

弹性伸缩

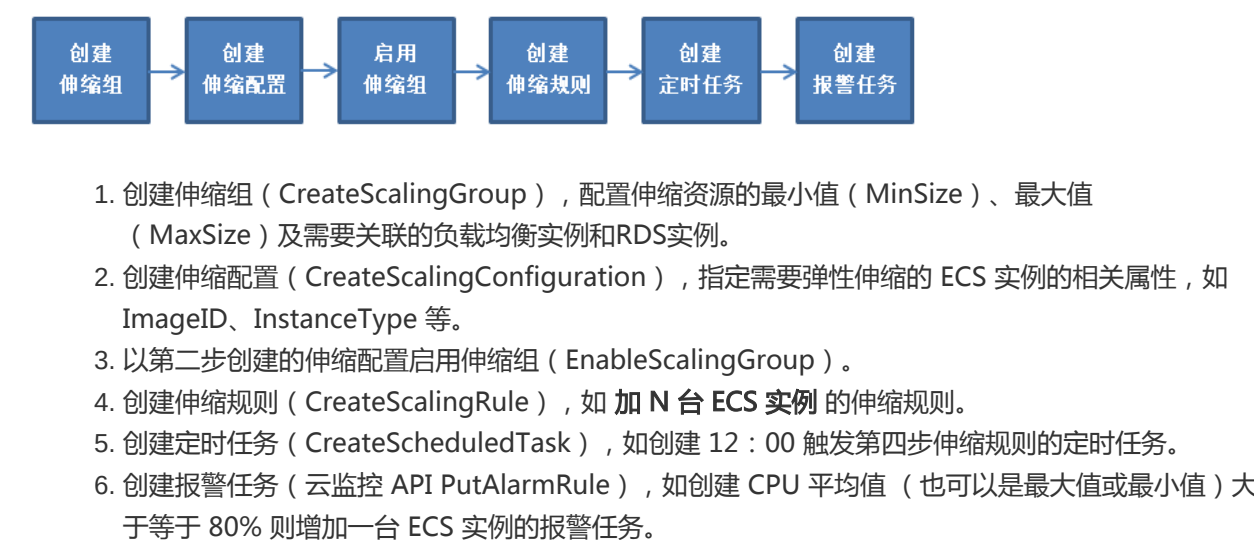
用户指南

用户指南

工作原理

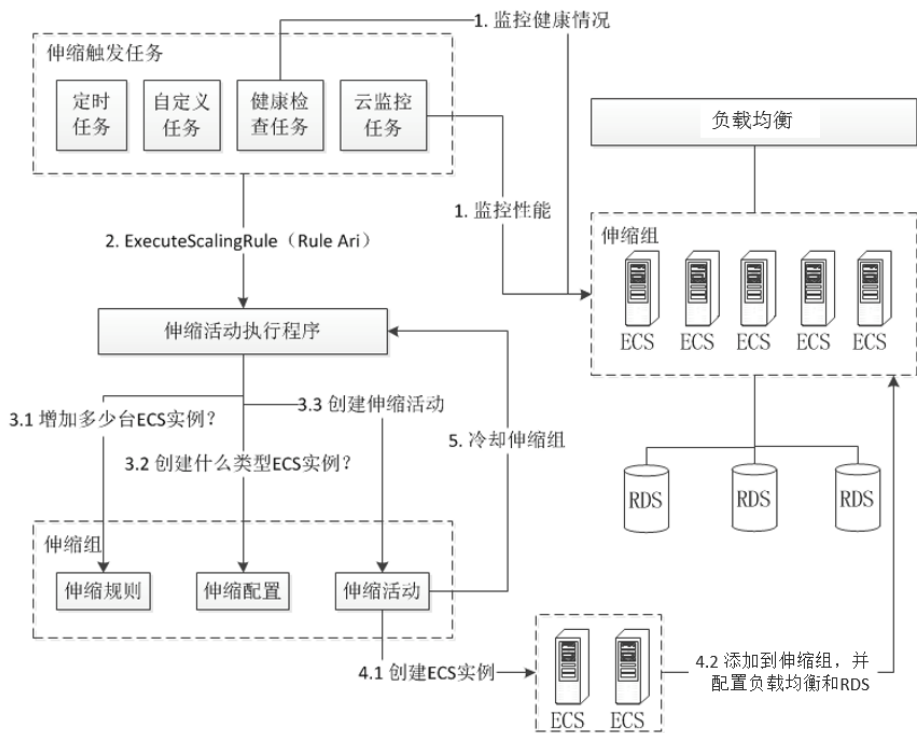
使用流程

创建完整的弹性伸缩方案，您需要通过以下步骤来完成：



工作流程

下图是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）：

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。

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。

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。

- 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。
- 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）。
- 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。

