

弹性伸缩服务ESS

什么是弹性伸缩ESS

- 背景
 - 配置如何取舍，是一步到位满足峰值吗？
 - 采购数量过大，资源浪费如何避免？
 - 传统服务器架构无法承担增加的计算需求：只能购买新的服务器资源
 - 估算峰值-采购硬件-安装硬件-运维=成本控制
- 概念
 - 弹性扩展、弹性收缩、弹性自愈
 - 可以调整弹性计算资源大小，可满足业务变化。可以根据设置伸缩策略，业务需求增长的时候自动增加计算能力，在业务需求下降的时候自动减少ECS实例
- 特点
 - 按需应变
 - 恰到好处分配资源
 - 无需担心预测的准确性
 - 无需担心突增的业务变化
 - 自动化
 - 无需人工干预
 - 自动创建和释放ECS实例
 - 自动配置SLB和RDS访问白名单
 - 丰富的伸缩模式
 - 多模式兼容，可同时支持定时、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
 - 满足各种复杂的应用场景需求

相关概念和计费模式

- 名词解释
 - 弹性伸缩：可以调整弹性计算资源大小，可满足业务变化。可以根据设置伸缩策略，业务需求增长的时候自动增加计算能力，在业务需求下降的时候自动减少ECS实例；一种服务
 - 伸缩组：具有相同应用场景的ECS实例集合。伸缩组定义了组内ECS实例最大和最小值；以及其负载均衡实例和RDS实例。
 - 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：扩展/收缩；
 - 伸缩活动：伸缩规则触发后，产生伸缩活动，用来描述ECS实例变化情况
 - 冷却时间：同一个伸缩组中，一个伸缩活动执行完成后属性的时间。这段时间内，这个伸缩组不执行其他伸缩活动。
- 概念
 - 伸缩组（管理生命周期）
 - 伸缩配置、伸缩规则、伸缩活动
 - 删除伸缩组会删除
 - 伸缩触发任务
 - 定时任务：定时任务独立于伸缩组。不依赖伸缩组生命周期管理，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，不依赖伸缩组生命周期管理，删除伸缩组不会删除定时任务。
 - ESS伸缩规则
 - 扩展/收缩 操作，根据定义的MinSize和MaxSize
 - 例子：Max Size=3，实例为2，伸缩规则的定义为加3台ECS实例，则实际执行过程中，只会增加一台
 - 例子：Max Size=2，实例为3，伸缩规则的定义为减少5台ECS实例，则实际执行过程中，只会减少一台
 - 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
 - 特点：创建过程中不能够停止。
 - 伸缩活动有ECS实例加入伸缩组失败时，需要保持ECS实例级完整程度。例如伸缩组
 - 冷却时间
 - 收费：弹性伸缩借助阿里云的RAM服务，通过ECS Open API代替用户弹性伸缩ECS实例资源，所以回滚的资源仍然会被扣费。
 - 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请求；定时任务可以绕过冷却时间。
 - 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。
 - 最后一个ECS实例接入/移除完成后，整个伸缩组冷却时间才开始倒计时。
 - 停用伸缩组，然后再启用，冷却时间效果将会失效。
- 概念之间的关系
 - 弹性伸缩免费：通过弹性伸缩自动创建或手工加入的ECS实例，需要按照ECS相关实例类型进行付费。
 - 按量付费ECS关机后仍收取费用，只有释放后不再收费
 - 大型活动期间，存在库存不足的问题

ESS的伸缩模式和限制

- 伸缩模式
 - 定时模式：自定义伸缩时间和频率
 - 动态模式：基于云监控性能指标（CMS），自动增加/减少实例
 - 固定数量模式：通过设置最小的实例数量，来保证整体业务可用
 - 自定义模式：通过API调用已有的监控系统，手动添加伸缩。
 - 健康模式：检测ECS状态，如果为非运行/非正常状态，ESS可以释放ECS，同时添加一台正常的ECS实例
 - 多模式并行：以上模式可以组合进行。
- ESS的限制
 - 伸缩组内部署在ECS实例的应用必须是无状态且可横向扩展。
 - 弹性伸缩是无法自动添加ECS实例到开放缓存（数据库中的产品）实例报名单中，需要自行添加
 - ESS不支持纵向扩展；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。
 - 数量限制：伸缩组：20个，每个伸缩组内可创建十个伸缩配置以及50个伸缩规则
 - ECS实例：一千台（只包含自动创建的ECS实例，不包含手工添加的ECS实例）
 - 定时任务：每个账户只能拥有20个。

