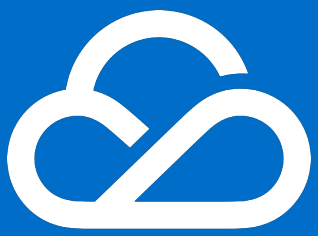
腾讯云弹性伸缩

扩缩容

产品文档



腾讯云

【版权声明】

©2013-2017 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

文档目录

文档声明 2
扩缩容 4
管理定时任务 4
管理告警触发策略 6
实例健康检查 10
手动扩容 11
缩容处理 14
查看伸缩活动 15
暂停及恢复扩缩容 16
指定实例免于缩容 18
伸缩活动取消 19
伸缩活动失败 20
冷却时间说明 21

版权所有：腾讯云计算（北京）有限责任公司

扩缩容

管理定时任务

定时任务简介

定时任务，顾名思义，即为设定时间计划，使您的业务根据可预测的负载变化，定时扩展或缩减所使用的云服务器实例数量。

例如，您的 Web 应用程序的流量会在每周的星期三开始增加，并在星期四保持高流量状态，然后在星期五开始下降。这种情况下，您可以根据 Web 应用程序的可预测流量模式来计划扩展活动。

要创建计划的扩展操作，请指定希望扩展操作生效的开始时间，以及用于扩展操作的新的最小大小（最小实例数）、最大大小（最大实例数）和所需大小（期望实例数）。在指定的时间，AS
将依据这些设定值来更新伸缩组中的实例数量。

您可以创建仅用于一次扩展的预先计划操作，或者创建用于按经常性计划进行扩展的预先计划操作。

管理定时任务

1. 打开控制台，选择导航条中的【伸缩组】。

2. 选择要修改的伸缩组，点击伸缩组ID进入伸缩组基本信息页面。

3. 在上方的导航条中选择【定时任务】，在该页面管理与伸缩组相关联的定时任务：

- 点击【新建】可添加新的定时任务；
- 选择某条定时任务，点击【修改】，可在弹出页面中修改任务名称，调整执行时间，设置是否周期执行，修改执行活动；
- 点击【删除】删除该条定时任务。

