弹性伸缩

产品简介

产品文档

腾讯云

文档目录

产品简介
产品概述

最近更新时间：2024-01-08 17:53:29

什么是弹性伸缩 AS？

弹性伸缩 AS（Auto Scaling）可以根据您的业务需求和策略，自动调整 CVM 计算资源，确保您拥有适量的 CVM 实例来处理您的应用程序负载。对于您的 Web 服务而言，智能的扩展和收缩是成本控制和资源管理的重要组成部分。Web 应用程序开始获得更多请求流量时，您将添加更多的服务器来应对额外负载。同时，当 Web 应用程序的流量开始减少时，您将终止未充分利用的服务器。

如果使用 AS 进行容量调整，您只需事先设置好扩容条件及缩容条件。AS 会在达到条件时自动增加使用的服务器数量以维护性能；在需求下降时，AS 会根据您的缩容条件减少服务器数量，最大限度地帮助您降低成本。

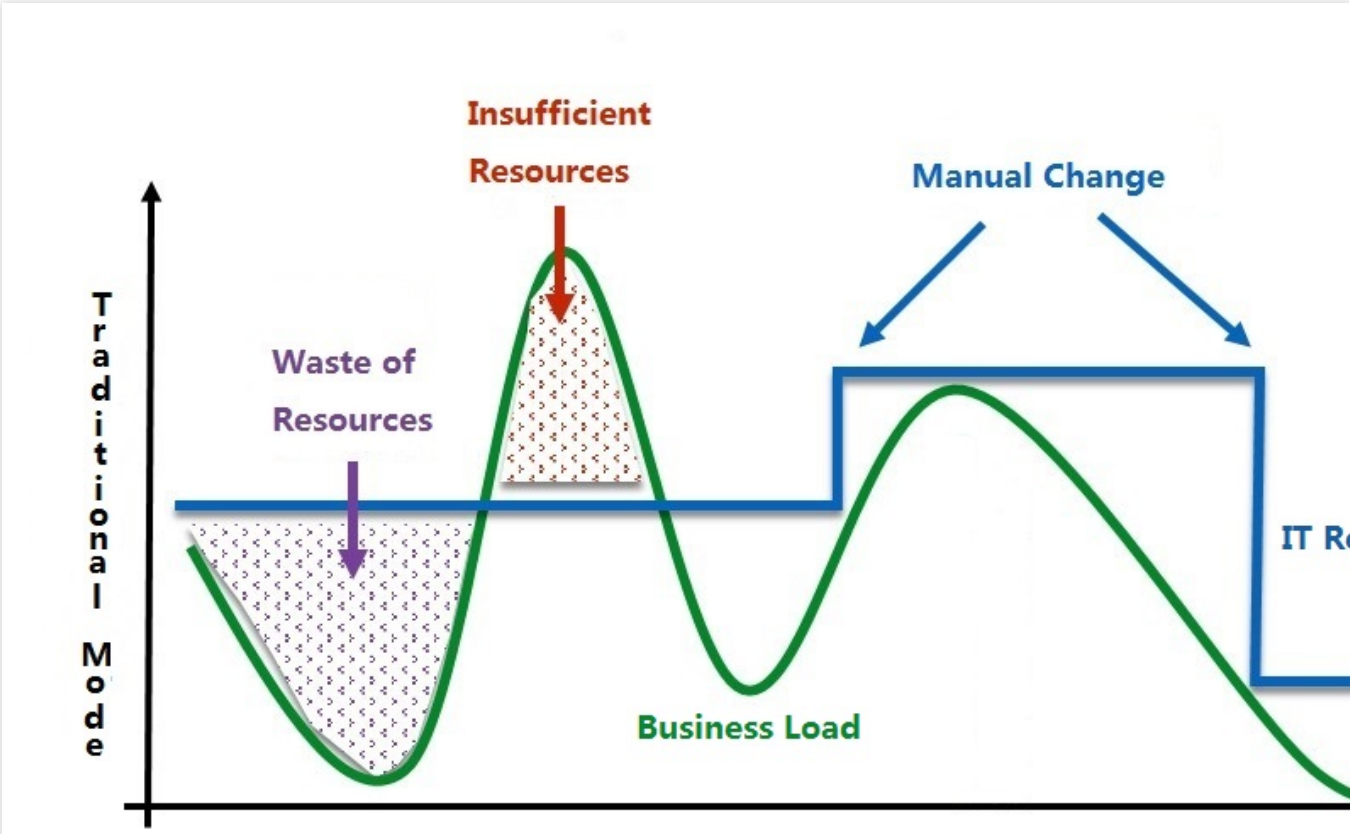
如下图对比所示，通过使用弹性伸缩 AS，您的集群可以永远保留恰到好处的资源量，并处于健康状态。您将告别传统模式下的多种烦恼：

