
| 1 | 2018.05 | 支持绑定多个负载均衡监听器 | 绑定多个负载均衡器，可以同时多个程序的流量分流，提高应用程序的容错能力。 | 添加负载均衡到伸缩组 |
|---|---------|---------------|--------------------------------------|----------------------------|

2 产品介绍

2.1 什么是弹性伸缩服务

弹性伸缩服务（Auto Scaling，CT-AS）是根据用户的业务需求，通过策略自动调整其弹性计算资源的管理服务。用户通过管理控制台设定弹性伸缩组策略，弹性伸缩服务将根据预设规则自动调整伸缩组内的云主机数量，在业务需求上升时自动增加云主机实例，业务需求下降时自动减少云主机实例，降低人为反复调整资源以应对业务变化和高峰压力的工作量，帮助用户节约资源和人力成本。



2.2 弹性伸缩相关术语解释

- 什么是最大或最小实例数？

当伸缩策略条件满足时，根据最大实例数和最小实例数自动调整需要添加或移除的云主机数量。例如，按照伸缩策略要求，需要将云主机数量增加到 10 台，但最大实例数是 8，那么系统会按照 8 台云主机数量进行弹性伸缩活动。

- **什么是期望实例数？**

默认的云主机数量，当没有伸缩规则满足时，系统保留的云主机数量。手工修改该值，会触发一次弹性伸缩活动。

- **什么是伸缩组？**

伸缩组是具有相同属性和应用场景的云主机和伸缩策略的集合。伸缩组是启停伸缩策略和进行伸缩活动的基本单位。

- **什么是伸缩配置？**

伸缩配置即伸缩活动中添加的云主机的规格。

- **什么是伸缩策略？**

触发伸缩活动的条件和执行的动作，当满足条件或者执行的动作时，会触发一次伸缩活动。

- **什么是伸缩活动？**

伸缩活动是指在伸缩组内由于伸缩条件满足而触发的云主机实例数量变更的活动，可能是增加或减少几台云主机实例。

- **什么是冷却时间？**

冷却时间是指冷却伸缩活动的时间，在每次触发伸缩活动之后，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等）触发的伸缩活动不受限制。

- **什么是默认配额**

默认配额是指每个用户在每个区域节点资源的数量，如默认配额无法满足用户需求，可通过控

制台申请调整配额。

2.3 弹性伸缩服务的优势

免费服务

弹性伸缩服务本身不收取费用，用户可免费设置使用，但需要注意的是，伸缩组内云主机实例会按照云主机计费标准收取费用。

跨可用区（AZ）使用

弹性伸缩服务不仅支持在一个 AZ 内进行云资源的伸缩，同时还可在同一 Region 内跨多个 AZ 使用，满足多 AZ 资源部署的业务场景。

自动化扩展

无需人工干预，弹性伸缩服务根据用户预先设定的规则自动完成云主机实例增减操作并保证云主机健康可用。

灵活策略配置

支持按照定时、周期及监控情况进行灵活的策略配置，满足不同业务场景。

可视化

控制中心可查看伸缩组内整体的监控图表及伸缩变更视图，帮助用户进行业务预测和运维管理。

智能通知

弹性伸缩服务执行结果会通过邮件的方式发送给用户，保证用户随时掌握伸缩服务的动态变化情况。

2.4 弹性伸缩服务的功能

弹性伸缩服务组件

弹性伸缩服务包括伸缩组、伸缩策略、伸缩配置、伸缩活动、云主机实例等组件。

管理弹性伸缩服务实例数量

支持设置弹性伸缩组内的期望实例数、最小实例数和最大实例数，并支持设置弹性伸缩策略的执行动作为增加、减少特定数量的云主机或将云主机数量设定为固定值。

支持多种伸缩策略

弹性伸缩服务不仅支持动态伸缩策略，基于用户设置的告警规则按照云主机的 CPU、内存利用率等指标的变化动态增减云主机数量；还支持定时伸缩策略，执行单一时点的定时云主机数量增减任务或周期性的定时任务。

定义伸缩组云主机规格

支持定义伸缩组内待添加的云主机规格，包括云主机 CPU、内存、镜像、登录方式等内容，当伸缩策略触发增加云主机操作时，将按照预定义的云主机规格创建新的云主机并添加至伸缩组内。

设置冷却时间

用户可自定冷却伸缩活动的时间，在冷却时间内，伸缩组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。每次触发伸缩活动后，系统自动计算冷

却时间。

支持绑定负载均衡

支持非负载均衡场景和负载均衡场景下的弹性伸缩服务。

支持健康检查

天翼云弹性伸缩服务提供健康检查功能，保证使用健康云主机负载业务。

支持云主机健康检查和负载均衡健康检查两种检查方式，自动替换不健康实例。

支持查看伸缩数据

用户可按照图形或表格的方式查看伸缩组的云主机实例数量变化情况，并能看到伸缩组内所有云主机实例的资源利用情况平均值变化曲线。

2.5 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：

- 弹性伸缩的云主机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主机，所以弹性伸缩组内的云主机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）。如果应用中需要云主机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下表所示。

表 1-1 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。 | 200 |

2.6 与其他服务的关联关系

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 1-2](#)所示。

图1-1 弹性伸缩服务与其他服务的关系示意图

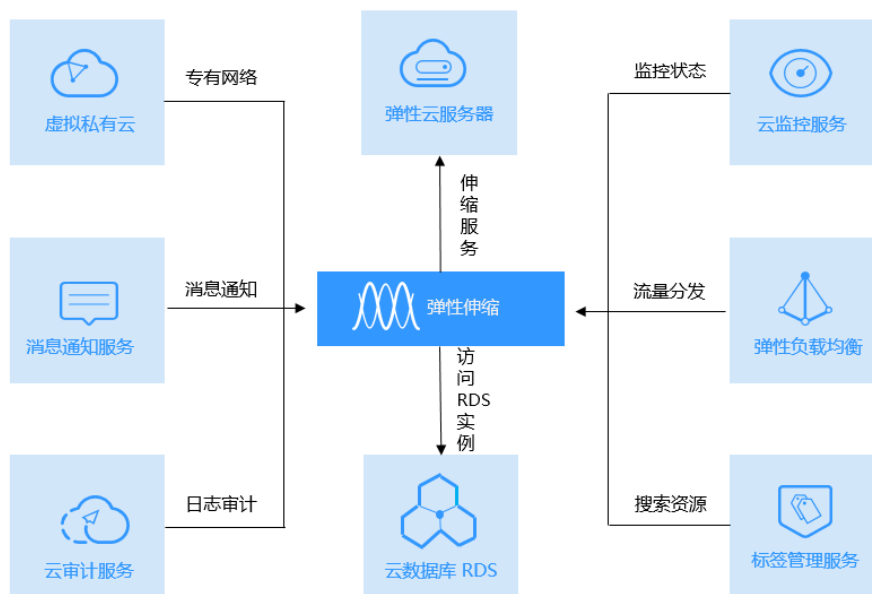


表1-1 弹性伸缩与其他服务的关系

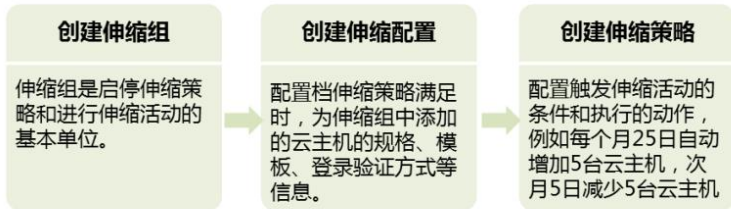
| 服务名称 | 说明 | 交互功能 | 相关内容 |
|-------------------------------|--|-----------------------|--|
| 弹性负载均衡 (Elastic Load Balance) | 当配置了负载均衡服务后，弹性伸缩组在添加或移除云主机时，自动会为云主机绑定或解绑负载均衡监听器。 | 使伸缩组中每一个实例均可分配到应用程序流量 | 添加负载均衡器到伸缩组 |
| 云监控服务 (Cloud Eye) | 弹性伸缩配置了告警触发策略时，会根据云监控的告警条件触发弹性伸缩活动。 | 通过监控伸缩组内实例的状态指标调节资源。 | 弹性伸缩支持的监控指标 |
| 弹性云主机 (Elastic Cloud Server) | 弹性伸缩活动中添加的云主机可以通过弹性云主机进行管理和维护。 | 自动调整弹性云主机数量 | 动态扩展资源 和 按计划扩展资源 |

| 服务名称 | 说明 | 交互功能 | 相关内容 |
|-----------------------------|---|------|------------------------|
| 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 |

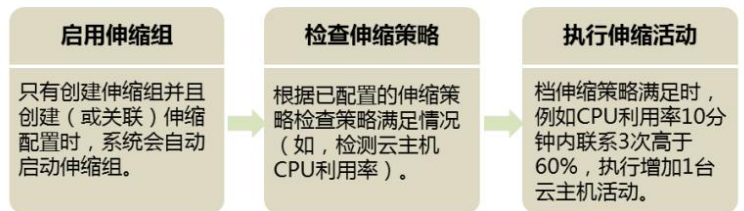
2.7 弹性伸缩使用流程

使用弹性伸缩服务, 首先需要用户进行手工配置, 包括创建伸缩组、创建伸缩配置、创建伸缩策略, 之后系统将根据用户的设置自动启动伸缩组、检查伸缩策略、执行伸缩活动。使用弹性伸缩服务的全流程如下图所示:

①用户手工配置



②系统自动执行



3 快速入门

3.1 产品价格

目前天翼云免费提供，但弹性伸缩组内的弹性云主机、云硬盘、虚拟私有云、弹性 IP、带宽等云产品需按照相关云产品订购页面列明公示信息支付相应服务费用。

3.2 快速创建弹性伸缩

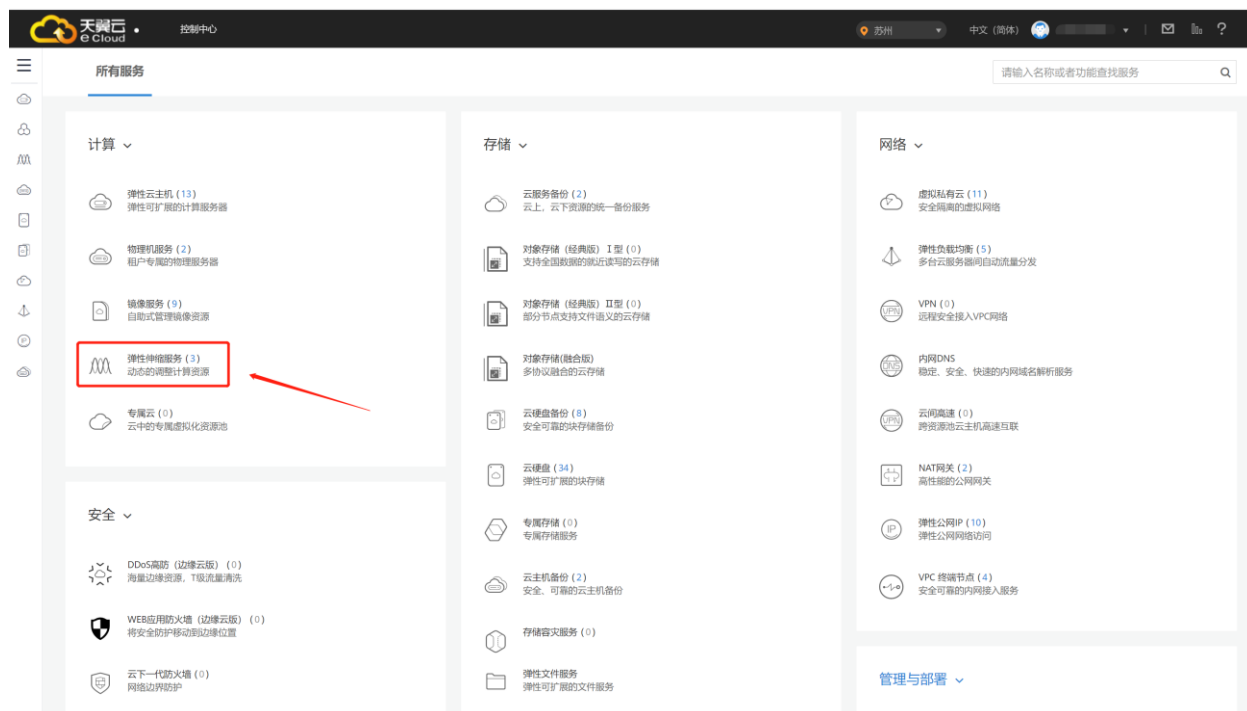
首次使用弹性伸缩服务，建议选择向导式创建弹性伸缩，具体操作请参考本章节快速创建弹性伸缩。

前提条件：

1. 已经创建所需的 VPC、安全组、弹性负载均衡器等；
2. 如果使用证书鉴权，还需要准备好 SSH 密钥。鉴权方式是指弹性伸缩活动中添加的云主机的鉴权方式，伸缩组内的其他云主机鉴权方式不受影响；

操作步骤：

1. 注册并登录天翼云 <https://www.ctyun.cn>;
2. 选择控制中心，切换到需要创建弹性伸缩的节点，选择【弹性伸缩服务】；



3. 在弹性伸缩服务页面，单击【创建弹性伸缩组】按钮；
4. 在弹性伸缩组创建页面，填写弹性伸缩组的基本信息，例如，伸缩组名称、最大实例数、最小实例数、VPC、子网、负载均衡等信息；

<
创建弹性伸缩组

*** 区域** ▼

苏州

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1 可用区2 ▼ C

可用区3

*** 名称** C

as-group-demo

*** 最大实例数(台)**

3

*** 期望实例数(台)** ?

1

*** 最小实例数(台)**

0

以下信息是伸缩后实例所使用的虚拟私有云、子网，负载均衡负责自动分发访问流量。

*** 虚拟私有云** ?

(192.168.0.0/16) C 新建虚拟私有云

*** 子网**

subn- (192.168.2.0/24) 本网作为云主机的主网卡

自动分配IPv6地址

+ 增加一个子网 您还可以增加 4 个子网 C 新建子网

负载均衡 ?