ESS操作使用

- 创建伸缩组
 - 指定的SLB实例必须是已启用状态
 - 指定的SLB实例所配置的监听端口必须开启健康检查，否则创建失败
 - 加入SLB的ECS实例的权重默认为50
 - 指定的RDS实例必须是运行状态中的
 - 指定的RDS访问白名单IP个数不能到达上限
 - 创建成功后不会立即生效，只有启用伸缩组后才能接受伸缩组规则触发和执行相关的伸缩活动
- 修改伸缩组
 - 伸缩组、负载均衡、RDS数据库实例不可修改
 - 伸缩组为生效、失效、和删除三种状态，只有在生效和失效状态下才可以修改伸缩组属性
 - 当伸缩组的ECS实例数不满足修改后的MAX和MIN，弹性伸缩服务会自动加入或移出RCS实例，使得ECS实例数等于MAX或者小于MIN
- 查询伸缩组
- 启用伸缩组
 - 伸缩组为失效状态，才可以启用
 - 当伸缩组失效的伸缩配置时，才可以启用伸缩组
 - 一个伸缩组在同一个时刻只能有一个生效的伸缩配置
- 停用伸缩组
 - 没有任何伸缩活动时，才能停用
 - 伸缩组为生效状态，才能够停用
- 删除伸缩组
 - 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、伸缩请求信息等
 - 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
 - 控制台删除，采用强制删除方式
 - 通过Open API方式删除伸缩组时，可以看到强制删除的属性，如果伸缩组ECS实例或者正在进行伸缩活动，是强制删除伸缩组并移出释放ECS实例。

伸缩组配置创建和维护

- 创建伸缩配置
 - 一旦创建成功不能更改；只能通过创建新的来替换
 - 被替换的伸缩配置而添加进组的ECS实例可以正常进行
 - 一个伸缩组最多只能创建十个伸缩配置，且只能有一个配置处于激活状态
- 删除伸缩配置
 - 生效状态中不允许删除
 - 某个伸缩配置创建的ECS实例仍存在于伸缩组中，这个伸缩配置不允许删除。

设置伸缩规则和查看伸缩活动

- 创建伸缩规则
 - 伸缩规则定义了伸缩组具体的扩张还是收缩操作规则
 - 设置MinSize和MaxSize
- 修改、删除伸缩规则
 - 支持按照比例扩展、删减
 - 伸缩活动结束后，冷却时间
- 弹性裸金属：就是一个ECS大规格的服务器，允许再虚拟化

事件通知和生命周期挂钩

- 事件通知
 - 定义：事件通知，当伸缩组发生指定的伸缩事件或资源变化，弹性伸缩会发送消息通知云监控或者信息服务MNS。
 - 消息服务MNS还没有覆盖所有地域
 - 事件通知有数量限制，为6
 - 使用新的事件通知功能后，原有的告警失效
 - 事件通知类型
 - 成功的弹性扩展活动
 - 成功的弹性收缩活动
 - 失败的弹性扩展活动
 - 失败的弹性收缩活动
 - 拒绝弹性收缩活动（冷却时间没到的时候）
- 生命周期挂钩
 - 当伸缩组发生伸缩活动时，生命周期挂钩会暂停伸缩活动一段时间，伸缩活动进入等待状态
 - 应用场景
 - 扩张活动时：可以在等待期间初始化ECS实例配置
 - 收缩活动时：获取ECS实例数据
 - 作用时机
 - 扩张活动：加入RDS、SLB实例IP白名单之前
 - 收缩活动：移除RDS、SLB实例IP白名单之前
 - 每个伸缩组最多创建6个生命周期挂钩

定时任务和报警任务

- 定时任务
 - 每个账户最多能创建20个定时任务
 - 由于伸缩组停用导致定时任务出发执行伸缩失败后，在重试过期时间内，定制任务会自动重新出发，否则会放弃本次触发
 - 同时定义了多个定时任务，执行了同一个伸缩组的伸缩规则后，则最早触发的定时任务会先执行伸缩活动；注意：同一个伸缩组同一时刻只能有一个伸缩活动，排在后面的定时任务会在重试过期时间内自动重试定时触发。过期则不触发
 - 如果指定了同一时间的定时任务，则会优先执行先创建的
- 报警任务
 - 报警任务是基于云监控服务（CMS）运行的
 - 如果报警任务名称已经存在，则用新的值对以前的值进行替换
 - 在使用报警任务之前，需要在ECS镜像里面安装新版本的云监控Agent

阿里云ESS 和API

- 服务地址：ess.aliyuncs.com
- 通信协议：支持HTTP/HTTPS，推荐HTTPS
- 请求方式：支持HTTP GET方式发送请求
- 字符编码：请求及返回结果都是用UTF-8
- 返回格式：XML和JSON，默认JSON