在伸缩活动中，自动创建 ECS 实例并配置负载均衡和 RDS。

- i. 按照实例配置信息创建指定数量的 ECS 实例。
- ii. 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。

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。

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。

使用须知

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。

ECS实例生命周期的管理

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。

冷却时间

冷却时间是指，在同一个伸缩组内，当有一个伸缩活动被成功执行后的一段锁定时间内，该伸缩组不执行其他的伸缩活动。

关于冷却时间，有以下说明：

伸缩组在冷却时间内，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。

冷却时间锁定的是同一个伸缩组的伸缩活动，对于多伸缩组内的不同伸缩活动，不做以上要求。

伸缩规则的冷却时间和伸缩组默认冷却时间同时存在时，以伸缩规则的冷却时间为伸缩活动的执行标准。

当伸缩组加入或移出多个 ECS 实例时，最后一个 ECS 实例加入或移出完成后，整个伸缩组冷却时间才开始计时。

- 伸缩活动只要有一个以上 ECS 实例成功加入或移出伸缩组，仍按最后一个 ECS 实例成功加入或移出伸缩组后，启动冷却。
- 伸缩活动如果没有一个 ECS 实例成功加入或移出伸缩组，则不启动冷却。

停用伸缩组，然后再启用伸缩组后，冷却时间的效果将会失效。例如一个伸缩活动在 0 点 0 分执行完成，冷却时间是 15 分钟。假如这时用户停用再启用伸缩组，冷却时间效果失效，即如果在 0 点 3 分云监控触发伸缩活动，请求将被接受执行。

伸缩组状态

伸缩组状态

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

状 态	Open 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

伸缩活动的流程

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

判断伸缩组的健康状态和边界条件。

分配 ActivityId 和执行伸缩活动。

创建 ECS 实例。

修改 Total Capacity。

分配 IP。

添加 RDS 白名单。

启动 ECS 实例，等待启动完成。

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**。

伸缩活动完成，启动 cooldown。

自动收缩

判断伸缩组的健康状态和边界条件。

分配 ActivityId 和执行伸缩活动。

从负载均衡移除 ECS 实例。

停止 ECS 实例，等待停止完成。

从 RDS 白名单移除。

释放 ECS 实例。

修改 Total Capacity。

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

判断伸缩组的健康状态、边界条件和 ECS 实例的状态、类型。

分配 ActivityId 和执行伸缩活动。

加入 ECS 实例。

修改 Total Capacity。

添加 RDS 白名单。

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”。

注意：目前弹性伸缩要求手工加入实例时，实例的规格必须与伸缩组中已激活的伸缩配置上的规格相同，所以此处使用了伸缩配置上指定的“负载均衡权重”。

伸缩活动完成，启动 cooldown。

手工移出

判断伸缩组的健康状态和边界条件。

分配 ActivityId 和执行伸缩活动。

负载均衡停止向该 ECS 实例转发流量。

等待 60 秒后，从负载均衡移除 ECS 实例。

从 RDS 白名单移除。

修改 Total Capacity。

从伸缩组移出。

伸缩活动完成，启动 cooldown。

伸缩活动的状态

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。

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。

移出不健康 ECS 实例

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。

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。

强制干预

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 权限授权予弹性伸缩。	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。
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 权限授权予弹性伸缩。	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。

数量限制

每个用户所能创建的伸缩组、伸缩配置、伸缩规则、伸缩 ECS 实例、定时任务的数量都有一定的限制。

一个用户最多能创建 20 个伸缩组。

- 一个伸缩组内最多只能创建 10 个伸缩配置。
- 一个伸缩组内最多只能创建 10 个伸缩规则。

对于所有地域和所有伸缩组，一个用户最多能弹性伸缩 100 台 ECS 实例。此数量只包含自动创建的 ECS 实例，不包含手工添加的 ECS 实例。

一个用户最多能创建 20 个定时任务。

伸缩组

创建伸缩组

伸缩组（Scaling Group）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。

ECS 实例移出伸缩组的策略参考移出策略。

参数描述

根据传入参数创建伸缩组。

伸缩最大实例数（MaxSize）、伸缩最小实例数（MinSize）定义了伸缩组内 ECS 实例个数的最大值和最小值。

- 当伸缩组的当前 ECS 实例数（Total Capacity）小于 MinSize 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 MinSize。
- 当伸缩组的当前 ECS 实例数（Total Capacity）大于 MaxSize 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 MaxSize。

