

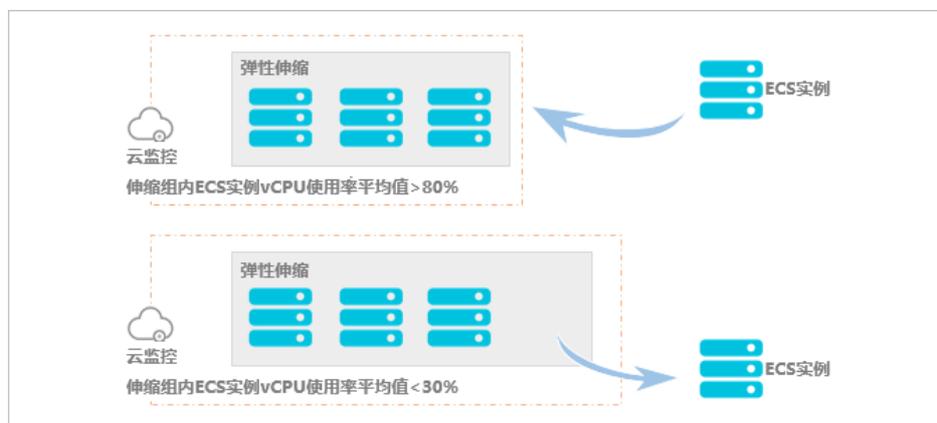
弹性伸缩

产品简介

产品简介

什么是弹性伸缩

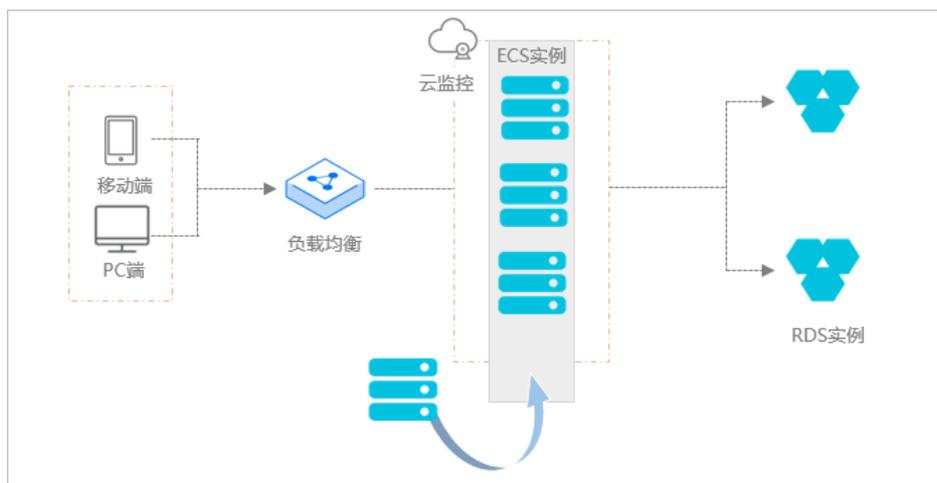
弹性伸缩自动为您调整弹性计算资源大小，以满足您业务需求的变化。弹性伸缩根据您设置的伸缩规则，在业务需求增长时自动为您增加ECS实例以保证计算能力，在业务需求下降时自动减少ECS实例以节约成本。