业务突增或 CC 攻击导致机器数量不足，以致您的服务无响应。

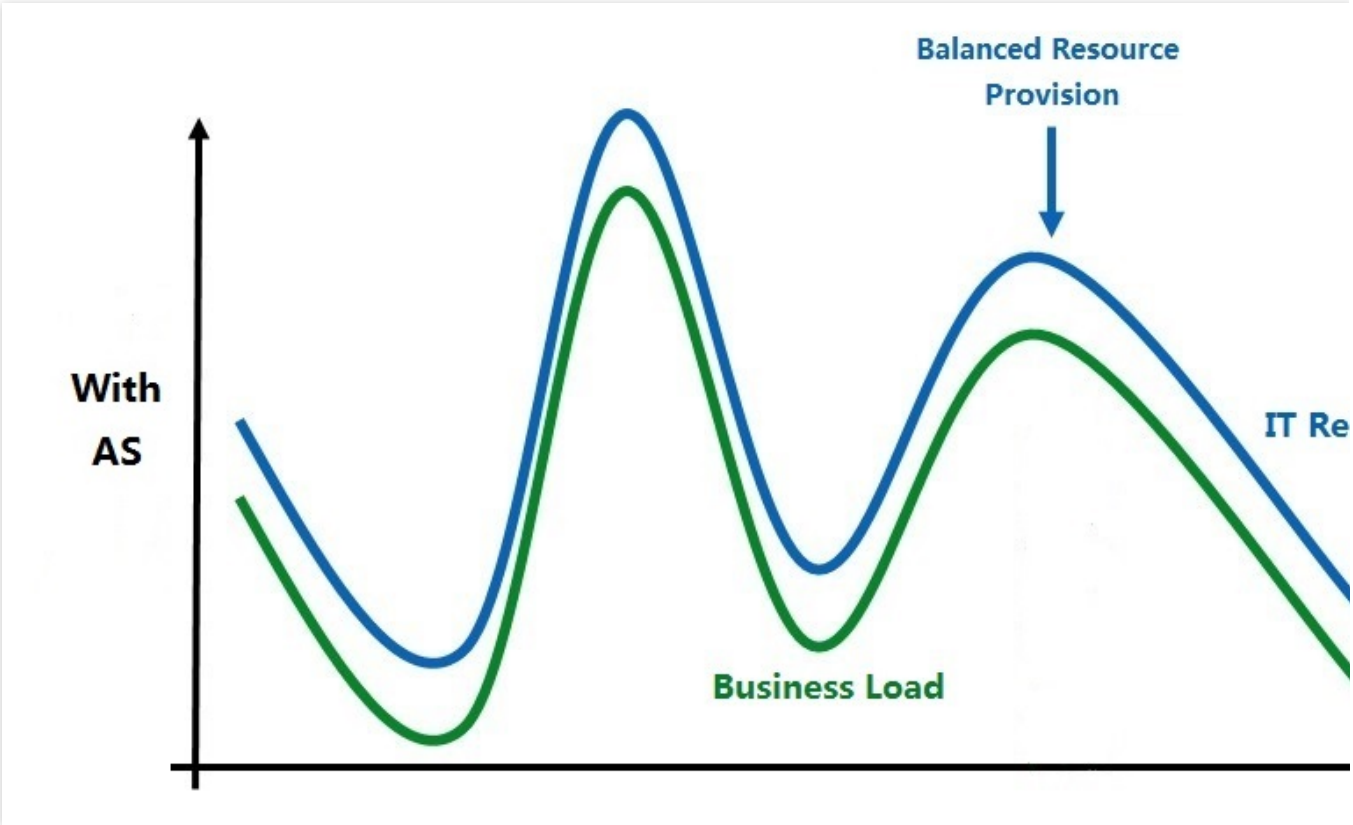
按高峰访问量预估资源，而平时访问量很少达到高峰，造成投入资源浪费。

人工守护及频繁处理容量告警，需要多次手动变更。

传统模式下的集群维护：


Insufficient
Resources
Manual Change
Waste of
Resources
Traditional Mode
Business Load
IT R