不使用 使用经典型 使用增强型 C 新建负载均衡

伸缩组中的云主机自动挂载到您关联的负载均衡下。

负载均衡器 elb-apj 3... 后端云主机组 --请选择--

后端端口 ? 权重 ? 1

+ 新增一个负载均衡器 您还可以增加2个负载均衡器。

*** 实例移除策略**

根据较早创建的配置较早创建的实例 ▼

弹性IP

释放 不释放

若选择“释放”，在伸缩组进行缩的活动时，则会将云主机上的弹性IP释放，否则仅做解绑定操作，保留弹性IP资源。

*** 健康检查方式** ?

负载均衡健康检查 ▼

受保护的实例状态异常时，会被健康检查移除，并重新创建新的实例。
实例所在安全组则需要配置放行100.125.0.0/16，并配置负载均衡用于健康检查的协议和端口，否则会导致健康检查失败。

*** 健康检查间隔** ?

5分钟 ▼

*** 企业项目** ?

default ▼ C

高级配置

暂不配置 现在配置

在高级配置中可对通知方式和标签进行配置。

下一页

各参数说明如下：

- **伸缩组名称：**创建伸缩组的名称。
- **最大/最小实例数：**指伸缩组中云主机个数的最大值/最小值。如果最小值不为 0，则创建伸缩

组后会触发一次伸缩活动，使伸缩组内实例数达到最小值。

- 虚拟私有云：弹性云主机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主机均属于该 VPC。
- 子网：默认情况下，一个 VPC 子网内的弹性云主机均可以进行通信。一个伸缩组只能选择 1 个子网。
- 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主机，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。伸缩组关联负载均衡实例后，无论是伸缩组自动创建实例，还是用户向伸缩组手动添加实例，实例都会自动加入到负载均衡实例的后端服务器组。

📖 说明

一个伸缩组可最多添加 10 组负载均衡监听器。添加多个负载均衡监听器后，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”，弹性云主机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主机替换掉。当选择使用弹性负载均衡时，需配置 4 个参数：负载均衡器、后端主机组、后端端口和权重。

- 实例移除策略：实例优先被移除的策略。当伸缩组进行收缩活动时，会触发实例移除策略，包括如下四种方式：
 - A. 根据较早创建的配置较早创建的实例：先筛选出较早创建的配置所创建出的实例，再筛选较早创建的实例移出。
 - B. 较早创建的实例：根据时间筛选较早创建的实例。
 - C. 根据较早创建的配置较晚创建的实例：指根据较早创建的配置中较晚创建的“实例”优先被移除伸缩组。。
 - D. 较晚创建的实例：根据时间筛选较晚创建的实例。

- 按量付费实例回收模式：指按量付费的实例实例被移出后的处理策略。对于包年包月的云主机，如果被移出伸缩组，也不删除不更改状态，不适用此策略。
 - A. 释放模式：将移出的按量付费的云主机释放；
 - B. 停机回收模式：将移出的云主机停机，不释放。随着扩张活动仍会优先再加入。
- 健康检查方式：健康检查会将异常的云主机从伸缩组中移除，并重新创建新的云主机，伸缩组的健康检查方式包括以下两种。
 - A. 云主机健康检查：是指对云主机的运行状态进行检查，如关机、删除都是云主机异常状态。默认为此选项，伸缩组会定期使用云主机健康检查结果来确定每个云主机的运行状况。如果未通过云主机健康检查，则伸缩组会将该云主机移出伸缩组，并重新创建云主机。
 - B. 弹性负载均衡健康检查：是指根据 ELB 对云主机的健康检查结果进行的检查。当伸缩组使用弹性负载均衡器时，您可以选择此方式来做健康检查，所有监听器下检测到的云主机状态必须均为正常，否则伸缩组会将该弹性云主机移出伸缩组，并重新创建云主机。
- 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（5 分钟、15 分钟、1 小时、3 小时），以确保其包含云主机中应用程序的预期启动时间。

5. 单击【下一步】，进入创建伸缩配置页面。

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。后续可以在伸缩组创建完成后，为伸缩组修改伸缩配置。如需要创建新的伸缩配置，需填写伸缩配置信息，例如配置名称、云主机镜像、云主机类型、登录方式等；

< 创建弹性伸缩组

伸缩组创建完成后，您还可以根据业务需求更换伸缩配置。

伸缩配置 使用已有 现在创建

* 名称
使用该配置创建的云主机名称为伸缩配置名称加八位随机码。

* 配置模板 使用新模板 使用已有云主机规格为模板

| | |
|--------|--|
| 配置模板ID | 083f8c33-8255-4747-ad32-29cdd8a6818d |
| 规格 | s3.large.2 2vCPUs 4GB |
| 镜像 | CentOS7.9 64位 |
| 系统盘 | 高IO(可用区3下已售罄) 40 GB |
| 数据盘 | -- |
| 安全组 | test-cc-e-node-ufz91 (入方向:ICMP; TCP/443, 80, 30000-32767, 3389, 22; UDP/30000-32767 出方向: -) |

弹性IP 不使用 自动分配
不使用弹性IP的云主机不能与互联网互通，仅可作为私有网络中部署业务或者集群所需云主机进行使用。

* 登录方式 密码 密钥对

用户名

* 密码
请妥善保管密码，系统无法获取您设置的密码内容。

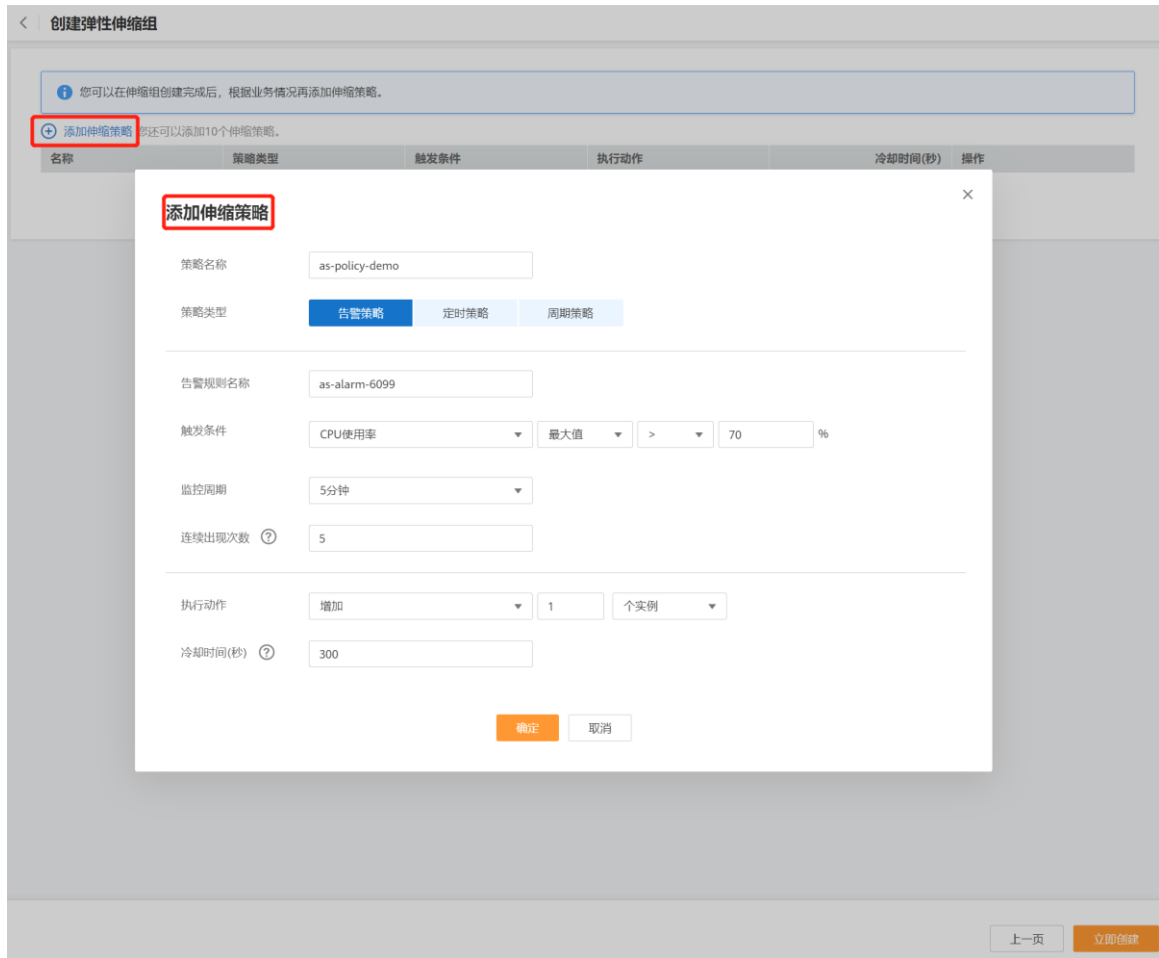
* 确认密码

6. 单击【下一页】，进入创建伸缩策略页面。

在伸缩策略页面，您可以创建新的伸缩策略，也可以在伸缩组创建完成后，根据业务需要再为伸缩组创建伸缩策略。

7. 在伸缩策略页面，单击【创建伸缩策略】按钮。

在弹出的创建伸缩策略窗口中，填写策略名称、策略类型、触发条件、执行动作、冷却时间等参数；



8. 单击【确认】;
9. 单击【立即创建】按钮;
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】;

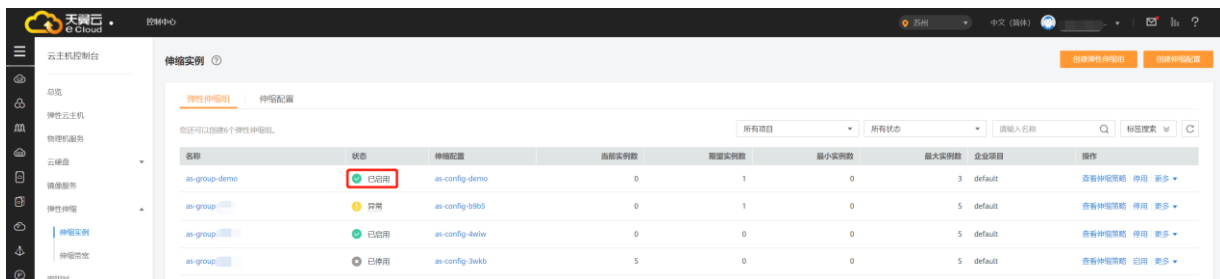
< | 创建弹性伸缩组

| 资源 | 基本信息 | |
|--------|----------------|---|
| 伸缩组 | 名称 | as-group-demo |
| | 最大实例数 | 3 |
| | 期望实例数 | 1 |
| | 最小实例数 | 0 |
| | 可用区 | 可用区1 可用区2 |
| | 区域 | 苏州 |
| | VPC | dualstack(192.168.0.0/16) |
| | 子网 | subnet-dualstack(192.168.2.0/24) |
| | 增强型后端云主机组 | server_group-demo(后端端口: 8,089, 权重: 1) |
| | 实例移除策略 | 根据较早创建的配置较早创建的实例 |
| | 弹性IP | 释放 |
| | 企业项目 | default |
| | 健康检查方式 | 负载均衡健康检查 |
| 健康检查间隔 | 5分钟 | |
| 伸缩配置 | 伸缩配置 | 创建新伸缩配置 |
| | 名称 | as-config-demo |
| | 安全组 | test-cc-e-ufz91 (入方向: ICMP, TCP/443, 80, 30000-32767, 3389, 22; UDP/30000-32767 出方向: -) |
| | 配置模板ID | 083f8c33-8255-4747-ad32-29cdd8a6818d |
| | 规格 | s3.large.2 2vCPUs 4GB |
| | 镜像 | CentOS7.9 64位 |
| | 系统盘 | 高IO(可用区3下已售罄) 40 GB |
| | 数据盘 | -- |
| | 弹性IP | 暂时不使用弹性IP |
| | 登录方式 | 密码 |
| 策略名称 | as-policy-demo | |
| 策略类型 | 告警策略 | |

上一页

提交

11. 创建伸缩组成功后，伸缩组状态变为“启用”。



4 操作指南

4.1 伸缩组管理

伸缩组是具有相同属性和应用场景的云主机和伸缩策略的集合。伸缩组是启停伸缩策略和进行伸缩活动的基本单位。

4.1.1 创建伸缩组

创建伸缩组配置伸缩组的基本信息，包括配置最大实例数、最小实例数和关联的负载均衡服务等。

1. 登录天翼云控制中心，切换到需要创建弹性伸缩的节点，选择【弹性伸缩服务】；
2. 在弹性伸缩服务页面，单击【创建弹性伸缩组】按钮；
3. 在弹性伸缩组创建页面，选择区域和可用区；
4. 在弹性伸缩组创建页面，填写弹性伸缩组的基本信息，例如，伸缩组名称、最大实例数、最小实例数、VPC、子网、负载均衡等信息；

< 创建弹性伸缩组

*** 区域** ▼

苏州

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1 ✕ 可用区2 ✕ ▼ C

可用区3 ✕

*** 名称** as-group-demo

*** 最大实例数(台)** 3

*** 期望实例数(台)** 1 ?

*** 最小实例数(台)** 0

以下信息是伸缩后实例所使用的虚拟私有云、子网，负载均衡负责自动分发访问流量。

*** 虚拟私有云** ?

(192.168.0.0/16) ▼ C 新建虚拟私有云

*** 子网**

subn- (192.168.2.0/24) ▼ 本网作为云主机的主网卡

自动分配IPv6地址

+ 增加一个子网 您还可以增加 4 个子网 C 新建子网

负载均衡 ?

不使用 使用经典型 使用增强型 C 新建负载均衡

伸缩组中的云主机自动挂载到您关联的负载均衡下。

负载均衡器 elb-app- 3... ▼ **后端云主机组** --请选择-- ▼

后端端口 ? 权重 ? 1 ▼

+ 新增一个负载均衡器 您还可以增加2个负载均衡器。

*** 实例移除策略** 根据较早创建的配置较早创建的实例 ▼

弹性IP

释放 不释放

若选择“释放”，在伸缩组进行缩的活动时，则会将云主机上的弹性IP释放，否则仅做解绑定操作，保留弹性IP资源。

*** 健康检查方式** ?

负载均衡健康检查 ▼

受保护的实例状态异常时，会被健康检查移除，并重新创建新的实例。
实例所在安全组则需要配置放行100.125.0.0/16，并配置负载均衡用于健康检查的协议和端口，否则会导致健康检查失败。

*** 健康检查间隔** ?

5分钟 ▼

*** 企业项目** ?

default ▼ C

高级配置

暂不配置 现在配置

在高级配置中可对通知方式和标签进行配置。

下一页

各参数说明如下：

- **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。

如果您选择多个可用区，为提高应用的高可用性，您的云主机会被均匀的创建在不同的可用区内。

- 多可用区扩展策略：可选择“均衡分布”或“选择优先”。

A.均衡分布：云主机扩容时优先保证选择的可用区列表中各可用区下云主机数量均衡，当无法在目标可用区下完成云主机扩容时，按照选择优先原则选择其他可用区。

B.选择优先：云主机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。

- 伸缩组名称：创建伸缩组的名称。
- 最大/最小实例数：指伸缩组中云主机个数的最大值/最小值。如果最小值不为 0，则创建伸缩组后会触发一次伸缩活动，使伸缩组内实例数达到最小值。
- 虚拟私有云：弹性云主机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主机均属于该 VPC。
- 子网：默认情况下，一个 VPC 子网内的弹性云主机均可以进行通信。一个伸缩组只能选择 1 个子网。
- 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主机，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。伸缩组关联负载均衡实例后，无论是伸缩组自动创建实例，还是用户向伸缩组手动添加实例，实例都会自动加入到负载均衡实例的后端服务器组。

