

弹性伸缩服务

快速入门

文档版本 02
发布日期 2021-10-19



版权所有 © 华为云计算技术有限公司 2024。保留一切权利。

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

商标声明



HUAWEI和其他华为商标均为华为技术有限公司的商标。

本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限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，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

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

目录

1 快速创建和使用弹性伸缩.....	1
2 快速创建弹性伸缩（旧）.....	8
3 动态扩展云服务器.....	13
4 定时扩展云服务器.....	14
5 弹性伸缩向导式使用流程.....	15

1 快速创建和使用弹性伸缩

操作场景

弹性伸缩 AS (Auto Scaling) 可以根据用户的业务需求，通过策略自动调整用户的资源，帮助用户节约资源和人力维护的成本。

弹性伸缩服务本身不收取费用，但伸缩组扩容时自动创建的实例（云服务器），实例使用的弹性公网IP、磁盘、镜像等资源，需要支付相应的费用。

创建伸缩配置和创建弹性伸缩组是使用AS服务必备的操作。

操作流程

操作步骤	说明
步骤一：创建伸缩配置	配置云服务器的规格、镜像、磁盘等参数。
步骤二：创建弹性伸缩组	配置最大、最小、期望实例数等参数。
步骤三：创建伸缩策略	通过创建伸缩策略调整客户业务资源。

步骤一：创建伸缩配置

本步骤仅针对本示例进行参数设置及介绍，更多创建伸缩配置的详细内容请参见[伸缩配置](#)。

1. 登录控制台，进入[创建伸缩配置](#)页面。
2. 设置“伸缩配置”参数。

图 1-1 “伸缩配置”参数设置

★ 计费模式 按需计费 竞价计费 ?

★ 区域 中国-香港

不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。

★ 名称 as-config-3817

使用该配置创建的云服务器名称为伸缩配置名称加八位随机码。

★ 配置模板 使用新模板 使用已有云服务器规格为模板

CPU架构 x86计算 鲲鹏计算 ?

★ 规格 最新系列 vCPUs 全部 内存(GiB) 全部 规格名称 请输入规格名称 Q

通用计算型 通用计算增强型 内存优化型 超大内存型 高性能计算型 磁盘增强型 超高I/O型 GPU加速型 通用入门型 Flexus云服务器X实例

[了解如何选择弹性云服务器类型](#)

实例类型	规格名称	vCPUs 内存(GiB)	CPU	基准 / 最大带宽	内网收发包
<input type="checkbox"/> 通用计算型s7n	s7n.small.05 (可用区3, 可用...	1vCPUs 0.5GiB	Intel Ice L...	0.1/0.8 Gbit/s	100,000
<input type="checkbox"/> 通用计算型s7n	s7n.small.1 (可用区7, 可用...	1vCPUs 1GiB	Intel Ice L...	0.1/0.8 Gbit/s	100,000
<input type="checkbox"/> 通用计算型s7n	s7n.medium.2 (可用区7, 可用...	1vCPUs 2GiB	Intel Ice L...	0.1/0.8 Gbit/s	100,000
<input type="checkbox"/> 通用计算型s7n	s7n.medium.4 (可用区7, 可用...	1vCPUs 4GiB	Intel Ice L...	0.1/0.8 Gbit/s	100,000
<input type="checkbox"/> 通用计算型s7n	s7n.large.025 (可用区3, 可用...	2vCPUs 0.5GiB	Intel Ice L...	0.2/1.5 Gbit/s	150,000
<input checked="" type="checkbox"/> 通用计算型s7n	s7n.large.2 (可用区7, 可用...	2vCPUs 4GiB	Intel Ice L...	0.2/1.5 Gbit/s	150,000
<input type="checkbox"/> 通用计算型s7n	s7n.large.4 (可用区7, 可用...	2vCPUs 8GiB	Intel Ice L...	0.2/1.5 Gbit/s	150,000
<input type="checkbox"/> 通用计算型s7n	s7n.xlarge.2 (可用区7, 可用...	4vCPUs 8GiB	Intel Ice L...	0.35/2 Gbit/s	250,000

已选规格 优先使用当前选中的规格进行伸缩，您可以点击已选规格查看规格信息。您还可以选择9个规格。

s7n.large.2 X

通用计算型 | s7n.large.2 | 2vCPUs | 4GiB

*** 镜像** 公共镜像 私有镜像 共享镜像 市场镜像

CentOS ▼ CentOS 8.2 64bit(40GiB) ▼ 🔍

CentOS 8已于2021年12月31日停止服务。您可以选择替代方案

*** 磁盘** 云硬盘

系统盘 通用型SSD ▼ - 100 + GiB IOPS上限3,000, IOPS突发上限8,000 ?

⊕ 增加一块数据盘 您还可以增加 23 块磁盘 (云硬盘)。

*** 安全组** default (入方向:TCP | 出方向:TCP) X ▼ 🔍 新建安全组 ?

安全组类似防火墙功能, 是一个逻辑上的分组, 用于设置网络访问控制。 [如何配置安全组?](#)

入方向:TCP | 出方向:TCP

弹性公网IP 不使用 自动分配 ?

