

阿里云 弹性伸缩 用户指南

文档版本：20190813

法律声明

阿里云提醒您在阅读或使用本文档之前仔细阅读、充分理解本法律声明各条款的内容。如果您阅读或使用本文档，您的阅读或使用行为将被视为对本声明全部内容的认可。

1. 您应当通过阿里云网站或阿里云提供的其他授权通道下载、获取本文档，且仅能用于自身的合法合规的业务活动。本文档的内容视为阿里云的保密信息，您应当严格遵守保密义务；未经阿里云事先书面同意，您不得向任何第三方披露本手册内容或提供给任何第三方使用。
2. 未经阿里云事先书面许可，任何单位、公司或个人不得擅自摘抄、翻译、复制本文档内容的部分或全部，不得以任何方式或途径进行传播和宣传。
3. 由于产品版本升级、调整或其他原因，本文档内容有可能变更。阿里云保留在没有任何通知或者提示下对本文档的内容进行修改的权利，并在阿里云授权通道中不时发布更新后的用户文档。您应当实时关注用户文档的版本变更并通过阿里云授权渠道下载、获取最新版的用户文档。
4. 本文档仅作为用户使用阿里云产品及服务的参考性指引，阿里云以产品及服务的“现状”、“有缺陷”和“当前功能”的状态提供本文档。阿里云在现有技术的基础上尽最大努力提供相应的介绍及操作指引，但阿里云在此明确声明对本文档内容的准确性、完整性、适用性、可靠性等不作任何明示或暗示的保证。任何单位、公司或个人因为下载、使用或信赖本文档而发生任何差错或经济损失的，阿里云不承担任何法律责任。在任何情况下，阿里云均不对任何间接性、后果性、惩戒性、偶然性、特殊性或刑罚性的损害，包括用户使用或信赖本文档而遭受的利润损失，承担责任（即使阿里云已被告知该等损失的可能性）。
5. 阿里云网站上所有内容，包括但不限于著作、产品、图片、档案、资讯、资料、网站架构、网站画面的安排、网页设计，均由阿里云和/或其关联公司依法拥有其知识产权，包括但不限于商标权、专利权、著作权、商业秘密等。非经阿里云和/或其关联公司书面同意，任何人不得擅自使用、修改、复制、公开传播、改变、散布、发行或公开发表阿里云网站、产品程序或内容。此外，未经阿里云事先书面同意，任何人不得为了任何营销、广告、促销或其他目的使用、公布或复制阿里云的名称（包括但不限于单独为或以组合形式包含“阿里云”、“Aliyun”、“万网”等阿里云和/或其关联公司品牌，上述品牌的附属标志及图案或任何类似公司名称、商号、商标、产品或服务名称、域名、图案标示、标志、标识或通过特定描述使第三方能够识别阿里云和/或其关联公司）。
6. 如若发现本文档存在任何错误，请与阿里云取得直接联系。

通用约定

格式	说明	样例
	该类警示信息将导致系统重大变更甚至故障，或者导致人身伤害等结果。	 禁止： 重置操作将丢失用户配置数据。
	该类警示信息可能导致系统重大变更甚至故障，或者导致人身伤害等结果。	 警告： 重启操作将导致业务中断，恢复业务所需时间约10分钟。
	用于补充说明、最佳实践、窍门等，不是用户必须了解的内容。	 说明： 您也可以通过按Ctrl + A选中全部文件。
>	多级菜单递进。	设置 > 网络 > 设置网络类型
粗体	表示按键、菜单、页面名称等UI元素。	单击 确定 。
<code>courier</code> 字体	命令。	执行 <code>cd /d C:/windows</code> 命令，进入Windows系统文件夹。
<code>##</code>	表示参数、变量。	<code>bae log list --instanceid Instance_ID</code>
<code>[]或者[a b]</code>	表示可选项，至多选择一个。	<code>ipconfig [-all -t]</code>
<code>{ }或者{a b}</code>	表示必选项，至多选择一个。	<code>swich {stand slave}</code>

目录

法律声明.....	I
通用约定.....	I
1 使用须知.....	1
1.1 ECS实例生命周期的管理.....	1
1.2 冷却时间.....	2
1.3 伸缩组状态.....	3
1.4 伸缩活动的流程.....	3
1.5 伸缩活动的状态.....	5
1.6 伸缩活动中失败实例的回滚.....	5
1.7 移出不健康ECS实例.....	6
1.8 通知.....	6
1.9 强制干预.....	6
1.10 数量限制.....	8
1.11 注意事项.....	8
1.12 使用流程.....	9
1.13 工作流程.....	10
2 管理伸缩组的实例模板配置.....	12
2.1 ECS 实例模板概述.....	12
2.2 创建伸缩配置.....	13
2.3 选用伸缩配置.....	16
2.4 修改伸缩配置.....	16
2.5 删除伸缩配置.....	17
2.6 导出伸缩配置.....	18
2.7 导入伸缩配置.....	19
3 实现自动伸缩.....	21
3.1 使用自定义伸缩配置创建伸缩组.....	21
3.2 使用实例启动模板创建伸缩组.....	25
3.3 创建伸缩规则.....	30
3.4 创建生命周期挂钩.....	33
3.5 创建预测规则.....	35
3.6 执行伸缩规则.....	37
3.7 定时任务.....	40
3.7.1 创建定时任务.....	40
3.7.2 管理定时任务.....	44
3.8 报警任务.....	45
3.8.1 弹性伸缩报警任务.....	45
3.8.2 系统监控报警任务.....	46
3.8.3 自定义监控项报警任务.....	47
3.8.4 创建报警任务.....	49
3.8.5 查看修改删除报警任务.....	50

4 维护自动伸缩	53
4.1 查看预测规则效果	53
4.2 修改生命周期挂钩	55
4.3 删除生命周期挂钩	56
4.4 修改伸缩规则	57
4.5 删除伸缩规则	59
4.6 设置实例备用状态	59
4.7 实例转为保护状态	61
4.8 实例移出保护状态	61
4.9 移出策略	62
4.10 更改伸缩组状态	62
4.11 修改伸缩组	63
4.12 删除伸缩组	64
5 手动伸缩	66
5.1 添加ECS实例	66
5.2 移出ECS实例	67
6 事件通知	69
6.1 事件通知概述	69
6.2 创建事件通知	71
6.3 管理事件通知	72
7 查询ECS实例列表	74
8 查看伸缩活动	75
9 排查伸缩活动异常	76
10 弹性伸缩FAQ	84
11 弹性伸缩关联产品	85
11.1 在弹性伸缩中使用负载均衡服务	85
12 管理伸缩组的伸缩活动	88
12.1 伸缩组管理概述	88

1 使用须知

1.1 ECS实例生命周期的管理

本文主要介绍ECS实例生命周期管理。

加入伸缩组的 ECS 实例有两种类型：自动创建的 ECS 实例、手工添加的 ECS 实例。

自动创建的 ECS 实例

自动创建的 ECS 实例是指根据用户的伸缩配置和伸缩规则，由弹性伸缩服务自动创建的 ECS 实例。

弹性伸缩服务负责该类型 ECS 实例的全生命周期管理，即在伸缩组扩展活动中，负责该 ECS 实例的创建，在伸缩组的收缩活动中，负责停止和释放该 ECS 实例。

手工添加的 ECS 实例

手工添加的 ECS 实例是指不是由弹性伸缩服务创建，但由用户手工添加到伸缩组中的 ECS 实例。

弹性伸缩服务不负责该类型 ECS 实例的全生命周期管理，即该类型的 ECS 实例是在弹性伸缩服务之外创建，只能手工加入伸缩组，在伸缩组的收缩活动中或用户手工移出该类型 ECS 实例时，弹性伸缩服务只负责将该 ECS 实例移出伸缩组，不会停止和释放该 ECS 实例。

实例状态

ECS 实例在伸缩组中的生命周期，通过以下几种状态描述：

- Pending – 表示 ECS 实例正在加入伸缩组，包括创建实例、加入负载均衡、添加 RDS 访问名单等过程。
- InService – 表示 ECS 实例已成功加入伸缩组，并正常提供服务。
- Removing – 表示 ECS 实例正在移出伸缩组。

实例的健康状态

ECS 实例在伸缩组中的健康状态为：

- Healthy（健康）
- Unhealthy（不健康）

本期仅通过 ECS 实例为非 运行中（Running） 状态来判断该实例不健康。弹性伸缩会自动移出伸缩组中不健康的 ECS 实例。

- 对于自动创建的 ECS 实例，弹性伸缩会停止和释放该 ECS 实例。

- 对于手工添加的 ECS 实例，弹性伸缩不会停止和释放该 ECS 实例。

1.2 冷却时间

本文主要介绍弹性伸缩的冷却时间。

冷却时间指成功执行伸缩活动后的一段锁定时间。您可以通过以下两种方式配置冷却时间：

- 在伸缩组中配置默认冷却时间，默认冷却时间不能为空，请参照 [创建伸缩组](#)。
- 在伸缩规则中配置冷却时间，如果不配置，将使用默认冷却时间，请参照 [创建伸缩规则](#)。



说明：

如果同时在伸缩组和伸缩规则中配置了冷却时间，弹性伸缩服务将优先使用伸缩规则中的冷却时间。

冷却时间规则

伸缩组成功执行伸缩活动后，弹性伸缩服务开始计算冷却时间。如果在伸缩活动中，多台 ECS 实例加入或者移出伸缩组，则从最后一台 ECS 实例加入或者移出伸缩组后开始计算冷却时间，请参照 [示例场景](#)。如果在伸缩活动中，没有 ECS 实例成功加入或者移出伸缩组，则不会开始计算冷却时间。

在冷却时间内，伸缩组会拒绝由云监控报警任务触发的伸缩活动请求。但其他类型任务（手动执行任务、定时任务）触发的伸缩活动可以立即执行，绕过冷却时间。

如果您停用伸缩组，然后再次启用伸缩组，冷却时间将会失效，请参照 [示例 2](#)。



说明：

冷却时间仅锁定同一个伸缩组的伸缩活动，不会对其他伸缩组的伸缩活动产生影响。

示例场景

示例 1

一个伸缩组 asg-uf6f3xewn3dvz4bsy7r1，默认冷却时间是 10 分钟，伸缩组内存在一条伸缩规则 add3，冷却时间为 15 分钟。

根据 add3 成功执行一次伸缩活动后，扩张了 3 台 ECS 实例，从第 3 台 ECS 实例加入伸缩组开始计算冷却时间，15 分钟内不会接受由云监控报警任务触发的伸缩活动请求。

示例 2

一个伸缩组 asg-m5efkz67re9x7a571bjh，默认冷却时间是 10 分钟，伸缩组内存在一条伸缩规则 remove1，没有设置冷却时间。

在 18:00 根据 `remove1` 成功执行一次伸缩活动后，收缩了 1 台 ECS 实例。正常情况下，在 18:10 前伸缩组不会接受由云监控报警任务触发的伸缩活动请求。此时停用伸缩组，然后在 18:05 再次启用伸缩组，冷却时间将失效。如果在 18:05 至 18:10 间云监控报警任务触发了伸缩活动请求，伸缩组仍会接受并执行请求。

1.3 伸缩组状态

本文主要介绍弹性伸缩的伸缩组状态。

伸缩组具有以下三种状态：Active、Inactive 和 Deleting。

状态	API 对应的状态
创建中	Inactive
已创建	Inactive
启用中	Inactive
运行中	Active
停用中	Inactive
已停止	Inactive
删除中	Deleting

1.4 伸缩活动的流程

本文主要介绍弹性伸缩的伸缩活动流程。

伸缩活动的生命周期为判断伸缩组的健康状态和边界条件步骤与启动 `cooldown` 步骤之间的所有活动。

伸缩组自动伸缩

自动扩展

1. 判断伸缩组的健康状态和边界条件。
2. 分配 `ActivityId` 和执行伸缩活动。
3. 创建 ECS 实例。
4. 修改 `Total Capacity`。
5. 分配 IP。
6. 添加 RDS 白名单。
7. 启动 ECS 实例，等待启动完成。
8. 挂载负载均衡，将权重设为创建伸缩配置时指定的 负载均衡权重。

9. 伸缩活动完成，启动 cooldown。

自动收缩

1. 判断伸缩组的健康状态和边界条件。
2. 分配 ActivityId 和执行伸缩活动。
3. 从负载均衡移除 ECS 实例。
4. 停止 ECS 实例，等待停止完成。
5. 从 RDS 白名单移除。
6. 释放 ECS 实例。
7. 修改 Total Capacity。
8. 伸缩活动完成，启动 cooldown。

用户手工加入和移出既有的 ECS 实例

手工加入

1. 判断伸缩组的健康状态、边界条件和 ECS 实例的状态、类型。
2. 分配 ActivityId 和执行伸缩活动。
3. 加入 ECS 实例。
4. 修改 Total Capacity。
5. 添加 RDS 白名单。
6. 挂载负载均衡，将权重设为当前伸缩组中已激活的伸缩配置上指定的 负载均衡权重。
7. 伸缩活动完成，启动 cooldown。

手工移出

1. 判断伸缩组的健康状态和边界条件。
2. 分配 ActivityId 和执行伸缩活动。
3. 负载均衡停止向该 ECS 实例转发流量。
4. 等待 60 秒后，从负载均衡移除 ECS 实例。
5. 从 RDS 白名单移除。
6. 修改 Total Capacity。
7. 从伸缩组移出。
8. 伸缩活动完成，启动 cooldown。

1.5 伸缩活动的状态

本文主要介绍弹性伸缩的伸缩活动状态。

伸缩活动在请求阶段被拒绝，为 Rejected 状态。

伸缩活动执行过程中，为 InProgress 状态。

伸缩活动执行完成后，有 3 种状态：

- 成功 (Successful)：根据 MinSize、MaxSize 调整后的伸缩规则，伸缩活动成功将所有 ECS 实例加入或移出伸缩组。



说明：

加入伸缩组成功表明创建 ECS 实例、加入负载均衡实例、配置 RDS 访问白名单每一步骤都成功。以上任何步骤失败都认为是该 ECS 实例加入伸缩组失败。

- 部分成功 (Warning)：根据 MinSize、MaxSize 调整后的伸缩规则，伸缩活动至少有一台 ECS 实例加入或移出伸缩组不成功。
- 全部失败 (Failed)：根据 MinSize、MaxSize 调整后的伸缩规则，伸缩活动没有一台 ECS 实例成功加入或移出伸缩组。

举例

伸缩规则：伸缩规则定义增加 5 台 ECS 实例，当前伸缩组的 Total Capacity = 3，MaxSize = 5，则执行该伸缩规则时会调整成增加 2 台 ECS 实例。伸缩活动执行完成后，3 种状态对应的情况如下：

- 成功：2 台 ECS 实例全部创建成功，并正确加入负载均衡实例和配置 RDS 访问白名单。
- 部分成功：2 台 ECS 实例全部创建成功，但只有 1 台正确加入负载均衡实例和配置 RDS 访问白名单，另外一台因无法加入负载均衡实例和配置 RDS 访问白名单而被回滚操作释放。
- 全部失败：2 台 ECS 实例全部创建失败，或者 2 台 ECS 实例全部创建成功，但这 2 台 ECS 实例都因无法加入负载均衡实例和配置 RDS 访问白名单而被回滚操作释放。

1.6 伸缩活动中失败实例的回滚

本文主要介绍伸缩活动中失败实例的回滚。

伸缩活动有 ECS 实例加入伸缩组失败时，需要保持 ECS 实例级事务的完整性，而非伸缩活动级事务的完整性，即只进行 ECS 实例级回滚，而不是伸缩活动级回滚。

例如，当伸缩组创建了 20 台 ECS 实例，但只有 19 台 ECS 实例成功加入负载均衡时，则只对不成功的 1 台 ECS 实例进行自动释放操作。

由于弹性伸缩是借助阿里云的 RAM（Resource Access Management）服务，通过 ECS Open API 代替用户弹性伸缩 ECS 实例资源，所以回滚的 ECS 实例仍然会被扣费。

1.7 移出不健康ECS实例

本文主要介绍移出不健康实例的操作。

当 ECS 实例成功加入伸缩组后，弹性伸缩服务会定期扫描该 ECS 实例的运行状态，如果发现该 ECS 实例为非 运行中（Running）状态，则会将该 ECS 实例移出伸缩组。

- 对于弹性服务自动创建的 ECS 实例：将立即启动移出和释放该 ECS 实例操作。
- 对于用户手工添加的 ECS 实例：将立即启动移出该 ECS 实例操作，但不会停止和释放该 ECS 实例。

移出不健康 ECS 实例，不受 MinSize 的限制，即移出不健康 ECS 实例后，有可能导致 Total Capacity 低于 MinSize，此时弹性伸缩服务会自动创建差额的 ECS 实例使得 Total Capacity 等于 MinSize。

1.8 通知

本文主要介绍伸缩活动通知。

需要短信息和邮件通知的伸缩活动条件如下：

- 只有定时任务、云监控报警任务、健康检查触发的伸缩活动才通知。
- 只有创建了 ECS 实例或释放了 ECS 实例才通知。

如满足以上条件，则一个伸缩活动对应一条通知。

1.9 强制干预

本文主要介绍强制干预相关内容。

弹性伸缩服务不会阻止用户从 ECS 控制台删除自动创建的 ECS 实例等强制干预行为。对于用户的强制干预，弹性伸缩服务的处理方法如下。

资源	强制干预类型	弹性伸缩服务的处理方法
ECS	通过 ECS 控制台或 OpenAPI 强制删除伸缩组里的 ECS 实例	会按照 健康检查 来判断该 ECS 实例为不健康状态，并将该 ECS 实例移出伸缩组。（该 ECS 实例在 RDS 实例访问白名单中的内网 IP 无法自动删除。）如移出该 ECS 实例后，使得 Total Capacity 低于 MinSize，弹性伸缩会自动创建差额的 ECS 实例使得 Total Capacity 等于 MinSize。
ECS	客户取消将 ECS Open API 权限授予弹性伸缩	弹性伸缩将拒绝所有伸缩活动请求。
负载均衡	通过负载均衡控制台或 OpenAPI 将 ECS 实例强制移出负载均衡实例	弹性伸缩不会自动感知用户的移出行为也不会处理该类异常状态，该 ECS 实例仍存在伸缩组中。当发生收缩活动时，如果按照移出策略挑选到该 ECS 实例，仍然会把该 ECS 实例释放掉。
负载均衡	通过负载均衡控制台或 OpenAPI 强制删除负载均衡实例或关闭负载均衡实例的健康检查。	对于配置该负载均衡实例的伸缩组，不会再往伸缩组加入 ECS 实例。伸缩任务可以触发伸缩规则从伸缩组移出 ECS 实例，健康检查判断某个 ECS 实例不健康，也可将该 ECS 实例移出伸缩组。
负载均衡	系统导致负载均衡实例不可用（如欠费、故障）	除了用户手工触发的移出 ECS 实例的伸缩活动，其它所有伸缩活动执行失败。
负载均衡	客户取消将负载均衡 Open API 权限授予弹性伸缩。	对于配置了负载均衡实例的伸缩组，弹性伸缩将拒绝所有伸缩活动请求。
RDS	通过 RDS 控制台或 OpenAPI 将 ECS 实例的 ip 强制移出 RDS 访问白名单	弹性伸缩不会自动感知用户的移出行为也不会处理该类异常状态，该 ECS 实例仍存在伸缩组中。当发生收缩活动时，如果按照移出策略挑选到该 ECS 实例，仍然会把该 ECS 实例释放掉。