📖 说明

一个伸缩组可最多添加 10 组负载均衡监听器。添加多个负载均衡监听器后，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”，弹性云主机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主机替换掉。当选择使用弹性负载均衡时，需配置 4 个参数：负载均衡器、后端主机组、后端端口和权重。

- 实例移除策略：实例优先被移除的策略。当伸缩组进行收缩活动时，会触发实例移除策略，包括如下四种方式：
 - A. 根据较早创建的配置较早创建的实例：先筛选出较早创建的配置所创建出的实例，再筛选较早创建的实例移出。
 - B. 较早创建的实例：根据时间筛选较早创建的实例。
 - C. 根据较早创建的配置较晚创建的实例：先筛选出较早创建的配置所创建出的实例，再筛选较晚创建的实例移出。
 - D. 较晚创建的实例：根据时间筛选较晚创建的实例。
- 按量付费实例回收模式：指按量付费的实例实例被移出后的处理策略。对于包年包月的云主机，如果被移出伸缩组，也不删除不更改状态，不适用此策略。
 - A. 释放模式：将移出的按量付费的云主机释放；
 - B. 停机回收模式：将移出的云主机停机，不释放。随着扩张活动仍会优先再加入。
- 健康检查方式：健康检查会将异常的云主机从伸缩组中移除，并重新创建新的云主机，伸缩组的健康检查方式包括以下两种。
 - A. 云主机健康检查：是指对云主机的运行状态进行检查，如关机、删除都是云主机异常状态。默认为此选项，伸缩组会定期使用云主机健康检查结果来确定每个云主机的运行状况。如果未通过云主机健康检查，则伸缩组会将该云主机移出伸缩组，并重新创建云主机。
 - B. 弹性负载均衡健康检查：是指根据 ELB 对云主机的健康检查结果进行的检查。当伸

缩组使用弹性负载均衡器时，您可以选择此方式来做健康检查，所有监听器下检测到的云主机状态必须均为正常，否则伸缩组会将该弹性云主机移出伸缩组，并重新创建云主机。

- 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（5 分钟、15 分钟、1 小时、3 小时），以确保其包含云主机中应用程序的预期启动时间。
5. 单击【下一步】，进入创建伸缩配置页面。创建一个伸缩组时，必须同时创建或选择一个伸缩配置。

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。后续可以在伸缩组创建完成后，为伸缩组修改伸缩配置。如需要创建新的伸缩配置，需填写伸缩配置信息，例如配置名称、云主机镜像、云主机类型、登录方式等；

< 创建弹性伸缩组

伸缩组创建完成后，您还可以根据业务需求更换伸缩配置。

伸缩配置

* 名称
使用该配置创建的云主机名称为伸缩配置名称加八位随机码。

* 配置模板

配置模板ID 083f8c33-8255-4747-ad32-29cdd8a6818d
规格 s3.large.2 | 2vCPUs | 4GB
镜像 CentOS7.9 64位
系统盘 高IO(可用区3下已售罄) | 40 GB
数据盘 --
安全组 test-cc-e-node-ufz91
(入方向:ICMP; TCP/443, 80, 30000-32767, 3389, 22; UDP/30000-32767 | 出方向:-)

弹性IP
不使用弹性IP的云主机不能与互联网互通，仅可作为私有网络中部署业务或者集群所需云主机进行使用。

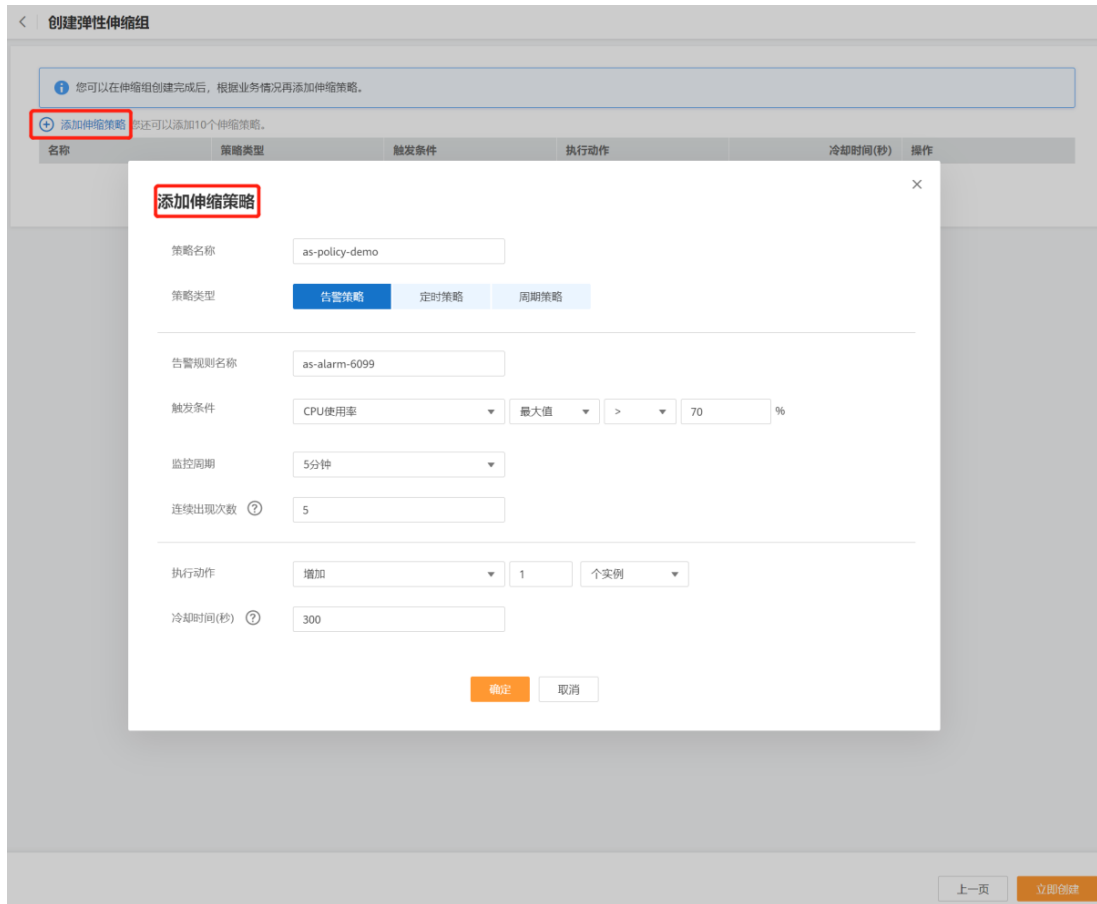
* 登录方式
用户名 root
* 密码 请妥善保管密码，系统无法获取您设置的密码内容。

* 确认密码

6. 单击【下一页】，进入创建伸缩策略页面。创建一个伸缩组时，可以同时创建伸缩策略，也可以在伸缩组创建完成后，根据业务需要再为伸缩组创建伸缩策略。

如果要同时创建伸缩策略，则在伸缩策略页面，单击【创建伸缩策略】按钮。

在弹出的创建伸缩策略窗口中，填写策略名称、策略类型、触发条件、执行动作、冷却时间等参数；



7. 单击【确认】;
8. 单击【立即创建】按钮;
9. 请核对弹性伸缩组、伸缩配置和伸缩策略的信息，单击【立即创建】。

4.1.2 (可选) 添加负载均衡器到伸缩组

弹性负载均衡 (Elastic Load ，简称 CT-ELB) 是将访问流量根据转发策略分发到后端多台云主机流量分发控制服务。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

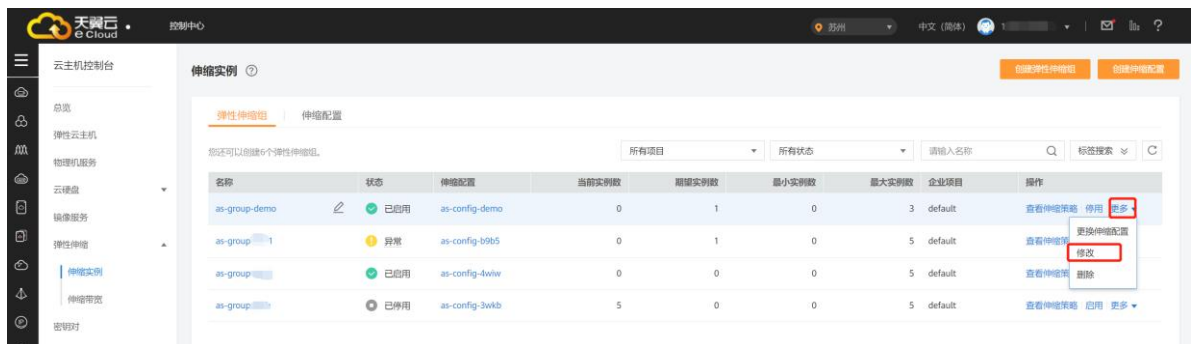
负载均衡器请参见《弹性负载均衡用户指南》。为伸缩组添加负载均衡器方法如下：

- 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，可以通过修改伸缩组的负载均衡配置，为伸缩组添加负载均衡器。操作可参考[修改伸缩组](#)。

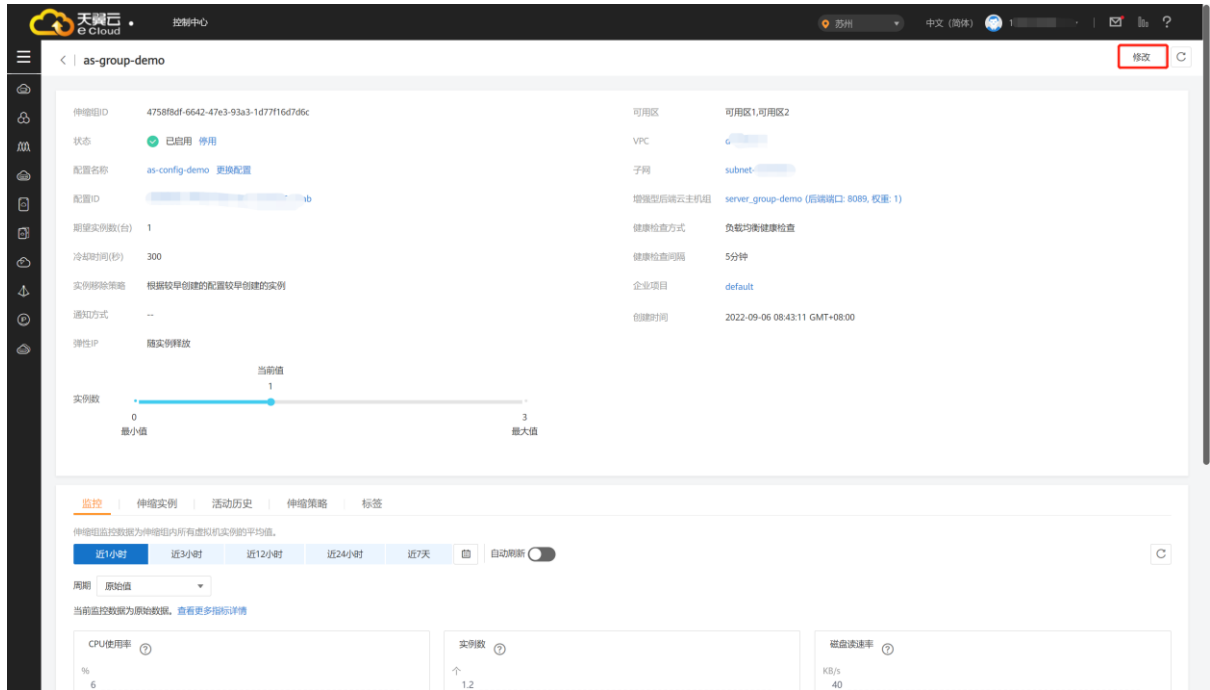
4.1.3 修改伸缩组信息

修改伸缩组的属性和参数。可以修改的参数有：伸缩组名称、最大实例数、最小实例数、健康检查方式、健康检查周期、实例移除策略和实例回收模式。如果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，还可以设置其他高级选项，包括子网和负载均衡监听器。

1. 在伸缩组列表中，伸缩组所在行的【操作】列下，单击【更多】-【修改】；



2. 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【修改】；



3. 修改相关数据;

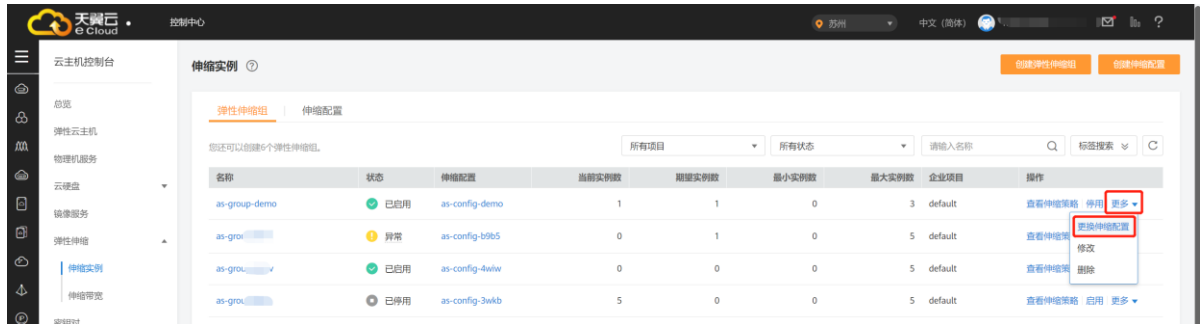


4. 单击【确认】，完成修改。

4.1.4 更换伸缩组的伸缩配置

更换伸缩组使用的伸缩配置：

1. 在伸缩组列表中，伸缩组所在行的【操作】列下，单击【更多】 - 【更换伸缩配置】；



2. 您也可以单击伸缩组的名称，进入该伸缩组基本信息页面，单击配置名称右侧的【更换配置】按钮；



3. 在弹出的【更换伸缩配置】对话框中，重新选择伸缩组使用的伸缩配置；

更换伸缩配置

| 名称 | 规格 | 镜像 | 系统盘 | 数据盘 | 登录方式 | 创建时间 |
|---|-----------------------|-------------------|-----------|----------------|------|------------------------|
| <input checked="" type="radio"/> as-config-demo | s3.large.2 2vCP... | CentOS7.9 64位 | 高IO(可用... | -- | 密码 | 2022-09-06 08:43:13 GM |
| <input type="radio"/> as-config-5f57 | s3.medium.2 1v... | CentOS7.6 64位 | 普通IO(可... | -- | 密码 | 2022-04-21 18:12:04 GM |
| <input type="radio"/> as-config-b9b5 | s3.medium.2 1v... | CentOS7.6 64位 | 普通IO(可... | -- | 密码 | 2022-04-11 11:29:49 GM |
| <input type="radio"/> as-config-4wiw | s3.large.2 2vCP... | Windows 2016 D... | 普通IO(可... | -- | 密码 | 2019-06-25 13:27:12 GM |
| <input type="radio"/> as-config-3wkb | c1.large 2vCPUs ... | Windows 2008 E... | 普通IO(可... | 普通IO(可用区1, ... | 密码 | 2017-07-24 17:57:29 GM |