弹性扩张

当您的业务升级时，弹性伸缩为您自动完成底层资源升级，避免访问延时和资源超负荷运行。

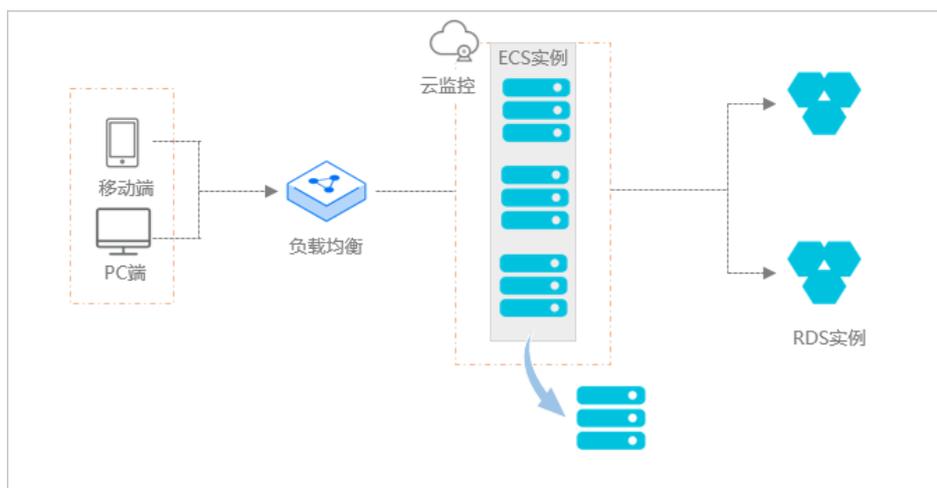
您可以配置云监控实时关注您的ECS实例使用情况。当云监控检测到伸缩组内的ECS实例vCPU使用率突破80%时，弹性伸缩根据您配置的伸缩规则弹性扩张ECS资源，自动创建合适数量的ECS实例，并自动添加ECS实例到负载均衡实例和RDS实例的访问白名单中。更多详情，请参阅 [创建伸缩组](#) 和 [云监控 监控弹性伸缩](#)。



弹性收缩

当您的业务需求下降时，弹性伸缩为您自动完成底层资源释放，避免资源浪费。

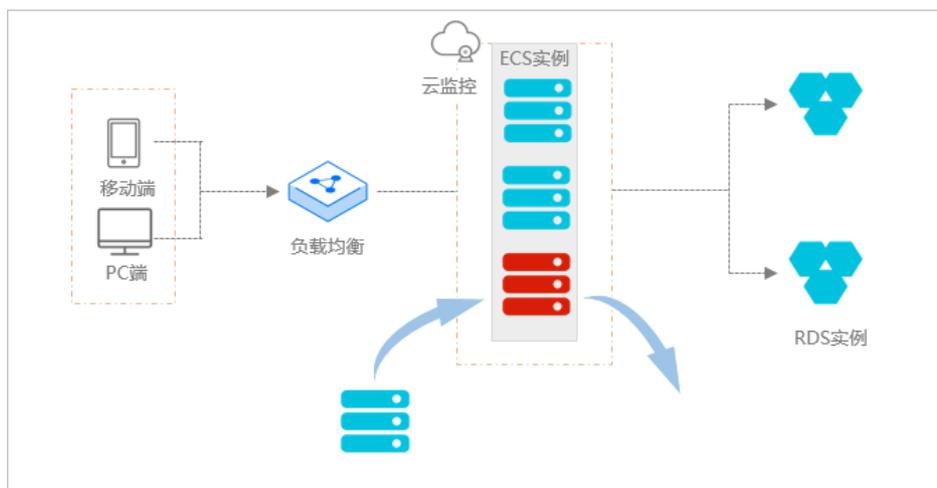
您可以配置云监控实时关注您的ECS实例使用情况。当云监控检测到伸缩组内的ECS实例vCPU使用率低于30%时，弹性伸缩根据您配置的伸缩规则弹性收缩ECS资源，自动释放合适数量的ECS实例，并自动从负载均衡实例和RDS实例的访问白名单中移除ECS实例。更多详情，请参阅 [移出策略](#) 和 [云监控 监控弹性伸缩](#)。



弹性自愈

弹性伸缩提供健康检查功能，自动监控伸缩组内的ECS实例的健康状态，避免伸缩组内健康ECS实例低于您设置的最小值。

当检测到某台ECS实例处于不健康状态时。弹性伸缩自动释放不健康ECS实例并创建新的ECS实例，自动添加新ECS实例到负载均衡实例和RDS实例的访问白名单中。更多详情，请参阅 [移出不健康ECS实例](#)。



相关链接

- 什么是云服务器ECS
- 什么是RDS
- 什么是负载均衡
- 什么是云监控

产品优势

功能概述

- 根据客户业务需求自动调整ECS实例数量。
- 自动向负载均衡的后端服务器组中添加或移除相应的ECS实例。
- 自动向RDS访问白名单中添加或移除ECS实例的IP。

产品特点

- 随需应变：根据需求“恰到好处”地分配资源，无需您提前预测需求变化，实时应对需求突增。
- 自动化：无需人工干预，自动创建和释放ECS实例，自动配置负载均衡和RDS访问白名单。
- 伸缩模式丰富：多模式兼容，可同时配置定时、动态、自定义、固定、健康模式，可通过API对接外在监控系统。
- 智能：智能调度云计算资源，应对各种复杂场景。