采用AS后的效果：


Balanced Resource
Provision
With
AS
IT Re
Business Load

AS的工作方式

在常见的 Web 应用服务中，您的集群通常运行应用程序的多个副本来满足客户流量。例如接入层的前端服务器集群、逻辑层的应用服务器集群、后端的缓存服务器集群。每个实例都可以处理客户请求。
这些类似或相同的实例，数量通常是可调节的。您可以将这些相同或类似的机器归到一个伸缩组中管理起来：
您可以指定每个伸缩组中最少的实例数量，AS 会确保伸缩组中的实例永远不会低于这个数量；
您可以指定每个伸缩组中最大的实例数量，AS 会确保伸缩组中的实例永远不会高于这个数量；
您可以指定伸缩策略，则 AS 会在应用程序需求增加或降低时启动或终止实例。伸缩策略有两类：
a. 告警触发策略：根据指定条件动态扩展（例如：伸缩组的机器的CPU 利用率超过60%时扩展）
b. 定时伸缩策略：根据指定的时间扩展（例如：每晚21：00扩展）
设置完策略后，您还可以设置伸缩活动通知。AS 会在发生伸缩活动时通过邮件、短信、站内信方式告知您。您不需要时刻关注您的业务请求量变化，只需要留意 AS 的通知即可。
您也可以在任何时候一键指定所需要的机器数量，或者把已有的机器加入到伸缩组中一起管理。

AS的基本概念

弹性伸缩产品有以下基本概念：
伸缩组
启动配置
伸缩策略
冷却时间

1. 伸缩组

伸缩组是遵循相同规则、面向同一场景的云服务器实例的集合。伸缩组定义了组内 CVM 实例数的最大值、最小值及其相关联的负载均衡实例等属性。

2. 启动配置

启动配置是自动创建云服务器的模版，其中包括镜像ID、云服务器实例类型、系统盘及数据盘类型和容量、密钥对、安全组等。
创建伸缩组时必须指定启动配置，启动配置一经创建后其属性将不能编辑。

3. 伸缩策略

即执行伸缩动作的条件。触发条件可以是时间或云监控的报警，动作可以是移出或加入 CVM。

伸缩策略有以下两种：

定时伸缩策略

到达某个固定时间点，自动增加或减少 CVM 实例，支持周期性重复。

告警伸缩

基于云监控指标（如CPU、内存、网络流量），自动增加或减少 CVM 实例。

4. 冷却时间

冷却时间是指在同一个伸缩组内，一个伸缩活动（添加或移出 CVM 实例）执行完成后的一段锁定时间。在这段时间内，该伸缩组不执行伸缩活动。冷却时间可指定范围为 0 - 999999（秒）。

产品优势

最近更新时间：2024-01-08 17:53:29

优势	使用弹性伸缩 AS	不使用弹性伸缩 AS
自动化	**自动伸缩实例，无需人工干预** 弹性伸缩根据业务负载情况动态实时自动创建和释放CVM实例，帮助您以最合适的实例数量应对业务情况，全程无需人工干预，为您免去人工部署负担。 例如，您可以设置一个伸缩策略，当 CPU 利用率较高时，就向伸缩组添加新的 CVM 实例，新增的 CVM 实例秒级计费；同样，您也可以设置一个策略，在CPU使用率较低时从伸缩组删除实例；如果您的负载变化情况是可以预知的，则可以设置定时任务，对您的扩展活动进行规划。 新增实例还可直接关联已有负载均衡CLB，以使伸缩组新增的实例承担分发流量，提高服务可用性；您还可以向管理员发送告警，帮助您及时关注异常情况。	**繁琐的手动操作** 手动创建、销毁资源，需手动配置负载均衡；手动操作容易出错，影响业务。
省成本	**适量伸缩实例，节省成本** 弹性伸缩帮助您以最合适的实例数量应对业务情况，当业务需求增加时，为您无缝地自动增加适量 CVM 实例，当业务需求下降时，为您自动削减不需要的 CVM 实例，提高设备利用率，为您节省部署和实例成本。	**资源闲置带来浪费** 需预留过量的 CVM 以防资源不足影响业务。
容错性	**系统自动检测，及时容错** 弹性伸缩自动检测实例的健康状况，一旦发现异常，即自动复制出健康的实例，以替换状态异常的实例，确保您的应用程序获得预期的计算容量，为您的业务保驾护航。	**无法及时容错** 通常在发现业务中断后才能处理异常实例，影响业务可用性。