注：伸缩配置的创建方法请参见 3.2 伸缩配置管理；

4. 单击【确认】按钮。

4.1.5 启用伸缩组

仅当伸缩组状态为“已停用”才可以启用伸缩组。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主机数量小于“最小实例数”或大于“最大实例数”时，系统自动添加或减少相应数量的云主机，相应的便会有有一次弹性伸缩活动。启用伸缩组的步骤如下：

1. 在伸缩组列表中，伸缩组所在行的【操作】列下，单击【启用】；

弹性伸缩组 | 伸缩配置您还可以创建0个弹性伸缩组。所有项目所有状态请输入名称Q标签搜索C

| 名称 | 状态 | 伸缩配置 | 当前实例数 | 期望实例数 | 最小实例数 | 最大实例数 | 企业项目 | 操作 |
|---------------|-----|----------------|-------|-------|-------|-------|---------|--|
| as-group-demo | 已启用 | as-config-demo | 1 | 1 | 0 | 3 | default | 查看伸缩策略 停用 更多 |
| as-group- | 异常 | as-config-b9b5 | 0 | 1 | 0 | 5 | default | 查看伸缩策略 停用 更多 |
| as-group- | 已启用 | as-config-4wiw | 0 | 0 | 0 | 5 | default | 查看伸缩策略 停用 更多 |
| as-group-kplb | 已停用 | as-config-3wkb | 5 | 0 | 0 | 5 | default | 查看伸缩策略 启用 更多 |

2. 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】。



4.1.6 停用伸缩组

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，不会触发任何弹性伸缩活动。只有当伸缩组状态是“已启用”时，才能停用伸缩组。如果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。停用伸缩组步骤如下：

1. 在伸缩组列表中，伸缩组所在行的【操作】列下，单击【停用】；



2. 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】。



4.1.7 删除伸缩组

当您不再需要某个伸缩组时，可以删除伸缩组。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。删除伸缩组，将释放组内由伸缩配置创建的实例、移除手动移入的实例，删除伸缩策略，但不删除伸缩配置。在伸缩组所在行的【操作】列下，单击【删除】。



4.2 伸缩配置管理

伸缩配置即伸缩活动中添加的云主机的规格。

4.2.1 创建伸缩配置

伸缩配置用于定义伸缩组内待创建的云主机的规格数据。包括云主机的操作系统镜像、系统盘大小等。

1. 登录天翼云控制中心，切换到需要创建伸缩配置的节点，选择【弹性伸缩服务】；
2. 切换至【伸缩配置】页面，单击【创建伸缩配置】；

伸缩实例 ②

创建弹性伸缩组
创建伸缩配置

弹性伸缩组 | 伸缩配置

删除 您还可以添加0~5个配置。

名称

| <input type="checkbox"/> | 名称 | 状态 | 规格 | 镜像 | 系统盘 | 数据盘(个) | 登录方式 | 创建时间 | 操作 |
|--------------------------|----------------|-----|---------------------------|--------------------------|-------------------|--------|------|-------------------------------|-------|
| <input type="checkbox"/> | as-config-demo | 绑定 | s3.large.2 2vCPUs 4GB | CentOS7.9 64位 | 高IO(可用区3下已售罄) ... | 0 | 密码 | 2022-09-06 08:43:13 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> | as-config-5 | 未绑定 | s3.medium.2 1vCPU ... | CentOS7.6 64位 | 普通IO(可用区1, 可用... | 0 | 密码 | 2022-04-21 18:12:04 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> | as-config- | 绑定 | s3.medium.2 1vCPU ... | CentOS7.6 64位 | 普通IO(可用区1, 可用... | 0 | 密码 | 2022-04-11 11:29:49 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> | as-config- | 绑定 | s3.large.2 2vCPUs 4GB | Windows 2016 DataCe... | 普通IO(可用区1, 可用... | 0 | 密码 | 2019-06-25 13:27:12 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> | as-config- | 绑定 | c1.large 2vCPUs 2GB | Windows 2008 Enterpri... | 普通IO(可用区1, 可用... | 2 | 密码 | 2017-07-24 17:57:29 GMT+08:00 | 复制 删除 |

3. 填写伸缩配置信息，例如配置名称、云主机类型、云主机的镜像等。

< 创建伸缩配置

* 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。

* 名称
使用该配置创建的云主机名称为伸缩配置名称加八位随机码。

* 配置模板 使用新模板 使用已有云主机规格为模板

* 规格

最新系列 vCPUs 内存(GiB) 规格名称

通用计算型 通用独享型 内存优化型 高性能计算型 超高I/O型 GPU加速型

| 规格名称 | vCPUs 内存(GiB) <input type="checkbox"/> | 基准 / 最大带宽 <input type="checkbox"/> | 内网收发包 <input type="checkbox"/> |
|--|--|------------------------------------|--------------------------------|
| <input checked="" type="radio"/> s3.small.1 (可用区1, 可用区3下已售罄) | 1vCPU 1GB | 0.06/0.3 Gbit/s | 30,000 |
| <input type="radio"/> s3.medium.2 (可用区3下已售罄) | 1vCPU 2GB | 0.06/0.3 Gbit/s | 30,000 |
| <input type="radio"/> s3.medium.4 (可用区3下已售罄) | 1vCPU 4GB | 0.06/0.3 Gbit/s | 30,000 |
| <input type="radio"/> s3.large.2 (可用区3下已售罄) | 2vCPUs 4GB | 0.12/0.5 Gbit/s | 60,000 |
| <input type="radio"/> s3.large.4 (可用区3下已售罄) | 2vCPUs 8GB | 0.12/0.5 Gbit/s | 60,000 |
| <input type="radio"/> s3.xlarge.2 (可用区3下已售罄) | 4vCPUs 8GB | 0.25/1 Gbit/s | 90,000 |
| <input type="radio"/> s3.xlarge.4 (可用区3下已售罄) | 4vCPUs 16GB | 0.25/1 Gbit/s | 90,000 |
| <input type="radio"/> s3.2xlarge.2 (可用区3下已售罄) | 8vCPUs 16GB | 0.5/1.5 Gbit/s | 120,000 |

当前规格 通用计算型 | s3.small.1 | 1vCPU | 1GB

立即创建

The screenshot shows a configuration interface with the following sections:

- * 镜像**: Buttons for "公共镜像", "私有镜像", and "共享镜像". Below are dropdowns for "CentOS" and "CentOS7.9 64位(40GB)".
- * 磁盘**: Button for "云硬盘". Below is a "系统盘" dropdown set to "高IO(可用区3下已售罄)", a "40" input field, and "GB IOPS上限1,440, IOPS突发上限5,000". A link says "增加一块数据盘 您还可以增加 23 块磁盘 (云硬盘) ."
- * 安全组**: A dropdown menu set to "Sys-default (入方向:ICMP, TCP/3389, 2...)", a "新建安全组" link, and descriptive text about security groups.
- 弹性IP**: Buttons for "不使用" and "自动分配". Below is a note: "不使用弹性IP的云主机不能与互联网互通, 仅可作为私有网络中部署业务或者集群所需云主机进行使用."
- * 登录方式**: Buttons for "密码" and "密钥对".
- 用户名**: Input field with "root".
- * 密码**: Input field with "*****" and a note: "请妥善保管密码, 系统无法获取您设置的密码内容."
- * 确认密码**: Input field with "*****".

An "立即创建" button is located at the bottom right of the configuration area.

- 配置名称：创建伸缩配置的名称。
- 配置规格：
 - A. 选择“使用现有云主机规格 > 请选择云主机” 创建配置，云主机类型、vCPU、内存、镜像、磁盘参数、安全组信息将默认与选择的云主机规格保持一致。
 - B. 选择“使用新规格”，接下来配置服务器类型、vCPU、内存、镜像、磁盘参数、安全组信息。
- 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主机绑定，以实现虚拟私有云内的弹性云主机通过固定的公网 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：
 - A. 不使用：弹性云主机不能与互联网互通，仅可作为私有网络中部署业务或者集群所需弹性云主机进行使用。

- B. 自动分配：自动为每台弹性云主机分配独享带宽的弹性 IP，带宽值可以由您设定。
- 登录方式：公有云提供两种弹性云主机鉴权方式。
 - A. 密钥对指使用密钥作为弹性云主机的鉴权方式。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主机。
 - B. 账户密码指使用设置 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主机的鉴权方式，如果选择此方式，您可以通过用户名密码方式登录弹性云主机。
 - 高级配置：高级配置可对文件注入、用户数据注入进行配置。可选择“暂不配置”和“现在配置”。
 - 文件注入：可选配置，主要用于创建弹性云主机时向弹性云主机注入脚本文件或其他文件。配置文件注入后，系统在创建弹性云主机时自动将文件注入到指定目录下。
 - A. Linux 系统请输入注入文件保存路径，例如 “/etc/foo.txt”。
 - B. 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。
 - 用户数据注入：可选配置，主要用于创建弹性云主机时向弹性云主机注入用户数据。配置用户数据注入后，弹性云主机首次启动时会自行注入数据信息。

说明

对于 Linux 云主机，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。

4. 单击【确认】，系统会提示伸缩配置创建成功，并返回弹性伸缩服务页面的【伸缩配置】标签页。您可以在【伸缩配置】标签页看到新创建的伸缩配置。

多规格配置说明：

伸缩配置支持多规格可极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。

一个伸缩配置内可选的实例规格最多为 10 个。

适用场景：

- 对于伸缩组创建的实例规格无特殊要求，不限于于某一种规格。
- 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。

4.2.2 使用已有云主机创建伸缩配置

操作场景：

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

操作步骤:

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩服务”。
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。配置数据说明如下表：

表：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。 选择“使用已有云主机规格、镜像、磁盘、安全组为模板，快速创建伸缩配置”。 | - |
| 选择云主机 | 选择使用已有云主机规格、镜像、磁盘、安全组为模板，快速创建伸缩配置。 | - |
| 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主机绑定，以实现虚拟私有云内的弹性云主机通过固定的公网 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"> • 不使用： 弹性云主机不能与互联网互通，仅可作为私有网络中部署业务或者集群所需弹性云主机进行使用。 • 自动分配： 自动为每台弹性云主机分配独享带宽的弹性 IP，您可以在“带宽”中设置带宽值。 | 自动分配 |
| 登录方式 | 云平台提供两种弹性云主机鉴权方式。 <ul style="list-style-type: none"> • 密钥对 指使用密钥作为弹性云主机的鉴权方式。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主机。 <ul style="list-style-type: none"> • 账户密码 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主机的鉴权方式，如果选择此方式， | Admin@123 |

| 参数 | 解释 | 取值样例 |
|--------|---|------|
| | 您可以通过用户名密码方式登录弹性云主机。 | |
| 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”。 | - |
| 文件注入 | 可选配置，主要用于创建弹性云主机时向弹性云主机注入脚本文件或其他文件。配置文件注入后，系统在创建弹性云主机时自动将文件注入到指定目录下。 <ul style="list-style-type: none"> 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。 | - |
| 用户数据注入 | 可选配置，主要用于创建弹性云主机时向弹性云主机注入用户数据。配置用户数据注入后，弹性云主机首次启动时会自行注入数据信息。 说明 对于 Linux 云主机，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。 | - |

- 单击“立即申请”，系统会提示伸缩配置创建成功，并返回弹性伸缩页面的“伸缩配置”页签。您可以在“伸缩配置”页签看到新创建的伸缩配置。
- 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。

4.2.3 使用新模板创建伸缩配置

操作场景：

若您对扩展的云主机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，使得伸缩组内云主机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 > 弹性伸缩服务”。
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、云主机的镜像、云主机类型等。

配置数据说明如下表：

表：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。 | - |
| 类型规格 | 公有云提供了多种类型的弹性云主机供您选择，针对不同的应用场景，可以选择不同规格的弹性云主机。 根据您选择的“云主机类型”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 | <ul style="list-style-type: none"> • 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 用户基于弹性云主机创建的个人镜像，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主机，可以节省您重复配置弹性云主机的时间。 • 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 |
| 磁盘 | 包括系统盘和数据盘。 <ul style="list-style-type: none"> • 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。 通用型 SSD：是指由 GPSSD 存储提供资源的磁盘类型。 极速型 SSD：是指由 ESSD 存储提供资源的磁盘类型。 • 数据盘 您可以为云主机添加多块数据盘，或者从指定的数据盘镜像导出数据到某个数据盘。 | “系统盘” 选为“普通 IO” |

| 参数 | 解释 | 取值样例 |
|--------|---|-----------|
| 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主机的访问控制，加强弹性云主机的安全保护。用户可以在安全组中定义各种访问规则，当弹性云主机加入该安全组后，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主机绑定，以实现虚拟私有云内的弹性云主机通过固定的公网 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"> 不使用：弹性云主机不能与互联网互通，仅可作为私有网络中部署业务或者集群所需弹性云主机进行使用。 自动分配：自动为每台弹性云主机分配独享带宽的弹性 IP，带宽值可以由您设定。 | 自动分配 |
| 登录方式 | 云平台提供两种弹性云主机鉴权方式。 <ul style="list-style-type: none"> 密钥对 指使用密钥作为弹性云主机的鉴权方式。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主机。 账户密码 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主机的鉴权方式，如果选择此方式，您可以通过用户名密码方式登录弹性云主机。 | Admin@123 |
| 高级配置 | 高级配置可对文件注入、用户数据注入进行配置。 可选择“暂不配置”和“现在配置”。 | - |
| 文件注入 | 可选配置，主要用于创建弹性云主机时向弹性云主机注入脚本文件或其他文件。配置文件注入后，系统在创建弹性云主机时自动将文件注入到指定目录下。 <ul style="list-style-type: none"> 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。 | - |
| 用户数据注入 | 可选配置，主要用于创建弹性云主机时向弹性云主机注入用户数据。配置用户数据注入后，弹性云主机首次启动时会自行注入数据信息。 | - |

| 参数 | 解释 | 取值样例 |
|----|---|------|
| | 说明 对于 Linux 云主机，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。 | |

- 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。
- 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。
- (可选) 启动伸缩组。
如果伸缩组状态时“未启用”状态，需要启动伸缩组。