如果您想创建定时重复的任务，则可以指定开始时间，AS
会在该时间执行操作，然后根据重复计划执行操作；如果您指定结束时间，AS 在该时间后不再执行操作。

管理告警触发策略

简介

AS 支持根据监控的指标动态扩展伸缩组中的实例数量，您需定义告警触发策略，即触发扩展的监控指标状态以及如何按照需求变化进行扩展。

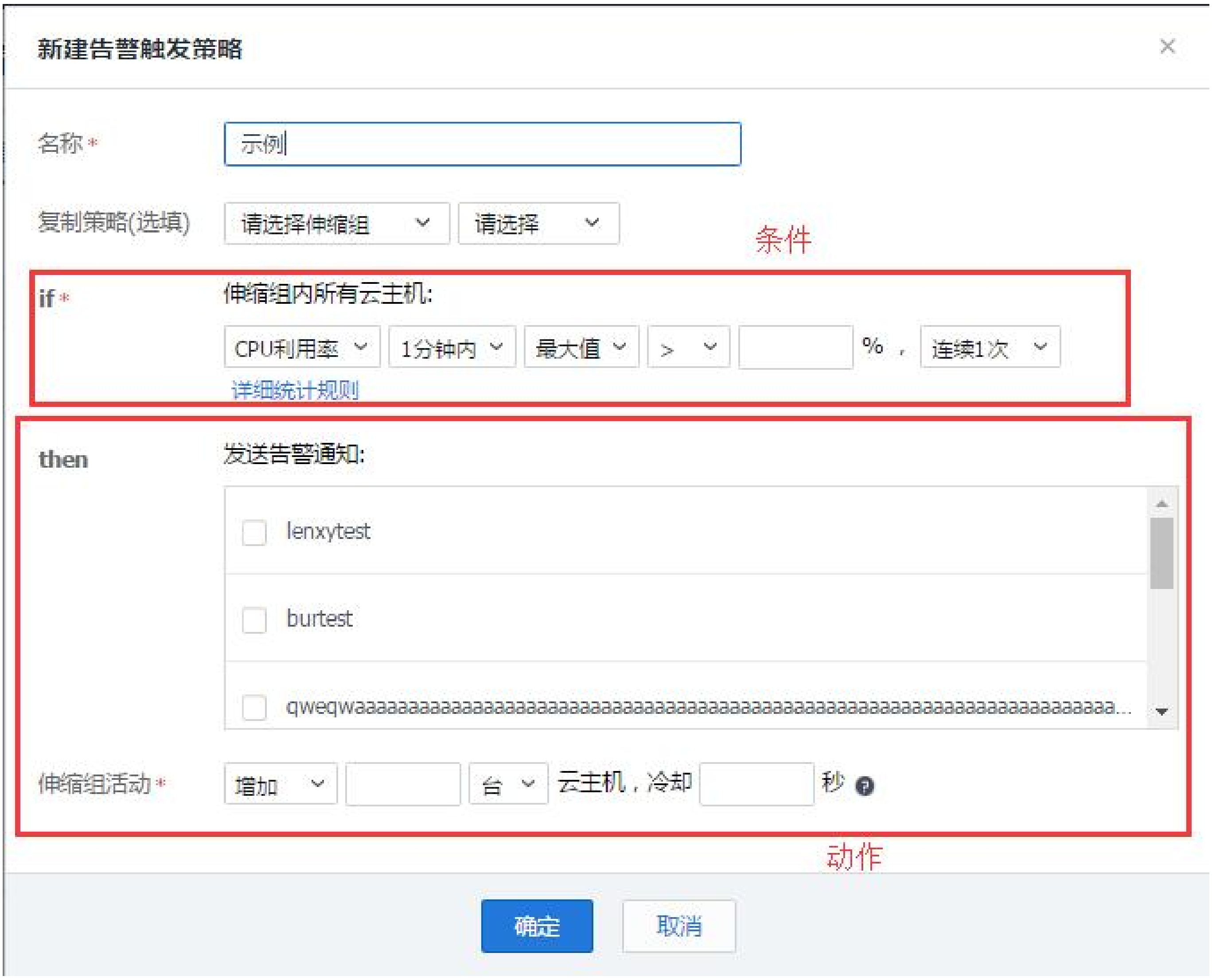
创建告警策略需指定条件和动作：

- 条件格式为：某个指标 + 阈值 + 周期 + 连续达到阈值的周期数。即指标在连续N个周期都达到了阈值。
- 执行动作为：发送通知 + 增加/减少 指定数量的云服务器。

建议您为每个伸缩组各创建两个策略：一个策略用于扩展，另一个策略用于收缩。当业务量达到了告警策略指定的条件后， AS 将执行关联的策略对伸缩组进行收缩（通过终止实例）或扩展（通过启动实例）。

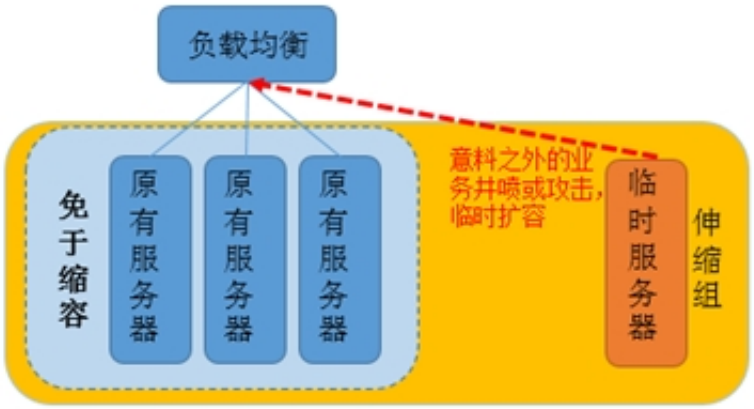
如下图所示：

版权所有：腾讯云计算（北京）有限责任公司



场景示例

比如您有一个电商网站应用程序，当前使用了五个实例。您做了一个运营活动，担心访问量远大于您的预估，您可以设置当前实例上的负载上升到 70% 时额外启动两个新的实例，然后在负载下降到 40% 时终止多余的实例。您可以配置伸缩组，根据这些条件自动扩展。