资源	强制干预类型	弹性伸缩服务的处理方法
RDS	通过 RDS 控制台或 OpenAPI 强制删除 RDS 实例	对于配置该 RDS 实例的伸缩组，不会再往伸缩组加入 ECS 实例。伸缩任务可以触发伸缩规则从伸缩组移出 ECS 实例，健康检查判断某个 ECS 实例不健康，也可将该 ECS 实例移出伸缩组。
RDS	系统导致 RDS 实例不可用（如欠费、故障）	除了用户手工触发的移出 ECS 实例的伸缩活动，其它所有伸缩活动执行失败。
RDS	客户取消将 RDS Open API 权限授予弹性伸缩。	对于配置了 RDS 实例的伸缩组，弹性伸缩将拒绝所有伸缩活动请求。

1.10 数量限制

本文为您介绍弹性伸缩服务涉及的数量限制。

目前，弹性伸缩服务的功能项存在以下数量限制：

- 同一个账号一个地域下最多创建 50 个伸缩组。
 - 一个伸缩组内最多创建 10 个伸缩配置。
 - 一个伸缩组内最多创建 50 个伸缩规则。
 - 一个伸缩组内最多创建 6 个事件通知。
 - 一个伸缩组内最多创建 6 个生命周期挂钩。
 - 一个伸缩组最多同时关联 5 个负载均衡实例。
- 一个账号下最多创建 20 个定时任务。
- 一个伸缩配置下，最多可以选择 10 个实例规格。

1.11 注意事项

本文主要介绍弹性伸缩相关注意事项。

伸缩规则

在计算和执行过程中，伸缩规则可以根据伸缩组的 MinSize、MaxSize 进行自动调整其需要增加或减少的 ECS 实例数（例如：伸缩规则中指定将伸缩组的 ECS 实例数调整至 50 台，但伸缩组 MaxSize 只有 45 台，则整个伸缩规则会按调整至 45 台来计算和执行。）

伸缩活动

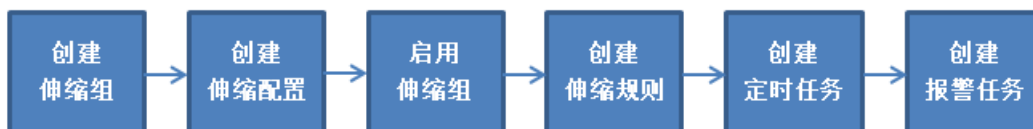
- 同一伸缩组内、同一时刻只能有一个伸缩活动在执行。
- 伸缩活动不可以中断。例如，某个创建 20 台 ECS 实例的伸缩活动正在执行中，当创建到第 5 台 ECS 实例时，您无法强行终止该伸缩活动。
- 伸缩活动有 ECS 实例加入伸缩组失败时，需要保持 ECS 实例级事务的完整性，而非伸缩活动级事务的完整性，即只进行 ECS 实例级回滚，而不是伸缩活动级回滚。例如，当伸缩组创建了 20 台 ECS 实例，但只有 19 台 ECS 实例成功加入负载均衡时，则只对不成功的 1 台 ECS 实例进行自动释放操作。
- 由于弹性伸缩是借助阿里云的 RAM（Resource Access Management）服务，通过 ECS Open API 代替用户弹性伸缩 ECS 实例资源，所以回滚的 ECS 实例仍然会被扣费。

冷却时间

- 在冷却时间内，伸缩组只会拒绝云监控报警任务类型的伸缩活动请求，其他类型的触发任务（如用户手工执行伸缩规则、定时任务等）可以绕过冷却时间立即执行伸缩活动。
- 每个伸缩活动的最后一个 ECS 实例加入或移出伸缩组成功后，整个伸缩组冷却时间才开始计时。

1.12 使用流程

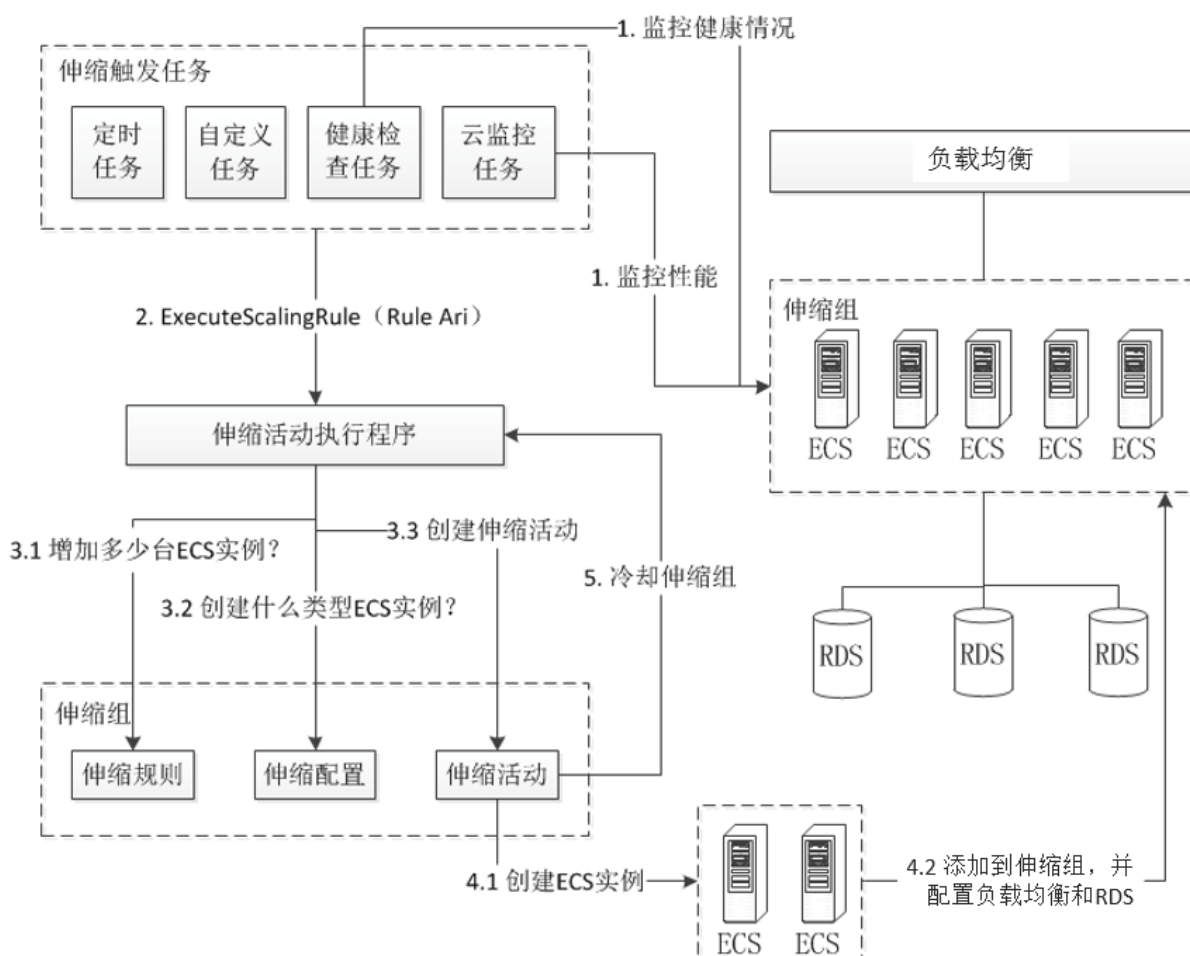
本文主要介绍创建完整弹性伸缩方案的步骤。



1. 创建伸缩组（CreateScalingGroup），配置伸缩资源的最小值（MinSize）、最大值（MaxSize）及需要关联的负载均衡实例和RDS实例。
2. 创建伸缩配置（CreateScalingConfiguration），指定需要弹性伸缩的 ECS 实例的相关属性，如 ImageID、InstanceType 等。
3. 以第二步创建的伸缩配置启用伸缩组（EnableScalingGroup）。
4. 创建伸缩规则（CreateScalingRule），如加 N 台 ECS 实例的伸缩规则。
5. 创建定时任务（CreateScheduledTask），如创建 12:00 触发第四步伸缩规则的定时任务。
6. 创建报警任务（云监控 API PutAlarmRule），如创建 CPU 平均值（也可以是最大值或最小值）大于等于 80% 则增加一台 ECS 实例的报警任务。

1.13 工作流程

本文主要介绍弹性伸缩的工作流程。



创建好伸缩组、伸缩配置、伸缩规则、伸缩触发任务后，系统会自动化执行以下流程（以增加 ECS 实例为例）：

1. 伸缩触发任务会按照各自触发生效的条件来触发伸缩活动。

- 云监控任务会实时监控伸缩组内 ECS 实例的性能，并根据用户配置的报警规则（如伸缩组内所有 ECS 实例的 CPU 平均值大于 60%）触发执行伸缩规则请求。
- 定时任务会根据用户配置的时间来触发执行伸缩规则请求。
- 您可以根据自己的监控系统及相应的报警规则（如在线人数、作业队列）来触发执行伸缩规则请求。
- 健康检查任务会定期检查伸缩组和 ECS 实例的健康情况，如发现有不健康的 ECS 实例（如 ECS 为非 Running 状态）会触发执行移出该 ECS 实例的请求。

2. 系统自动通过 ExecuteScalingRule 接口触发伸缩活动，并在该接口中指定需要执行的伸缩规则的阿里云资源唯一标识符（Ari）。

如果是用户自定义的任务，则需要用户在自己的程序中调用 ExecuteScalingRule 接口来实现。

3. 根据步骤 2 传入的伸缩规则 Ari（Rule Ari）获取伸缩规则、伸缩组、伸缩配置的相关信息，并创建伸缩活动。
 - a. 通过伸缩规则 Ari 查询伸缩规则以及相应的伸缩组信息，计算出需要增加的 ECS 实例数量，并获得需要配置的负载均衡和 RDS 信息。
 - b. 通过伸缩组查询到相应的伸缩配置信息，即获得了需要创建的 ECS 实例的配置信息（CPU、内存、带宽等）。
 - c. 根据需要增加的 ECS 实例数量、ECS 实例配置信息、需要配置的负载均衡实例和 RDS 实例创建伸缩活动。
4. 在伸缩活动中，自动创建 ECS 实例并配置负载均衡和 RDS。
 - a. 按照实例配置信息创建指定数量的 ECS 实例。
 - b. 将创建好的 ECS 实例的内网 IP 添加到指定的 RDS 实例的访问白名单当中，将创建好的 ECS 实例添加到指定的负载均衡实例当中。
5. 伸缩活动完成后，启动伸缩组的冷却功能。待冷却时间完成后，该伸缩组才能接收新的执行伸缩规则请求。

2 管理伸缩组的实例模板配置

2.1 ECS 实例模板概述

在您的业务需求增长时，弹性伸缩可以根据 ECS 实例模板进行弹性扩张，自动创建具有指定配置的 ECS 实例并添加到伸缩组中。目前，ECS 实例模板分为两类：自定义伸缩配置 和 实例启动模板。

自定义伸缩配置

伸缩配置是您为伸缩组创建的一类 ECS 实例模板。在创建伸缩配置时，您可以指定 ECS 实例的参数，例如实例规格、镜像类型、存储大小、登录实例用的密钥对等。您也可以修改现有伸缩配置，快速满足新的业务需求。



说明：

伸缩配置依托伸缩组，因此要确保已经存在一个伸缩组。一个伸缩组内可以创建的伸缩配置数量有限，请参阅 [使用限制](#)。

目前，您可以进行以下伸缩配置相关的操作：

- [创建伸缩配置](#)
- [选用伸缩配置](#)
- [修改伸缩配置](#)
- [删除伸缩配置](#)
- [导出伸缩配置](#)
- [导入伸缩配置](#)

实例启动模板

实例启动模板是云服务器 ECS 中一项持久化 ECS 实例配置的功能，您可以直接使用现有实例启动模板配置伸缩组。更多信息，请参阅 [实例启动模板](#)。



说明：

如果您选择使用实例启动模板，可以在创建伸缩组后直接启用伸缩组。

自定义伸缩配置 vs. 实例启动模板

对比项	自定义伸缩配置	实例启动模板
伸缩组支持的参数	伸缩组支持所有自定义伸缩配置的参数。	伸缩组对实例启动模板的部分参数尚不支持，使用启动模板创建出来的实例可能缺失模板的部分配置信息。
参数校验	支持。缺少镜像等必要参数时无法创建自定义伸缩配置，因此在使用时不会由于缺少必要参数导致创建 ECS 实例失败。	实例启动模板不对参数做额外的校验，所有参数都是可选的。因此，如果指定模板中未包含镜像等必要参数，会导致无法使用该模板成功地创建实例。
配置顺序	在伸缩组内配置，您需要先创建伸缩组。	云服务器 ECS 中的一项功能，不依托伸缩组。
功能入口	只能在伸缩组内创建。	可以通过 ECS 售卖页、ECS 启动模板页和现有 ECS 实例设置页等方式创建。
演进方式	对单个自定义伸缩配置，您只能手动修改，且修改不可追溯。您可以针对不同需求创建多个自定义伸缩配置。	不支持修改，但可以创建新的版本，通过版本管理体现演进过程，更加清晰快捷。
多实例规格	支持。适用于只关注性能，而不限定于某一种特定规格的情况，弹性扩张的成功率更高。	不支持。

具体操作步骤，请参阅 [使用自定义伸缩配置创建伸缩组](#) 和 [使用实例启动模板创建伸缩组](#)。

2.2 创建伸缩配置

本文介绍如何为伸缩组创建伸缩配置，指定弹性扩张时的ECS实例模板。

背景信息

创建伸缩配置的流程和创建ECS实例类似，但由于伸缩配置是伸缩组弹性扩张时的ECS实例模板，因此具有个性化的功能，例如多实例规格，不过不支持配置某些项目，例如地域、资源组。实际操作时请以弹性伸缩控制台页面为准。页面上的各配置项均配有简洁说明，如果您想深入了解各配置项的含义，请参见[使用向导创建实例](#)。

准备工作

- 如果您在控制台[创建伸缩组](#)过程中创建伸缩配置，请从[步骤 3](#)开始阅读。

- 伸缩组创建完成后，专有网络不可修改。在创建伸缩配置时需要指定一个安全组，请确保该安全组和伸缩组所在的专有网络一致。如果伸缩组所在专有网络下无安全组，您需要[创建安全组](#)。
- 如果您需要使用本文中的API示例，请[设置API访问权限](#)。

控制台操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组的操作列下单击管理。
3. 前往组内实例配置信息来源页面，单击创建伸缩配置。
4. 在基础配置页面中，配置计费方式、实例、镜像、存储、公网带宽和安全组，然后单击下一步：系统配置。



说明：

在伸缩配置中：

- 计费方式：仅支持[按量付费实例](#)和[抢占式实例](#)。
- 实例：支持多个实例规格，当特定实例规格库存不足时，会切换到至备选实例规格，提高弹性扩张的成功率。

5. 在系统配置页面中，配置标签（可选）、登录凭证、实例名称（可选）和高级选项（可选），然后单击下一步：确认配置。



说明：

仅专有网络下伸缩组的伸缩配置中支持高级选项，包括实例RAM角色和实例自定义数据。

6. 在确认配置页面中，检查所选配置，填写伸缩配置名称，然后单击确认创建。
7. 在创建成功对话框中，单击启用配置。

API使用示例

伸缩配置是伸缩组用于弹性创建ECS实例的模板，如果您需要调用API创建伸缩配置，请确保请求中包含伸缩组ID、ECS实例所属安全组的ID、ECS实例镜像ID和ECS实例规格。

推荐根据开发平台选用对应语言的[阿里云SDK](#)，调用API时更加方便快捷。以Java SDK为例：

```
import com.aliyuncs.CommonRequest;
import com.aliyuncs.CommonResponse;
import com.aliyuncs.DefaultAcsClient;
import com.aliyuncs.IAcsClient;
import com.aliyuncs.exceptions.ClientException;
import com.aliyuncs.exceptions.ServerException;
import com.aliyuncs.http.MethodType;
import com.aliyuncs.profile.DefaultProfile;
/*
pom.xml
<dependency>
  <groupId>com.aliyun</groupId>
```

```
<artifactId>aliyun-java-sdk-core</artifactId>
<version>4.0.3</version>
</dependency>
*/
public class CommonRpc {
    public static void main(String[] args) {
        DefaultProfile profile = DefaultProfile.getProfile("cn-
hangzhou", "<accessKeyId>", "<accessSecret>");
        IAcsClient client = new DefaultAcsClient(profile);

        CommonRequest request = new CommonRequest();
        request.setMethod(MethodType.POST);
        request.setDomain("ess.aliyuncs.com");
        request.setVersion("2014-08-28");
        request.setAction("CreateScalingConfiguration");
        request.putQueryParameter("ScalingGroupId", "asg-bp1a4xzjr1
ypd601****");
        request.putQueryParameter("SecurityGroupId", "sg-bp147qpndp
7iyj08****");
        request.putQueryParameter("ImageId", "centos6u5_64_20G_aliaeg
****.vhd");
        request.putQueryParameter("InstanceType", "ecs.t1.xsmall");
        try {
            CommonResponse response = client.getCommonResponse(request
);
            System.out.println(response.getData());
        } catch (ServerException e) {
            e.printStackTrace();
        } catch (ClientException e) {
            e.printStackTrace();
        }
    }
}
```

通过Java SDK或者其它方式发起调用后，请求正文可能如下所示：

```
https://ess.aliyuncs.com/?Action=CreateScalingConfiguration
&ImageId=centos6u5_64_20G_aliaeg****.vhd
&InstanceType=ecs.t1.xsmall
&ScalingGroupId=asg-bp1a4xzjr1ypd601****
&SecurityGroupId=sg-bp147qpndp7iyj08****
&Version=2014-08-28
```

其中：

- centos6u5_64_20G_aliaeg****.vhd为ECS实例镜像ID。
- ecs.t1.xsmall为ECS实例规格。
- AG6CQdPU80KdwLjgZc****为伸缩组ID。
- sg-280****为ECS实例所属安全组ID。
- 2014-08-28为API版本。

您还可以个性化设置多实例规格、ECS实例磁盘等相关属性，详细的API说明，请参见[CreateScalingConfiguration](#)。

2.3 选用伸缩配置

在伸缩组中，您可以创建多个伸缩配置，但只能选用一个伸缩配置。

背景信息

伸缩配置包含的配置项较多，您可能会忘记具体的配置项。因此弹性伸缩还提供了查看详情功能，供您了解各个伸缩配置，确保选用合适的ECS实例模板。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 单击伸缩组操作列下的管理。
3. 前往组内实例配置信息来源页面。
4. 在伸缩配置页面中，单击待选用伸缩配置操作列下的查看详情。
5. 如果确定选用该伸缩配置，单击操作列下选用。