4.2.4 复制伸缩配置

根据原有伸缩配置，新增一条伸缩配置。

- 在【弹性伸缩】页面中，单击【伸缩配置】标签进入伸缩配置列表页；
- 在需要复制的伸缩配置所在行的【操作】列下，单击【复制】按钮；



- 在【复制配置】页面，可以修改配置名称、云主机规格和镜像等参数；

复制伸缩配置 ×

* 名称
使用该配置创建的云主机名称为伸缩配置名称加八位随机码。

* 规格

最新系列 vCPUs 全部 内存(GiB) 全部 规格名称

通用计算型 通用独享型 内存优化型 高性能计算型 超高I/O型 GPU加速型

| 规格名称 | vCPUs 内存(GiB) | 基准 / 最大带宽 | 内网收发包 |
|--|-----------------|-----------------|---------|
| <input type="radio"/> s3.small.1 (可用区1, 可用区3下已售罄) | 1vCPU 1GB | 0.06/0.3 Gbit/s | 30,000 |
| <input type="radio"/> s3.medium.2 (可用区3下已售罄) | 1vCPU 2GB | 0.06/0.3 Gbit/s | 30,000 |
| <input type="radio"/> s3.medium.4 (可用区3下已售罄) | 1vCPU 4GB | 0.06/0.3 Gbit/s | 30,000 |
| <input checked="" type="radio"/> s3.large.2 (可用区3下已售罄) | 2vCPUs 4GB | 0.12/0.5 Gbit/s | 60,000 |
| <input type="radio"/> s3.large.4 (可用区3下已售罄) | 2vCPUs 8GB | 0.12/0.5 Gbit/s | 60,000 |
| <input type="radio"/> s3.xlarge.2 (可用区3下已售罄) | 4vCPUs 8GB | 0.25/1 Gbit/s | 90,000 |
| <input type="radio"/> s3.xlarge.4 (可用区3下已售罄) | 4vCPUs 16GB | 0.25/1 Gbit/s | 90,000 |
| <input type="radio"/> s3.2xlarge.2 (可用区3下已售罄) | 8vCPUs 16GB | 0.5/1.5 Gbit/s | 120,000 |

当前规格 通用计算型 | s3.large.2 | 2vCPUs | 4GB

4. 单击【确认】按钮完成复制。

4.2.5 删除伸缩配置

删除指定伸缩配置。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。

1. 在【弹性伸缩】页面中，单击【伸缩配置】标签进入伸缩配置列表页；
2. 在需要删除的伸缩配置所在行的【操作】列下，单击【删除】按钮；

弹性伸缩组 | 伸缩配置

删除 您还可以添加95个配置。

| 名称 | 状态 | 规格 | 镜像 | 系统盘 | 数据盘(个) | 登录方式 | 创建时间 | 操作 |
|---|-----|---------------------------|--------------------------|-------------------|--------|------|-------------------------------|-------|
| <input type="checkbox"/> as-config-demo | 绑定 | s3.large.2 2vCPUs 4GB | CentOS7.9 64位 | 高IO(可用区3下已售罄) ... | 0 | 密码 | 2022-09-06 08:43:13 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> as-config-5f57 | 未绑定 | s3.medium.2 1vCPU ... | CentOS7.6 64位 | 普通IO(可用区1, 可用... | 0 | 密码 | 2022-04-21 18:12:04 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> as-config-... | 绑定 | s3.medium.2 1vCPU ... | CentOS7.6 64位 | 普通IO(可用区1, 可用... | 0 | 密码 | 2022-04-11 11:29:49 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> as-config-... | 绑定 | s3.large.2 2vCPUs 4GB | Windows 2016 DataCe... | 普通IO(可用区1, 可用... | 0 | 密码 | 2019-06-25 13:27:12 GMT+08:00 | 复制 删除 |
| <input type="checkbox"/> as-config-... | 绑定 | c1.large 2vCPUs 2GB | Windows 2008 Enterpri... | 普通IO(可用区1, 可用... | 2 | 密码 | 2017-07-24 17:57:29 GMT+08:00 | 复制 删除 |

4.3 伸缩策略管理

伸缩策略用于定义伸缩活动的触发模式、触发条件和触发动作。例如，一般情况下，凌晨 0 点到 8 点业务访问量较小，因此，为节约资源，您可以添加两条策略：每天凌晨 0 点移除一台云主机；每天 8 点添加一台云主机。

弹性伸缩的策略模式有以下几种：

- 告警策略：基于云监控系统告警数据（例如 CPU 利用率等），自动增加、减少或设置指定数量的云主机；
- 周期策略：按照配置周期（天、星期、月），定期增加、减少或设置指定数量的云主机；
- 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主机；

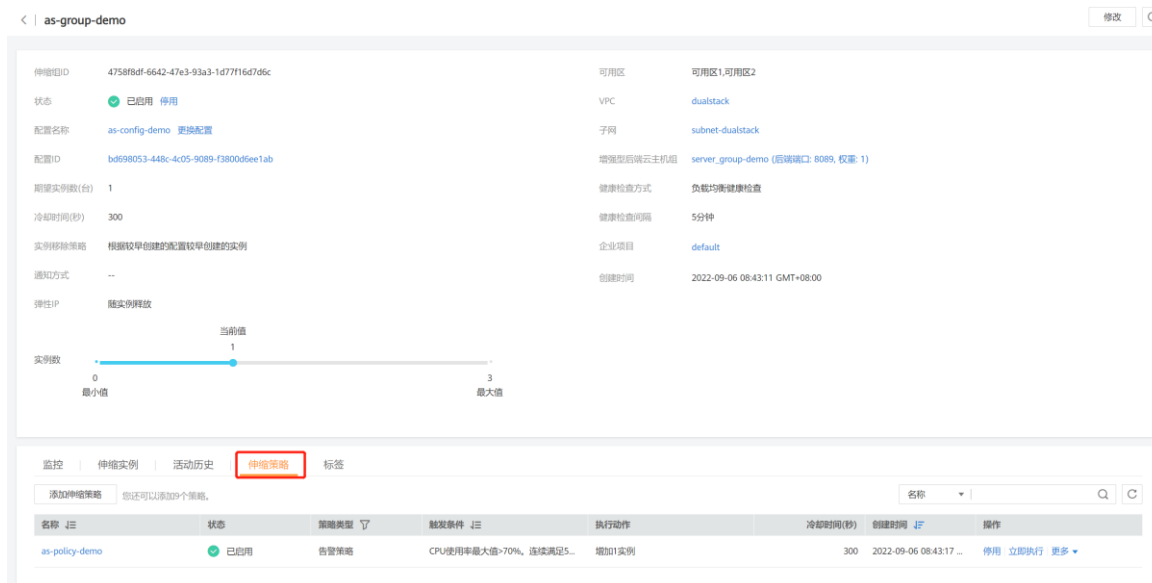
4.3.1 创建伸缩策略

1. 登录天翼云控制中心，切换到需要创建伸缩策略的节点，选择【弹性伸缩服务】；
2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；



| 名称 | 状态 | 伸缩配置 | 当前实例数 | 期望实例数 | 最小实例数 | 最大实例数 | 企业项目 | 操作 |
|---------------|-----|----------------|-------|-------|-------|-------|---------|--------------|
| as-group-demo | 已启用 | as-config-demo | 1 | 1 | 0 | 3 | default | 查看伸缩策略 停用 更多 |
| as-grou | 异常 | as-config-b5b5 | 0 | 1 | 0 | 5 | default | 查看伸缩策略 停用 更多 |
| as-grou | 已启用 | as-config-4wiw | 0 | 0 | 0 | 5 | default | 查看伸缩策略 停用 更多 |
| as-grou | 已停用 | as-config-3wkb | 5 | 0 | 0 | 5 | default | 查看伸缩策略 启用 更多 |

3. 在【伸缩组详情页】，单击【伸缩策略】标签，进入【伸缩策略标签页】；



- 在【伸缩策略】标签页，单击【添加伸缩策略】按钮；



- 在【创建策略】页面，配置弹性伸缩组的策略，例如策略名称、策略类型、执行动作、冷却时间；

添加伸缩策略

✕

| | |
|----------------------|--|
| 策略名称 | <input type="text" value="as-policy-demo"/> |
| 策略类型 | 告警策略 定时策略 周期策略 |
| 告警规则 | 现在创建 使用已有 |
| 告警规则名称 | <input type="text" value="as-alarm-d13a"/> |
| 触发条件 | <input type="text" value="CPU使用率"/> <input type="text" value="最大值"/> <input type="text" value=">"/> <input type="text" value="70"/> % |
| 监控周期 | <input type="text" value="5分钟"/> |
| 连续出现次数 [?] | <input type="text" value="5"/> |
| 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="300"/> |

确定 取消

- 策略名称：创建伸缩策略的名称。
- 策略类型：计划扩展资源的策略类型可选择告警策略。

A. 告警策略：

添加伸缩策略

✕

| | | | |
|----------------------|---|------------------------------------|---|
| 策略名称 | <input type="text" value="as-policy-demo"/> | | |
| 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-d13a"/> | | |
| 触发条件 | <input type="text" value="CPU使用率"/> ▼ | <input type="text" value="最大值"/> ▼ | <input type="text" value=">"/> ▼ <input type="text" value="70"/> % |
| 监控周期 | <input type="text" value="5分钟"/> ▼ | | |
| 连续出现次数 [?] | <input type="text" value="5"/> | | |
| 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="300"/> | | |

- a) 创建新的告警规则，该规则同时将用于对该伸缩组进行告警。选择【现在创建】，配置告警规则名称、触发条件、监控周期、连续出现次数。
- b) 使用已有的告警规则，在现有的针对当前伸缩组的告警规则中选择一条，触发条件、监控周期、连续出现次数均与此告警规则一致。

添加伸缩策略

| | |
|----------------------|---|
| 策略名称 | <input type="text" value="as-policy-demo"/> |
| 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 type="radio"/> 现在创建 <input checked="" type="radio"/> 使用已有 |
| 告警规则名称 | <input type="text" value="as-alarm-6099"/> <input type="button" value="新建告警规则"/> |
| 触发条件 | CPU使用率最大值>70% |
| 监控周期 | 5分钟 |
| 连续出现次数 [?] | 5 |
| 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="300"/> |

配置冷却时间，冷却时间是指冷却伸缩活动的时间，在每次触发伸缩活动之后，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等）触发的伸缩活动不受限制，但会重新开始计算冷却时间。

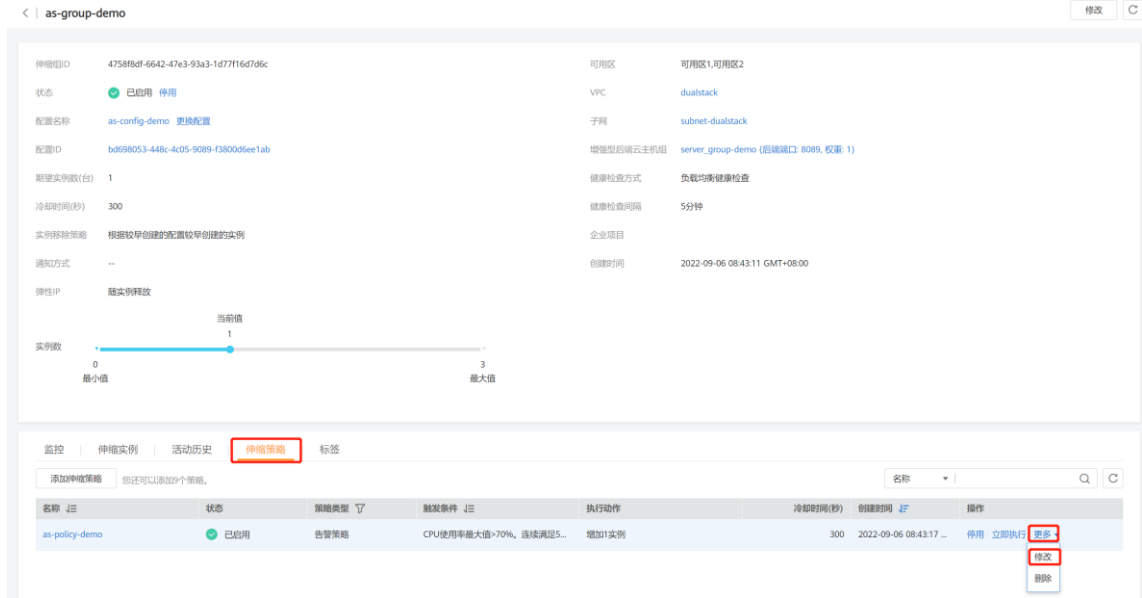
- 执行动作：设置伸缩活动执行动作及实例的个数。执行动作包括：
 - A. 增加：当执行伸缩活动时，向伸缩组增加实例。
 - B. 减少：当执行伸缩活动时，从伸缩组中减少实例。
 - C. 设置为：将伸缩组中的实例数设置为固定值。
- 6. 单击【确定】按钮，新创建的伸缩策略会在“伸缩策略”页签中显示。

4.3.2 修改伸缩策略

创建完伸缩策略后，您还可以根据需要调整触发条件、策略名称、冷却时间等。

1. 登录天翼云控制中心，切换到需要修改伸缩策略的节点，选择【弹性伸缩服务】；

2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击【伸缩策略】标签，进入【伸缩策略标签页】；
4. 在伸缩策略所在行的【操作】列下，单击【修改】；



5. 可修改策略名称、执行动作等，或者创建新告警规则并使用。

修改伸缩策略

✕

策略名称

策略类型 告警策略 定时策略 周期策略

告警规则 现在创建 使用已有

告警规则名称 新建告警规则

触发条件 CPU使用率最大值>70%

监控周期 5分钟

连续出现次数

执行动作

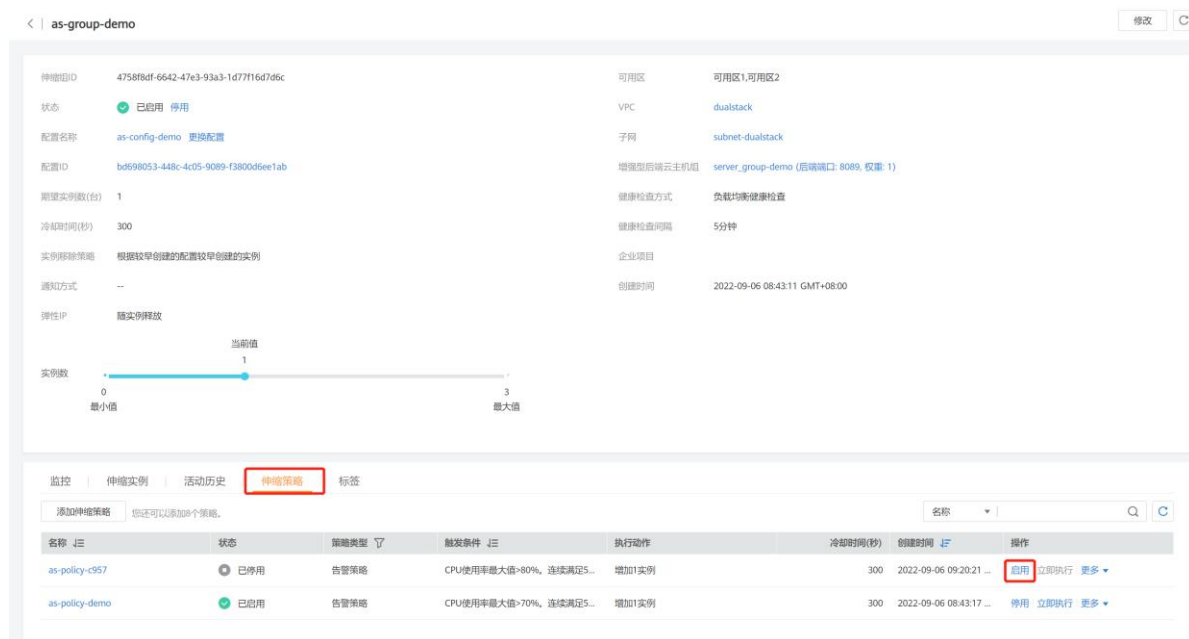
冷却时间(秒)