默认冷却时间（DefaultCooldown）是伸缩组的默认冷却时间。

- 一个伸缩活动（添加或移出 ECS 实例）执行完成后的一段冷却时间内，该伸缩组不执行其它的伸缩活动。
- 目前仅针对云监控报警任务触发的伸缩活动有效。

移出策略（RemovalPolicy）是指当需要从伸缩组移出 ECS 实例并且有多种选择时，则按该策略选择需要移出的 ECS 实例。

负载均衡注意事项

如果在伸缩组中指定了负载均衡实例，伸缩组会自动将加入伸缩组的 ECS 实例添加到指定的负载均衡实例当中。

- 指定的负载均衡实例必须是已启用状态。
- 指定的负载均衡实例所有配置的监听端口必须开启健康检查，否则创建失败。
- 加入负载均衡的 ECS 实例的权重默认为 50。

RDS注意事项

如果在伸缩组中指定了 RDS 实例，伸缩组会自动将加入伸缩组的 ECS 实例的内网 IP 添加到指定的 RDS 实例的访问白名单当中。

- 指定的 RDS 实例必须是运行中状态。
- 指定的 RDS 实例访问白名单的 IP 个数不能达到上限。
- 伸缩组创建成功后，伸缩组不会立即生效，只有启用伸缩组，才能接受伸缩规则的触发和执行相关的伸缩活动。

限制

伸缩组、负载均衡实例和 RDS 实例必须在同一个地域。

一个用户最多只能创建 20 个伸缩组。

操作步骤

在 **伸缩组列表** 页，单击 **创建伸缩组**。

弹性伸缩服务 ESS		伸缩组管理列表				刷新	创建伸缩组
伸缩组管理		深圳	北京	杭州	青岛		
自动触发任务管理	伸缩组	状态	伸缩配置	总实例数	最小实例数	最大实例数	默认冷却时间(秒)
整体方案创建管理	33	启用	33	0	0	0	300
帮助手册	22	启用	22	0	0	0	300
常见问题	11	启用	11	0	0	0	300
	多组	停用	111	63	0	0	300
							管理 修改 停用 删除

在 **创建伸缩组** 页面，选择伸缩组所在的地域，并填写伸缩组名称。

如果将伸缩 **最大实例数** 和 **伸缩最小实例数** 设置为 **1**，则在整体伸缩方案创建完成后，将自动创建 1

台 ECS 实例。

根据需要选择负载均衡实例和选择 RDS 数据库实例。

创建伸缩组

* 伸缩组名称 :

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"-"

* 伸缩最大实例数 (台) ? :

1

最小为0，最大为200

* 伸缩最小实例数 (台) ? :

1

最小为0，最大为200

* 默认冷却时间 (秒) ? :

300

最小为0，必须为整数

移出策略 ? :

先筛选

最早伸缩配置对应的实例

在结果中再筛选

最早创建的实

移出

如何保证手工添加的ECS实例不被移出伸缩组

负载均衡 ? :

选择负载均衡

数据库 ? :

选择数据库

提交

取消

选择完成后，单击 **确认**，然后单击 **提交**。

修改伸缩组

修改伸缩组的属性。

在创建伸缩组的属性中，有以下属性不可以修改：

- 地域
- 负载均衡
- RDS 数据库实例

当伸缩组为生效 (Active) 和失效 (Inactive) 状态，才可以执行此功能。

当伸缩组的 ECS 实例数 (Total Capacity) 不满足修改后的 MaxSize 或 MinSize，弹性伸缩服务会自动加入或移出 ECS 实例，使得伸缩组的 ECS 实例数等于 MaxSize 或 MinSize。

示例

修改伸缩组

*伸缩组名称：

33

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"."