应用场景

最近更新时间：2024-01-08 17:53:29

提前部署扩缩容

用户明确何时需要扩缩容，则可提前设置 Auto Scaling 定时策略。到相应时间时，系统将自动添加或减少 CVM 实例，无需人工等待。

低成本应对业务浪涌

当客户面临访问峰值，需要提前准备服务器，预防CPU增长造成的服务器压力过大；待压力过去后再根据实际负载缩减服务器。客户可提前设置 Auto Scaling 监控策略，系统将根据设定好的业务监控指标自动判定是否需要 CVM 平行扩展。如果监控指标达到阈值，则实时自动增加或减少 CVM 实例，并自动完成负载均衡配置。既节约了成本，也无需客户时刻为手动扩容作准备。

自动替换不健康 CVM

为避免不健康的云服务器继续运行对业务造成影响，用户需要时刻关注系统中 CVM 的运行情况，并随时准备处理。使用 Auto Scaling，系统将定时对 CVM 进行健康检查，若扫描出运行异常的 CVM 实例，则自动平行扩展一台实例替换异常实例。该操作记录将被保留供用户查看。

使用限制

最近更新时间：2024-01-08 17:53:29

目前弹性伸缩已在除边缘节点地域外的所有地域上线，使用限制如下表：

限制类型	说明
一个用户在一个地域下	最多可创建50个启动配置。 最多可创建50个伸缩组。 最多能创建的 CVM 实例数量受用户 CVM 配额影响，详情请参见 购买按量计费云服务器实例限制。
一个伸缩组下	只能对应1个启动配置。 最多能弹性伸缩2000台 CVM 实例。 最多可创建100条伸缩策略，且最多可创建10个定时任务。 最多创建5个通知。 最多创建10个生命周期挂钩。
其他	伸缩组的子机数量不能超过私有网络 VPC 子网能提供的 IP 上限。 弹性伸缩目前不支持纵向扩展，即无法自动升降 CVM 的 CPU、内存和带宽。 弹性伸缩、启动配置均为地域概念，仅能在同一地域下启动或销毁 CVM 实例。 伸缩组关联的负载均衡实例（跨地域负载均衡实例则为其后端实例私有网络 VPC）必须与伸缩组在同一个网络环境（私有网络 VPC 或同一地域的基础网络）中。

访问管理

最近更新时间：2024-01-08 17:53:29

概述

访问管理（Cloud Access Management，CAM）是腾讯云提供的 Web 服务，主要用于帮助用户对腾讯云账户下资源的访问权限的安全管理。您可以通过 CAM 创建、管理和销毁用户或用户组，并使用身份管理和策略管理控制其他用户使用腾讯云资源的权限。策略能够授权或者拒绝用户使用指定资源完成指定任务，当您在使用 CAM 时，可以将策略与一个用户或一组用户关联起来进行权限控制。
弹性伸缩已接入 CAM，您可以使用 CAM 对弹性伸缩服务的相关资源进行权限控制。

相关概念

CAM 用户

CAM 用户 是您在腾讯云中创建的一个实体，每一个 CAM 用户仅同一个腾讯云账户关联。您注册的腾讯云账号身份为**主账号**，您可以通过 用户管理 来创建拥有不同权限的**子账号**进行协作。子账号的类型分为 子用户、协作者 以及 消息接收人。

策略

策略 是用于定义和描述一条或多条权限的语法规范，腾讯云的策略类型分为预设策略和自定义策略。
预设策略： 由腾讯云创建和管理的策略，是被用户高频使用的一些常见权限集合，如资源全读写权限等。预设策略操作对象范围广，操作粒度粗，且为系统预设，不可被用户编辑。
自定义策略：由用户创建的策略，允许进行细粒度的权限划分。例如，为子账号关联一条使用策略，使其有权管理弹性伸缩的伸缩组，而无权管理云数据库实例。