确定 取消

4.3.3 启用伸缩策略

只有当伸缩策略处于启用状态时，伸缩策略才能触发伸缩活动。伸缩策略可以启用一个，也可以启用多个。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，启用伸缩策略步骤如下：

1. 登录天翼云控制中心，切换到需要修改伸缩策略的节点，选择【弹性伸缩服务】；
2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击【伸缩策略】标签，进入【伸缩策略标签页】；
4. 在伸缩策略所在行的【操作】列下，单击【启用】；



4.3.4 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。仅当伸缩策略状态为“已启用”时，才可以停用伸缩策略。具体步骤如下：

1. 登录天翼云控制中心，切换到需要修改伸缩策略的节点，选择【弹性伸缩服务】；
2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击【伸缩策略】标签，进入【伸缩策略标签页】；
4. 在伸缩策略所在行的【操作】列下，单击【停用】。



| 名称 | 状态 | 策略类型 | 触发条件 | 执行动作 | 冷却时间(秒) | 创建时间 | 操作 |
|----------------|-----|------|------------------------|-------|---------|-------------------------|------------|
| as-policy-c957 | 已停用 | 告警策略 | CPU使用率最大值>80%。连续满足5... | 增加1实例 | 300 | 2022-09-06 09:20:21 ... | 启用 立即执行 更多 |
| as-policy-demo | 已启用 | 告警策略 | CPU使用率最大值>70%。连续满足5... | 增加1实例 | 300 | 2022-09-06 08:43:17 ... | 停用 立即执行 更多 |

4.3.5 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，步骤如下：

1. 登录天翼云控制中心，切换到需要修改伸缩策略的节点，选择【弹性伸缩服务】；
2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击【伸缩策略】标签，进入【伸缩策略标签页】；
4. 在伸缩策略所在行的【操作】列下，单击【立即执行】。



| 名称 | 状态 | 策略类型 | 触发条件 | 执行动作 | 冷却时间(秒) | 创建时间 | 操作 |
|----------------|-----|------|------------------------|-------|---------|-------------------------|------------|
| as-policy-c957 | 已停用 | 告警策略 | CPU使用率最大值>80%。连续满足5... | 增加1实例 | 300 | 2022-09-06 09:20:21 ... | 启用 立即执行 更多 |
| as-policy-demo | 已启用 | 告警策略 | CPU使用率最大值>70%。连续满足5... | 增加1实例 | 300 | 2022-09-06 08:43:17 ... | 停用 立即执行 更多 |

4.3.6 删除伸缩策略

某个伸缩策略不再用于触发伸缩活动，可选择删除指定伸缩策略。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。步骤如下：

1. 登录天翼云控制中心，切换到需要修改伸缩策略的节点，选择【弹性伸缩服务】；

2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击【伸缩策略】标签，进入【伸缩策略标签页】；
4. 在伸缩策略所在行的【操作】列下，单击【删除】。



4.4 伸缩活动管理

伸缩活动是指在伸缩组内由于伸缩条件满足而触发的云主机实例数量变更的活动，可能是增加或减少几台云主机实例。

4.4.1 查询伸缩活动

以下场景会触发伸缩活动：

- 系统检测到伸缩策略中配置的条件满足要求。
- 手工修改伸缩组的最大/最小实例数，导致当前组内实例数超过该最大/最小的限制。
- 手动添加或移除实例。

下面介绍如何查看伸缩组的伸缩活动：

1. 登录天翼云控制中心，切换到需要查询伸缩活动的节点，选择【弹性伸缩服务】；
2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击在【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。



4. 【活动历史】标签，进入【伸缩活动标签页】，查看伸缩活动详情，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。

| 活动ID | 状态 | 活动类型 | 描述 | 开始时间 | 结束时间 |
|--------------------------------------|----|------|-----------------|-------------------------------|-------------------------------|
| aa114b8a-e887-49fb-b438-591a7aa1b269 | 成功 | 实例伸缩 | 期望实例数触发，增加1个实例。 | 2022-09-06 09:13:41 GMT+08:00 | 2022-09-06 09:14:41 GMT+08:00 |
| 6c329fa6-8d4e-407f-b207-dc3a18661600 | 成功 | 实例伸缩 | 健康检查触发，减少1个实例。 | 2022-09-06 09:13:11 GMT+08:00 | 2022-09-06 09:13:31 GMT+08:00 |
| 66925d2b-e8bc-440c-bd18-3b03c5867638 | 成功 | 实例伸缩 | 期望实例数触发，增加1个实例。 | 2022-09-06 08:58:41 GMT+08:00 | 2022-09-06 08:59:51 GMT+08:00 |
| d8fcb5e7-febd-46f6-9635-c853a99f88b2 | 成功 | 实例伸缩 | 健康检查触发，减少1个实例。 | 2022-09-06 08:58:11 GMT+08:00 | 2022-09-06 08:58:31 GMT+08:00 |
| a704c181-9538-4849-8df9-510e93a73f94 | 成功 | 实例伸缩 | 期望实例数触发，增加1个实例。 | 2022-09-06 08:43:22 GMT+08:00 | 2022-09-06 08:44:31 GMT+08:00 |

4.4.2 将实例移入伸缩组

伸缩组处于“已启用”状态时，您可以手动将云主机移入指定伸缩组。手动移入的云主机将加入到伸缩组关联的负载均衡服务器组。将云主机成功移入指定伸缩组必须满足如下条件：

- 云主机不能存在于其它伸缩组中；
- 被移入的云主机所在的 VPC 必须和伸缩组所在的 VPC 相同；
- 移入的云主机必须是运行中状态。
- 移入后实例数不能大于伸缩组的最大实例数。

将云主机移入伸缩组的步骤如下。

1. 登录天翼云控制中心，切换到需要修改伸缩策略的节点，选择【弹性伸缩服务】；

2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击【伸缩实例】标签，进入【伸缩实例标签页】；
4. 在【伸缩实例】页签，单击【移入伸缩组】；



| 名称 | 生命周期状态 | 健康状态 | 伸缩配置 | 实例加入方式 | 实例保护 | 加入时间 | 操作 |
|--|--------|------|----------------|--------|--------------------------|----------------------------|----------------|
| <input type="checkbox"/> ecs-8889 | 已启用 | 正常 | -- | 手动加入 | <input type="checkbox"/> | 2022-09-06 09:25:24 GMT... | 移出伸缩组 移出伸缩组并删除 |
| <input type="checkbox"/> as-config-demo-XMA6PL16 | 已启用 | 正常 | as-config-demo | 自动伸缩 | <input type="checkbox"/> | 2022-09-06 09:13:41 GMT... | 移出伸缩组 移出伸缩组并删除 |

5. 选择待移入的实例名称，单击【确认】按钮。

移入伸缩组

移入伸缩组 ×

提示

- 单次最大添加实例数为10。
- 您只能选择与伸缩组VPC相同、AZ包含于伸缩组AZ、且未被其它伸缩组使用的实例。

待选实例

| <input type="checkbox"/> | 名称 | ID | 可用区 |
|-------------------------------------|-------------|----------------------|------|
| <input checked="" type="checkbox"/> | test-98363 | 21a5f07c-2078-47... | 可用区1 |
| <input type="checkbox"/> | test-10649 | eceac32e-7fc4-4cb... | 可用区1 |
| <input type="checkbox"/> | ecs-3206-as | 2d53daa6-d9fc-4a... | 可用区1 |

| 名称 | ID | 可用区 | 操作 |
|------------|-----------------|------|-------------------------------------|
| test-98363 | 21a5f07c-207... | 可用区1 | <input type="button" value="取消勾选"/> |

4.4.3 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。当您选择“移出伸缩组”时，系统仅将其移出伸缩组，不释放、不删除、不改变云主机状态，云主机实例可以作为其他用途；当您选择“移出伸缩组并释放”时，系统会将伸缩组创建的云主机移出伸缩组并且将其释放，但该操作对手动移入伸缩组的云主机不生效。

将云主机成功移出指定伸缩组必须满足如下条件：

- 伸缩组没有正在进行的伸缩活动；
- 伸缩实例状态为“已启用”；

- 移出后实例数不能小于伸缩组的最小实例数时。

将云主机移出伸缩组的步骤如下：

1. 登录天翼云控制中心，切换到需要修改伸缩策略的节点，选择【弹性伸缩服务】；
2. 在伸缩组所在行，单击伸缩组名称，进入【伸缩组详情页】；
3. 在【伸缩组详情页】，单击【伸缩实例】标签，进入【伸缩实例标签页】；
4. 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】；



| 名称 | 生命周期状态 | 健康状态 | 伸缩配置 | 实例加入方式 | 实例保护 | 加入时间 | 操作 |
|-------------------------|--------|------|----------------|--------|------|----------------------------|----------------|
| ecs-8889 | 已启用 | 正常 | -- | 手动加入 | 关闭 | 2022-09-06 09:25:24 GMT... | 移出伸缩组 移出伸缩组并删除 |
| as-config-demo-XMA6PL16 | 已启用 | 正常 | as-config-demo | 自动伸缩 | 关闭 | 2022-09-06 09:13:41 GMT... | 移出伸缩组 移出伸缩组并删除 |

5. 如果您要移出所有实例，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】。

4.4.4 实例移除策略

当您的伸缩组自动移除实例时，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：

- 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。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。
- 根据较晚创建的配置较晚创建的实例：根据“较晚创建的配置”较晚创建的“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。
- 较早创建的实例：创建时间较早的实例优先被移除伸缩组。当您为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。

说明：当实例移出策略配置为“较早/较晚创建的实例”时，手动移入伸缩组的云主机也有可能被移除。如果手动移入的实例是按量付费的实例，则移出后将遵守您配置的“按量付费实例回收模式”规则；如果手动移入的实例是包周期的实例，则移出后不删除、不更改其状态。如果您想保证手动移入的实例不被移除，可以再【伸缩组详情页】-【伸缩实例】标签页中，对实例开启保护。

4.4.5 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

8. 登录管理控制台。
9. 选择“计算 > 弹性伸缩 > 弹性伸缩组”。
10. 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。
11. 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。
2. 选择“计算 > 弹性伸缩 > 弹性伸缩组”。
3. 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。
4. 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。



4.4.6 动态资源扩展

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。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。

4.4.7 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。

4.5 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、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 > 弹性伸缩服务 > 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。
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”。

< 创建伸缩带宽策略

区域 苏州
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。

策略名称 as-policy-bandwidth-demo

弹性IP 117.80.148.165 C 购买弹性公网IP
当前带宽大小 5 Mbit/s 当前您仅可对按流量计费模式的弹性IP进行伸缩。

策略类型 告警策略 定时策略 周期策略

告警规则 现在创建 使用已有

告警规则名称 as-alarm-b58c

触发条件 下行带宽 最大值 > 10 bit/s

监控周期 5分钟

连续出现次数 5

执行动作 设置为 1 Mbit/s
由于带宽在不同的取值范围内步长不同，最终调整后的带宽会根据实际步长自动调整为就近值。

冷却时间(秒) 300

立即创建

表1-2 “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。 | - |
| | 需要进行伸缩管理的公网 IP。 | - |
| 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。 若选择新建告警，需配置如下参数： <ul style="list-style-type: none"> • 告警规则名称 新建告警的名称，例如 as-alarm-demo。 | - |

| 参数 | 解释 | 取值样例 |
|------|---|-------|
| | <ul style="list-style-type: none"> • 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 1-4 所示。例如平均值 > 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： <ul style="list-style-type: none"> • 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。 • 设置为 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。 <ul style="list-style-type: none"> • 小于等于 300Mbit/s：默认步长为 1Mbit/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1-3 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。 |

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 > 弹性伸缩服务 > 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 1-5](#)进行配置。
5. 参数配置完成后，单击“立即创建”。

< 创建伸缩带宽策略

区域 苏州
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。

策略名称

弹性IP 117.80.148.165 [C 购买弹性公网IP](#)
当前带宽大小 5 Mbit/s 当前您仅可对按流量计费的弹性IP进行伸缩。

策略类型 告警策略 定时策略 周期策略

时区 GMT+08:00

触发时间 2022-09-10 00:00:00 X | 自
请确保触发时间晚于默认时间，并且晚于当前时间。

执行动作 设置为 Mbit/s
由于带宽在不同的取值范围内步长不同，最终调整后的带宽会根据实际步长自动调整为就近值。

冷却时间(秒) 300

立即创建

表1-4 “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。 策略名称只能由英文字母、数字、下划线、和中划线组成。 | as-policy-p6g5 |
| | 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。 | - |
| 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"> • 生效时间 选择伸缩策略触发的时间段。 • 重复周期 <ul style="list-style-type: none"> - 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： <ul style="list-style-type: none"> • 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。 • 设置为 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。 <ul style="list-style-type: none"> • 小于等于 300Mbit/s：默认步长为 1Mbit/s。 | - |

| 参数名称 | 参数说明 | 取值样例 |
|------|--|-------|
| 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 |

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 > 弹性伸缩服务 > 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。



The screenshot shows a policy configuration page for 'as-policy-bandwidth-demo'. It includes fields for policy name, ID, elastic IP, cooldown time, and creation time. It also lists the policy type, alarm rule name, trigger conditions, and actions. Below this is a 'Policy Execution Log' table with columns for execution type, status, time, original value, target value, and actions.

| 策略名称 | as-policy-bandwidth-demo | 策略类型 | 告警策略 |
|---------|--------------------------------------|--------|--------------------------------------|
| ID | d6d3cfbb-d250-4e36-a587-9a96e5a1af8f | 告警规则名称 | as-alarm-b58c |
| 弹性IP | 117.80.148.165 | 触发条件 | 入网带宽最大值>10bit/s, 连续满足5次后触发, 监控周期5分钟。 |
| 冷却时间(秒) | 300 | 执行动作 | 设置为1Mbit/s |
| 创建时间 | 2022-09-06 10:20:18 GMT+08:00 | 限制值 | -- |

| 策略执行类型 | 策略执行状态 | 策略执行时间 | 伸缩资源 | 伸缩原始值(Mbit/s) | 伸缩目标值(Mbit/s) | 操作 |
|----------|--------|-------------------------------|----------------|---------------|---------------|------|
| 自动触发(告警) | 成功 | 2022-09-06 10:25:01 GMT+08:00 | 117.80.148.165 | 5 | 1 | 查看详情 |

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。

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。

📖 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 > 弹性伸缩服务 > 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”。

The screenshot shows the '伸缩带宽' (Elastic Bandwidth) management page. It includes a table with columns for name, status, elastic IP, policy type, trigger conditions, actions, and cooldown time. The '操作' (Action) column contains buttons for '启用' (Enable), '立即执行' (Execute Immediately), and '更多' (More).