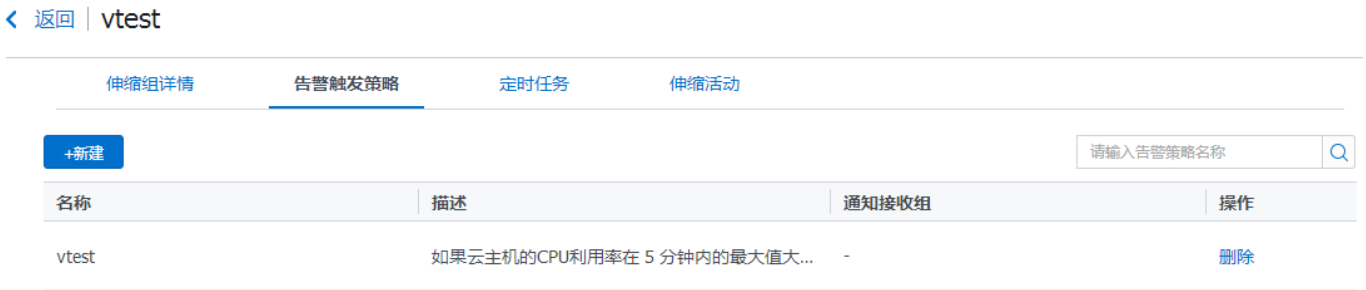
操作步骤

1. 打开控制台，选择导航条中的【伸缩组】。

2. 选择要修改的伸缩组，点击伸缩组ID进入伸缩组基本信息页面。

3. 在上方的导航条中选择【告警触发策略】，在该页面管理与伸缩组相关联的告警触发策略。

   - 点击【新建】可添加新的告警触发策略；
   - 点击【删除】删除该条告警触发策略；



指定某台服务器不受告警伸缩策略影响

使用 auto scaling

前，也许您的系统已经有常用的服务器，您出于以下考虑，不希望机器被告警伸缩策略移出：

- 一机多用
  ：集群中某台服务器除了做集群所做的事情外，还兼做其他用途。比如网站建设初期，您的某台服务器既作为缓存服务器使用，又作为文件服务器。在缓存服务器集群放入伸缩组时，您不希望它被告警伸缩策略移出。

- 存放数据
  ：该服务器是有状态的或自带其他服务器没有的数据。比如集群中其他服务器运行中产生的增量数据，都统一保存到该服务器里。

- 更新镜像/快照：固定使用该服务器定期做镜像和快照

设置方法：

step 1 :您可以在伸缩组列表里点击服务器所在的伸缩组，进入管理页面；

step 2：在管理页面下方的“云主机列表”，对所要设置的云主机点击“设置移出保护”；

实例健康检查

如果您在新建伸缩组时指定了【起始实例数】，创建启动配置和伸缩组后，伸缩组将新建与起始实例数相等的云服务器实例，同时，伸缩组会确保运行着大于【最小伸缩数】、小于【最大伸缩数】的实例。

注：

- 最小伸缩数：伸缩组中允许的实例最小数量。当伸缩组的CVM数量小于最小伸缩数时，弹性伸缩AS会增加实例，使得伸缩组当前实例数匹配最小伸缩数。
- 起始实例数：伸缩组刚创建时的云服务器数量。
- 最大伸缩数：伸缩组中允许的实例最大数量。当伸缩组的CVM数量大于最大伸缩数时，弹性伸缩AS会移出实例，使得伸缩组当前实例数匹配最大伸缩数。

为了保持伸缩组中的实例正常运行，AS 会对伸缩组内实例的运行状况执行定期检查。如果发现实例运行状况不佳，它将终止该实例，并启动一台新的云服务器实例。

- 实例健康检查

伸缩组定期检查实例运行状态来确定每个实例是否健壮，判断标准为该机器是否连续1分钟 ping 不可达。如果实例超过1分钟 ping 不可达，则 AS 会标记该实例运行状况不佳。

- 替换不健康实例

不健康的实例被标记为运行状况不佳之后，伸缩组将立即启动新的实例对它进行替换（设置了“移出保护”的机器除外）。

版权所有：腾讯云计算（北京）有限责任公司

手动扩容

AS除支持根据业务负载扩缩容外，还支持您手动介入，达到快速手动扩缩容的效果。您可以通过以下两种方式达到手动扩容效果：

- 将已有的CVM实例加到伸缩组中
- 通过修改伸缩组的期望实例数，实现一键扩容

将已有的CVM实例添加到伸缩组中

伸缩组为您提供了一个选项，可通过将一个或多个 CVM
实例附加到现有伸缩组，与伸缩组的其他机器一起观察负载和管理。

- 添加实例的条件

- 实例处于运行状态
- 实例与伸缩组位于同一地域
- 实例的网络属性必须与伸缩组一样，即同属基础网络或属于同一个 VPC

- 手动添加实例后的处理

- AS 会将该组的所需容量与要附加的实例数相加。比如您伸缩组目前的期望实例数是5，手动增加3台实例后，您伸缩组的期望实例数会变为
  5+3=8。（如果要附加的实例数加上所需容量超过组的最大大小，请求将失败）

- 如果伸缩组有关联一个或多个 CLB，手动添加的实例会自动注册到伸缩组的所有 CLB 中。

- 伸缩组缩容时会先移出自动创建的机器，没有自动创建的机器时，才会选择移出手动添加的机器。伸缩组移出手动添加的实例时，只是移出伸缩组和 CLB，不再通过伸缩组管理，而不会销毁您的实例。