资源

资源（resource） 是策略的元素，描述一个或多个操作对象，例如弹性伸缩的启动配置和伸缩组。

弹性伸缩预设策略介绍

预设策略名	授权范围描述
QcloudASFullAccess	关联后，获得弹性伸缩（AS）全读写访问权限
QcloudASReadOnlyAccess	关联后，获得弹性伸缩（AS）只读访问权限

可授权的资源类型

资源级权限指的是能够指定用户对哪些资源具有执行操作的能力。例如，您可以授权用户拥有广州地域伸缩组的操作权限。
在访问管理中对弹性伸缩可授权的资源类型如下：

资源类型	授权策略中的资源描述方法
启动配置相关	`qcs::as:$region:$account:launch-configuration/*`
伸缩组相关	`qcs::as:$region:$account:auto-scaling-group/*`

下表列出弹性伸缩支持资源级权限操作的各个 API，以及每个操作支持的资源路径。
设置资源路径时，您需要将 `$region`、`$account`、`$LaunchConfigurationId`、`$AutoScalingGroupId` 等变量参数修改为您实际的参数信息，同时您也可以在路径中使用 `*` 通配符。
访问管理策略中的 region、action、account、resource 等相关概念请参见 资源描述方式。

API 接口：action	资源路径：resource
CreateLaunchConfiguration	`qcs::as:$region:$account:launch-configuration/*`
DeleteLaunchConfiguration	`qcs::as:$region:$account:launch-configuration/$LaunchConfigurationId`
DescribeLaunchConfigurations	`qcs::as:$region:$account:launch-configuration/*` `qcs::as:$region:$account:launch-configuration/$LaunchConfigurationId`
ModifyLaunchConfigurationAttributes	`qcs::as:$region:$account:launch-configuration/$LaunchConfigurationId`
UpgradeLaunchConfiguration	`qcs::as:$region:$account:launch-configuration/$LaunchConfigurationId`
CreateAutoScalingGroup	`qcs::as:$region:$account:auto-scaling-group/*`
CreateAutoScalingGroupFromInstance	`qcs::as:$region:$account:auto-scaling-group/*`
DeleteAutoScalingGroup	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
DescribeAutoScalingGroups	`qcs::as:$region:$account:auto-scaling-group/*` `qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`