| 名称 | 状态 | 弹性IP | 策略类型 | 触发条件 | 执行动作 | 冷却时间(秒) | 操作 |
|--------------------------|-----|----------------|------|-----------------------|------------|---------|------------|
| as-policy-bandwidth-d... | 已停用 | 117.80.148.165 | 告警策略 | 入网带宽最大值>10bit/s, 连... | 设置为1Mbit/s | 300 | 启用 立即执行 更多 |

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 > 弹性伸缩服务 > 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 > 弹性伸缩服务 > 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”。

修改伸缩带宽策略

✕

策略名称

弹性IP [购买弹性公网IP](#)

当前带宽大小 1 Mbit/s 当前您仅可对按需付费模式的弹性IP进行伸缩。

策略类型 **告警策略** 定时策略 周期策略

告警规则 **现在创建** 使用已有

告警规则名称 [新建告警规则](#)

触发条件 入网带宽最大值>10bit/s

监控周期 5分钟

连续出现次数

执行动作 Mbit/s

由于带宽在不同的取值范围内步长不同，最终调整后的带宽会根据实际步长自动调整为就近值。

冷却时间(秒)

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 > 弹性伸缩服务 > 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 > 弹性伸缩服务 > 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。

- **云主机健康检查：**是指对云主机的运行状态进行检查，如关机、删除都是云主机异常状态。伸缩组的健康检查方式默认是“云主机健康检查”方式，指伸缩组会定期使用云主机健康检查结果来确定每个云主机的运行状况。如果未通过云主机健康检查，则伸缩组会将该云主机移出伸缩组。
- **弹性负载均衡健康检查：**是指根据 ELB 对云主机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主机状态为正常的情况下，才会认为该弹性云主机正常。否则只要有一个负载均衡器检测到云主机状态异常，伸缩组会将该弹性云主机移出伸缩组。

以上两种健康检查方式，检查的结果均是将异常的云主机从伸缩组中移除，移出伸缩组的实例，是否会删除云主机，如下所述：

弹性伸缩活动中自动添加的云主机，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 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 操作列表如下表 1 所示。

表1-1 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 |

| 操作名称 | 资源类型 | 事件名称 |
|----------|-----------------------|---------------------------------|
| 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 |

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"
}
```

4.6.3 监控指标说明

弹性伸缩支持的监控指标如下表 1-3 所示。

表1-5 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：伸缩组中的所有云主机的 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：伸缩组中的所有云主机内存使用率之和/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 |
| instance_num | 实例数 | 该指标用于统计弹性伸缩组中可用的云主机云主机数量。 计算公式：弹性伸缩组内生命周期状态为“已启用”的云主机云主机数量之和 | ≥0 | 测量对象：弹性伸缩组 测量维度：AutoScalingGroup | 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：伸缩组中所有云主机的带内网络流入速率之和 / 伸缩组实例数 | ≥0 Byte/s | 测量对象：弹性伸缩组 测量维度： | 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) |
|------------------------------------|----------|---|------------------|---------------------------------------|-------------|
| | | 单位: 字节/秒 | | AutoScalingGroup | |
| 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: 伸缩组中所有云主机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 | 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: 伸缩组中所有云主机的磁盘读速率之和 / 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 |
| 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: 伸缩组中所有云主机的磁盘写速率之和 / 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: 伸缩组中所有云主机的磁盘读操作速率之和 / 伸缩组实例数 单位: 请求/秒 | ≥0 requests/s | 测量对象: 弹性伸缩组 测量维度: AutoScalingGroup | 5 分钟 |
| disk_write_ | 磁盘写操 | 该指标用于统计每秒往弹 | ≥0 | 测量对 | 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) |
|---------------|------|--|-----------|--------------------------------------|-------------|
| requests_rate | 作速率 | 性伸缩组写数据的请求次数。 计算公式：伸缩组中的所有云主机的磁盘写操作速率之和 / 伸缩组实例数 单位：请求/秒 | request/s | 象：弹性伸缩组 测量维度： AutoScalingGroup | |

📖 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《弹性云主机用户指南》。

4.6.4 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主机运行状态，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，更好地了解弹性云主机的各项性能指标。

前提条件

弹性伸缩组中的弹性云主机正常运行。

📖 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、删除状态的弹性云主机，无法查看其 CPU 使用率、磁盘读速率等监控指标。当弹性云主机再次启动或恢复后，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“计算 > 弹性伸缩 > 弹性伸缩组”。
3. 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。


4. 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

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”、“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。

4.6.5 设置监控告警规则

操作场景

通过设置弹性云主机告警规则，用户可自定义监控目标与通知策略，及时了解弹性云主机运行状况，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“管理与部署 > 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

- 多关于设置告警规则的信息，请参见《云监控用户指南》。
- 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 最佳实践

5.1 搭建可自动伸缩的 WEB 服务

5.1.1 场景介绍

某电商平台为吸引用户，除定期推出优惠活动外，还会在节假日、会员日、购物节开展大型促销活动。为保证顺利承载活动带来的流量，运维人员可以分析活动历史数据，提前预估新活动所需的计算资源。但如果高峰期流量超出预估，仍需要临时手动创建 ECS 实例，不仅操作仓促，而且可能因操作不及时影响应用可用性。

本文重点介绍创建弹性伸缩实现云主机自动伸缩的过程。如何利用弹性伸缩搭建可自动伸缩的 Web 应用，快速响应业务的峰谷波动，稳定承载日常业务的同时，轻松应对活动期间突增的流量。当电商平台服务的负载增加时云主机的 CPU 使用率会增大，负载降低时 CPU 使用率会降低。我们配置两条监控 CPU 使用率的告警策略，分别在 CPU 使用率高于 50% 时增加一台云主机，在 CPU 使用率低于 10% 时减少一台云主机，保证服务始终有合适数量的云主机，实现自动伸缩云主机的功能。

5.1.2 前期准备

- 已注册天翼云账号，如还未注册，请先完成[账号注册](#)。
- 为应用的 ECS 实例创建了私有镜像，具体操作请参见[使用 ECS 创建私有镜像](#)。

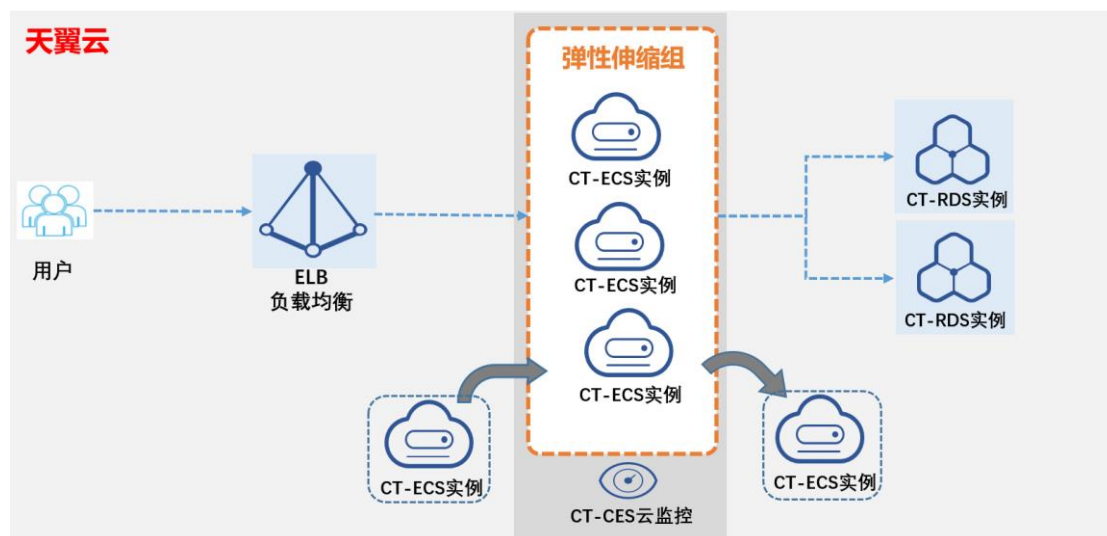
5.1.3 解决方案

弹性伸缩可以实现计算资源随业务峰谷自动伸缩，无需您提前预估和手动干预即可确保应用可用性。尤其针对双十一等大促活动，弹性伸缩具备几分钟内交付上千台 ECS 实例的能力，自动及时地应对突增流量，提升业务的可靠性。

您可以采用以下方案：

- 针对日常业务流量，购买包年包月 ECS 实例。
- 针对计划外突增流量，通过弹性伸缩监控负载变化并实现自动创建 ECS 实例。

示意图如下：



5.1.4 操作步骤

请根据您的业务架构评估业务模块，并执行以下操作实现指定业务模块的自动扩缩容：

- 步骤一：使用私有镜像创建 ECS 实例
- 步骤二：创建并启用伸缩组
- 步骤三：设置自动伸缩策略

5.1.4.1 步骤一：使用私有镜像创建 ECS 实例

创建指定数量的 ECS 实例，用于添加到伸缩组，满足业务模块的日常业务要求。

1. 登录 [ECS 管理控制台](#)。
2. 在左侧导航栏，选择“计算” > 弹性云主机。
3. 单击“创建弹性云主机”。
4. 选择“计费模式”：“包年/包月”或“按需付费”。
5. 配选择区域、可用区、规格
6. 镜像选择已创建好的“私有镜像”。
7. 下一步网络配置：选择规划的 VPC 及子网、安全组，设置密码等

请根据需要配置其它信息，详细信息请参见[创建云主机](#)。

5.1.4.2 步骤二：创建并启用伸缩组

为需要弹性扩缩容的业务模块创建伸缩组，并为伸缩配置选择 Web 应用实例的自定义镜像，确保自动创建出的 ECS 实例符合 Web 应用的要求。

1. 登录[弹性伸缩控制台](#)。
2. 选择控制中心，选择区域。
3. 创建一个弹性伸缩组。
 - 可用区依据业务诉求选择可用区。
 - 最小实例数设置为 2。
 - 期望实例数设置为 3。
 - 最大实例数设置为 5。
 - 网络选择规划的 VPC 网段。
 - 负载均衡选择不使用。
 - 实例移除策略设置为释放模式。
 - 弹性 IP 选择释放

请根据需要配置其它信息，详细信息请参见[创建伸缩组](#)。

<
创建弹性伸缩组

*** 区域** 苏州

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1 可用区2 可用区3

*** 名称** as-group-b971

*** 最大实例数(台)** 5

*** 期望实例数(台)** 3

*** 最小实例数(台)** 2

以下信息是伸缩后实例所使用的虚拟私有云、子网。负载均衡负责自动分发访问流量。

*** 虚拟私有云** VP

*** 子网** sub

本子网作为云主机的主网卡

自动分配IPv6地址

+ 增加一个子网 您还可以增加 4 个子网

负载均衡

不使用
使用经典型
使用增强型

*** 实例移除策略** 根据较早创建的配置较早创建的实例

弹性IP

释放
不释放

若选择“释放”，在伸缩组进行缩的活动时，则会将云主机上的弹性IP释放，否则仅做解绑定操作，保留弹性IP资源。

*** 健康检查方式** 云主机健康检查

受保护的实例状态异常时，会被健康检查移除，并重新创建新的实例。

*** 健康检查间隔** 5分钟

*** 企业项目** default

高级配置

暂不配置
现在配置

在高级配置中可对通知方式和标签进行配置。

下一步

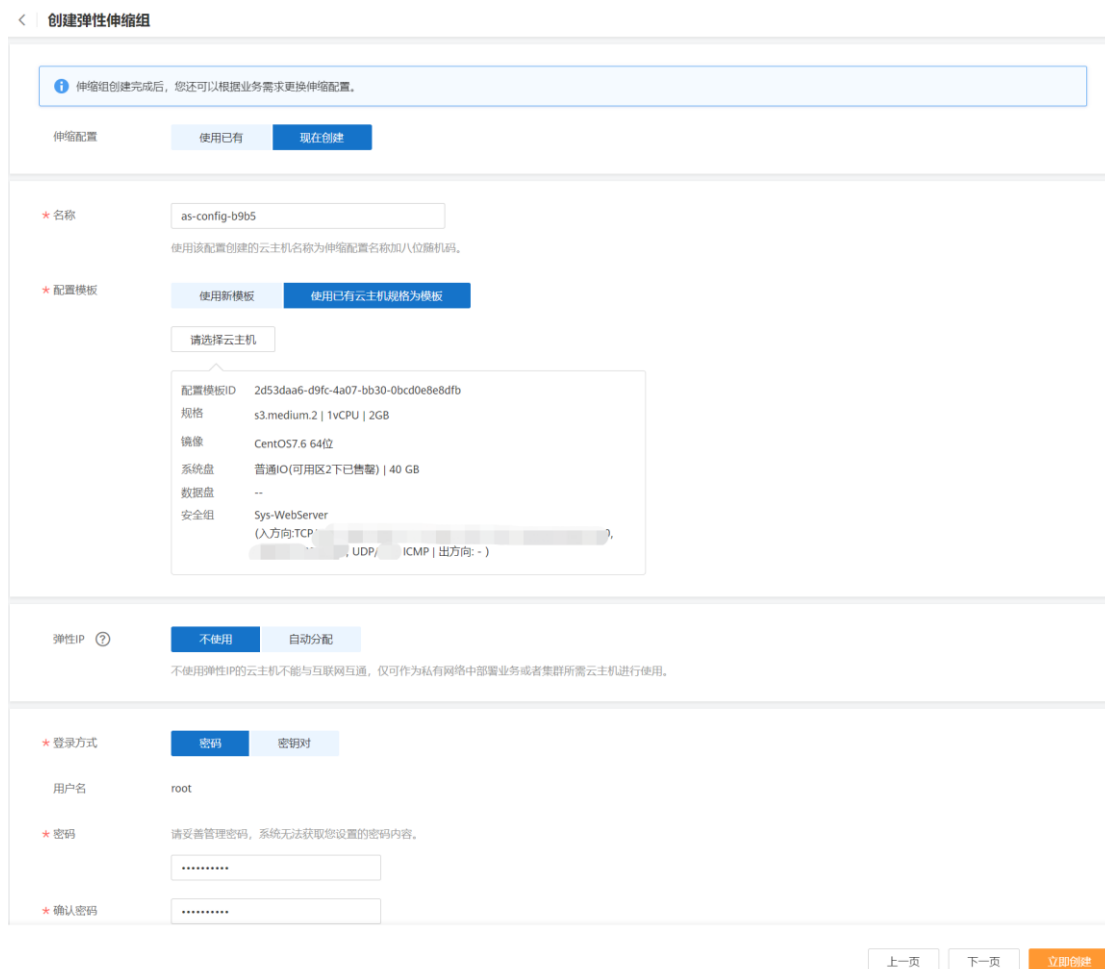
4. 点击下一步，创建伸缩配置。

5. 创建一个伸缩配置。

- 名称可以自定义。
- 配置模板选择为使用已有主机规格为模板。
- 登录方式设置为密码。

- 密码设置为 xxx

请根据需要配置其它信息，详细信息请参见[创建伸缩配置](#)。



6. 点击立即创建

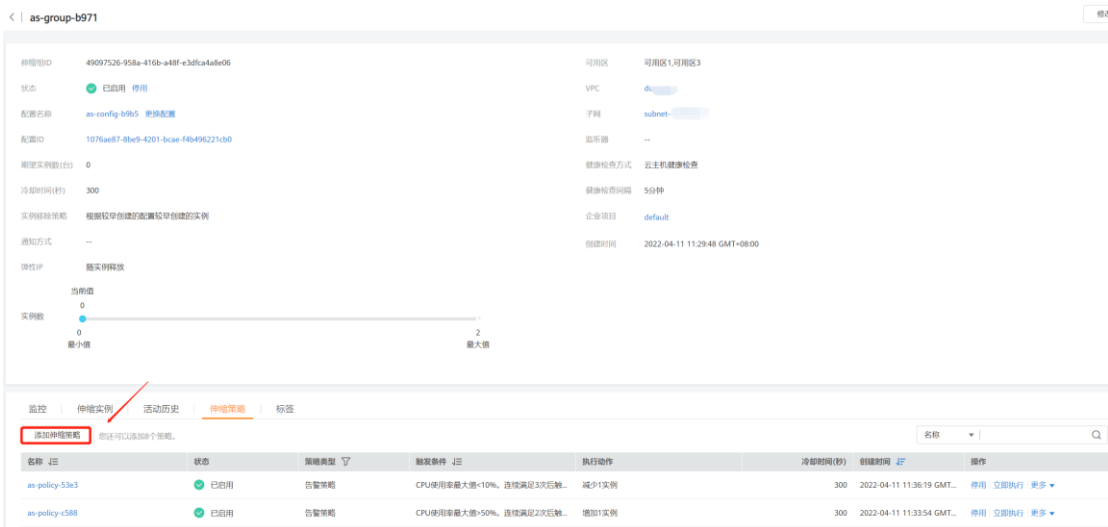
5.1.4.3 步骤三：设置自动伸缩策略

将按需或包周期 ECS 实例添加至伸缩组，并创建自动伸缩策略，实现根据业务峰谷自动伸缩，应对突增流量。

1. 前往伸缩实例列表界面，将创建好的按需或包年包月 ECS 实例添加至伸缩组。
2. 开启实例保护，将 ECS 实例转为保护状态，保证日常业务正常运行。
3. 前往基本信息界面，根据业务需求，修改伸缩组的最小实例数和最大实例数。

4. 在伸缩策略界面，添加 2 条伸缩策略，一条伸缩策略 CPU 使用率设置大于 50%。

- 伸缩策略类型选择为告警策略
- 触发条件设置为 CPU 使用率，最大值 大于 50%（大小可依据业务调整）
- 监控周期设置为 5 分钟
- 连续出现次数 3
- 执行动作增加一个实例



策略 ID: 49097526-958a-416b-a48f-e3dfca4a8e06

可用区: 可用区1可用区3

状态: 已启用 停用

VPC: dc-...

配置名称: as-config-b965 更新配置

子网: subnet-...

配置 ID: 1076ae87-8be8-4201-bcae-f6b496221cb0

健康检查方式: 云主机健康检查

期望实例数(台): 0

冷却时间(秒): 300

健康检查周期: 5分钟

实例释放策略: 根据权重创建的配置权重创建的实例

企业项目: default

通知方式: --

创建时间: 2022-04-11 11:29:48 GMT+08:00

弹性IP: 随实例释放

实例数: 0 (范围: 0 到 2)

最小值: 0 最大值: 2

| 名称 | 状态 | 策略类型 | 触发条件 | 执行动作 | 冷却时间(秒) | 创建时间 | 操作 |
|----------------|-----|------|--------------------------|-------|---------|----------------------------|------------|
| as-policy-53c3 | 已启用 | 告警策略 | CPU使用率最大=10%。连续满足3次后触... | 减少1实例 | 300 | 2022-04-11 11:36:19 GMT... | 停用 立即执行 更多 |
| as-policy-588 | 已启用 | 告警策略 | CPU使用率最大=50%。连续满足2次后触... | 增加1实例 | 300 | 2022-04-11 11:33:54 GMT... | 停用 立即执行 更多 |

添加伸缩策略

| | | | |
|----------------------|--|------------------------------------|---|
| 策略名称 | <input type="text" value="as-policy-c588"/> | | |
| 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-a0fd"/> | | |
| 触发条件 | <input type="text" value="CPU使用率"/> ▾ | <input type="text" value="最大值"/> ▾ | <input type="text" value=">"/> ▾ <input type="text" value="50"/> % |
| 监控周期 | <input type="text" value="5分钟"/> ▾ | | |
| 连续出现次数 [?] | <input type="text" value="3"/> | | |
| 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="300"/> | | |

另外一条弹性伸缩策略，CPU 使用率设置为最大值 小于 10%

- 伸缩策略类型选择为告警策略
- 触发条件设置为 CPU 使用率，最大值 小于 10%
- 监控周期设置为 5 分钟
- 连续出现次数 3
- 执行动作减少一个实例

添加伸缩策略

| | | | |
|----------------------|--|------------------------------------|---|
| 策略名称 | <input type="text" value="as-policy-53e3"/> | | |
| 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-53e4"/> | | |
| 触发条件 | <input type="text" value="CPU使用率"/> ▾ | <input type="text" value="最大值"/> ▾ | <input type="text" value="<"/> ▾ <input type="text" value="10"/> % |
| 监控周期 | <input type="text" value="5分钟"/> ▾ | | |
| 连续出现次数 [?] | <input type="text" value="3"/> | | |
| 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="300"/> | | |

请根据需要配置其它信息，详细信息请参见[创建伸缩策略](#)。

5.1.5 结果验证

当伸缩组中的云主机 ECS 实例的 CPU 使用率持续高于 50%，伸缩组会自动增加一台 ECS 实例分担流量。

当伸缩组中的云主机 CPU 使用率持续低于 10%，且伸缩组中至少存在两台云主机时，伸缩组会自动减少一台 ECS 实例节省成本，ECS 实例数量始终大于等于最小实例数，且小于等于最大实例数，保证满足业务需求且成本不会超出期望范围。

6 常见问题

6.1 通用类

6.1.1 弹性伸缩有什么限制？

弹性伸缩的云主机中运行的应用需要是无状态、可横向扩展的。因为 AS 会自动释放云主机，所以弹性伸缩组内的云主机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）。

如果应用中需要云主机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下表所示：

| 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。 | 200 |

6.1.2 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

6.1.3 用户默认能创建和使用多少个伸缩策略和配置？

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。

6.1.4 弹性伸缩是否能够自动升降云主机的 CPU、内存和带宽？

弹性伸缩目前支持自动升降带宽资源，支持自动增加或减少伸缩组内 ECS 实例的数量，不支持自动升降服务器的 CPU、内存等配置。

6.1.5 伸缩带宽对弹性公网 IP 带宽和共享带宽有什么要求？

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。

6.2 计费类

6.2.1 弹性伸缩服务是否收取费用？

A：弹性伸缩服务本身不收取费用。但会按伸缩组内的云主机实例收取云主机的费用。

6.3 开通类

6.3.1 天翼云所有节点都支持弹性伸缩服务么？

A：当前贵州/福州/杭州/深圳/广州 4/苏州/郑州/青岛/西安 2/上海 4/芜湖/南宁/长沙 2/南昌/成都 3/乌鲁木齐/昆明/海口/重庆/武汉 2/兰州/西宁/太原/石家庄/中卫/长春/天津/北京 2/哈尔滨节点支持弹性伸缩服务。

6.3.2 退订弹性云主机后弹性伸缩服务是否继续使用？

A：当退订弹性云主机后，弹性伸缩服务也随之暂停服务。

6.3.3 弹性云主机关机后弹性伸缩服务会关闭么？

A：弹性云主机关机后，弹性伸缩服务不会暂停服务。

6.4 操作类

6.4.1 弹性伸缩一定要搭配弹性负载均衡、云监控才能使用吗？

A：弹性伸缩服务可以单独使用，也可以同弹性负载均衡（ELB），云监控（CES）一起使用。其中，云监控服务为免费服务，系统默认开通；负载均衡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。

6.4.2 弹性伸缩是否会因监控指标突变导致误伸缩?

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

6.4.3 能创建和使用多少个伸缩策略和配置?

A: 您可以创建 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。

6.4.4 按照伸缩策略增加的云主机在我不用时可以自动删除吗?

A: 可以，但是需要您增加一条删除云主机的伸缩策略。

6.4.5 伸缩组使用负载均衡健康检查，为何没有立即生效?

A: 通常，弹性伸缩活动中自动添加的实例需要足够的预热时间才能通过负载均衡健康检查。因此，如果您为伸缩组添加了负载均衡器并且健康检查方式选择负载均衡健康检查，那么伸缩组会等健康状况检查宽限期结束后才检查实例的运行状况。

当实例加入伸缩组并且进入已启用状态后，健康状况检查宽限期才会启动，默认为 10 分钟。

6.4.6 如何删除通过弹性伸缩创建的云主机？

处理方法

方法一：

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩”。
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”。
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主机的伸缩策略，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。
2. 选择“计算 > 弹性伸缩”。
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，单击右上方“修改”，进入“修改伸缩组”页面。
5. 手动修改伸缩组的期望实例数。

6.4.7 如何阻止伸缩组内的云主机被自动移除？

A: 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，当伸缩组发生自动减容活动时，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云主机实例运行。

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

6.4.8 如何自动初始化弹性伸缩新增的云主机数据盘？

A: 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。以下操作步骤仅支持初始化一个数据盘，以 centos 6.6 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。

具体操作步骤如下：

1. 以 root 用户登录已有云主机；
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，创建并编辑脚本。