说明：

选用一个伸缩配置后，其他伸缩配置将被置于失效状态。

预期结果

选用一个伸缩配置后，在满足弹性扩张条件时，伸缩组会以该伸缩配置为模板自动创建ECS实例。

2.4 修改伸缩配置

如果业务需求发生变更，您可以修改现有伸缩配置的部分项，无须重新创建一个伸缩配置，提高运维效率。

背景信息

如果您想深入了解各配置项的含义，请参阅[使用向导创建实例](#)。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 单击伸缩组操作列下的管理。
3. 前往组内实例配置信息来源页面。
4. 单击待修改伸缩配置操作列下的修改。
5. 在基础配置页面中，根据需要修改配置项，然后单击下一步：系统配置。
6. 在系统配置页面中，根据需要修改配置项，然后单击下一步：确认配置。

7. 在确认配置页面中，单击确认修改。



说明：

修改一个伸缩配置后，如果存在以该伸缩配置为模板自动创建的ECS实例，这些ECS实例仍然可以正常运行。

2.5 删除伸缩配置

您可以删除不再需要的伸缩配置，空出伸缩配置的配额。

前提条件

在删除一个伸缩配置前，请确保满足以下条件，否则删除操作会失败：

- 该伸缩配置必须处于失效状态。
- 伸缩组中不存在根据该伸缩配置自动创建的ECS实例。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 单击伸缩组操作列下的管理。
3. 前往组内实例配置信息来源页面。

4. 在伸缩配置页面中，单击待删除伸缩配置操作列下的删除。



如果您需要删除多个伸缩配置，也可以勾选待删除伸缩配置左侧的复选框，然后单击伸缩配置名称/ID列下的删除。



5. 在删除伸缩配置对话框中，单击确认。

2.6 导出伸缩配置

您可以将伸缩组下的伸缩配置导出为.csv文件，用于本地备份或者快速导入其它伸缩组中。

操作步骤

1. 登录**弹性伸缩控制台**。
2. 单击伸缩组操作列下的**管理**。
3. 前往组内实例配置信息来源页面。

4. 在伸缩配置页面中，单击导出伸缩配置，下载并保存.csv文件到本地。



2.7 导入伸缩配置

您可以将伸缩配置文件导入到其它伸缩组中，提高创建伸缩配置的效率，但伸缩组的网络类型必须一致。

操作步骤

- 1. 登录[弹性伸缩控制台](#)。
- 2. 单击伸缩组操作列下的管理。
- 3. 前往组内实例配置信息来源页面。
- 4. 在伸缩配置页面中，单击导入伸缩配置。



5. 单击选择文件，浏览并选择待导入的.csv文件。

6. 在导入预览中，勾选需要导入的伸缩配置，然后单击开始导入。

导入伸缩配置

选择文件

导入的配置数目少于10条，导入为非覆盖式导入，且需确保源数据所属伸缩组与当前伸缩组的网络相关配置尽量相同（即网络类型及专有网络、交换机等设置），否则可能使得导入的配置无法正常伸缩实例

选择文件..

导入预览

☐	检查	伸缩配置	标签	实例规格组	状态	镜像	宽带计费	系统盘类型	数据盘	登录凭证
☒	1			ecs.g5.large (2vcpu 8GB)		centos_7_06_64_20G_a...	流量计费	高效云盘	-	未设置
☒	2			ecs.sn2.xlarge (16vcpu 64GB)		debian_9_02_64_20G_a...	流量计费	高效云盘	-	未设置
☐	3			ecs.sn2.large (4vcpu 16GB)		centos_7_03_64_20G_a...	流量计费	高效云盘	-	未设置

已选择导入 2 条记录

☐ 为导入配置的名称添加统一后缀



说明:

- 您可以为导入的伸缩配置统一添加名称后缀，避免与伸缩组现有伸缩配置的名称重复。
- 如果您无法勾选导入预览中的伸缩配置，请检查导出该伸缩配置时所属伸缩组的网络类型，可能与当前伸缩组的网络类型不一致。

7. 查看导入伸缩配置的结果，单击确定。

3 实现自动伸缩

3.1 使用自定义伸缩配置创建伸缩组

在使用弹性伸缩调整弹性计算资源大小之前，您首先需要创建一个伸缩组。

伸缩组是具有相同应用场景的ECS实例的集合。您可以在伸缩组内定义最大实例数、最小实例数和冷却时间等属性，还可以为这些ECS实例关联负载均衡和数据库实例，便于进行综合管理。



说明：

一个账号能创建的伸缩组数量有上限。更多信息，请参阅[使用限制](#)。

如果您希望使用现有的实例启动模板，请参阅[使用实例启动模板创建伸缩组](#)。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组管理页面中，单击创建伸缩组。
3. 设置伸缩组属性。
 - a. 设置伸缩组名称。
 - b. 设置伸缩最大实例数（台）。



说明：

超过上限时，弹性伸缩会自动移出ECS实例，使得伸缩组的当前ECS实例数等于上限。

- c. 设置伸缩最小实例数（台）。



说明：

低于下限时，弹性伸缩会自动添加ECS实例，使得伸缩组的当前ECS实例数等于下限。

- d. 设置默认冷却时间（秒）。



说明：

伸缩组发生伸缩活动后的默认冷却时间。更多信息，请参阅[冷却时间](#)。

- e. 设置移出策略。



说明：

伸缩组内ECS实例数超过上限时，移出ECS实例的策略。更多信息，请参阅[移出策略](#)。

- f. 设置组内实例配置信息来源，选择自定义伸缩配置。
- g. 设置网络类型，如果您选择专有网络，还需要指定以下信息：

A. 指定专有网络和虚拟交换机。

B. 指定多可用区扩缩容策略，具体的策略说明，请参阅[多可用区扩缩容策略](#)。

C. 指定回收模式，具体的模式说明，请参阅[回收模式](#)。

* 网络类型：☐ 经典网络 ☒ 专有网络 专有网络 ☒ 伸缩组可支持选择多个虚拟交换机

* 专有网络： 专有网络ID： [创建专有网络](#)

虚拟交换机：

多可用区扩缩容策略：☒ 优先级策略 ☐ 均衡分布策略 ☐ 成本优化策略

回收模式：☒ 释放模式 ☐ 停机回收模式

- h. (可选) 设置负载均衡。



说明：

一个伸缩组最多同时关联5个负载均衡实例。伸缩组还支持选择一个负载均衡实例下的[默认服务器组](#)和[虚拟服务器组](#)，一个伸缩组最多支持同时选择5个虚拟服务器组。更多信息，请参阅[在弹性伸缩中使用负载均衡服务](#)。

负载均衡： [管理我的负载均衡](#)

只有配置过监听的负载均衡才能被伸缩组使用

负载均衡配置详情

伸缩组下负载均衡最多可配置 1 / 5 [滚动查看全部](#)

负载均衡 ID：

负载均衡名称：

服务器组	端口(1-65535)	权重(1-100)	
默认服务器组	-	在伸缩配置里指定	×
☒			×

+ 默认服务器组 + 虚拟服务器组

伸缩组下负载均衡虚拟服务器组最多可配置 1 / 5

- i. (可选) 设置数据库，目前支持RDS。



说明：

只能添加和伸缩组位于同一地域的RDS实例，伸缩组弹性扩张出ECS实例后，会自动将ECS实例的内网IP添加到RDS实例的访问白名单，允许两者内网通信。

- 4. 单击提交。

5. 在伸缩组创建成功对话框中，单击创建伸缩配置，继续创建用于弹性扩张的ECS实例模板。



说明：

有关伸缩配置，请参阅[创建伸缩配置](#)。

6. 确定伸缩配置后，在启用伸缩组对话框中，单击确定。

多可用区扩缩容策略

策略名称	说明
优先级策略	根据您定义的虚拟交换机扩缩容。当优先级较高的虚拟交换机所在可用区无法创建ECS实例时，自动使用下一优先级的虚拟交换机创建ECS实例。
均衡分布策略	在伸缩组指定的多可用区（即指定多个专有网络交换机）之间均匀分配ECS实例。如果由于库存不足等原因可用区之间变得不平衡，您可以进行再均衡操作来平衡资源的可用区分布。  说明： 当您设置了多个专有网络交换机时，该策略才能生效。

策略名称	说明
成本优化策略	伸缩组网络类型为专有网络： · 当伸缩配置选择抢占式实例时，可以通过成本优化策略来保证业务的稳定性。 · 当伸缩配置选择多实例规格时，可以通过成本优化策略来降低ECS实例使用成本。成本优化策略按vCPU单价从低到高进行尝试创建。 · 当伸缩配置设置了抢占式计费方式的多实例规格时，优先创建对应抢占式计费实例。 · 当抢占式计费实例规格由于库存等原因都无法创建时，自动尝试以按量付费的方式创建。

回收模式

模式名称	说明
释放模式	在弹性收缩时，根据定时或报警任务自动释放合适数量的ECS实例。 在弹性扩张时，根据定时或报警任务创建新的ECS实例加入伸缩组。

模式名称	说明
停机回收模式	停机回收模式可以提高扩缩容的时间效率。在此模式下： - 当弹性收缩时，自动创建的ECS实例将进入停止状态，此状态下实例的CPU和内存不收费，云盘（包括系统盘、数据盘）、弹性公网IP和带宽仍旧收费，公网IP将被回收，待启动时重新分配（弹性公网IP仍旧保留），这些停机的实例形成了停机实例池。 - 当弹性扩张时，停机实例池内的实例会优先变为运行状态，如果停机实例池数量不足则会重新启动实例。  说明： - 只有VPC实例的伸缩组才能开启此模式。 - 所有本地盘实例（包括但不限于d1、d1ne、ga1、gn5、i1和i2）不支持此模式。 - 扩容时，停机实例池的实例不能保证成功进入运行状态。如果由于库存等其他原因，停止的实例不能正常进入运行状态，该实例会被释放，并重新创建实例，保证伸缩规则的执行结果达到预期。 - 伸缩组设置为停机回收模式后，暂不支持修改。

3.2 使用实例启动模板创建伸缩组

在使用弹性伸缩调整弹性计算资源大小之前，您首先需要创建一个伸缩组。

伸缩组是具有相同应用场景的ECS实例的集合。您可以在伸缩组内定义最大实例数、最小实例数和冷却时间等属性，还可以为这些ECS实例关联负载均衡和数据库实例，便于进行综合管理。



说明：

一个账号能创建的伸缩组数量有上限。更多信息，请参阅[使用限制](#)。

如果您希望使用自定义伸缩配置，请参阅[使用自定义伸缩配置创建伸缩组](#)。

操作步骤

1. 登录[弹性伸缩控制台](#)。

2. 在伸缩组管理页面中，单击创建伸缩组。



3. 设置伸缩组属性。

a. 设置伸缩组名称。

b. 设置伸缩最大实例数（台）。



说明：

超过上限时，弹性伸缩会自动移出ECS实例，使得伸缩组的当前ECS实例数等于上限。

c. 设置伸缩最小实例数（台）。



说明：

低于下限时，弹性伸缩会自动添加ECS实例，使得伸缩组的当前ECS实例数等于下限。

d. 设置默认冷却时间（秒）。



说明：

伸缩组发生伸缩活动后的默认冷却时间。更多信息，请参阅[冷却时间](#)。

e. 设置移出策略。



说明：

伸缩组内ECS实例数超过上限时，移出ECS实例的策略。更多信息，请参阅[移出策略](#)。

f. 设置组内实例配置信息来源，选择启动模板。



说明：

启动模板通过版本管理，请参阅[启动模板版本策略](#)。

* 组内实例配置信息来源 ? : ☐ 自定义伸缩配置 ☒ 启动模板

* 选择启动模板 :

* 选择启动模板版本 : ☐ 始终使用默认版本 ☒ 始终使用最新版本 ☐ 使用自定义版本

g. 设置网络类型，如果您选择专有网络，还需要指定以下信息：

A. 指定专有网络和虚拟交换机。

B. 指定多可用区扩缩容策略，具体的策略说明，请参阅[多可用区扩缩容策略](#)。

C. 指定回收模式，具体的模式说明，请参阅[回收模式](#)。

* 网络类型 : ☐ 经典网络 ☒ 专有网络 专有网络 ? 伸缩组可支持选择多个虚拟交换机

* 专有网络 : 专有网络ID : [创建专有网络](#)

虚拟交换机 :

多可用区扩缩容策略 ? : ☒ 优先级策略 ☐ 均衡分布策略 ☐ 成本优化策略

回收模式 ? : ☒ 释放模式 ☐ 停机回收模式

h. (可选) 设置负载均衡。



说明：

一个伸缩组最多同时关联5个负载均衡实例。伸缩组还支持选择一个负载均衡实例下的[默认服务器组](#)和[虚拟服务器组](#)，一个伸缩组最多支持同时选择负载均衡实例下的5个虚拟服务器组。更多信息，请参阅[在弹性伸缩中使用负载均衡服务](#)。

负载均衡 ? : [管理我的负载均衡](#)

只有配置过监听的负载均衡才能被伸缩组使用

负载均衡配置详情

伸缩组下负载均衡最多可配置 1 / 5 [滚动查看全部](#)

负载均衡 ID :

负载均衡名称 :

服务器组	端口(1-65535)	权重(1-100)	
默认服务器组 ?	-	在伸缩配置里指定	×
☒			×

+ 默认服务器组 + 虚拟服务器组

伸缩组下负载均衡虚拟服务器组最多可配置 1 / 5

i. (可选) 设置数据库，目前支持RDS。



说明：

只能添加和伸缩组位于同一地域的RDS实例，伸缩组弹性扩张出ECS实例后，会自动将ECS实例的内网IP添加到RDS实例的访问白名单，允许两者内网通信。

- 4. 单击提交。
- 5. 在启用伸缩组对话框中，单击确定。

 说明:
伸缩组创建成功后，不支持修改网络配置信息（专有网络和虚拟交换机）。在修改启动模板版本时，请确保网络配置信息一致，否则会操作失败。



启动模板版本策略

版本管理策略	说明
始终使用默认版本	伸缩组将始终使用实例启动模板的默认版本来创建ECS实例。
始终使用最新版本	伸缩组将始终使用实例启动模板的最新版本来创建ECS实例。
使用自定义版本	伸缩组将使用指定的实例启动模板版本来创建ECS实例。  说明: 选择该策略后，伸缩组会根据启动模板中的网络信息自动规划网络配置。

多可用区扩缩容策略

策略名称	说明
优先级策略	根据您定义的虚拟交换机扩缩容。当优先级较高的虚拟交换机所在可用区无法创建ECS实例时，自动使用下一优先级的虚拟交换机创建ECS实例。
均衡分布策略	在伸缩组指定的多可用区（即指定多个专有网络交换机）之间均匀分配ECS实例。如果由于库存不足等原因可用区之间变得不平衡，您可以进行再均衡操作来平衡资源的可用区分布。  说明： 当您设置了多个专有网络交换机时，该策略才能生效。
成本优化策略	伸缩组网络类型为专有网络： · 当伸缩配置选择抢占式实例时，可以通过成本优化策略来保证业务的稳定性。 · 当伸缩配置选择多实例规格时，可以通过成本优化策略来降低ECS实例使用成本。成本优化策略按vCPU单价从低到高进行尝试创建。 · 当伸缩配置设置了抢占式计费方式的多实例规格时，优先创建对应抢占式计费实例。 · 当抢占式计费实例规格由于库存等原因都无法创建时，自动尝试以按量付费的方式创建。

回收模式

模式名称	说明
释放模式	在弹性收缩时，根据定时或报警任务自动释放合适数量的ECS实例。 在弹性扩张时，根据定时或报警任务创建新的ECS实例加入伸缩组。

模式名称	说明
停机回收模式	停机回收模式可以提高扩缩容的时间效率。在此模式下： - 当弹性收缩时，自动创建的ECS实例将进入停止状态，此状态下实例的CPU和内存不收费，云盘（包括系统盘、数据盘）、弹性公网IP和带宽仍旧收费，公网IP将被回收，待启动时重新分配（弹性公网IP仍旧保留），这些停机的实例形成了停机实例池。 - 当弹性扩张时，停机实例池内的实例会优先变为运行状态，如果停机实例池数量不足则会重新启动实例。  说明： - 只有VPC实例的伸缩组才能开启此模式。 - 所有本地盘实例（包括但不限于d1、d1ne、gal、gn5、i1和i2）不支持此模式。 - 扩容时，停机实例池的实例不能保证成功进入运行状态。如果由于库存等其他原因，停止的实例不能正常进入运行状态，该实例会被释放，并重新创建实例，保证伸缩规则的执行结果达到预期。 - 伸缩组设置为停机回收模式后，暂不支持修改。

3.3 创建伸缩规则

创建伸缩组后，您还需要在伸缩组内创建伸缩规则，用于管理伸缩组的具体伸缩动作。本文介绍伸缩规则的创建步骤。

限制条件

- 一个伸缩组内可以创建的伸缩规则数量有限，请参见[数量限制](#)。
- 目标追踪规则只能由自动创建的关联报警任务执行。

- 如果执行伸缩规则后，伸缩组内服务中实例数将大于伸缩最大实例数或者小于伸缩最小实例数，弹性伸缩服务会自动调整扩张或收缩的ECS实例数量，保证弹性伸缩结果不会超过数量限制。例如：
 - 您拥有一个伸缩组asg-bp19ik2u5w7esjcu****，伸缩最大实例数为3，伸缩规则add3为增加3台实例。如果当前服务中实例数为2，执行伸缩规则add3时，仅会增加1台ECS实例。
 - 您拥有一个伸缩组asg-bp19ik2u5w7esjcu****，伸缩最小实例数为2，伸缩规则reduce2为减少2台实例。如果当前服务中实例数为3，执行伸缩规则reduce2时，仅会减少1台ECS实例。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在指定伸缩组操作列下，单击管理。
3. 前往伸缩规则页面，单击创建伸缩规则。
4. 配置伸缩规则参数，然后单击创建伸缩规则。

请根据需要选择伸缩规则类型，参数说明请参见[简单规则](#)和[目标追踪规则](#)。



说明：

如果创建了目标追踪规则，会自动创建一项关联报警任务，只能由该报警任务执行目标追踪规则。

伸缩规则	伸缩规则类型	执行的时间	执行的操作	操作
	目标追踪规则	是将 平均CPU使用率 维持在 80.000% 所必需的 每次执行先新增的实例需要 300 秒预热 已关联的报警任务： TargetTracking-asr-b...	根据需要添加实例,而不减少实例	查看详情 修改 删除