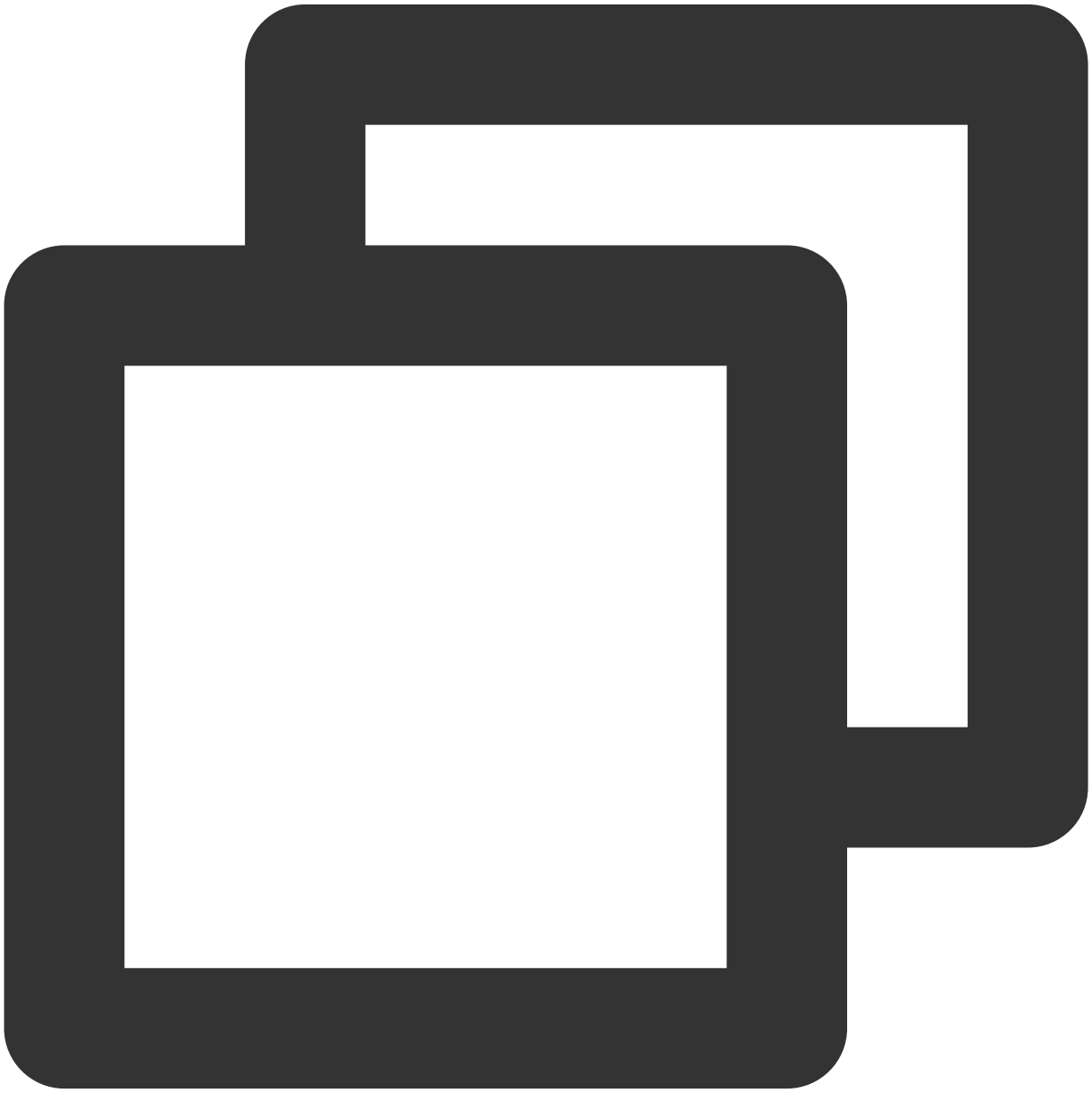
ModifyAutoScalingGroup	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
ModifyLoadBalancers	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
EnableAutoScalingGroup	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
DisableAutoScalingGroup	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
ModifyDesiredCapacity	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
DescribeAutoScalingActivities	`qcs::as:$region:$account:auto-scaling-group/*` `qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
AttachInstances	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
DetachInstances	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
RemoveInstances	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
DescribeAutoScalingInstances	`qcs::as:$region:$account:auto-scaling-group/*` `qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
SetInstancesProtection	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
CreateScheduledAction	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
DeleteScheduledAction	`qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
DescribeScheduledActions	`qcs::as:$region:$account:auto-scaling-group/*` `qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId`
ModifyScheduledAction	`qcs::as:$region:$account:auto-scaling-`

	group/$AutoScalingGroupId
CreateScalingPolicy	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
DeleteScalingPolicy	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
DescribeScalingPolicies	qcs::as:$region:$account:auto-scaling-group/* qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
ModifyScalingPolicy	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
ExecuteScalingPolicy	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
CreateNotificationConfiguration	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
DeleteNotificationConfiguration	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
DescribeNotificationConfigurations	qcs::as:$region:$account:auto-scaling-group/* qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
ModifyNotificationConfiguration	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
CreateLifecycleHook	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
DeleteLifecycleHook	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
DescribeLifecycleHooks	qcs::as:$region:$account:auto-scaling-group/* qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
UpgradeLifecycleHook	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId
CompleteLifecycleAction	qcs::as:$region:$account:auto-scaling-group/$AutoScalingGroupId