```
vi 脚本名称
```

例如：

```
vi fdisk_mount.sh
```

以下脚本为云主机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'
    
```

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

```

Vi fdisk.ini

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
    
```

6. 按“Esc”键，输入:wq，按“Enter”保存并退出编辑。

7. 执行以下命令，打开配置文件“rc.local”。

```
Vi /etc/rc.local
```

8. 按“i”，在“rc.local”文件中添加如下内容。

```
/home/fdisk_mount.sh
```

配置“rc.local”后，云主机首次启动时会自动执行初始化数据盘脚本。

9. 按“Esc”键，输入:wq，按“Enter”保存并退出编辑。
10. 通过已有云主机创建私有镜像。
11. 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。
12. 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云主机，会按照私有镜像中的配置自动初始化数据盘。

6.5 伸缩组类

6.5.1 伸缩组启用失败如何处理？

伸缩组启用失败情况及其处理方法：

- 情况描述：云主机配额不足、云硬盘配额不足、弹性 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、安全组不存在、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、安全组相关参数，之后重新启用伸缩组。
- 情况描述：负载均衡监听器不存在、后端云主机组不存在、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主机超过最大限制。
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 个后端云主机。
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主机，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：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主机资源已售罄。
可能原因：您为伸缩组选择的可用区下该类型云主机资源售罄，或者该可用区不支持该类型云主机资源。该类型云主机指的是伸缩配置里选择的云主机规格。
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：伸缩配置中的云主机类型和磁盘类型不匹配导致创建云主机失败。
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：伸缩配置中的云主机类型和镜像不匹配导致创建云主机失败。
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：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。

6.5.2 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：云主机配额不足、云硬盘配额不足、弹性 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、安全组不存在、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、安全组相关参数，之后重新启用伸缩组。
- 情况描述：负载均衡监听器不存在、后端云主机组不存在、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主机超过最大限制。
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 个后端云主机。
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主机，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：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主机资源已售罄。
 - 可能原因：您为伸缩组选择的可用区下该类型云主机资源售罄，或者该可用区不支持该类型云主机资源。该类型云主机指的是伸缩配置里选择的云主机规格。
 - 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：伸缩配置中的云主机类型和磁盘类型不匹配导致创建云主机失败。
 - 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：伸缩配置中的云主机类型和镜像不匹配导致创建云主机失败。
 - 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：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。

6.5.3 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。

6.5.4 已知弹性云主机实例 ID，能否查看当前实例所在的伸缩组信息？以及如何查看伸缩组及其中的实例信息？

不能。

查看伸缩组及其中的实例信息步骤如下：

步骤 1 进入“伸缩实例”页面。

步骤 2 点击“弹性伸缩组”页签下的伸缩组名称，进入伸缩组概览页，可查看伸缩组信息。

步骤 3 点击“伸缩实例”页签，可查看该伸缩组下的实例信息。

----结束

6.6 伸缩策略类

6.6.1 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。

6.6.2 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。

6.6.3 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，并且启动后可以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。

6.6.4 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主机的内存使用率、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。

如果弹性云主机类型是 IO 优化实例，无论是否安装了 vmttools，监控弹性云主机的监控指标不包

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。

6.6.5 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。

6.7 实例类：

6.7.1 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 台云主机 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。

6.7.2 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云主机实例，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云主机、数据库等。

6.7.3 按照伸缩策略增加的云主机，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主机的伸缩策略。

6.7.4 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主机的个数，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主机。

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。

6.7.5 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主机发起心跳检查的方式来实现的，而负载均衡系统和云主机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主机，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. 检查云主机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：需要放行后端云主机业务端口。
 - 若配置了不同于云主机业务端口的健康检查端口：需要放行云主机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。

6.7.6 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。

6.7.7 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。

6.8 其他：

6.8.1 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，您需要在源云主机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。
3. 创建一个新的伸缩配置，确保该伸缩配置选择的镜像为应用的私有镜像。
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。

6.8.2 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主机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 进行对云主机的初始化配置。

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主机会出现如下情况：

- 如使用未安装 Cloudbase-init 的 Windows 私有镜像创建的弹性云主机，在获取弹性云主机密码时，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云主机。若忘记镜像本身密码，可以通过云主机页面“重置密码”功能，自助完成云主机的密码重置。
- 如使用未安装 Cloud-Init 的 Linux 私有镜像创建的云主机，使用创建时输入的密码或密钥将无法登录云主机。您只能通过镜像本身的密码或密钥登录此云主机。若忘记镜像本身密码，或镜像本身的密钥丢失，可以通过云主机页面“重置密码”功能，自助完成云主机的密码重置。
- 如使用未安装 Cloud-Init/Cloudbase-init 的私有镜像创建云主机时，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：

1. 登录管理控制台。
2. 选择“计算 > 弹性伸缩”。
3. 选择“伸缩配置”页签，进入伸缩配置列表页面。
4. 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。
5. 在伸缩组中将伸缩配置修改为新创建的伸缩配置。

6.8.3 为什么使用密钥文件无法正常登录云主机？

问题描述

用户使用密钥文件登录弹性伸缩组中的云主机时，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，用户将无法对云主机进行自定义配置，只能使用镜像原有密码或密钥登录云主机。

处理方法

1. 判断是否需要继续登录该云主机。
 - 是，请使用镜像原有密码或密钥登录云主机。
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，执行步骤 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。

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主机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主机。

6.8.4 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。