应用场景

- 某视频公司：春晚或每周五热门节目来临时，负载激增，需及时、自动扩展云计算资源。
- 某视频直播公司：业务负载变化难以预测，需要阿里云自动根据CPU利用率、应用负载、带宽利用率作为衡量指标进行弹性伸缩。
- 某游戏公司：每天中午12点及晚上6点到9点间需求增长，需要定时扩容。

伸缩模式

- 定时模式：您自定义自动伸缩发生的时间和频率，如每天 13:00 增加 ECS 实例。
- 动态模式：基于云监控性能指标（如 CPU 利用率），自动增加或减少 ECS 实例。
- 固定数量模式：通过设置 **最小实例数**（MinSize），即健康运行的ECS 实例最小数量，以保证可用性。

自定义模式：通过 API 调用您的自有监控系统，您可以执行手工伸缩。

- 手工执行伸缩规则。
- 手工添加或移出既有的 ECS 实例。
- 自定义 MinSize、MaxSize，弹性伸缩会自动创建或释放 ECS 实例，将当前 ECS 实例数维持在 MinSize 与 MaxSize 之间。

健康模式：如 ECS 实例为非 **running** 状态，弹性伸缩将自动移出或释放不健康的 ECS 实例。

- 多模式并行：以上所有模式都可以组合配置。例如设置了每天 13:00 ~ 14:00 创建 20 个 ECS 实例以应对业务高峰，但实际需求有可能需要多于20个实例，则您可以选择其他伸缩模式，与定时模式配合一起使用。

限制条件

伸缩组内部署在 ECS 实例的应用必须无状态并且可横向扩展。

弹性伸缩会自动释放ECS实例，所以建议伸缩组内ECS实例不要保存应用状态信息和相关数据等信息，例如会话记录（Session）、数据库或者日志等。如果有需要，您可以保存状态信息到独立的状态云服务器ECS、保存数据库到 云数据库RDS 或者集中日志存储到 日志服务。

弹性伸缩无法自动添加ECS实例到开放缓存 Memcache 实例访问白名单，需要您自行添加。

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规格、内存和带宽等配置。

您能创建的伸缩组、伸缩配置、伸缩规则、ECS实例、定时任务有一定的限制数量。

发展历史

2015年8月27日 弹性伸缩产品上线开放

2014年10月15日 弹性伸缩产品内测

名词解释

弹性伸缩

弹性伸缩是根据用户的业务需求和策略，自动调整其弹性计算资源的管理服务。其能够在业务增长时自动增加ECS实例，并在业务下降时自动减少ECS实例。

伸缩组

伸缩组是具有相同应用场景的ECS实例的集合。伸缩组定义了组内ECS实例数的最大值、最小值及其相关联的负载均衡实例和RDS实例等属性。

伸缩配置

伸缩配置定义了用于弹性伸缩的ECS实例的配置信息。

伸缩规则

伸缩规则定义了具体的扩展或收缩操作，例如加入或移出N个ECS实例。

伸缩活动

伸缩规则成功触发后，就会产生一条伸缩活动。伸缩活动主要用来描述伸缩组内 ECS 实例的变化情况。

伸缩触发任务

用于触发伸缩规则的任务，如定时任务、云监控的报警任务。

冷却时间

冷却时间是指，在同一伸缩组内，一个伸缩活动执行完成后的一段锁定时间。在这段锁定时间内，该伸缩组不执行其他的伸缩活动。

备注

- 伸缩组包含伸缩配置、伸缩规则、伸缩活动。
- 伸缩配置、伸缩规则、伸缩活动依赖伸缩组的生命周期管理，删除伸缩组的同时会删除与伸缩组相关的伸缩配置、伸缩规则和伸缩活动。
- 伸缩触发任务有定时任务、云监控报警任务等类型。
- 定时任务独立于伸缩组存在，不依赖伸缩组的生命周期管理，删除伸缩组不会删除定时任务。
- 云监控报警任务独立于伸缩组存在，不依赖伸缩组的生命周期管理，删除伸缩组不会删除报警任务。