DescribeAccountLimits	*

弹性伸缩访问管理策略示例

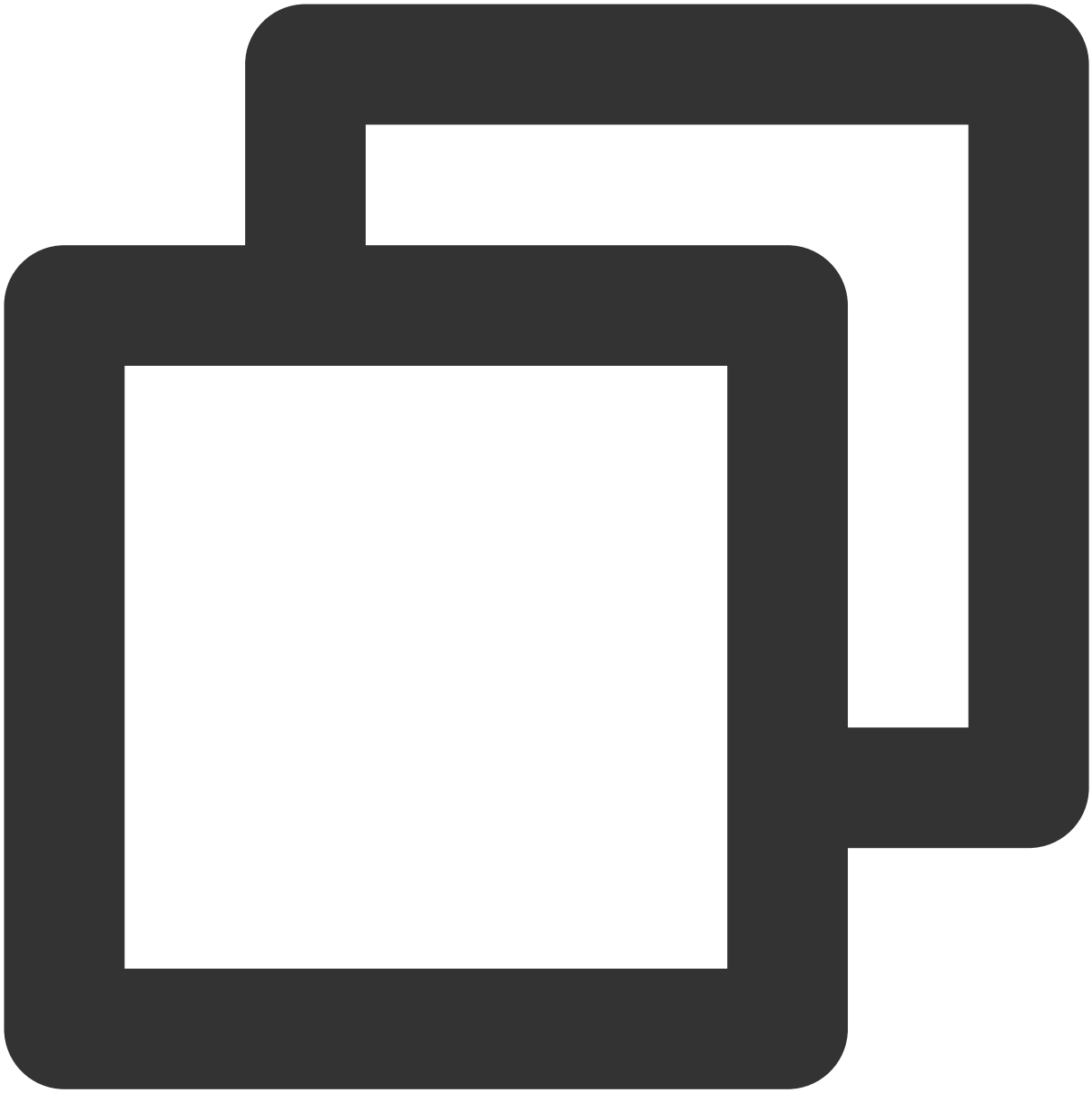
下面以具体的示例展示如何通过访问管理对弹性伸缩资源进行权限控制：
创建策略：广州地域允许对所有伸缩组的访问权限。



```
# 其中$account需要替换成账号信息
{
```

```
    "version": "2.0",
    "statement": [
        {
            "effect": "allow",
            "action": [
                "name/as:*"
            ],
            "resource": [
                "qcs::as:ap-guangzhou:$account:auto-scaling-group/*"
            ]
        }
    ]
}
```