简单规则

简单规则直接指定增加/减少实例的数量，或者将实例数量调整至指定值。简单规则参数如下：

名称	说明
规则名称	伸缩规则的名称。
伸缩规则类型	伸缩规则的类型，创建成功后不可修改。

名称	说明
执行的操作	视需要选择一种操作： · 调整至N台：执行伸缩规则后，伸缩组内实例的数量变为N台。最多能触发伸缩500台实例。 · 增加N台：执行伸缩规则后，伸缩组内实例的数量在当前基础上增加N台。增加的数量不可超过500。 · 增加N%：执行伸缩规则后，伸缩组内实例的数量在当前基础上增加N%。最多能触发伸缩500台实例。 · 减少N台：执行伸缩规则后，伸缩组内实例的数量在当前基础上减少N台。减少的数量不可超过500。 · 减少N%：执行伸缩规则后，伸缩组内实例的数量在当前基础上减少N%。最多能触发伸缩500台实例。
冷却时间	可选项，单位为秒。不填写则默认使用伸缩组的冷却时间，更多信息请参见 冷却时间 。

目标追踪规则

目标追踪规则指定一项云监控指标的目标值，弹性伸缩会自动计算所需的实例数量并进行扩缩容，从而将云监控指标维持在目标值附近。目标追踪规则的参数如下：

名称	说明
规则名称	伸缩规则的名称。
伸缩规则类型	伸缩规则的类型，创建成功后不可修改。
指标类型	视需要选择一项云监控指标： · 平均CPU使用率 · 内网入流量的平均值 · 内网出流量的平均值 · 公网入流量的平均值 · 公网出流量的平均值
目标值	云监控指标的目标值，目标追踪规则会将云监控指标值维持在目标值附近。
实例预热时间	单位为秒。在预热时间内，执行目标伸缩规则创建成功实例后，已启动的实例不会对云监控指标值产生影响，以免云监控指标值反复变化。
禁用缩容	启用后，目标伸缩规则不会减少伸缩组内实例的数量。

3.4 创建生命周期挂钩

本文主要介绍生命周期挂钩的定义和创建。

在 [执行伸缩规则](#) 中，您已经成功实现弹性扩张或收缩 ECS 资源。但是，这些 ECS 资源只进行了简单的初始配置，面对复杂的业务情况，您可能希望在启用这些 ECS 资源前进行自定义操作，这时就需要使用生命周期挂钩。

什么是生命周期挂钩

如果您为伸缩组创建了生命周期挂钩，在伸缩组进行伸缩活动时，正在加入或正在移出伸缩组的实例将被挂钩挂起并置于等待状态。生命周期挂钩仅在自动创建或移出 ECS 实例时生效，手动添加或移出 ECS 实例时不受其影响。



说明：

一个伸缩组内可以创建的生命周期挂钩数量有限，请参阅 [使用限制](#)。

示例

您拥有一个伸缩组 sg-yk201808201449，伸缩最小实例数为 0，伸缩组内有 1 个适用于弹性扩张活动的生命周期挂钩，目前伸缩组内没有 ECS 实例。

将伸缩最小实例数修改为 1，由于 ECS 实例数没有达到下限，弹性扩张活动被触发，伸缩组自动创建一台 ECS 实例。但是，由于存在生命周期挂钩，该 ECS 实例不会马上进入 `服务中` 状态，而是进入 `加入挂起中` 状态。

在超时时间结束前，您可以登录该 ECS 实例，进行预装软件或其他自定义操作。

功能特性

在实例保持等待状态的时间内，当前伸缩组将具有以下特性：

- 伸缩组不再运行其它的伸缩活动。
- 为您保留指定时长的操作时间（即挂钩的超时时间），您可以在挂起期间执行自定义操作，例如，初始化 ECS 实例配置或者获取 ECS 实例数据。
- 您可以删除生命周期挂钩来恢复执行伸缩活动。
- 您可以调用相关 API 来 [结束生命周期活动](#) 或者 [删除生命周期挂钩](#)。

操作步骤

请参照以下步骤创建一个生命周期挂钩：

1. 登录 [弹性伸缩控制台](#)。
2. 在 `伸缩组管理` 页面中，单击指定伸缩组 `操作` 列下的 `管理`。

- 3. 前往 生命周期挂钩 页面，单击 创建生命周期挂钩。
- 4. 在 创建生命周期挂钩 对话框中，依次指定名称、适用的伸缩活动类型、超时时间、执行策略、通知方式、MNS 主题和通知标识，然后单击 创建生命周期挂钩。

创建生命周期挂钩

名称：

名称为2-40个字符，以大小写字母，数字或中文开头，可包含".","_"或"-"

*适用的伸缩活动类型：

☐ 弹性收缩活动

☒ 弹性扩张活动

超时时间（秒）：

3600

最小为30，最大为21600，必须为整数

执行策略：

☒ 继续

☐ 拒绝

?

通知方式：

☐ MNS主题

☒ MNS队列

MNS队列：

通知标识 ?：

通知标识的长度不超过 128 个字符

创建生命周期挂钩

取消

生命周期挂钩属性

以下表格描述了具体的生命周期挂钩属性含义和示例。

名称	说明	示例
名称	生命周期挂钩名称（2-40个字符），以大小写字母、数字或中文开头，可包含英文句号（.）、下划线（_）或短横线（-）。挂钩名字一旦确定，不能修改。	hz_yk0626
适用的伸缩活动类型	适用的伸缩活动类型。	弹性收缩活动

名称	说明	示例
超时时间	默认情况下，实例保持等待状态的时间。最小为 30 秒，最大为 21600 秒（6 小时），必须为整数。	600
执行策略	执行策略包括 继续 和 拒绝。 · 继续：继续执行弹性扩张活动或者弹性收缩活动。 · 拒绝：针对弹性扩张活动，会直接释放创建出来的 ECS 实例；针对弹性收缩活动，直接移出 ECS 实例。	继续
通知方式	通知方式包括 MNS 主题 和 MNS 队列。选择通知方式后，您还需要选择具体的 MNS 主题或者 MNS 队列。 如果尚未创建 MNS 主题或者 MNS 队列，请 创建主题 或者 创建队列 。	MNS 主题
通知标识	弹性伸缩每次推送消息到通知对象时，会同时发送您预先指定的通知标识，便于管理和标记不同类别的通知信息。	General information

3.5 创建预测规则

预测规则基于机器学习，可以通过分析伸缩组的历史监控数据预测未来监控指标值，并支持自动创建定时任务，智能设置伸缩组边界值。本文介绍如何创建预测规则。

背景信息

在创建伸缩组时，您可以设置伸缩组边界值，即伸缩最大实例数和伸缩最小实例数。但如果您并不了解业务运行情况，设置的伸缩组边界值可能与实际需求存在偏差。伸缩最小实例数过高可能会导致购入过多计算资源，伸缩最大实例数过低又可能导致计算资源不足，影响服务稳定。

预测规则会获取至少24小时的历史监控数据，利用机器学习能力预测未来48小时的监控指标值，然后计算出伸缩组每小时需要的实例数，即预测值。预测结果每天更新一次，并为未来48小时创建48个预测任务。预测任务执行后会修改伸缩组边界值，而非组内实例数量。

预测规则可以与目标追踪规则和简单规则同时使用。当与目标追踪规则同时使用时，推荐设置相同的指标类型和目标值，防止因为指标差异造成伸缩组实例数抖动。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在指定伸缩组操作列下，单击管理。
3. 前往伸缩规则页面，单击创建伸缩规则。
4. 设置伸缩规则参数，然后单击创建伸缩规则。

一个伸缩组内只能创建一条预测规则。预测规则的参数如下：

参数	说明
规则名称	伸缩规则的名称。
伸缩规则类型	伸缩规则的类型，选择预测规则。
是否配合已有的目标追踪规则	和选择规则结合使用。
选择规则	勾选是否配合已有的目标追踪规则后出现，您可以继续选择一条目标追踪规则，自动沿用目标追踪规则的指标类型和目标值。
指标类型	视需要选择一项云监控指标： · 平均CPU使用率（%） · 内网入流量的平均值（KB/Min） · 内网出流量的平均值（KB/Min）
目标值	所选云监控指标的目标值，预测规则将综合目标值等因素计算出合适的预测值。 如果您修改了目标值，伸缩组会清空现存的预测任务，并在一小时内创建新的预测任务。
预测模式	视需要选择预测模式： · 预测并伸缩：产生预测结果并创建预测任务。 · 只预测不伸缩：产生预测结果，但不会创建预测任务。 建议您先选择只预测不伸缩，如果预测效果符合预期，再修改为预测并伸缩。您可以在伸缩规则详情中 查看预测规则效果 。
初始最大值	伸缩组实例数上限，和最大值处理方式结合使用。 默认值为当前的伸缩最大实例数。

参数	说明
最大值处理方式	当预测值大于初始最大值时，预测值的处理方式分为以下三种： · 预测值会覆盖初始最大值：预测值大于初始最大值时，预测任务的最大值采用预测值。 · 初始最大值会覆盖预测值：预测值大于初始最大值时，预测任务的最大值采用初始最大值。 · 预测值会附加一定比例：和增加比例结合使用。 默认值为初始最大值会覆盖预测值。
增加比例	最大值处理方式为预测值会附加一定比例时生效，预测值会按照该比例增加，当增加后的值大于初始最大值时，会采用增加后的值。 默认值为0，最大值为100。
预启动时间（分钟）	预测规则自动创建的预测任务默认均在整点执行，您可以设置预启动时间提前执行预测任务，预先准备资源。 默认值为0，最大值为60。

3.6 执行伸缩规则

本文为您介绍如何执行伸缩规则，实现手动或自动伸缩ECS实例资源。

前提条件

执行伸缩规则前，请注意以下前提条件：

- 伸缩规则所在的伸缩组的状态必须为启用。
- 伸缩规则所在的伸缩组内不能存在执行中的伸缩活动。
- 目标追踪规则只能由自动创建的关联报警任务执行，更多信息请参见[创建伸缩规则](#)。
- 伸缩组内不限制ECS实例的数量，但是您需要遵守[ECS实例的使用限制](#)。

手动执行

如果您临时需要伸缩ECS实例资源，可以直接手动执行伸缩规则。



说明：

当伸缩组内不存在执行中的伸缩活动时，手动执行可以绕过[冷却时间](#)。

1. 在伸缩规则页面中，单击待执行伸缩规则操作列的执行。
2. 在执行伸缩规则对话框中，单击确定。

3. 如果执行成功，页面右上角会出现成功提示。



如果执行失败，会弹出错误提示。



4. 您可以前往伸缩活动页面，查看具体的执行信息。

通过定时任务执行

针对ECS实例资源需求量规律的业务，您可以选用定时任务，在**创建定时任务**时指定一条伸缩规则，弹性伸缩服务会在指定的时间点自动执行该伸缩规则。

创建定时任务

*任务名称：

名称为2-40个字符，以大小写字母，数字或中文开头，可包含"，"，"_"或"-"

描述：

最少需要2个字符

*执行时间？：

2019-01-04

16

:

23

*伸缩规则？：

伸缩组：

伸缩规则：

重试过期时间(秒)？：

600

☐ 重复周期设置（高级）

38

文档版本：20190813

通过报警任务执行

如果业务的ECS资源需求量并没有明显的规律，您可以选用基于监控项的报警任务，在[创建报警任务](#)时指定一条伸缩规则，弹性伸缩服务会在满足报警条件时自动执行该伸缩规则。

报警任务包括系统监控报警任务和自定义监控报警任务，满足不同场景下的监控需求。详细信息请参见[弹性伸缩报警任务概述](#)。

创建报警任务

在使用报警任务之前，需要在ECS的镜像里安装新版本的云监控Agent。 [查看帮助文档](#)

*任务名称：

名称为2-40个字符，以大小写字母，数字或中文开头，可包含"，"_"或"-"

描述：

最少需要2个字符

*监控资源：

监控类型

☒ 系统监控 ☐ 自定义监控 ?

*监控项：

CPU

统计周期(分钟) ?：

1

*统计办法 ?：

平均值

>=

阈值

0

%

重复几次后报警 ?：

3次

*报警触发规则 ?：

创建报警任务

在使用报警任务之前，需要在ECS的镜像里安装新版本的云监控Agent。 [查看帮助文档](#)

*任务名称：

名称为2-40个字符，以大小写字母，数字或中文开头，可包含".", "_"或"-"

描述：

最少需要2个字符

*监控资源：

监控类型

☐ 系统监控

☒ 自定义监控 ?

*应用分组：

*监控项：

纬度：

统计周期(分钟) ?：

1

*统计办法 ?：

平均值

>=

阈值

0

重复几次后报警 ?：

3次

*报警触发规则 ?：

3.7 定时任务

3.7.1 创建定时任务

本文主要介绍定时任务的定义和创建。

如果您可以准确预料业务量的变化时间，就可以提前添加定时任务，在业务高峰到达前准备好足量的计算资源，或者在业务高峰后释放空闲的计算资源。

什么是定时任务

定时任务是在指定时间执行指定伸缩规则的一种预设任务，从而自动扩张或者收缩计算资源，满足业务需求的同时又能控制成本。您还可以为定时任务指定重复周期，通过灵活的规则应对业务量的变化。

**说明:**

一个账号下可以创建的定时任务数量有限，请参阅 [使用限制](#)。

由于在同一时刻一个伸缩组中只能存在一个伸缩活动，定时任务还提供了自动重试时间设置功能，避免因单次执行伸缩规则失败影响定时任务的结果。如果同一分钟内存在多个待执行的定时任务，弹性伸缩会执行最新创建的定时任务。

操作步骤

请参照以下步骤创建一个定时任务：

1. 登录 [弹性伸缩控制台](#)。
2. 前往 自动触发任务管理 下的 定时任务 页面，单击 创建定时任务。



3. 在 **创建定时任务** 对话框中，依次指定任务名称、执行时间、伸缩规则、重试过期时间（可选）和重复周期（可选），您还可以添加描述以便日后查看，然后单击 **提交**。

创建定时任务

*任务名称：

st-yk

名称为 2-64 个字符，以小写字母、数字或中文开头，可包含 "."、"_" 或 "-"。

描述：

月初访问量最大，增加 3 台实例。

最少需要 2 个字符

*执行时间：

2018-09-02 00:00

*伸缩规则：

伸缩组：sg-yk

伸缩规则：add3

重试过期时间(秒)：

600

*重复周期：

☒ 重复周期设置（高级）

按月

每月第 2 天 - 第 3 天执行

*重复结束时间：

2018-09-30 00:00

提交

取消



说明：

有关定时任务的属性，请参阅 [定时任务属性](#)。

定时任务属性

名称	说明	示例
任务名称	名称为 2-64 个字符，以大小写字母、数字或中文开头，可包含英文句号（.）、下划线（_）或短横线（-）。	st-yk20180830****
描述	描述定时任务的目的、作用等信息，便于日后查看。	月初访问量最大，增加 3 台实例。
执行时间	定时任务的触发时间。	2018 年 9 月 2 日 00:00
伸缩规则	伸缩规则的名称，指定任务触发时执行的伸缩动作。	add3

名称	说明	示例
重试过期时间	时间范围是 0 秒 ~ 21600 秒（6 小时）。如果未能在执行时间触发伸缩动作，弹性伸缩将在重试过期时间内继续执行定时任务。	600
重复周期	重复执行定时任务的周期，您可以选择按日、周、月的方式重复。如果要求多样，也可以使用 Cron 表达式 。	按月 每月第 2 天 - 第 3 天执行
重复结束时间	停止重复执行定时任务的时间。	2018年9月30日 00:00

Cron 表达式

Cron 表达式采用 UTC+0 时区，在中国换算成系统本地时间时需要加 8 小时。另外，第一次执行 Cron 表达式的时间必须小于重复结束时间，否则会导致定时任务创建失败。

Cron 表达式是一个字符串，字符串以若干个空格隔开，分为 5 到 7 个域。目前，弹性伸缩定时任务支持 5 域的 Cron 表达式，分别是分钟、小时、日、月和星期，取值范围如下表所示。

域	是否必需	取值范围
分钟	是	[0, 59]
小时	是	[0, 23]
日	是	[1, 31]
月	是	[1, 12]
星期	是	[0, 7], 星期日 = 0 或 7

在同一个域内可以填写多个数值，方法如下：

- 利用逗号 (,) 指定多个值，例如：1,3,4,7,8。
- 利用短横线 (-) 指定值的范围，例如：1-6，结果等同于1,2,3,4,5,6。
- 利用星号 (*) 指定任何可能的值，例如：小时域中填写星号代表每个整点，结果等同于0-23。
- 利用正斜线 (/) 指定间隔频率，例如：小时域中填写0-23/2表示每 2 小时执行一次。同时正斜线可以和星号一起使用，例如：小时域中填写*/3表示每 3 小时执行一次。

3.7.2 管理定时任务

定时任务创建完成后，您可以根据需要随时停用、启用、修改和删除定时任务，灵活应对业务变化。

停用或启用定时任务

定时任务创建完成后自动启用，您可以手动停用暂时不需要执行的定时任务。

1. 登录[弹性伸缩控制台](#)。
2. 依次单击自动触发任务管理 > 定时任务。
3. 在定时任务页面中，单击定时任务操作列下的停用。
4. 在停用定时任务对话框中，单击确定。



说明：

停用定时任务后，操作列下的停用变为启用。您也可以再度启用被停用的定时任务，步骤与停用定时任务类似。

修改定时任务

如果定时任务的需求发生变更，比如执行时间、伸缩规则或者重复周期等，您可以直接修改现有定时任务，无需重新创建。



说明：

本步骤仅介绍修改定时任务的入口，如果需要了解配置项的含义，请参阅[创建定时任务](#)。

1. 登录[弹性伸缩控制台](#)。
2. 依次单击自动触发任务管理 > 定时任务。
3. 在定时任务页面中，单击定时任务操作列下的修改。
4. 在修改定时任务页面中，根据需要修改配置项，然后单击提交。

删除定时任务

您可以删除不再需要的定时任务，空出定时任务配额。

1. 登录[弹性伸缩控制台](#)。
2. 依次单击自动触发任务管理 > 定时任务。
3. 在定时任务页面中，单击定时任务操作列下的删除。
4. 在删除定时任务对话框中，单击确定。

3.8 报警任务

3.8.1 弹性伸缩报警任务

本文主要介绍弹性伸缩报警任务。

弹性伸缩（Auto Scaling）报警任务是弹性伸缩与云监控服务（CMS）深度合作，提供的一种动态管理伸缩组的方式，类似于弹性伸缩定时任务，弹性伸缩报警任务通过触发您指定的伸缩规则来执行伸缩活动，达到调整伸缩组内实例个数的目的。

定时任务可以在您指定的时间执行您指定的伸缩规则，当业务场景在时间上可预料时，能够提前做出响应，但是，在面对突发或者时间上不可预料的业务场景时，定时任务就显得捉襟见肘，此时，就需要报警任务来提供更灵活的触发伸缩规则的方式，在业务高峰期增加伸缩组内实例数量来缓解业务压力，在业务低谷时释放伸缩组内实例，减小生产成本。