*伸缩最大实例数（台）：

0

最小为0，最大为200

*伸缩最小实例数（台）：

0

最小为0，最大为200

⚠️最小实例数设置为0将存在移出所有实例的可能（包括手工加入的ECS实例）

*默认冷却时间（秒）：

300

最小为0，必须为整数

移出策略：

先筛选

最早伸缩配置对应的实例

在结果中再筛选

最早创建的实例

移出

如何保证手工添加的ECS实例不被移出伸缩组

负载均衡：

-

数据库：

-

提交

取消

查询伸缩组

伸缩组有三种状态：

- 启用（Enable）：该状态下伸缩组接收伸缩规则请求并触发伸缩活动。
- 停用（Disable）：该状态下伸缩组不接收任何伸缩规则请求。
- 删除中（Deleting）：该状态下伸缩组不接收任何伸缩规则请求。

操作步骤

- 登录 弹性伸缩管理控制台。
- 在左侧导航栏中，单击 伸缩组管理。
- 选择地域。
- 查看伸缩组状态。



启用伸缩组

启用一个指定的伸缩组。

启用伸缩组成功以后（ active 状态 ），如果当前 ECS 实例数量小于 MinSize ，则弹性伸缩服务会自动创建差额的按量付费的 ECS 实例。

例如，创建伸缩组时，指定 MinSize = 5 ，启用前伸缩组已有 2 台 ECS 实例，则启动后会自动创建 3 台 ECS 实例。

当伸缩组为失效（ Inactive ）状态，才可以执行此功能。

当伸缩组有生效的伸缩配置时，才可以执行此功能。

一个伸缩组在同一时刻只能有一个 **生效** 的伸缩配置。

示例

伸缩组管理列表							刷新 创建伸缩组	
深圳 北京 杭州 青岛								
伸缩组	状态	伸缩配置	总实例数	最小实例数	最大实例数	默认冷却时间(秒)	操作	
33	启用	33	0	0	0	300	管理	修改 停用 删除
22	启用	22	0	0	0	300	管理	修改 停用 删除
11	启用	11	0	0	0	300	管理	修改 停用 删除
多组	停用	111	63	0	0	300	管理	修改 启用 删除
单组3	停用	111	40	0	0	300	管理	修改 启用 删除
单组2	停用	111	40	0	0	300	管理	修改 启用 删除
单组1	停用	111	4	0	0	300	管理	修改 启用 删除
暴力	停用	伸缩配置	0	0	100	300	管理	修改 启用 删除
共有8条， 每页显示：20条							<< < 1 > >>	

停用伸缩组

停用一个指定的伸缩组。

- 只有在当前伸缩组没有任何伸缩活动进行时，才能停用伸缩组。
- 当伸缩组为生效（Active）状态，才可以执行此功能。

示例

伸缩组管理列表 深圳 北京 杭州 青岛 刷新 创建伸缩组

伸缩组	状态	伸缩配置	总实例数	最小实例数	最大实例数	默认冷却时间(秒)	操作
33	启用	33	0	0	0	300	管理 修改 停用 删除
22	启用	22	0	0	0	300	管理 修改 停用 删除
11	启用	11	0	0	0	300	管理 修改 停用 删除

删除伸缩组

删除一个指定的伸缩组。

强制删除（ForceDelete）属性表示如伸缩组存在 ECS 实例或正在进行伸缩活动，是否强制删除伸缩组并移出和释放 ECS 实例。

- 只有 Open API 才会看到此属性
- 控制台删除一个伸缩组时，会默认采用强制删除的模式

如果 Forcedelete 属性为 false，必须满足以下两个条件，才能删除伸缩组：

伸缩组没有任何伸缩活动正在执行。

伸缩组当前的 ECS 实例数量（Total Capacity）为 0。

满足以上条件，会先停止伸缩组，最后再删除伸缩组。

当 Forcedelete 属性为 true 时：

先停止伸缩组，拒绝接收新的伸缩活动请求。

然后等待已有的伸缩活动完成。

最后将伸缩组内所有 ECS 实例移出伸缩组，并删除伸缩组。您手工添加的 ECS 实例会被移出伸缩组，弹性伸缩自动创建的 ECS 实例会被自动删除。

删除伸缩组，包含删除相关联的伸缩配置、伸缩规则、伸缩活动、伸缩请求的信息。

删除伸缩组，不会删除以下任务或实例：定时任务、云监控报警任务、负载均衡实例、RDS 实例。

示例

伸缩组管理列表

深圳北京杭州青岛

刷新

创建伸缩组

伸缩组	状态	伸缩配置	总实例数	最小实例数	最大实例数	默认冷却时间(秒)	操作
33	启用	33	0	0	0	300	管理 修改 停用 删除
22	启用	22	0	0	0	300	管理 修改 停用 删除
11	启用	11	0	0	0	300	管理 修改 停用 删除
多组	停用	111	63	0	0	300	管理 修改 启用 删除
单组3	停用	111	40	0	0	300	管理 修改 启用 删除

