

阿里云云计算ACP认证培训

弹性伸缩服务AutoScaling

课程目标

学习完本课程后，你将能够：

1. 了解弹性伸缩服务的概念、应用场景
2. 掌握阿里云AutoScaling的功能、组成
3. 掌握AutoScaling的基本操作

目录

- 1. 弹性伸缩服务的产生背景**
2. 阿里云弹性伸缩服务AutoScaling介绍
3. AutoScaling的配置流程
4. AutoScaling的最佳实践
5. AutoScaling的典型应用案例

弹性伸缩服务的产生背景

场景1：某视频公司，春晚或每周五热门节目来临时，如临大敌，需要按负载自动弹性伸缩

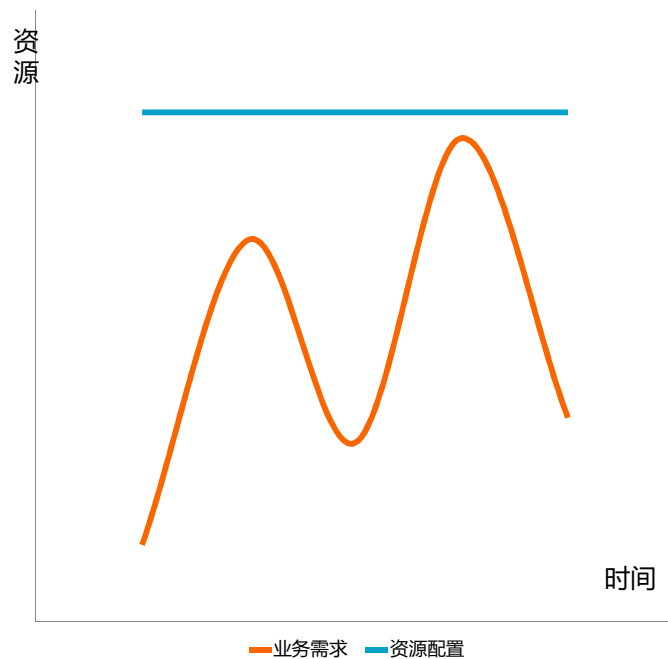
场景2：某视频直播公司，无法预估业务负载情况，需要根据CPU利用率、Load、带宽利用率，自动弹性伸缩