报警任务通过监控特定的监控指标，对数据指标进行实时的统计，当统计值满足您指定的报警条件时，触发报警，执行您指定的伸缩规则。使用报警任务，您可以实时的根据业务的变化来不断调整伸缩组内的实例数量，保证您监控的指标维持在您期望的范围内。

弹性伸缩报警任务为您提供了一种通过监控特定监控指标来动态调整伸缩组内实例数量的方法，让您能够根据业务的变化实时的执行指定的伸缩规则，调整伸缩组内的实例数量。

弹性伸缩报警任务升级版

弹性伸缩报警任务已经全新升级，从监控范围、监控方式、监控响应速度等三个方面做出了全方位的优化，升级后的弹性伸缩报警任务将为您提供一种更全面，更可靠的利用报警任务动态管理伸缩组的方式。

升级内容主要包括：

- 增加了对系统磁盘，网卡，TCP连接数等监控指标的报警任务支持。
- 报警任务最小统计周期升级到 1 分钟，提供更灵敏的监控报警。
- 增加自定义监控，为用户自有监控系统接入弹性伸缩报警任务提供标准化方式。

弹性伸缩报警任务升级版扩展了原有的监控指标，并在原有监控项的基础上，支持用户接入自定义的监控项，提供定制化的报警任务，大大增强了弹性伸缩报警任务的可用性和实用性，满足用户具体的，多样化的需求。

3.8.2 系统监控报警任务

本文主要介绍系统监控报警任务。

系统监控报警任务的监控指标是云监控为用户采集的 ECS 实例的相关数据指标，以伸缩组作为监控粒度，即以伸缩组内所有实例的监控指标的统计平均值作为伸缩组的指标值。当伸缩组内实例数量发生变化时，监控指标也会同时进行更新。

支持的监控指标

目前，系统监控报警任务支持的监控指标如下表所示。

监控指标	单位	适用网络
CPU	%	经典网络、专有网络
内存	%	经典网络、专有网络
系统平均负载	无	经典网络、专有网络
内网出流量	KB/min	经典网络、专有网络
内网入流量	KB/min	经典网络、专有网络
TCP总连接数	个	经典网络、专有网络
TCP已建立连接数	个	经典网络、专有网络
系统盘读BPS	Byte/s	经典网络、专有网络
系统盘写BPS	Byte/s	经典网络、专有网络
系统盘读IOPS	个/s	经典网络、专有网络
系统盘写IOPS	个/s	经典网络、专有网络
内网网卡发包数	个/s	经典网络、专有网络
内网网卡收包数	个/s	经典网络、专有网络
外网出流量	KB/min	经典网络、专有网络
外网入流量	KB/min	经典网络、专有网络
外网网卡发包数	个/s	经典网络
外网网卡收包数	个/s	经典网络

注意事项

- 伸缩组在同一时刻只能执行一个伸缩活动，当伸缩组内存在正在执行的伸缩活动时，由报警任务触发伸缩规则产生的伸缩活动将被拒绝。
- 弹性伸缩报警任务在触发伸缩规则时受伸缩规则冷却时间影响，伸缩规则在冷却时间内时，将拒绝执行伸缩规则。伸缩组内新增加的 ECS 实例从加入伸缩组到完成系统启动配置，部署用户业

务，获取到监控数据需要一定的时间（通常需要几分钟），因此您应该根据具体的业务场景，设置合适的冷却时间，防止在新增实例的监控数据缺失的这段时间中，重复触发伸缩规则。

- 弹性伸缩报警任务，默认设置了一分钟的沉默时间，即触发报警之后，一分钟内不会再次触发伸缩规则。
- 部分系统监控项（内存、负载、网卡发包数、TCP 连接数）指标的采集需要为您安装云监控客户端。默认情况下，当您针对需要云监控客户端采集的监控项设置报警任务时，将为报警任务关联的伸缩组内的所有实例安装云监控客户端，同时，将为您在云监控控制台开启 新购ECS自动安装云监控，为您所有新购的 ECS 实例安装云监控客户端。

3.8.3 自定义监控项报警任务

本文主要介绍弹性伸缩自定义监控项报警任务。

弹性伸缩自定义监控项报警任务的监控对象为用户自主上报到云监控中的监控指标。在一些场景下，系统监控项可能不包含您所需要的监控指标，您可能拥有自己的一套监控系统，并且关心的是与您特定业务相关的某些指标，自定义监控报警任务，为您自有的监控系统，或者与业务相关的自有监控指标提供了设置报警任务的接入点。

弹性伸缩自定义监控报警任务是针对阿里云云监控服务的自定义监控项设置报警的，用户在使用弹性伸缩自定义监控报警任务之前需要首先向云监控上报自定义监控数据，即自定义监控项。云监控自定义监控是提供给用户自由定义监控项及报警规则的一项服务，通过此服务，用户可以针对自己关心的业务指标进行监控，将采集到监控数据上报至云监控，由云监控来进行数据的处理，并可以对其设置报警规则。

上报监控数据到云监控

云监控的自定义监控服务为您提供了上报监控数据的方式，您可以将自己采集到的时序数据上报到云监控，此类数据称作时间序列。云监控提供了OpenAPI、Java SDK和阿里云命令行工具（CLI）三种方式上报数据，本文给出使用Java SDK的方式上报监控数据的示例，更多信息请参见[上报监控数据](#)。

使用Java sdk之前您需要首先在项目中引入相应的jar包，如果您使用maven管理项目，您只需要在项目中加入以下依赖：

```
<dependency>
  <groupId>com.aliyun</groupId>
  <artifactId>aliyun-java-sdk-core</artifactId>
  <version>3.2.6</version>
</dependency>
<dependency>
  <groupId>com.aliyun.openservices</groupId>
  <artifactId>aliyun-cms</artifactId>
  <version>0.2.4</version>
```

```
</dependency>
```

您可以按照如下的方式向云监控上报自定义监控项：

```
static String endPoint      = "https://metrichub-cms-cn-hangzhou.
aliyuncs.com";
CMSClient cmsClient = new CMSClient(endPoint, accAutoScalingKey,
accAutoScalingSecret);
CustomMetricUploadRequest request = CustomMetricUploadRequest.builder
()
    .append(CustomMetric.builder()
        .setMetricName("myCustomMetric")//自定义指标名
        .setGroupId(54504L)//设置分组id
        .setTime(new Date())//时间
        .setType(CustomMetric.TYPE_VALUE)//类型为原始值
        .appendValue(MetricAttribute.VALUE, number)//原始
值, key只能为这个
        .appendDimension("key1", "value1")//添加维度
        .appendDimension("key2", "value2")
        .build())
    .build();
CustomMetricUploadResponse response = cmsClient.putCustomM
etric(request);//上报
```

在上述的代码片段中，我们上报了一个数据点到云监控。上报时，必须指定GroupId参数，即云监控应用分组id，该分组id可以是您在云监控中已创建的应用分组，也可以是一个不存在的应用分组。您可以在[云监控控制台](#)的应用分组中创建应用分组和查看应用分组信息。您上报的自定义监控项（时间序列），可以在自定义监控中查看。

我们建议您向一个云监控中已经存在的应用分组中推送自定义监控数据，云监控的应用分组是针对多种云产品的逻辑分组，向一个已存在的应用分组推送自定义监控数据将使您在需要使用云监控等相关功能时保留扩展的能力。当然，您也可以完全不必理会应用分组的概念，选择向任意一个分组id推送数据。

您上报到云监控的监控数据，云监控会自动帮您进行聚合。当然，当您需要推送的数据量太大时，您也可以选择在本地聚合之后再推送到云监控，详细信息请参见[上报监控数据](#)。

限制

云监控对用户上报监控数据设置了以下限制：

- 单云账号QPS限制为100。
- 单次最多上报100条数据，body最大为256KB。
- metricName字段只支持字母、数字、下划线。需要以字母开头，非字母开头会替换为大写“A”，非法字符替换为“_”。
- dimensions字段不支持“=”、“&”、“,”等，非法字符会被替换为“_”。
- metricName和dimensions的Key-value最大均为64字节，超过64字节会被截断。
- 其他限制请关注[计费方式](#)说明。

3.8.4 创建报警任务

本文为您介绍如何创建系统监控项报警任务和自定义监控项报警任务。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 单击左侧导航栏中的自动触发任务管理 > 报警任务。
3. 在报警任务页面中，单击创建报警任务。
4. 在创建报警任务页面中，配置各项信息。
 - a. 填写任务名称。
 - b. 填写任务描述。
 - c. 选择监控资源，即您需要监控的伸缩组。
 - d. 选择监控类型。
 - 如果选择系统监控，您需要选择一个系统定义的监控项。详细的系统监控项说明，请参阅[系统监控报警任务](#)。
 - 如果选择自定义监控项，您需要选择在云监控中预先配置的应用分组、监控项和维度。详细的自定义监控项使用方法，请参阅[自定义监控项报警任务](#)。
 - e. 配置统计周期。



说明：

您可以选择1分钟、2分钟、5分钟或15分钟。弹性伸缩会根据统计周期收集、汇总和计算数据，粒度越小，越容易触发报警，请根据业务需要选择合理的统计周期。

- f. 配置统计方法。



说明：

即判定为超出阈值范围的规则，您可以围绕平均值、最大值或最小值设置规则。比如，统计方法为CPU使用率大于80%时：

- 平均值：针对伸缩组内所有ECS实例，CPU使用率平均值大于80%时触发报警。
- 最大值：针对伸缩组内CPU使用率最高的ECS实例，CPU使用率大于80%时触发报警。
- 最小值：针对伸缩组内CPU使用率最低的ECS实例，CPU使用率大于80%时触发报警。

- g. 配置重复几次后报警。



说明：

您可以选择1次、2次、3次或5次。根据统计方法判断超出阈值后，弹性伸缩会记录次数，满足重复次数后才会触发报警，自动执行指定的伸缩规则。

- h. 选择报警触发规则，即满足条件时执行的伸缩规则。
5. 单击提交。

3.8.5 查看修改删除报警任务

本文介绍如何查看、修改及删除报警任务。

查看监控项详情

成功创建报警任务之后，您可以在报警任务列表看到刚创建的报警任务。

1. 登录 [弹性伸缩控制台](#)。
2. 选择 系统监控 可查看创建的系统监控报警任务。

报警任务

刷新

创建报警任务

系统监控自定义监控							
报警任务	状态	监控资源	监控类型	统计周期	触发伸缩条件规则	触发规则数量	操作
test_cpu_alarm	正常	伸缩组：classic	系统监控	1分钟	CPU 使用率 平均值 连续 3 次 >= 50.0 %	1个	查看详情 停用 删除 更多操作

3. 选择 自定义监控 可查看创建的自定义报警任务。

弹性伸缩

伸缩组管理

自动触发任务管理

定时任务

报警任务

帮助

进一步了解弹性伸缩

常见问题

报警任务

华北1 华北2 华北3 华北5 华东1 华东2 华南1 香港 亚太东北 1 (东京) 亚太东南 1 (新加坡) 亚太东南 2 (悉尼) 亚太东南 3 (吉隆坡) 亚太东南 5 (雅加达) 亚太南部 1 (孟买) 美国东部 1 (弗吉尼亚) 美国西部 1 (硅谷) 中东东部 1 (迪拜) 欧洲中部 1 (法兰克福)

系统监控 自定义监控

报警任务	状态	监控资源	监控类型	统计周期	触发伸缩条件规则	触发规则数量	操作
test_custom_metric	正常	伸缩组: classic	自定义监控	1分钟	testMetric使用率平均值连续3次>=0.0	1个	查看详情 停用 删除 更多操作
age20	正常	伸缩组: classic	自定义监控	1分钟	testMetric使用率平均值连续1次>=23.0	1个	查看详情 停用 删除 更多操作
age10	报警	伸缩组: classic	自定义监控	1分钟	testMetric使用率平均值连续1次>=9.0	1个	查看详情 停用 删除 更多操作

刷新 创建报警任务

4. 单击报警任务的名称跳转到详情页，在详情页您可查看报警任务对应监控项的历史数据。



修改报警任务

您可以在报警任务的列表页修改报警任务，也可以进入报警任务的详情页修改报警规则。

报警任务							刷新	创建报警任务
系统监控	自定义监控							
报警任务	状态	监控资源	监控类型	统计周期	触发伸缩条件规则	触发规则数量	操作	
test_cpu_alarm	正常	伸缩组：classic	系统监控	1分钟	CPU 使用率 平均值 连续 3 次 >= 50.0 %	1个	查看详情	停用 删除
共有1条, 每页显示：10条							更多操作	
							修改报警任务	
							修改报警触发规则	

报警任务详情 [↑ 返回报警任务列表](#)

基本信息

修改基本信息

报警名称：	test_cpu_alarm	是否触发规则	☒
报警状态：	正常	统计周期	1m
监控资源：	伸缩组：: classic		
监控类型	系统监控		
触发伸缩规则的条件：	CPU 使用率 平均值 连续 3 次 >= 50.0 %		
描述：	cpu utilization		

报警触发规则

您最多可以建5条触发规则，目前还可以添加4条伸缩规则

报警触发规则

规则名称	ari:acs:ess:cn-qingdao:1406926474064770:scalingrule/asr-m5e9omkfub3ypueleq13
------	--

修改报警任务分为两个部分：修改报警任务的基本信息和修改报警规则的触发规则。

修改基本信息包括修改任务名称、监控项、统计周期、统计办法、重复次数等，我们不建议您修改报警任务的监控项，因为修改监控项意味着监控不同的指标，此时，新建一条针对新指标的报警任务是更合适的方法。

删除报警任务

您可以在 报警任务 页的操作列删除报警任务。

报警任务

刷新

创建报警任务

系统监控	自定义监控						
报警任务	状态	监控资源	监控类型	统计周期	触发伸缩条件规则	触发规则数量	操作
test_cpu_alarm	正常	伸缩组：classic	系统监控	1分钟	CPU 使用率 平均值 连续 3 次 >= 50.0 %	1个	查看详情 停用 删除 更多操作

4 维护自动伸缩

4.1 查看预测规则效果

您可以查看预测规则计算出的结果，评估是否符合预期。

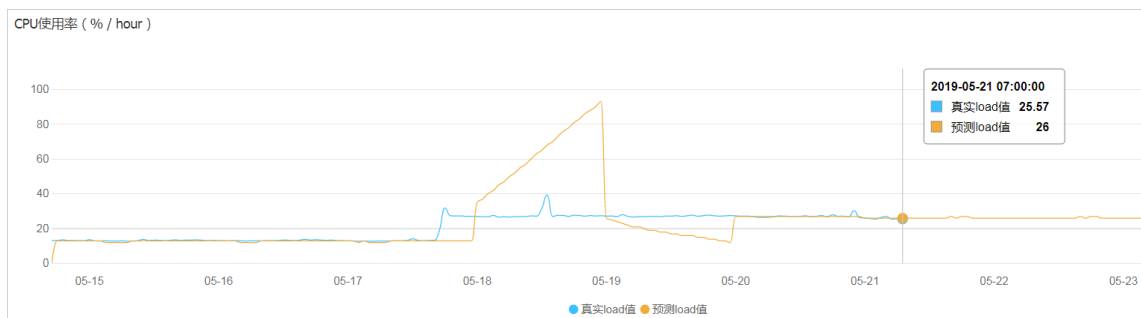
操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在指定伸缩组操作列下，单击管理。

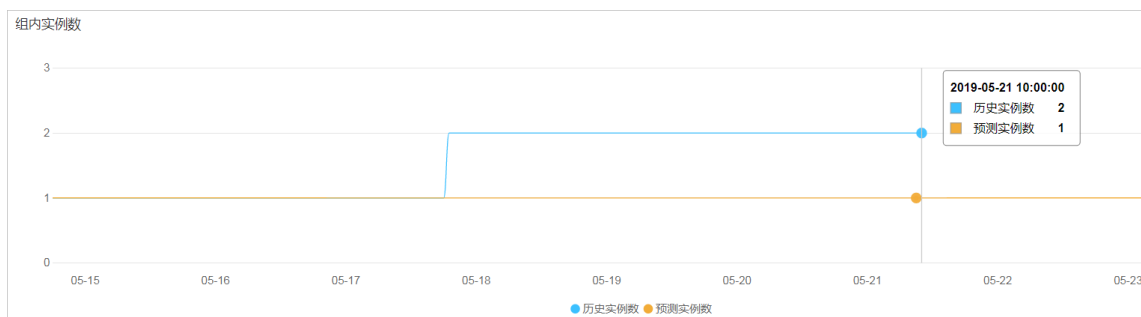
3. 前往伸缩规则页面，在指定预测规则的操作列下，单击查看详情。

规则详情页面提供了多个指标供您了解预测效果，您可以在确定预测效果符合预期后再开启预测并伸缩，详细步骤请参见[修改预测规则](#)。

- 对比真实和预测的CPU使用率评估预测准确性。



- 对比真实和预测的实例数评估预测准确性。

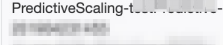
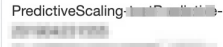
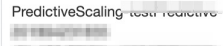
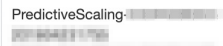


- 查看预测产生的定时计划评估效果是否符合预期。

预测产生的定时计划 刷新		
开始时间	最小实例	最大实例
2019年5月21日 09:50	1	100
2019年5月21日 10:50	1	100
2019年5月21日 11:50	1	100
2019年5月21日 12:50	1	100
2019年5月21日 13:50	1	100
2019年5月21日 14:50	1	100
2019年5月21日 15:50	1	100

后续步骤

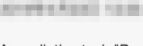
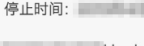
如果您开启了预测并伸缩，预测规则会根据预测产生的定时计划自动创建预测任务。预测任务是一种定时任务，您可以前往定时任务页面查看预测任务详情，预测任务的命名格式为：PredictiveScaling-伸缩规则名-执行时间。

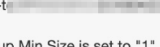
定时任务名称/ID	描述	状态	执行的操作	执行时间	重试过 期时间	重复 周期	重复停 止时间	操作
PredictiveScaling- 	PredictiveScali...	 运行中	由预测伸缩规则  创建，修改伸缩组  最小值为  ，最大值为 100		600 秒	暂无	-	停用 修改 删除
PredictiveScaling- 	PredictiveScali...	 运行中	由预测伸缩规则  创建，修改伸缩组  最小值为 1，最大值为 100		600 秒	暂无	-	停用 修改 删除
PredictiveScaling- 	PredictiveScali...	 运行中	由预测伸缩规则  创建，修改伸缩组  最小值为 1，最大值为 100		600 秒	暂无	-	停用 修改 删除
PredictiveScaling- 	PredictiveScali...	 运行中	由预测伸缩规则  创建，修改伸缩组  最小值为 1，最大值为 100		600 秒	暂无	-	停用 修改 删除

预测任务执行后会修改伸缩组边界值，执行成功之后会自行删除。您可以前往伸缩活动页面查看预测任务活动详情。

伸缩活动	变化后总实例数	开始时间	停止时间	描述	状态(全部)	操
	-			Group Max Size ...	成功	查看详
	-			Group Max Size ...	成功	查看详

伸缩活动ID:  状态: 成功

开始时间:  停止时间: 

活动起因: A predictive task "PredictiveScaling-t" is changing Group Max Size to "100" and Min Size to "1"

详细信息: Group Max Size is set to "100", Group Min Size is set to "1"

状态信息: Group Max Size and Min Size is changed

4.2 修改生命周期挂钩

本文主要介绍修改生命周期挂钩的步骤。

背景信息

[创建生命周期挂钩](#) 后，如果当前挂钩属性不再满足需求，您可以直接修改一项或者多项属性，无需重新创建。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在 [伸缩组管理](#) 页面中，单击指定伸缩组 操作 列下的 管理。
3. 前往 [生命周期挂钩](#) 页面，单击要修改的挂钩 操作 列下的 修改。

4. 在 **修改生命周期挂钩** 对话框中，根据需要修改属性，然后单击 **修改生命周期挂钩**。

修改生命周期挂钩

名称：

*适用的伸缩活动类型：

☐ 弹性收缩活动

☒ 弹性扩张活动

超时时间（秒）：

3600

最小为30，最大为21600，必须为整数

执行策略：

☐ 继续

☒ 拒绝

?