伸缩配置

查询伸缩配置

查询伸缩配置的信息。伸缩配置具有以下状态（LifecycleState）：

- Active：生效状态，伸缩组会采用生效状态的伸缩配置自动创建 ECS 实例。
- Inactive：失效状态，该伸缩配置存在于伸缩组中，但伸缩组不会采用失效状态的伸缩配置自动创建 ECS 实例。

示例

在伸缩配置列表页，单击 **查看详情**。



创建伸缩配置

在您的业务需求增长时，弹性伸缩会根据您自定义的条件，在伸缩组中为您自动添加ECS实例个数。您可以通过**创建伸缩配置**，自定义某一伸缩组中需要自动添加的ECS实例。

说明：

- 伸缩配置一旦创建成功，不可以编辑更改、不过您可以创建一个新的伸缩配置，来替换不需要的伸缩配置。
- 依据被替换的伸缩配置而添加进组的ECS实例仍可以正常运行。
- 一个伸缩组内最多只能创建 10 个伸缩配置，且仅能有1个伸缩配置处于激活状态。

操作步骤

登录弹性伸缩管理控制台。

在**伸缩组管理**中，选择您需要创建伸缩配置的伸缩组。

在左侧导航栏中，选择 **伸缩配置**，点击 **创建伸缩配置**。



在弹出的**创建伸缩配置**页面中，填写伸缩配置名称。

在列出的ECS实例的各参数中选择您需要的选项。

22

说明：

- 在**镜像类型**中，如果您需要实现自动启动 Web 服务器、自动下载代码和脚本等功能，请选择**自定义镜像**。
- 如果您需要使用云监控触发的报警任务，请为该镜像安装云监控插件。

请参见云监控插件安装指南。

删除伸缩配置

您可以删除一个指定的伸缩配置。操作比较直观，不再赘述。

注意以下情况：

- 伸缩配置在伸缩组中属于生效状态，则该伸缩配置不允许删除。
- 某个伸缩配置创建的任意一个 ECS 实例仍存在于伸缩组中，则该伸缩配置不允许删除。

伸缩规则

创建伸缩规则

伸缩规则（Scaling Rule）定义了具体的扩展或收缩操作，例如加入或移出 N 个实例。如果伸缩规则的执行会造成伸缩组的 ECS 实例数低于 MinSize 或高于 MaxSize 时，则弹性伸缩会自动调整需要加入或移出的 ECS 实例数，使之按照 **将伸缩组的实例数调整到 MinSize** 或 **将伸缩组的实例数调整到 MaxSize** 的规则执行。

例如：某个伸缩组，MaxSize = 3，当前实例数 Total Capacity = 2，伸缩规则指定 **加 3 台 ECS 实例**，则在实际执行过程中只会 **加 1 台 ECS 实例**。（伸缩规则当中的值不会调整）

例如：某个伸缩组，MinSize = 2，当前实例数 Total Capacity = 3，伸缩规则指定 **减去 5 台 ECS 实例**，则在实际执行过程中只会 **减 1 台 ECS 实例**。（伸缩规则当中的值不会调整）

说明

根据传入参数创建伸缩规则。

当调整类型（AdjustmentType）是 **调整至指定实例数（TotalCapacity）** 时，对应的 AdjustmentValue 值必须大于等于 0。

当调整类型（AdjustmentType）是 **按百分比调整（PercentChangeInCapacity）**，弹性伸缩服务以伸缩组当前实例数（Total Capacity）* AdjustmentValue/100，并使用四舍五入原则来确认增加或减少的 ECS 实例个数。

当伸缩规则指定了冷却时间（Cooldown），则执行该伸缩规则的伸缩活动完成后，会以伸缩规则中指定的冷却时间对伸缩组进行冷却处理，如果伸缩规则未指定冷却时间，则以伸缩组指定的冷却时间（DefaultCooldown）为准。