- 使用控制台手动添加实例示例

点击您要管理的伸缩组 ID，或点击其后的【管理】，进入伸缩组详情页，拉到网页下半部分，有 CVM

列表，点击【添加云主机】，在对话框中勾选对应的主机，然后点击【确定】即可。

修改期望实例数，一键扩容

如果您满足以下条件，建议您可以使用 AS 的手动一键扩容：

- 业务的波峰波谷较难预测，但不愿把扩缩容完全交给系统决定；（如果业务波峰波谷可预测，建议您使用定时任务扩容）
- 您的计算需求是项目性的，且每次用的机器都类似（比如社情舆论收集、基因测序、天气预测）。
  这种情况下，您将云服务器的模板作出启动配置，并配置对应的伸缩组。以后要扩容，直接更改伸缩组所需要的容量即可。

在控制台进行一键扩容

1. 创建自定义镜像。
   后续扩容的实例将依据此镜像部署好环境。
   自定义镜像创建的推荐思路：您可选择已有的一台 CVM 或新创建一台
   CVM，将您的业务部署好，并将业务设置成随操作系统一起启动，然后导出为自定义镜像。
   如果遇到问题，可参照 创建自定义镜像的详细方法。

2. 基于该自定义镜像创建启动配置。
   您可 参照创建启动配置的详细步骤 进行。

3. 创建伸缩组。
   创建时选择已创建的启动配置，最小伸缩数、最大伸缩数、起始实例数根据您需要的服务器数量的下限、上限以及当前数量来填写。完成步骤三，您就可以随时扩容了。

4. 扩容。

   业务需要扩容时（例如开始一次基因测序任务或开通请求类机器收集数据），您可直接修改伸缩组，提高最小伸缩数、最大伸缩数、期望实例数即可，AS将快速为您完成扩容。

总结：提前将 CLB 转发规则、机器配置、业务部署这类工作做好，需要扩容时只要一键修改伸缩组的参数即可，这样的业务才是敏捷的。

版权所有：腾讯云计算（北京）有限责任公司

缩容处理

对于每个伸缩组，您可以控制何时向其添加实例（即扩容）或从中删除实例（即缩容）。您可以通过添加或移出实例手动扩展组大小，也可以使用扩展策略让 AS 自动执行该过程。

伸缩组自动缩容时，需要知道哪些实例应首先终止，选择的依据是移出策略。

在缩容时，您可以通过使用实例保护防止 AS 终止特定的实例。

移出策略

伸缩组缩容时，会根据移出策略决定移出哪台机器。您可从以下两种移出策略中选择：

- 删除最旧机器：删除最早自动增加的机器；自动增加的机器删除完后，删除最早手动增加的机器。
- 删除最新机器：删除最新自动增加的机器；自动增加的机器删除完后，删除最新手动增加的机器。

  注：不管删除最新机器还是删除最旧机器，AS 都会先删除自动创建的CVM，然后再删除您手动加入的CVM。

在控制台设置和修改移出策略

有两种方法设置：

- 创建伸缩组时，选择您需要的移出策略。
- 在伸缩组详情页，点击【编辑】，可修改伸缩策略。

版权所有：腾讯云计算（北京）有限责任公司

查看伸缩活动

查看伸缩活动

打开控制台，选择导航条中的【伸缩组】。

选择要修改的伸缩组，点击伸缩组ID进入伸缩组基本信息页面。

在上方的导航条中选择【伸缩活动】，在该页面可查看该伸缩组根据伸缩策略已执行过的伸缩活动信息。

暂停及恢复扩缩容

使用场景

如果您需要排查配置或与 Web 应用程序相关的其他问题（例如关机重置密码、升级业务等），希望在不触发自动伸缩流程的前提下对应用程序进行更改，那么您可以暂停 伸缩组，完成后再恢复。