不使用弹性公网IP的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云服务器进行使用。

*** 登录方式** 密钥对 密码

请妥善保管密钥对的私钥文件, 登录、重装和切换云服务器操作系统时, 均需要使用该文件。

*** 密钥对** KeyPair-2325 ▼ 🔍 新建密钥对

我确认已获取密钥对私钥文件KeyPair-2325.pem, 否则无法登录弹性云服务器。

高级配置 暂不配置 现在配置

表 1-1 “伸缩配置” 参数设置说明

参数	示例	说明
计费模式	按需计费	按云服务的实际使用时长计费, 可以随时开通或删除。更多信息, 请参见 弹性云服务器ECS 计费说明 。
区域	中国-香港	请就近选择靠近您业务的区域, 可减少网络时延, 提高访问速度。更多信息, 请参见 区域和可用区 。
名称	as-config-3817	伸缩配置的名称。
配置模板	使用新模板	需要配置伸缩实例的规格、镜像、磁盘等参数。
CPU架构	x86计算	x86采用复杂指令集(CISC)。
规格	s7n.xlarge.2	请根据业务需要选择合适的规格。更多信息, 请参见 弹性云服务器规格清单 。

参数	示例	说明
镜像	CentOS 8.2 64bit(40GiB)	华为云提供的Linux类型公共镜像，该镜像免费。
磁盘	通用型SSD 100GiB	伸缩实例自带系统盘，用于存储云服务器的操作系统。
安全组	default	使用系统默认创建的安全组。
弹性公网IP	不使用	如需访问外网，则可以为伸缩实例配置和绑定弹性公网IP。
登录方式	密钥对	选择“登录凭证”方式为“密钥对”。
密钥对	KeyPair-2325	使用已有的密钥对或新建密钥对，并确认已获取密钥对的私钥文件。
高级配置	暂不配置	-

3. 单击“立即创建”，完成伸缩配置的创建。
4. 单击“返回伸缩配置列表”，查看已创建的伸缩配置。

图 1-2 查看伸缩配置



步骤二：创建弹性伸缩组

本步骤仅针对本示例进行参数设置及介绍，更多创建弹性伸缩组的详细内容请参见[伸缩组](#)。

1. 登录控制台，进入[创建弹性伸缩组](#)页面。
2. 设置“弹性伸缩组”参数。

图 1-3 “弹性伸缩组” 参数设置

* 区域
不同区域的云服务产品之间内网互不相通；请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。

* 可用区
?

* 多可用区扩展策略 均衡分布 选择优先 ?

* 名称

* 最大实例数(台)

* 期望实例数(台) ?

* 最小实例数(台)

选择伸缩配置作为您创建的伸缩组内伸缩实例的模板；选择子网后将向伸缩组中的每个实例分配IP地址。

* 伸缩配置 +

* 虚拟私有云 ? [新建虚拟私有云 ?](#)

* 子网 本子网作为云服务器的主网卡
 源/目的检查 ?
+ 增加一个子网 您还可以增加 4 个子网 [新建子网](#)

负载均衡 不使用 使用弹性负载均衡

* 实例移除策略

弹性公网IP 释放 不释放
若选择“释放”，在伸缩组进行缩的活动时，则会将云服务器上的弹性公网IP释放，否则仅做解绑定操作，保留弹性公网IP资源。

数据盘 删除 不删除
若选择“删除”，在伸缩组进行缩的活动时，则会将云服务器上的数据盘删除，否则仅做解绑定操作，保留数据盘资源。

* 健康检查方式 ?
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 ?

* 健康状况检查宽限期(秒) ?

* 企业项目 ?

标签 如果您需要使用同一标签标识多种云资源，即所有服务均可在标签输入框下拉选择同一标签，建议在TMS中创建预定义标签。 [查看预定义标签](#) ?

您还可以添加10个标签。

委托 ? [新建委托 ?](#)

表 1-2 “弹性伸缩组” 参数设置说明

参数	示例	说明
区域	中国-香港	请就近选择靠近您业务的区域，可减少网络时延，提高访问速度。 更多信息，请参见 区域和可用区 。
可用区	可用区1、可用区2、可用区3、可用区7	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	均衡分布	保证选择的可用区列表中各可用区下云服务器数量均衡。
名称	as-group-test1	伸缩组的名称。
最大实例数(台)	1	伸缩组中运行的最大云服务器的个数。
期望实例数(台)	0	伸缩组中期望运行的云服务器的个数。
最小实例数(台)	0	伸缩组中运行的最小云服务器的个数。
伸缩配置	as-config-3817	步骤一中创建的伸缩配置。
虚拟私有云	vpc-default-smb	使用系统默认创建的VPC和子网。 更多信息，请参见 虚拟私有云和子网规划建议 。
子网	subnet-default-smb	
负载均衡	不使用	可选参数，更多信息，请参见 添加负载均衡器到伸缩组 。
实例移除策略	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。
弹性公网IP	释放	当云服务器从伸缩组中移除时，会将云服务器上的弹性公网IP释放。
数据盘	删除	当云服务器从伸缩组中移除时，云服务器挂载的数据盘将被删除。
健康检查方式	云服务器健康检查	对云服务器的运行状态进行检查，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	5分钟	伸缩组执行健康检查的周期。