通知方式：

☐ MNS主题

☒ MNS队列

MNS队列：

通知标识 ?：

通知标识的长度不超过 128 个字符

修改生命周期挂钩

取消



说明：

- 不能修改挂钩名称，其他属性均可修改。
- 有关生命周期挂钩的属性，请参阅 [创建生命周期挂钩](#) 下的 生命周期挂钩属性。

4.3 删除生命周期挂钩

本文主要介绍删除生命周期挂钩的步骤。

背景信息

如果不再需要一个生命周期挂钩，您可以将其删除，空出挂钩配额。

操作步骤

- 登录[弹性伸缩控制台](#)。
- 在 **伸缩组管理** 页面中，单击指定伸缩组 **操作** 列下的 **管理**。

3. 前往 生命周期挂钩 页面，单击要删除的生命周期挂钩 操作 列下的 删除。



说明:

您也可以勾选生命周期挂钩左侧的复选框，然后单击下方的 删除 按钮进行批量删除。

4. 在 删除生命周期挂钩 对话框中，单击 确定 完成删除操作。



说明:

如果生命周期挂钩已经挂起一个 ECS 实例，在删除生命周期挂钩时，该 ECS 实例的等待状态会提前结束。

4.4 修改伸缩规则

创建伸缩规则后，如果当前伸缩规则配置不再满足需求，您可以修改一项或者多项属性，无需重新创建。本文介绍修改伸缩规则的步骤。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在指定伸缩组的操作列下，单击管理。
3. 前往伸缩规则页面，在指定伸缩规则的操作列下，单击修改。

伸缩规则	伸缩规则类型	执行的时间	执行的操作	操作
弹性伸缩伸缩规则	预测规则	为了将 平均CPU使用率 维持在 80.000%，根据预测数据来调整伸缩组的最大值最小值，从而影响伸缩组的实例伸缩 在预测结果产生后，生成每小时执行定时任务计划	只展示预测结果及相应定时任务计划，不伸缩	查看详情 修改 删除
弹性伸缩伸缩规则	简单规则	手动执行，未关联报警任务	将伸缩组内实例数目调整至 1 台	查看详情 执行 修改 删除
弹性伸缩伸缩规则	目标追踪规则	是将 平均CPU使用率 维持在 80.000% 所必需的 每次执行完新增的实例需要 300 秒预热 已关联的报警任务： TargetTracking- 报警任务	根据需要添加实例,而不减少实例	查看详情 修改 删除

4. 修改参数，然后单击修改伸缩规则。

参数说明请参见[简单规则参数](#)、[目标追踪规则参数](#)和[预测规则参数](#)。



说明:

不支持修改伸缩规则类型，预测规则不支持修改初始最大值。

修改伸缩规则

*规则名称：

名称为2-64个字符，以大小写字母，数字或中文开头，可包含".", "_或"-"

* 伸缩规则类型 ?：

简单规则

* 执行的操作：

调整至

1

台

"增加"或"减少"选项最多只能支持500台，否则报错
"调整至N台"以及"增加或减少N%"一次只能触发500台实例的伸缩活动

冷却时间（秒） ?：

修改伸缩规则

取消

修改伸缩规则

*规则名称：

名称为2-64个字符，以大小写字母，数字或中文开头，可包含".", "_或"-"

* 伸缩规则类型 ?：

目标追踪规则

* 指标类型 ?：

平均CPU使用率

* 目标值 ?：

80

%

* 实例预热时间（秒） ?：

300

* 禁用缩容 ?：

☒

修改伸缩规则

取消

修改伸缩规则

*规则名称：

名称为2-64个字符，以大小写字母，数字或中文开头，可包含".", "_"或"-"

* 伸缩规则类型 ?：

预测规则

是否配合已有的目标追踪规则 ?：

☐

* 指标类型 ?：

平均CPU使用率

* 目标值 ?：

80

%

预测模式 ?：

预测并伸缩

初始最大值 ?：

100

最大值处理方式 ?：

预测值会覆盖初始最大值

预启动时间（分钟） ?：

10

修改伸缩规则

取消

4.5 删除伸缩规则

本文主要介绍删除伸缩规则的步骤。

背景信息

如果不再需要一条伸缩规则，您可以将其删除，空出伸缩规则配额。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在 [伸缩组管理](#) 页面中，单击指定伸缩组 操作 列下的 管理。
3. 前往 [伸缩规则](#) 页面，单击要删除的伸缩规则 操作 列下的 删除。
4. 在 [删除伸缩规则](#) 对话框中，单击 确定。

4.6 设置实例备用状态

本文介绍如何设置实例备用状态。

弹性伸缩支持设置ECS实例备用状态（Standby）。我们不会检查处于备用状态的ECS实例健康状态，也不会释放ECS实例。

功能特性

- 设置ECS实例备用状态后：
 - ECS实例保持备用状态直至您使其退出备用状态。
 - 弹性伸缩不会管理ECS实例的生命周期，而是由您管理。
 - 如果对应伸缩组配置了负载均衡，则备用状态ECS实例的负载均衡权重会被置零。
 - 您可以 [停止](#) 实例、[重启](#) 实例或者做其他变配操作，例如 [更改ECS实例规格](#)、[#unique_96](#)、[初始化云盘](#) 或者 [迁移经典网络到专有网络VPC](#) 等。
 - 伸缩组内发生伸缩活动时，不会移除处于备用状态的ECS实例。
 - ECS实例被停止或者重启，不会更新ECS实例的健康检查状态。
 - 您需要预先从伸缩组内移除该ECS实例才能将其释放。
 - 如果您删除伸缩组，ECS实例会自动解除备用状态并随伸缩组一起释放。
- 解除ECS实例备用状态后：
 - ECS实例恢复运行状态。
 - 如果伸缩组设置了负载均衡实例，则将负载均衡实例中该ECS实例的权重设置为定义值。
 - ECS实例被停止或者重启时，会重新更新健康检查状态。
 - 弹性伸缩继续接管ECS实例的生命周期。

设置备用状态

1. 登录 [弹性伸缩管理控制台](#)。
2. 选择地域，如华东2。
3. 找到并单击目标伸缩组。
4. 在左侧导航栏，单击 ECS实例列表。
5. 勾选目标ECS实例，单击 转为备用状态。

解除备用状态

1. 登录 [弹性伸缩管理控制台](#)。
2. 选择地域，如华东2。
3. 找到并单击目标伸缩组。
4. 在左侧导航栏，单击 ECS实例列表。
5. 勾选目标ECS实例，单击 移出备用状态。

API操作

- 设置备用状态：[EnterStandby](#)
- 解除备用状态：[ExitStandby](#)

相关链接

- [什么是负载均衡](#)
- [移出不健康ECS实例](#)

4.7 实例转为保护状态

本文介绍如何将ECS实例转为保护状态。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在顶部状态栏处，选择地域。
3. 找到实例所在伸缩组，在操作列中，单击管理。
4. 在左侧导航栏中，单击ECS实例列表。
5. 找到待操作的实例，在操作列中，单击更多 > 转为保护状态。
6. 单击确定。

预期结果

ECS实例转为保护状态后：

- 如果ECS实例所在伸缩组配置了负载均衡，不会影响该ECS实例的负载均衡权重。
- 该ECS实例保持保护状态，直至您将其移出保护状态。
- 如果伸缩组内ECS实例数量的变化和监控任务触发自动缩容的伸缩活动，不会移除处于保护中状态的ECS实例。您需要自行移出ECS实例后才能释放ECS实例。
- 该ECS实例被停止或者重启时，不会更新ECS实例的健康检查状态。

4.8 实例移出保护状态

本文介绍如何将ECS实例移出保护状态。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在顶部状态栏处，选择地域。
3. 找到实例所在伸缩组，在操作列中，单击管理。
4. 在左侧导航栏中，单击ECS实例列表。
5. 找到待操作的实例，在操作列中，单击移出保护状态。
6. 单击确定。

4.9 移出策略

本文主要介绍移出策略。

有两种移出策略：默认策略和自定义策略。

默认移出策略

先按最早伸缩配置(OldestScalingConfiguration)对应的 ECS 实例进行一级筛选，再按最早创建的 ECS 实例(OldestInstance)进行二级筛选。

- 先在伸缩组里选最早伸缩配置(OldestScalingConfiguration)对应的 ECS 实例，然后在筛选的结果中，选最早创建的 ECS 实例(OldestInstance)。如果还有多台 ECS 实例，则随机选择一台移出伸缩组。
- 由于用户手工加入的既有 ECS 实例不与伸缩配置绑定，所以不会先选出该类型的 ECS 实例。
- 如果伸缩配置关联的 ECS 实例已经全部移出，但还需移出 ECS 实例时，则选最早用户手工移入的 ECS 实例。

自定义设置释放策略

根据用户设置的多条策略依次筛选 ECS 实例并移出伸缩组。

释放策略类型：

- OldestInstance：取最早创建的 ECS 实例，如果作为一级筛选条件，不区分是用户手工加入的 ECS 实例还是自动创建的 ECS 实例，直接按最早创建的实例进行移出。
- NewestInstance：取最新创建的实例，如果作为一级筛选条件，不区分是用户手工加入的 ECS 实例还是自动创建的 ECS 实例，直接按最新创建的 ECS 实例进行移出。
- OldestScalingConfiguration：取最早伸缩配置创建的 ECS 实例；对于用户手工移入的 ECS 实例，自动跳过此条件；如果与伸缩配置相关联的 ECS 实例已经全部移出，还要继续移出 ECS 实例时，则在用户手工加入的 ECS 实例中（没有与伸缩配置关联的 ECS 实例），随机选一台。

4.10 更改伸缩组状态

本文主要介绍更改伸缩组状态的步骤。

背景信息

[创建伸缩组](#)后，您可以控制伸缩组的状态。如果暂时不需要使用一个伸缩组，停用伸缩组即可，在需要使用时再启用。

操作步骤

1. 登录[弹性伸缩控制台](#)。

2. 在 伸缩组管理 页面中，单击待停用伸缩组 操作 列下的 停用，然后单击待启用伸缩组 操作 列下的 启用。
3. 检查伸缩组 状态 列下的状态是否正确。

4.11 修改伸缩组

本文主要介绍修改伸缩组的步骤。

创建伸缩组后，如果当前伸缩组属性不再满足需求，您可以直接修改一项或者多项属性，无需重新创建。



说明：

如果修改伸缩最大实例数和伸缩最小实例数，导致伸缩组内的ECS实例数超过上限或低于下限，弹性伸缩会自动加入或移出ECS实例，保证ECS实例数量满足要求。

操作步骤

请参照以下步骤修改一个伸缩组的属性：

1. 在伸缩组管理页面中，单击待修改伸缩组操作列下的修改。
2. 在修改伸缩组对话框中，根据需要修改伸缩组的属性。
 - a. 设置伸缩组名称。
 - b. 设置伸缩最大实例数（台）。



说明：

超过上限时，弹性伸缩会自动移出ECS实例，使得伸缩组的当前ECS实例数等于上限。

- c. 设置伸缩最小实例数（台）。



说明：

低于下限时，弹性伸缩会自动添加ECS实例，使得伸缩组 的当前ECS实例数等于下限。

- d. 设置默认冷却时间（秒）。



说明：

伸缩组发生伸缩活动后的默认冷却时间。更多信息，请参阅[冷却时间](#)。

- e. 设置移出策略。



说明：

伸缩组内ECS实例数超过上限时，移出ECS实例的策略。更多信息，请参阅[移出策略](#)。

- f. 设置组内实例配置信息来源。
- g. （可选）您无法修改伸缩组的网络类型。如果待修改伸缩组的网络类型为专有网络，您可以修改虚拟交换机，但不能修改多可用区扩缩容策略和回收模式。

The screenshot shows the 'Network' configuration section of the Elastic Scaling console. It includes a dropdown for '专有网络' (VPC) with a '创建专有网络' (Create VPC) link, a dropdown for '虚拟交换机' (VSwitch), and two radio buttons for '多可用区扩缩容策略' (Multi-AZ scaling strategy) and '回收模式' (Recovery mode).

- h. （可选）设置负载均衡。



说明:

一个伸缩组最多同时关联5个负载均衡实例。伸缩组还支持选择一个负载均衡实例下的[默认服务器组](#)和[虚拟服务器组](#)，一个伸缩组最多支持同时选择5个虚拟服务器组。更多信息，请参阅[在弹性伸缩中使用负载均衡服务](#)。

The screenshot shows the 'Load Balancer' configuration section of the Elastic Scaling console. It includes a dropdown for '负载均衡' (Load Balancer) with a '管理我的负载均衡' (Manage my load balancers) link. Below this is a table for '负载均衡配置详情' (Load Balancer configuration details) with columns for '服务器组' (Server Group), '端口(1-65535)' (Port), and '权重(1-100)' (Weight). The table shows the '默认服务器组' (Default Server Group) and a '虚拟服务器组' (Virtual Server Group). At the bottom, there are buttons for '+ 默认服务器组' and '+ 虚拟服务器组'.

- i. （可选）设置数据库，目前支持RDS。



说明:

只能添加和伸缩组位于同一地域的RDS实例，伸缩组弹性扩张出ECS实例后，会自动将ECS实例的内网IP添加到RDS实例的访问白名单，允许两者内网通信。

4.12 删除伸缩组

本文主要介绍删除伸缩组的步骤。

背景信息

如果不再需要某个伸缩组，您可以将其删除，空出伸缩组配额。

**说明:**

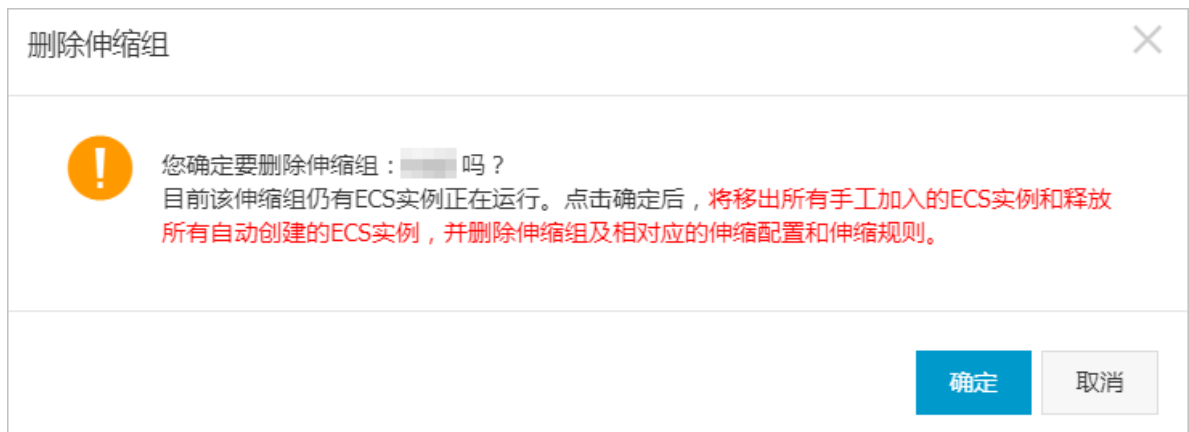
删除伸缩组会同时删除组内的伸缩配置和伸缩规则。如果组内存在运行中的 ECS 实例，弹性伸缩会先停止 ECS 实例，然后移出所有手动添加的 ECS 实例并释放所有自动创建的实例。

操作步骤

1. 在伸缩组管理页面中，单击要删除的伸缩组操作列下的更多 > 删除。



2. 在删除伸缩组对话框中，单击确定。



3. 在伸缩组管理页面中查看状态，进入删除中，您可以刷新页面确认删除已经完成。



5 手动伸缩

5.1 添加ECS实例

本文介绍如何手动向伸缩组添加一个ECS实例。

前提条件

手动添加ECS实例前，请确保待添加的ECS实例满足以下条件：

- 实例必须与伸缩组处于同一个地域。
- 实例不能已加入到其它伸缩组中。
- 实例的状态必须是运行中。
- 实例的网络类型可以为经典网络或专有网络，但有以下限制：
 - 当伸缩组的网络类型为经典网络时，只能添加网络类型为经典网络的ECS实例。
 - 当伸缩组的网络类型为专有网络时，只能添加同一专有网络下的ECS实例。