一个伸缩组内最多只能创建 10 条伸缩规则。

创建成功后，会生成伸缩规则唯一标识符（ScalingRuleAri），主要可以被以下 Open API 接口所使用：

在执行伸缩规则（ExecuteScalingRule）的 ScalingRuleAri 参数中指定，您可以手工执行该伸缩规则。

在创建定时任务（CreateScheduledTask）的 ScheduledAction 参数中指定，您可以定时执行该伸缩规则。

在云监控新增报警规则 PutAlarmRule 的 AlarmActions 参数中指定，您可以根据各种性能指标动态的执行该伸缩规则。

操作步骤

在 **伸缩规则** 页，单击 **创建伸缩规则**，弹出创建伸缩规则对话框。



填写规则名称和名称。

在 创建伸缩规则 对话框中，单击 创建伸缩规则，将在 伸缩规则列表 页展示该伸缩规则。

创建伸缩规则

*规则名称：

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"-

*规则：

调整至

台

"增加"或"减少"选项最多只能支持100台，否则报错

"调整至N台"以及"增加或减少N%"一次只能触发100台实例的伸缩活动

冷却时间（秒）：

创建伸缩规则

取消

修改查询删除伸缩规则

修改、查询、删除伸缩规则的按钮如图所示。操作比较直观，不再赘述。

基本信息

ECS实例列表

伸缩活动

伸缩配置

伸缩规则

伸缩规则

创建伸缩规则

每个伸缩组中，您最多可以拥有10个伸缩规则。

伸缩规则	调整类型	调整值	冷却时间	操作
test	调整至	20台	300秒	查看详情 执行 修改 删除

共有1条， 每页显示：20条

触发任务

执行伸缩规则

执行一个指定的伸缩规则。执行此功能必须同时满足以下条件：

- 当伸缩组为生效（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。

关于执行的规则：

- 伸缩规则需要增加的 ECS 实例数加上当前伸缩组的实例数（Total Capacity）大于 MaxSize 时，则按 $Total\ Capacity = MaxSize$ 的规则进行执行。
- 如果当前伸缩组的实例数（Total Capacity）减去该伸缩规则需要减少的 ECS 实例数小于 MinSize 时，则按 $Total\ Capacity = MinSize$ 的规则进行执行。

对于所有地域和所有伸缩组，一个用户最多能弹性伸缩 100 台 ECS 实例。此数量只包含自动创建的 ECS 实例，不包含手工添加的 ECS 实例。

示例



添加ECS实例

往指定的伸缩组添加 ECS 实例。支持包年包月和按量付费两种类型。加入的 ECS 实例的限定条件如下：

- 必须与伸缩组在同一个地域。
- 规格（InstanceType）必须与生效伸缩配置的实例规格完全一致。
- 状态必须是 **运行（running）** 状态。
- 不能已加入到其它伸缩组中。

可以为经典网络类型或 VPC 类型，但有以下限制：

- 当伸缩组为经典网络类型时，只能添加经典网络类型的 ECS。
- 当伸缩组为 VPC 类型时，只能添加同一 VPC 下的 ECS。

执行此功能，必须同时满足以下条件：

- 当伸缩组为生效（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。

如果该功能指定的实例数加上当前伸缩组的实例数（Total Capacity）大于 MaxSize 时，则执行失败。

手工添加的 ECS 实例不与伸缩组生效的伸缩配置进行关联。

操作步骤

在 ECS 实例列表 页，单击 添加已有实例。



移出ECS实例

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 the 'ECS实例列表' (ECS Instance List) page. At the top, there are three summary cards: '实例总数: 2' (Total instances: 2), '正在服务中的实例数: 2' (Instances in service: 2), and '正在加入实例数: 0' (Instances being added: 0). Below these is a table with columns: '云服务器名称' (ECS Instance Name), '伸缩配置' (Scaling Configuration), '状态' (Status), '健康检查状态(全部)' (Health Check Status (All)), '加入时间' (Add Time), and '操作' (Actions). The table contains two rows of instances, both with 'Running' status and 'Healthy' health check status. A red box highlights the '移出伸缩组并释放' button in the '操作' column for the first instance.