参数	示例	说明
健康状况检查宽限期(秒)	600	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。
企业项目	default	伸缩组归属的企业项目，由伸缩组创建的弹性云服务器将归属于此企业项目。

3. 单击“立即创建”，完成弹性伸缩组的创建。
4. 单击“返回弹性伸缩组列表”，查看已创建的伸缩组。

图 1-4 查看伸缩组



步骤三：创建伸缩策略

弹性伸缩组创建成功后，您可以通过创建伸缩策略调整业务资源。

操作类型	您可以参考
使用AS动态扩展云服务器	动态扩展云服务器
使用AS定时扩展云服务器	定时扩展云服务器

2 快速创建弹性伸缩（旧）

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2-1](#)所示。

表 2-1 伸缩组参数说明

参数	解释	取值样例
区域	区域也称为Region。创建的伸缩组所在的区域。不同区域的资源之间内网不互通。请选择靠近您客户的区域，可以降低网络时延、提高访问速度。	-
可用区	可用区也称为AZ（Availability Zone）。可用区指在同一区域下，电力、网络隔离的物理区域，可用区之间内网互通，不同可用区之间物理隔离。 <ul style="list-style-type: none">• 如果您需要提高应用的高可用性，建议您将云服务器创建在不同的可用区内。• 如果您需要较低的网络时延，建议您将云服务器创建在相同的可用区内。	-

参数	解释	取值样例
多可用区扩展策略	<p>可选择“均衡分布”或“选择优先”。</p> <ul style="list-style-type: none"> 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。 <p>说明 当选择两个及以上可用区时，才需要配置该选项。</p>	均衡分布
名称	<p>创建的伸缩组的名称。</p> <p>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。</p>	-
最大实例数	最大实例数是指伸缩组中云服务器个数的最大值。	1台
期望实例数	<p>期望实例数是指伸缩组中期望的云服务器数量。</p> <p>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。</p>	0台
最小实例数	最小实例数是指伸缩组中云服务器个数的最小值。	0台
虚拟私有云	<p>弹性云服务器使用的网络是虚拟私有云（VPC）提供的。</p> <p>同一伸缩组内的弹性云服务器均属于该VPC。</p>	-
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。	-

参数	解释	取值样例
负载均衡	<p>可选参数。选择使用负载均衡器后，访问流量将自动分发到伸缩组内的所有弹性云服务器，扩展应用系统对外的服务能力，实现更高水平的应用程序容错性能。</p> <p>说明</p> <ul style="list-style-type: none"> 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。 <p>当使用弹性负载均衡时，需要配置以下四个参数：</p> <ul style="list-style-type: none"> 负载均衡器 后端云服务器组 后端端口：是指后端云服务器监听的端口。 权重：是指后端云服务器经分发得到的请求数量的比例。 <p>更多关于弹性负载均衡的信息，请参见《弹性负载均衡用户指南》。</p>	-
实例移除策略	<p>实例优先被移除的策略。当满足条件时，会触发实例移除活动，包括如下四种方式：</p> <ul style="list-style-type: none"> 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。 <p>说明</p> <ul style="list-style-type: none"> 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。 	-
弹性公网IP	<p>若伸缩组的伸缩配置使用了弹性公网IP，在进行扩容活动时，会给创建出来的云服务器绑定一个弹性公网IP。若选择“释放”，当进行缩容活动时，会将云服务器上的弹性公网IP释放，否则仅做解绑定操作，保留弹性公网IP资源。</p>	-

参数	解释	取值样例
健康检查方式	<p>健康检查会将异常的云服务器从伸缩组中移除，并重新创建新的云服务器，伸缩组的健康检查方式包括以下几种：</p> <ul style="list-style-type: none"> 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。 	-
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	<p>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。</p> <p>说明</p> <ul style="list-style-type: none"> “default”为默认企业项目，账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	<p>高级配置可对通知进行配置。</p> <p>可选择“暂不配置”或“现在配置”。</p>	-
通知	<p>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。</p> <ul style="list-style-type: none"> 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。 <ul style="list-style-type: none"> - 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。 	-

参数	解释	取值样例
标签	<p>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。</p> <p>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。</p> <ul style="list-style-type: none">● 键：<ul style="list-style-type: none">- 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<ul style="list-style-type: none">- 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”。

3 动态扩展云服务器

4 定时扩展云服务器

5 弹性伸缩向导式使用流程

使用弹性伸缩的流程如图5-1所示。

图 5-1 弹性伸缩向导式使用流程