暂停伸缩组

操作

打开控制台，选择导航条中的【伸缩组】，在伸缩组列表的右侧点击【停用】。

设置完后，在【状态】列中可看到“停用”字样。

注意事项

停用伸缩组后，伸缩组扩缩容自动触发活动不会进行，但是伸缩组的限制仍然生效。

自动触发的活动包括：

- 告警伸缩
- 定时任务
- 健康检查
- 手动造成期望实例数不匹配

伸缩组的限制包括：

- 若手动移出时小于最小实例数，不允许移出；
- 若手动加入超过最大实例数，不允许加入；
- 手动提高最小实例数或最大实例数，不触发伸缩活动。

恢复伸缩组

如您已完成暂停伸缩组活动期间的问题排查或操作，您可为您的业务恢复自动伸缩设置。

打开控制台，选择导航条中的【伸缩组】，在伸缩组列表的右侧点击【启用】即可。

指定实例免于缩容

简介

在伸缩组中，您可以指定某台子机在缩容活动时不被缩容掉；当缩容活动发生时，AS 在其他机器中选择要缩容的子机。

您可以对一个多个伸缩组实例启用 实例保护 设置。您可以随时更改伸缩组或实例保护设置。

如果伸缩组剩下的所有实例都受缩容保护，同时发生缩容事件，则 AS 会减少所需容量，而不会移出实例。

适用场景

正常来说，伸缩组的机器都是无状态的，所有的机器都可以随时被移走。但在实际实践中，有以下情况适用将指定实例设置免于缩容：

- 一机多用：
  基于成本考虑，个别机器除做集群中的事情外，还兼作其他用途，例如存储集群中产生的数据，那么这台机器实际上是有状态的。

- 避免误操作：若担心策略设置错误影响业务，则可以对部分机器设置“免于缩容”，这样 AS 永远不会缩容该机器，“请求-LB-子机”的通路可以保持畅通。

如何设置

在伸缩组的子机列表中，可直接设置：

伸缩活动取消

伸缩活动取消是指，定时任务时间到或者告警伸缩的条件达到，伸缩活动被触发，但是存在冲突，伸缩活动被迫取消。

冲突原因：

- 有进行中的伸缩活动；
- 伸缩组处于冷却时间中。

伸缩活动取消后是否会重试？

- 告警伸缩
  活动如果取消，不会再重试。但是如果告警伸缩的条件继续成立，会触发下一次告警伸缩活动。
- 定时任务 定义的是期望实例数、最大伸缩数、最小伸缩数，所以伸缩组会一直重试，使实际存在的实例数符合期望实例数。

  注：伸缩组被暂停，伸缩组会直接不尝试伸缩活动，所以在“伸缩活动”记录中，不会留下取消的伸缩活动。

版权所有：腾讯云计算（北京）有限责任公司

伸缩活动失败

伸缩活动取消 是符合预期的，伸缩活动失败 则是不符合预期的。

如何查看失败的伸缩活动？

您可查看伸缩活动详情。

要第一时间知道伸缩活动失败，您可配置通知策略。

为什么会发生失败的伸缩活动？

我们已经将伸缩活动的原因做好归类，请查阅 失败原因归类 >>

冷却时间说明

什么是冷却时间

弹性伸缩（AS）冷却时间是伸缩组的一个可配置设置，设置冷却时间，可以确保在上一扩展活动生效前 AS 不会启动或终止其他实例。伸缩组使用简单的扩展策略动态扩展后，AS 会等待冷却时间完成，然后再继续扩展活动。

手动扩展伸缩组时，默认为不等待冷却时间，但您可以设置冷却时间覆盖默认设置。请注意，如果监测出实例运行状况不佳，AS 会即时替换运行状况不佳的实例，而不会等待冷却时间完成。

为什么需要冷却时间

机器加入伸缩组后，需要一段时间才能将负载降下来。如果没有冷却时间，系统会在负载降下来前不断扩容，新加入的机器接管业务后，发现负载过低，然后又缩容。

在实例投入使用之前，这些实例使用配置脚本安装和配置软件，因此实例从启动到投入使用大约需要两到三分钟的时间。（当然，实际时间取决于诸多因素，如实例大小和是否有启动脚本要完成等。）

示例场景：

业务出现流量高峰，导致告警策略的警报触发。该警报触发时，AS 会启动一个实例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，并且启动后需要时间逐渐从 CLB 接收请求。在此期间，监控警报可能会继续触发，从而导致 AS 在警报每次出现时都另外启动一个实例。

但若您设置了冷却时间，AS 在启动一个实例后，将暂停所有简单扩展策略或手动扩展引起的扩展活动，直至经过了该指定时间量（默认值为 60 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，所有暂停的扩展操作都会恢复。如果警报再次触发，则 AS 将启动另一个实例，而冷却时间也会再次生效。不过，如果新增的实例足以将 CPU 使用率降为正常水平，则该组会保持其当前大小。

设置冷却时间

默认的冷却时间为60秒。

如需修改，请按以下步骤进行：

- 打开伸缩组的详情页；
- 点击【告警触发策略】，选择要设置的告警伸缩策略，选择【修改】，在修改框下方指定冷却时间的时长（可设置为 0-999999 秒）。

版权所有：腾讯云计算（北京）有限责任公司