定时任务

创建定时任务

根据传入参数创建定时任务。每个用户最多能创建 20 个定时任务。

由于伸缩组正在发生伸缩活动或者伸缩组停用等原因，导致定时任务触发执行伸缩规则失败后，在重试过期时间（`LaunchExpirationTime`）内，定时任务会自动重试触发，否则放弃本次定时触发。

如果多个定时任务在相近的时间内触发执行同一个伸缩组的伸缩规则，则最早触发的定时任务会先执行伸缩活动，由于同一个伸缩组同一时刻内只能有一个伸缩活动，排在后面的定时任务会在 `LaunchExpirationTime` 内自动重试定时触发。如上一个伸缩活动完成之后，排在后面的定时任务还在 `LaunchExpirationTime` 内重试，则执行该定时任务的伸缩规则并触发相应的伸缩活动。

如果指定了同一分钟的定时任务，会选取最近创建的定时任务执行。

操作步骤

单击 **自动触发任务管理** 目录下的 **定时任务** 子目录，显示定时任务列表页。



单击 **创建定时任务**，弹出 **创建定时任务** 对话框。

填写任务名称。

填写执行时间。如果未设置重复周期，则按指定的日期和时间执行一次。如果设置了重复周期，则此属性指定的时间点，默认为周期性任务的执行时间点。

填写重复周期。

选择定时任务需要触发的伸缩规则以及所属的伸缩组。

单击 **提交**，将在 **定时任务列表** 页展示该定时任务。

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：

2015-08-08

18

:

26

*伸缩规则：

伸缩组：

-- 请选择伸缩组 --

重试过期时间(秒)：

600

重复周期设置（高级）

重复周期：

按日

每

天执行一次

重复结束时间：

2015-09-06

18

:

26

提交

取消

修改、查询、删除定时任务

描述

修改定时任务的属性、查询定时任务的信息、删除一个指定的伸缩规则



报警任务

创建报警任务

根据传入参数创建报警任务。报警任务名称在用户账号下唯一。如果该报警任务名称已经存在，则用新的值进行替换。

操作步骤

单击 **自动触发任务管理** 目录下的 **报警任务** 子目录，显示报警任务列表页。



单击 **创建报警任务**，弹出创建报警任务对话框。

填写任务名称。

选择需要监控的伸缩组。

填写监控项。

填写统计周期。统计周期的粒度越小，则触发报警的机制越敏感。

填写统计方法。

填写重复几次后报警。

选择报警触发的伸缩规则。

单击 **提交**，将在 **报警任务列表** 页展示该报警任务。

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。 <http://jiankong.aliyun.com/readme.htm>

*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：

33

*监控项：

CPU

统计周期(分钟)：

2

*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：

test

提交

取消

修改查询删除报警任务

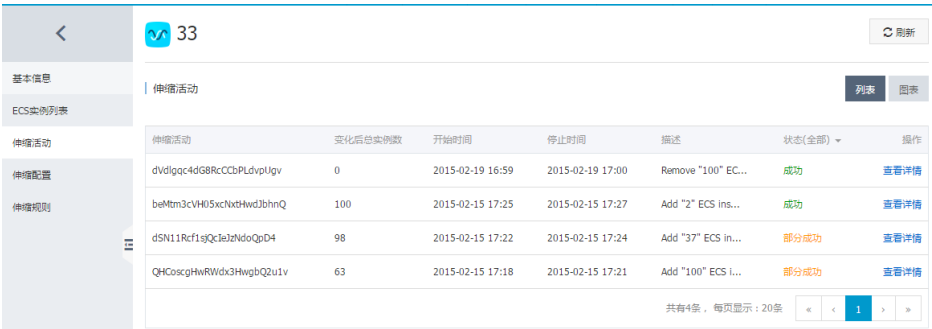
修改报警任务的属性、查询报警任务的信息、删除一个指定的伸缩规则。



查看伸缩活动

查询伸缩活动的信息，最多只能返回30天之内的伸缩活动。

示例



查询ECS实例列表

查看 ECS 实例列表

查询伸缩组内 ECS 实例列表。

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** 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。
- **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。

ECS 实例在伸缩组中的生命周期

ECS 实例在伸缩组中的生命周期，通过以下几种状态描述：

- 加入中（ 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。
- 服务中（ InService ） – 表示 ECS 实例已成功加入伸缩组，并正常提供服务。
- 移出中（ Removing ） – 表示 ECS 实例正在移出伸缩组。

ECS 实例的健康状态

ECS 实例在伸缩组中的健康状态为：

- 健康（ Healthy ）
- 不健康（ Unhealthy ）

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。

示例

下图是一个示例。