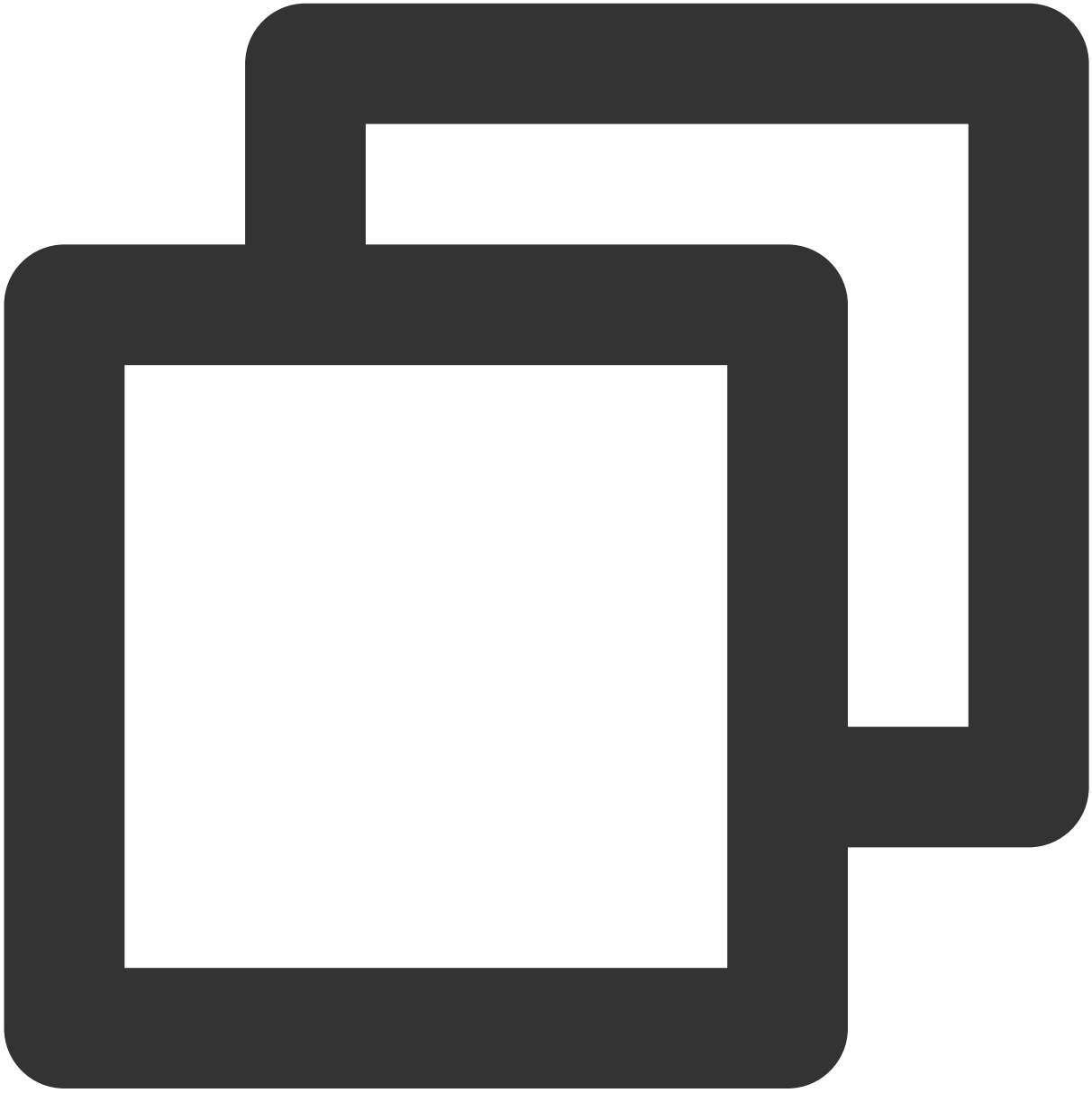
创建策略：广州地域禁止对某个伸缩组的访问权限。



```
# 其中$account需要替换成账号信息，$AutoScalingGroupId 需要替换成相应的AutoScalingGroupId
{
  "version": "2.0",
  "statement": [
      {
          "effect": "deny",
          "action": [
              "name/as:*"
          ],
          "resource": [
              "qcs::as:ap-guangzhou:$account:auto-scaling-group/$AutoScalingGroupId
```

```
            ]
        }
    ]
}
```

创建策略：对全部地域所有读接口拥有访问权限。



```
{
  "version": "2.0",
  "statement": [
      {
          "effect": "allow",
```

```
            "action": [
                "name/as:Describe*"
            ],
            "resource": [
                "*"
            ]
        }
    ]
}
```