同时，请确保伸缩组满足以下条件：

- 伸缩组的状态必须是启用。
- 伸缩组内不能存在执行中的伸缩活动。

背景信息

手动添加ECS实例的配置与当前伸缩配置没有关联，并且手动添加ECS实例时可以绕过[冷却时间](#)。

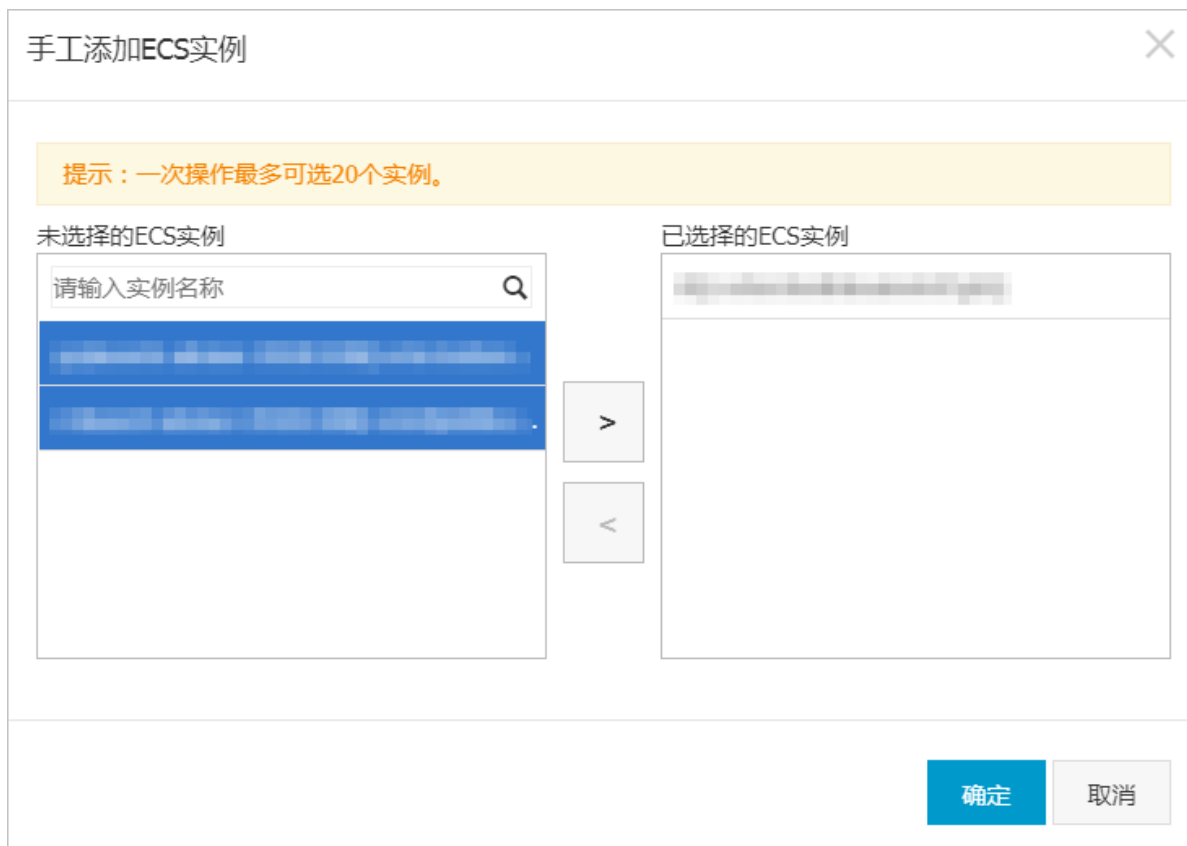
弹性伸缩服务尽力保证足额弹出待添加的ECS实例，但是，如果出现云服务器库存不足、待添加的ECS实例数超过伸缩组上限等问题，ECS实例会无法足额弹出。这种情况下，请您检查伸缩组相关配置定位问题。如果无法解决问题，请[提交工单](#)。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组管理页面，单击指定伸缩组操作列下的管理。
3. 前往ECS实例列表页面，单击添加已有实例。



4. 从左侧列表中选择可用的ECS实例，单击>添加到伸缩组，然后单击确定。



后续步骤

您可以前往手动添加页面查看结果。如果页面没有自动刷新，请单击页面右上角的刷新手动刷新。



5.2 移出ECS实例

本文主要介绍从指定的伸缩组里移出ECS实例。

从伸缩组移出弹性伸缩自动创建的 ECS 实例时，停止和释放该 ECS 实例。

从伸缩组移出用户手工加入的 ECS 实例时，不停止和释放该 ECS 实例。

执行该功能，必须同时满足以下条件：

- 伸缩组为生效（Active）状态。
- 当伸缩组没有伸缩活动正在执行。

当伸缩组没有伸缩活动正在执行时，该功能可以绕过冷却时间（Cooldown）直接执行。

执行功能返回成功，只是表示弹性伸缩服务接受了该功能的调用请求，伸缩活动可以执行，但不代表伸缩活动能够执行成功。用户需要通过返回的 ScalingActivityId 查看该伸缩活动的执行状态。

如果当前伸缩组的实例数（Total Capacity）减去该功能指定的实例数小于 MinSize 时，则执行失败。

示例

The screenshot shows a web interface for managing ECS instances. The top navigation bar includes a logo, the text 'phpwind-cms-pa...', a link to '返回伸缩组管理列表', and a '刷新' button. The left sidebar contains a menu with '基本信息', 'ECS实例列表' (highlighted), '伸缩活动', '伸缩配置', and '伸缩规则'. The main content area displays '实例总数: 2' and three boxes for '正在服务中的实例数' (2), '正在加入实例数' (0), and '正在移出实例数' (0). Below this is the 'ECS实例列表' section, which includes tabs for '自动创建' and '手动添加', and a '添加已有实例' button. The table lists instances with columns for selection, name, configuration, status, health, creation time, and actions. Two instances are listed, both in '服务中' (Running) status. At the bottom, a red box highlights the '移出伸缩组并释放' button, and the pagination shows '共有2条, 每页显示: 10条'.

☒	云服务器名称	伸缩配置	状态	健康检查状态(全部)	加入时间	操作
☒	ESS-sg-phpwind-cms-patch-ecs-i25pxu6uxw	test	服务中	健康	2014-10-07 17:11	移出伸缩组并释放
☒	ESS-sg-phpwind-cms-patch-ecs-i25dkmmajr	test	服务中	健康	2014-10-05 17:11	移出伸缩组并释放

6 事件通知

6.1 事件通知概述

事件通知是一种监控手段，可以自动发送消息至云监控或消息服务，帮助您及时掌握伸缩组的动态，进一步实现自动化管理。

事件通知方式

事件通知支持发送消息到云监控系统事件、MNS主题或MNS队列。

云监控提供了各类云产品系统事件的统计和查询入口，包括弹性伸缩，您可以通过云监控及时了解伸缩组的状态。如需详细了解云监控的事件监控功能，请参见[云产品系统事件监控](#)。

MNS主题和MNS队列是阿里云消息服务（Message Service）的两种服务模型。消息服务是一种分布式消息服务，能够帮助您在分布式组件之间自由的传递数据、通知消息，构建松耦合系统。如需详细了解MNS主题和MNS队列的特性，请参见[消息服务概述](#)。



说明：

消息服务涉及计费，详情请参见[消息服务定价](#)。

- 队列模型支持一对一发送和接收消息，旨在提供高可靠高并发的一对一消费能力，队列中的每一条消息都只能够被某一个消费者进行消费。
- 主题模型支持一对多发布和订阅消息，旨在提供一对多的发布订阅以及消息通知功能，您可以通过多种推送方式发布消息。

下面给出三种事件通知方式的示例，具体参数信息，请参见[创建事件通知](#)。

云监控事件通知示例

您创建了一个事件通知，通知方式为云监控，事件通知类型为成功的弹性扩张活动、弹性伸缩组扩容伸缩活动开始。在伸缩组成功执行弹性扩张活动后，云监控会收到事件通知并显示事件。下图为成功执行弹性扩张活动的效果，显示了弹性伸缩扩容伸缩活动开始和弹性伸缩扩容活动成功的事件。



除在**云监控控制台**查看伸缩组的动态外，您还可以**创建报警规则**，通过短信、邮箱等方式批量通知其它联系人，提高运维效率。

MNS主题事件通知示例

您创建了一个事件通知，通知方式为MNS主题，事件通知类型为成功的弹性收缩活动、弹性伸缩组缩容伸缩活动开始。在伸缩组成功执行弹性收缩活动后，MNS主题会收到事件通知并分发给订阅者。下图为成功执行弹性收缩活动的效果，MNS主题的消息数增加，您可以前往订阅者查看消息详情。

MNS主题不支持直接消费消息，您可以通过MNS队列、http、邮箱等方式订阅MNS主题，在MNS主题收到消息后，会自动推送到这些订阅者，达到统一分发、各自消费的效果，实现高效的自动化管理。



MNS队列事件通知示例

您创建了一个事件通知，通知方式为MNS队列，事件通知类型为失败的弹性扩张活动、弹性伸缩组扩容伸缩活动开始。在伸缩组执行弹性扩张活动失败后，MNS队列会收到事件通知并根据配置维护消息以供您消费。下图为执行弹性扩张活动失败的效果，MNS队列内的活跃消息数增加。

您可以自由消费、延迟、激活或者删除消息，灵活地通过事件通知实现自动化管理。



6.2 创建事件通知

本文为您介绍创建事件通知的限制条件和操作步骤。

限制条件

- 您可以同时拥有的事件通知数量有限，请参见[数量限制](#)。
- 同一个伸缩组内不可重复使用同一个接收端，比如不同的事件通知都试图使用云监控、同一个MNS主题或者同一个MNS队列。
- 如果您需要使用MNS主题或MNS队列，请提前创建MNS主题或MNS队列，并确保和伸缩组位于同一地域。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组操作列下，单击管理。
3. 在左侧导航栏，单击事件通知。
4. 单击创建事件通知。
5. 设置事件通知属性。

a. 配置通知方式。

通知方式	说明
云监控	发生指定事件时，通知云监控。详细信息，请参见 云产品系统事件监控 。
MNS主题	发生指定事件时，推送消息到MNS主题。
MNS队列	发生指定事件时，推送消息到MNS队列。

b. 配置事件通知类型。

您可以勾选多种事件通知类型。

事件通知类型	说明
成功的弹性扩张活动	成功向伸缩组添加了一台ECS实例。

事件通知类型	说明
成功的弹性收缩活动	成功从伸缩组移出了一台ECS实例。
失败的弹性扩张活动	触发了弹性扩张活动，但未能成功添加ECS实例。
失败的弹性收缩活动	触发了弹性收缩活动，但未能成功移出ECS实例。
拒绝弹性伸缩活动	伸缩组收到弹性伸缩请求，但不满足触发条件，请求被拒绝。
弹性伸缩组扩容伸缩活动开始	触发了弹性扩张活动，开始向伸缩组添加ECS实例。
弹性伸缩组缩容伸缩活动开始	触发了弹性收缩活动，开始从伸缩组移出ECS实例。
定时任务到期提醒	选择该类型后，在定时任务到期的前7天，伸缩组会每天发送到期提醒。  说明： 如果定时任务设置了重复周期，定时任务到期时间为最后一次执行任务的时间。



说明：

其中，伸缩活动成功包含了部分成功和全部成功两种情况，您可以查看事件通知中的伸缩活动详情来判断是部分成功还是全部成功。

6. 单击创建事件通知。

6.3 管理事件通知

您可以方便地管理事件通知，包括查看通知方式详情、修改事件通知类型和删除事件通知。

查看通知方式详情

弹性伸缩提供了控制台跳转链接，您可以方便的跳转到云监控和消息服务页面，查看收到的消息。

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组操作列下，单击管理。
3. 在左侧导航栏，单击事件通知。
4. 在通知方式列下，单击跳转链接。



说明：

云监控显示为云监控，MNS主题和MNS队列则显示为详细的主题或队列名称。

事件通知			创建事件通知
每个伸缩组中，您最多可以有6个事件通知。使用前必读			
事件通知类型	通知方式	操作	
成功的弹性扩张活动 失败的弹性扩张活动 成功的弹性收缩活动 失败的弹性收缩活动 拒绝弹性伸缩活动 定时任务到期提醒 弹性伸缩组扩容伸缩活动开始 弹性伸缩组扩容伸缩活动开始	云监控	修改 删除	
成功的弹性扩张活动 失败的弹性扩张活动 成功的弹性收缩活动 失败的弹性收缩活动 拒绝弹性伸缩活动 定时任务到期提醒 弹性伸缩组扩容伸缩活动开始 弹性伸缩组扩容伸缩活动开始	MNS队列：mnsqueue-as-event-notification	修改 删除	
成功的弹性扩张活动 失败的弹性扩张活动 成功的弹性收缩活动 失败的弹性收缩活动 拒绝弹性伸缩活动 定时任务到期提醒 弹性伸缩组扩容伸缩活动开始 弹性伸缩组扩容伸缩活动开始	MNS主题：mnstopic-as-event-notification	修改 删除	

修改事件通知类型



注意：

不支持修改已有事件通知的通知方式。

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组操作列下，单击管理。
3. 在左侧导航栏，单击事件通知。
4. 在事件通知操作列下，单击修改。
5. 配置事件通知类型。
6. 单击修改事件通知。

您可以通过[ModifyNotificationConfiguration](#)修改一个事件通知的通知类型。在修改通知类型前，您可以通过[DescribeNotificationConfigurations](#)查看已有事件通知的信息。

删除事件通知

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组操作列下，单击管理。
3. 在左侧导航栏，单击事件通知。
4. 在事件通知操作列下，单击删除。
5. 单击确定。

您可以通过[DeleteNotificationConfiguration](#)删除一个事件通知。在删除事件通知前，您可以通过[DescribeNotificationConfigurations](#)查看已有事件通知的信息。

7 查询ECS实例列表

本文介绍如何查询ECS实例列表。

本文仅通过 ECS 实例为非 运行中（Running）状态来判断该实例不健康，弹性伸缩会自动移出伸缩组中不健康的 ECS 实例。对于自动创建的 ECS 实例，弹性伸缩会停止和释放该 ECS 实例。对于手工添加的 ECS 实例，弹性伸缩不会停止和释放该 ECS 实例。

示例

下图是一个示例。

The screenshot shows the Elastic Scaling (ESS) console interface. At the top, there are navigation tabs: 弹性伸缩服务ESS, 伸缩组管理 (selected), 自动触发任务管理, and 整体方案创建管理. Below the tabs, the breadcrumb is 'phpwind-cms-pa...' and there is a '刷新' (Refresh) button. On the left sidebar, there are links: 基本信息, ECS实例列表 (selected), 伸缩活动, 伸缩配置, and 伸缩规则. The main content area shows summary statistics for the scaling group: 实例总数: 2, 正在服务中的实例数: 2, 正在加入实例数: 0, and 正在移出实例数: 0. Below this is a section titled 'ECS实例列表' with tabs for 自动创建 and 手动添加. A table lists the ECS instances:

☐	云服务器名称	伸缩配置	状态	健康检查状态(全部)	加入时间	操作
☐	ESS-sg-phpwind-cms-patch-ecs+25pxu6uxw	test	服务中	健康	2014-10-07 17:11	移出伸缩组并释放
☐	ESS-sg-phpwind-cms-patch-ecs+25dkmmajr	test	服务中	健康	2014-10-05 17:11	移出伸缩组并释放

At the bottom of the table, there is a button '移出伸缩组并释放' and a pagination bar showing '共有2条, 每页显示: 10条' and a page number '1'.

8 查看伸缩活动

弹性伸缩提供了查看伸缩活动功能，帮助您了解通过定时任务、报警任务等途径执行伸缩活动的结果。本文介绍查看伸缩活动的步骤。

操作步骤

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组管理列表中，在待查看伸缩组的操作列下单击管理。
3. 在左侧导航栏中单击伸缩活动。



说明：

您最多可以查看30天之内的伸缩活动信息。

4. 在指定伸缩活动的操作列下单击查看详情。



说明：

如果伸缩活动的状态为失败或拒绝，排查方法请参见[排查伸缩活动异常](#)。

伸缩活动						
伸缩活动	变化后总实例数	开始时间	停止时间	描述	状态(全部)	操作
伸缩活动ID: [redacted]	1	2019年2月21日 17:29	2019年2月21日 17:29	Add "1" ECS ins...	成功	查看详情
伸缩活动ID: [redacted]	0	2019年2月19日 14:27	2019年2月19日 14:27	Remove "2" ECS ...	成功	查看详情
伸缩活动ID: [redacted]	2	2019年2月19日 14:25	2019年2月19日 14:26	Add "2" ECS ins...	成功	查看详情
伸缩活动ID: [redacted]	0	2019年2月19日 14:23	2019年2月19日 14:24	Remove "2" ECS ...	成功	查看详情
伸缩活动ID: [redacted]	2	2019年2月19日 14:21	2019年2月19日 14:23	Add "2" ECS ins...	成功	查看详情
伸缩活动ID: [redacted]	0	2019年2月19日 14:05	2019年2月19日 14:08	Add "2" ECS ins...	失败	查看详情
伸缩活动ID: [redacted]	0	2019年2月19日 14:02	2019年2月19日 14:03	Remove "1" ECS ...	成功	查看详情

伸缩活动ID: [redacted] 状态: 成功

开始时间: 2019年2月21日 17:29 停止时间: 2019年2月21日 17:29

活动起因: A user requests to attach instance "[redacted]" from the specified scaling group.

详细信息: new ECS instances "[redacted]" are attached.

状态信息: "1" ECS instances are added

您也可以调用[DescribeScalingActivities](#)查看伸缩活动。

9 排查伸缩活动异常

执行伸缩活动时，可能因一些原因导致异常，无法成功完成。本文列出了常见的控制台说明信息，并给出原因和解决建议。

背景信息

伸缩活动异常包括伸缩活动被拒绝和伸缩活动执行失败，您可以参考[查看伸缩活动](#)了解相关的状态信息或详细信息。

伸缩活动被拒绝

如果伸缩活动被拒绝，您可以从状态信息了解原因，并采取对应的解决方案，如下表所示：

伸缩活动状态信息	原因及建议
DBInstanceIdentifier does not refer to an existing DB instance.(xxxx)	原因：RDS实例不存在。 建议：请确保伸缩组关联的RDS实例已存在。
The current status of the load balancer xxxxx does not support this action.	原因：负载均衡实例状态不是active。 建议：请确保伸缩组关联的负载均衡实例处于正常工作的状态。
To execute the specified scaling rule, the total capacity will not change.	原因：伸缩组内的实例数不能小于伸缩组设置的最小值，也不能大于伸缩组设置的最大值，执行此伸缩规则将导致伸缩组内实例小于最小值或大于最大值，因此拒绝执行。 建议：请检查目前实例数量或伸缩规则配置，确保执行伸缩规则后不会超出实例数量限制。
The current status of DB instance xxxxx does not support this action.	原因：RDS实例状态异常，暂不支持执行伸缩活动。 建议：如果伸缩组关联的RDS实例处于迁移状态，建议登录 RDS控制台 确认实例状态处于正常工作状态，然后重新执行伸缩活动。
You cannot delete a scaling group or launch a new scaling activity while there is scaling activity in progress for the specified scaling group.	原因：当前伸缩组有正在执行的伸缩活动。 建议：请等待当前的伸缩活动运行完成，然后执行新的伸缩活动。

伸缩活动状态信息	原因及建议
You cannot launch a new scaling activity during the cooldown period.	原因：伸缩组处于冷却时间，伸缩活动被拒绝执行。 建议：需要等待冷却时间结束，具体说明请参考冷却时间。
The current status of the specified scaling group does not support this action.	原因：当期伸缩组状态不支持此操作，伸缩组可能为停用状态或者删除中状态。 建议：如果伸缩组为停用状态，请先启用伸缩组。
The specified SecurityGroupId does not exist.	原因：伸缩配置对应的安全组不存在。 建议：请修改伸缩配置，设置对应的安全组。
To remove the instances, the total capacity will be lesser than the min size.	原因：伸缩组内的实例数不能小于伸缩组设置的最小值，如果移除伸缩组内现有的实例，会导致伸缩组内实例数小于伸缩组最小值，因此被拒绝执行。 建议：请检查目前实例数量或伸缩规则配置，确保执行伸缩规则后不会超出实例数量限制。
The specified LoadBalancerId does not exist.	原因：负载均衡实例不存在。 建议：请确保伸缩组关联的负载均衡实例已存在，伸缩组中设置负载均衡实例详情可参考在弹性伸缩中使用负载均衡服务。
Ess should be authorized in ram to operate user resource.	原因：主账号未授权当前子账号操作弹性伸缩服务相关资源。 建议：请登录访问控制控制台授权相关资源，然后继续操作。
To attach the instances, the total capacity will be greater than the max size.	原因：伸缩组内的实例数不能大于伸缩组设置的最大值，如果添加实例到伸缩组，会导致伸缩组内实例数大于伸缩组最大值，因此被拒绝执行。 建议：请检查目前实例数量或伸缩规则配置，确保执行伸缩规则后不会超出实例数量限制。

伸缩活动状态信息	原因及建议
The current status of virtual switch does not support this operation.	原因：虚拟交换机（Vswitch）的状态为pending或modifying，伸缩组暂时无法执行伸缩活动。 建议：请检查伸缩组关联的虚拟交换机的状态，确保处于可用状态，然后重试。
The specified virtual switch does not exist.	原因：虚拟交换机不存在。 建议：请确保伸缩组关联的虚拟交换机未被删除。
The specified launch template set is not found.	原因：实例启动模板不存在。 建议：请登录ECS控制台，确保伸缩组关联的实例启动模板已存在，然后重新设置伸缩组关联的实例启动模板。
Instance "xxxxx" and the specified scaling group are not in the same VPC.	原因：当前ECS实例所属的专有网络（VPC）与伸缩组对应的专有网络不一致。 建议：请确保专有网络伸缩组内的ECS实例与伸缩组属于同一个专有网络，或者ECS实例网络类型为经典网络。
The specified virtual switch and the instance in Load Balancer "xxxxx" are not in the same VPC.	原因：负载均衡实例与虚拟交换机（Vswitch）所属的专有网络（VPC）不一致。 建议：请确保Vswitch和负载均衡实例处于同一个专有网络下，负载均衡实例的添加限制请参考在弹性伸缩中使用负载均衡服务。
Instance "xxxx" is already attached to another scaling group.	原因：ECS实例只能添加到一个伸缩组，不能同时属于多个伸缩组。 建议：请确保待添加的ECS实例没有加入到其它伸缩组。
The specified image does not exist.	原因：镜像不存在。 建议：请检查伸缩配置，确保镜像可用。

伸缩活动状态信息	原因及建议
DataDiskSnapshotId "xxxx" does not exist .	原因：数据盘快照不存在。 建议：请检查伸缩配置，确保数据盘的快照可用。

伸缩活动执行失败

如果伸缩活动执行失败，您可以从详细信息了解原因，并采取对应的解决方案，如下表所示：

伸缩活动详细信息	原因及建议
Fail to scale instances for scaling group("Ignore to remove N instances("instances in standby/protected status")").	原因：指定实例处于备用/实例保护状态。 建议：请手动移除或释放实例。  说明： 具体说明请参考 实例保护 和 实例备用 。
Fail to remove Instances "i-xxx1, i-xxx2"("The current status of the specified scaling group does not support this action.").	原因：当前伸缩组状态不支持移除实例操作。 建议：请确保伸缩组状态为生效。移除实例相关限制请参考官方文档 RemoveInstances 。
Fail to create Instances into scaling group ("The specified InstanceType has been offline.").	原因：指定实例规格已下线。 建议：请在伸缩配置中选择在售的实例规格，更多说明请参考 实例规格族汇总 。
Fail to create Instances into scaling group ("The specified ImageId does not exist.").	原因：伸缩配置指定镜像ID不存在。 建议：请检查伸缩配置，确认指定镜像可用，或者重新选择可用的镜像。
Fail to create Instances into scaling group ("Your account has been in arrears.").	原因：您的账户已欠费。 建议：请保证账户额度充足，您可以在 费用中心 查看账户费用信息。
Fail to create Instances into scaling group ("Your account does not have enough balance.").	原因：您的账户可用额度不足。 建议：请保证账户额度充足，您可以在 费用中心 查看账户费用信息。
Fail to create Instances into scaling group ("Your account has not passed the real-name authentication yet.").	原因：您的账户尚未通过实名认证。 建议：请开启 实名认证 。

伸缩活动详细信息	原因及建议
Fail to create Instances into scaling group ("The specified parameter "spotPriceLimit" can't be lower than current public price.").	原因：指定抢占式实例的每小时最高价格低于当前市场价格。 建议：请检查伸缩配置，重新设置抢占式实例出价或者出价策略，确保满足创建抢占式实例的条件。更多信息请参考抢占式实例。
Fail to create Instances into scaling group("Living afterpay instances quota exceeded.").	原因：您持有的按量付费实例已超过限额。 建议：请联系客服或提交工单申请提高按量付费实例限额。
Fail to create Instances into scaling group ("The specified parameter "KeyPairName" does not exist in our records.").	原因：伸缩配置中指定密钥对不存在。 建议：请检查伸缩配置，确保当前密钥对可用，或者重新设置密钥对。
Fail to create Instances into scaling group ("The specified parameter 'SystemDisk.Category' is not support IoOptimized Instance. Valid Values: cloud_efficiency; cloud_ssd. ").	原因：当前系统盘类型不支持IO优化实例。 建议：请检查伸缩配置，修改系统盘类型，建议值： · 高效云盘（cloud_efficiency） · SSD云盘（cloud_ssd）
Fail to create Instances into scaling group ("The resource is out of stock in the specified zone. Please try other types, or choose other regions and zones.").	原因：指定资源库存不足。 建议：请检查伸缩配置和伸缩组信息，视情况选择其他实例规格或者可用区。
Fail to create Instances into scaling group("The RAM user does not have the privilege to pass a RAM role.").	原因：伸缩配置中指定RAM角色不存在。 建议：请检查伸缩配置，重新指定合法RAM角色。
Fail to create Instances into scaling group("No payment method has been registered on the account.").	原因：当前账户未设置默认支付方式。 建议：请前往账号中心设置默认支付方式。
Fail to create Instances into scaling group("Exceeding the allowed amount of instances of a security group.").	原因：指定安全组下的实例数量超过上限。 建议：请检查当前伸缩组，尝试将单个伸缩组扩展成多个伸缩组，并分别关联不同的安全组，以满足横向扩容的需要。

伸缩活动详细信息	原因及建议
Fail to create Instance into scaling group ("The current status of instance "pending" does not support this action.").	原因：当前实例状态不支持指定操作。 建议：请检查确实例状态，确保为Running后重试。
Fail to create Instances into scaling group("Abandon lifecycleActionResult parameter caused the instance to rollback.").	原因：伸缩组配置了扩容（ScaleOut）类型的生命周期挂钩（LifecycleHook），当伸缩组扩容时，生命周期挂钩被触发，并且生命周期挂钩最终的执行结果为拒绝，导致扩容伸缩活动回滚，扩容出来的 ECS 实例被释放。 建议：请检查生命周期挂钩的执行策略。  说明： 更多生命周期挂钩的信息，请参考创建生命周期挂钩。
Fail to create instance into scaling group ("The specified SecurityGroupId is invalid or does not exist.")	原因：安全组不存在。 建议：请检查伸缩配置，重新指定已存在的安全组。
Fail to create instance into scaling group ("The specified ImageId does not exist.")	原因：镜像不存在。 建议：请检查伸缩配置，重新指定已存在的镜像。
Fail to create instance into scaling group ("Your account has been in arrears.")	原因：您的账户可用额度不足。 建议：请保证账户额度充足。
Fail to scale instances for scaling group("Backend server quota exceeded in load balancer "lb-xxxxxxx").	原因：负载均衡实例的后端服务器组下可挂载ECS实例配额超出限制。 建议：请调整后端服务器的配额。更多说明请参考使用限制和管理配额。
Fail to create Instances into scaling group ("The specified imageName does not exist.").	原因：镜像名称不存在。 建议：请检查伸缩配置，重新指定已存在的镜像。

伸缩活动详细信息	原因及建议
Fail to create Instances into scaling group ("This resource type is not supported. Please try other resource types.").	原因：可用区下无指定的资源类型。 建议：建议在伸缩组中设置多可用区，并在伸缩配置中设置多实例规格，提高扩容成功率。更多说明请参考使用自定义伸缩配置创建伸缩组和创建伸缩配置。
Fail to create Instances into scaling group ("The specified parameter SnapshotId is not exist.").	原因：磁盘快照不存在。 建议：请检查伸缩配置，重新指定已存在的磁盘快照。
Fail to create Instances into scaling group ("Don't have enough private IPs in this switch").	原因：虚拟交换机没有足够的私网IP。 建议：请检查伸缩配置关联的虚拟交换机，调大虚拟交换机可用私网IP数量，或者在伸缩组配置中设置多可用区。更多说明请参考使用自定义伸缩配置创建伸缩组。
Fail to create Instances into scaling group ("The specified parameter SystemDisk.Size is less than the image size.").	原因：系统盘容量小于镜像文件大小。 建议：请检查伸缩配置，增大系统盘容量或者使用占用空间更小的镜像。
Fail to create Instances into scaling group ("Living postPaid disks quota exceeded.").	原因：当前账号下现存的按量付费磁盘限额超过限制。 建议：请联系客服或提交工单申请调整配额。
Fail to create Instances into scaling group ("The specified ram role is not authorized for ecs, please check your role policy.").	原因：弹性伸缩服务对应的RAM角色AliyunEcsDefaultRole没有包括ECS接口权限。 建议：请重新授权。  说明： 具体操作请参考AutoScaling授权指南。
Fail to create instance into scaling group ("The resource in the specified zone is no longer available for sale. Please try other regions and zones.")	原因：实例规格在指定可用区下已停止售卖。 建议：请检查伸缩组设置，建议设置多可用区。更多说明请参考使用自定义伸缩配置创建伸缩组。

伸缩活动详细信息	原因及建议
Fail to create instance into scaling group ("The resource is out of stock in the specified zone. Please try other types, or choose other regions and zones.")	原因：实例规格在指定可用区下已停止售卖。 建议：建议在伸缩组中设置多可用区，并在伸缩配置中设置多实例规格，提高扩容成功率。更多说明请参考使用自定义伸缩配置创建伸缩组和创建伸缩配置。
Fail to create instance into scaling group ("The specified parameter "KeyPairName" only support IoOptimized Instance.")	原因：实例规格为非I/O的优化实例规格，非I/O优化实例不支持密钥对。 建议：请检查伸缩配置，调整伸缩配置实例规格为I/O优化实例规格，或者删除KeyPair参数。
Fail to create instance into scaling group ("A RAM role can't be used for classic instance.")	原因：经典网络类型实例不支持RAM角色。 建议：请检查伸缩组的网络类型，建议您修改伸缩组网络类型为专有网络，或者移除伸缩配置中的RamRoleName参数。
Fail to create instance into scaling group ("The specified parameter "RAMRoleName" does not exist.")	原因：RAMRoleName不存在。 建议：请登录访问控制控制台检查当前RAM角色是否存在，或者指定其它RAM角色。
Fail to create instance into scaling group ("The specified image does not support the specified InstanceType.")	原因：镜像与实例规格不匹配。 建议：请登录ECS售卖页查看实例规格支持的镜像，然后重新调整伸缩配置。
Fail to create instance into scaling group ("The requested resource is sold out in the specified zone; try other types of resources or other regions and zones.")	原因：实例规格在伸缩组对应的可用区下库存不足。 建议：建议在伸缩组中设置多可用区，并在伸缩配置中设置多实例规格，提高扩容成功率。更多说明请参考使用自定义伸缩配置创建伸缩组和创建伸缩配置。

10 弹性伸缩FAQ

本文汇总了使用弹性伸缩时的常见问题。

- [如何避免单实例规格库存不足导致扩容失败？](#)
- [报警任务和定时任务之间有执行优先级吗？](#)
- [ECS实例是否可以加入到多个伸缩组中？](#)
- [伸缩组内ECS实例开启了释放保护，为什么仍然被自动释放了？](#)

如何避免单实例规格库存不足导致扩容失败？

建议您在创建伸缩组时设置多可用区（选择不同可用区下的虚拟交换机即可），并在创建伸缩配置时选择多实例规格，当某个ECS实例规格在某个可用区没有库存时，弹性伸缩服务会自动切换到有库存的实例规格及可用区进行扩容。更多说明请参见[使用自定义伸缩配置创建伸缩组](#)和[创建伸缩配置](#)。

报警任务和定时任务之间有执行优先级吗？

报警任务与定时任务相互独立，没有相对优先级。但由于伸缩组同一时间只能执行一个伸缩活动，先执行的任务触发的伸缩活动可以正常完成，后执行的任务触发的伸缩活动会失败。

如果报警任务失败，但仍然满足报警任务触发条件，待当前伸缩活动结束后报警任务仍会被执行。

您可以为定时任务设置重试时间，避免被拒绝执行后无法再次触发，详细操作请参见[创建定时任务](#)。

ECS实例是否可以加入到多个伸缩组中？

不支持。

伸缩组内ECS实例开启了释放保护，为什么仍然被自动释放了？

在弹性伸缩自动创建一台ECS实例后，如果您在ECS控制台的实例列表页面或者调用[ModifyInstanceAttribute](#)为ECS实例开启了释放保护，并不能阻止弹性伸缩自动释放实例。您可以将伸缩组内的ECS实例转为保护状态，避免被自动释放，具体操作请参见[实例转为保护状态](#)。

11 弹性伸缩关联产品

11.1 在弹性伸缩中使用负载均衡服务

伸缩组支持关联负载均衡实例，通过负载均衡实例将访问流量分发到伸缩组内的多个ECS实例，有效增强伸缩组的服务能力。

负载均衡服务简介

负载均衡服务通过设置虚拟服务地址，将添加的同一地域的多个ECS实例虚拟成一个高性能、高可用的应用服务池。简单来说，负载均衡服务通过组合负载均衡实例、监听和后端服务器，提供流量分发控制服务。更多信息，请参阅[什么是负载均衡](#)。

前提条件

在为伸缩组添加负载均衡实例前，请确保：

- 您持有一个或多个运行中的负载均衡实例。否则，请先[创建负载均衡实例](#)。
- 负载均衡实例和伸缩组必须位于同一地域。
- 如果负载均衡实例和伸缩组的网络类型均为专有网络，则必须位于同一专有网络。
- 当负载均衡实例的网络类型为经典网络，伸缩组的网络类型为专有网络时，如果负载均衡实例的后端服务器组中包含专有网络ECS实例，该ECS实例必须与伸缩组位于同一专有网络。
- 负载均衡实例配置至少一个监听。否则，请参阅[监听介绍](#)添加合适类型的监听。
- 负载均衡实例必须开启健康检查。否则，请参阅[配置健康检查](#)开启健康检查。

通过控制台管理负载均衡实例



说明：

本步骤重点介绍负载均衡实例相关的控制台操作，如需了解其它配置，请参阅[使用自定义伸缩配置创建伸缩组](#)或者[使用实例启动模板创建伸缩组](#)。

1. 登录[弹性伸缩控制台](#)。
2. 在伸缩组管理页面中，选择一种方式进入添加负载均衡实例的页面：
 - 创建伸缩组时，单击创建伸缩组。
 - 修改伸缩组时，单击待修改伸缩组操作列下的修改。
3. 配置网络类型。



说明：

伸缩组的网络类型在创建伸缩组时指定，不支持修改。

4. 关联负载均衡实例。



说明：

一个伸缩组最多同时关联5个负载均衡实例。如果没有加载出可选负载均衡实例，可能是您的负载均衡实例未满足[关联条件](#)，请单击管理我的负载均衡前往负载均衡控制台查看并更新配置。

5. 选择用于承载伸缩组中ECS实例的后端服务器组。



说明：

伸缩组支持选择一个负载均衡实例下的[默认服务器组](#)和[虚拟服务器组](#)，一个伸缩组最多支持同时选择5个虚拟服务器组。

负载均衡 ? : [管理我的负载均衡](#)

只有配置过监听的负载均衡才能被伸缩组使用

负载均衡配置详情

伸缩组下负载均衡最多可配置 1 / 5 [↑滚动查看全部↓](#)

负载均衡 ID : ×

负载均衡名称 :

服务器组	端口(1-65535)	权重(1-100)	
默认服务器组 ?	-	在伸缩配置里指定	×
⊗ ?			×

+ 默认服务器组 + 虚拟服务器组

伸缩组下负载均衡虚拟服务器组最多可配置 1 / 5

6. 根据需要配置其余选项。

通过API管理负载均衡实例

使用[CreateScalingGroup](#)创建伸缩组时，您可以通过LoadBalancerIds关联相应的负载均衡实例，并通过VServerGroup控制虚拟服务器组的属性。

修改伸缩组时，您可以通过AttachLoadBalancers和DetachLoadBalancers控制关联负载均衡实例的动作，然后通过AttachVServerGroups和DetachVServerGroups控制选择虚拟服务器组的动作。

负载均衡效果

伸缩组关联负载均衡实例后，无论是伸缩组自动创建ECS实例，还是您向伸缩组手动添加ECS实例，ECS实例都会自动加入到负载均衡实例的后端服务器组。负载均衡实例会根据流量分发、健康检查等策略灵活使用ECS实例资源，在资源弹性的基础上大大提高资源可用性。

**说明:**

这些ECS实例的权重默认为50，您可以根据需要在对应负载均衡实例中调整权重。

12 管理伸缩组的伸缩活动

12.1 伸缩组管理概述

本章节介绍与控制单个伸缩组有关的功能，包括伸缩组、伸缩规则和生命周期挂钩，这三项功能为进行伸缩活动提供了操作支持，您可以灵活控制伸缩组的活动，满足业务需要。

弹性伸缩服务支持的伸缩组、伸缩规则和生命周期挂钩数量有限，请参阅 [使用限制](#)。

假设出于业务需要，您需要在一段时间内使用弹性伸缩服务，实现按需增减 1 至 10 台 ECS 实例，并且在启动 ECS 实例前安装软件。

接下来将针对该场景，为您逐步展示通过弹性伸缩满足需求的过程，涉及到的操作有：

- [创建伸缩组](#)：创建一个伸缩组，用于承载伸缩活动。
- [创建伸缩规则](#)：创建一条伸缩规则，用于定义伸缩活动动作。
- [执行伸缩规则](#)：执行一条伸缩规则，用于启动伸缩活动。
- [创建生命周期挂钩](#)：创建一个生命周期挂钩，用于在伸缩活动前保留实例操作时间。
- [修改生命周期挂钩](#)：修改一个生命周期挂钩，用于修改挂钩属性，适应新场景。
- [删除生命周期挂钩](#)：删除一个不再使用的生命周期挂钩。
- [修改伸缩规则](#)：修改一条伸缩规则，用于修改伸缩规则属性，适应新场景。
- [删除伸缩规则](#)：删除一条不再使用的伸缩规则。
- [更改伸缩组状态](#)：用于控制伸缩组启用/停用状态。
- [修改伸缩组](#)：修改一个伸缩组，用于修改伸缩组属性，适应新场景。
- [删除伸缩组](#)：删除一个不再使用的伸缩组。