场景3：某游戏公司，每天中午12点，每天晚上6点~9点，需要定时扩容

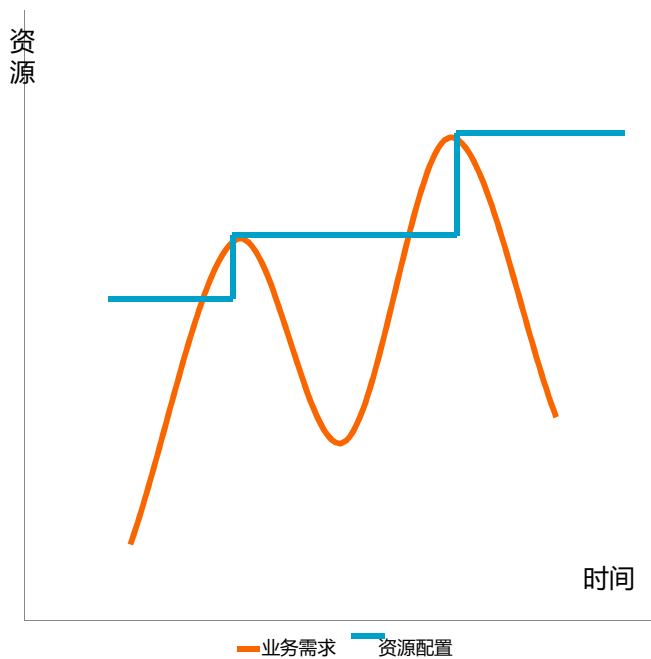


不同的解决方式

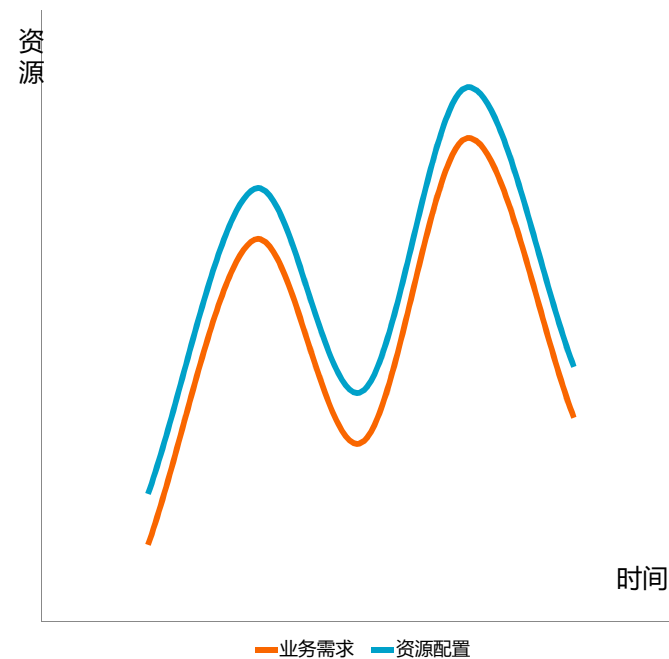
超配资源



人工伸缩

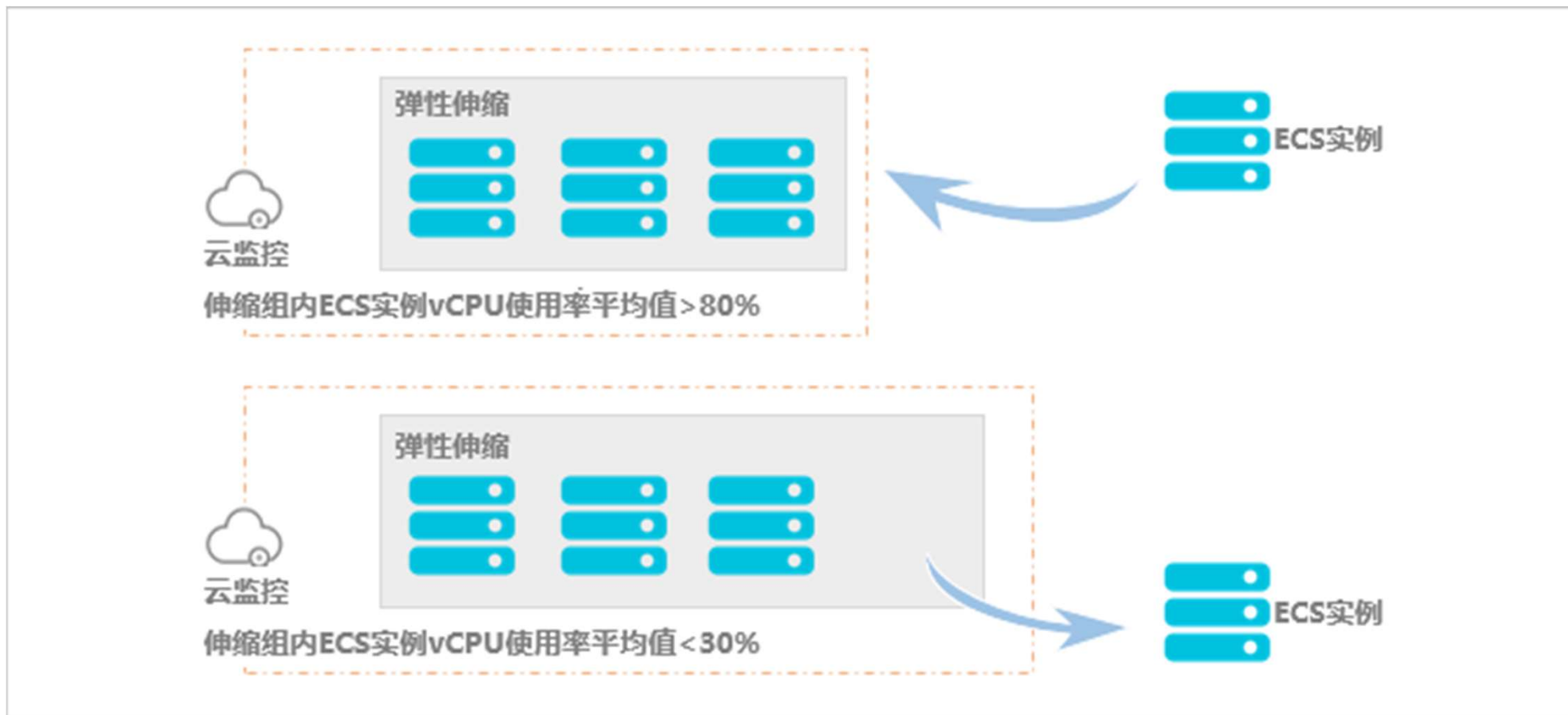


系统自动伸缩



问题：业务峰值如何高效应对？是否可以有一种服务，可以自动调整弹性计算资源大小，以满足业务需求的变化？

弹性伸缩服务的产生背景



目录

1. 弹性伸缩服务的产生背景

2. 阿里云弹性伸缩服务AutoScaling介绍

2.1 AutoScaling的概念

2.2 AutoScaling的功能

2.3 AutoScaling的组成

3. AutoScaling的配置流程

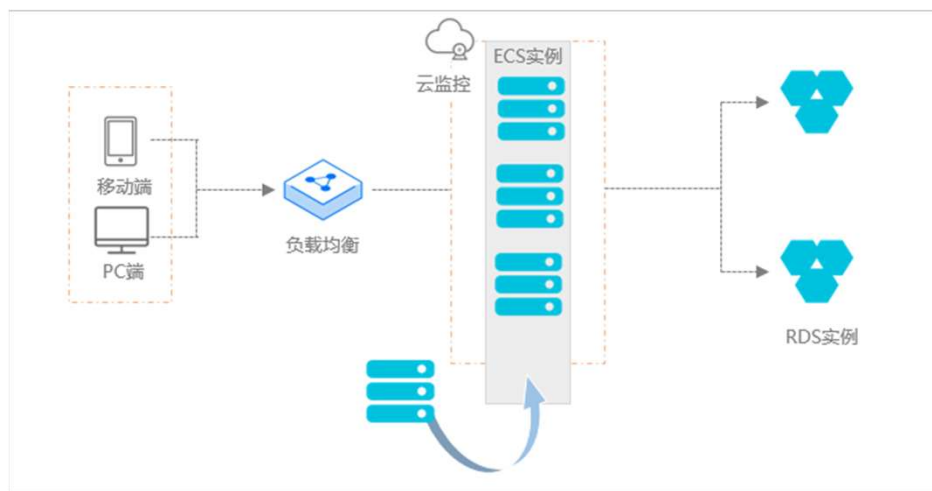
4. AutoScaling的最佳实践

5. AutoScaling的典型应用案例

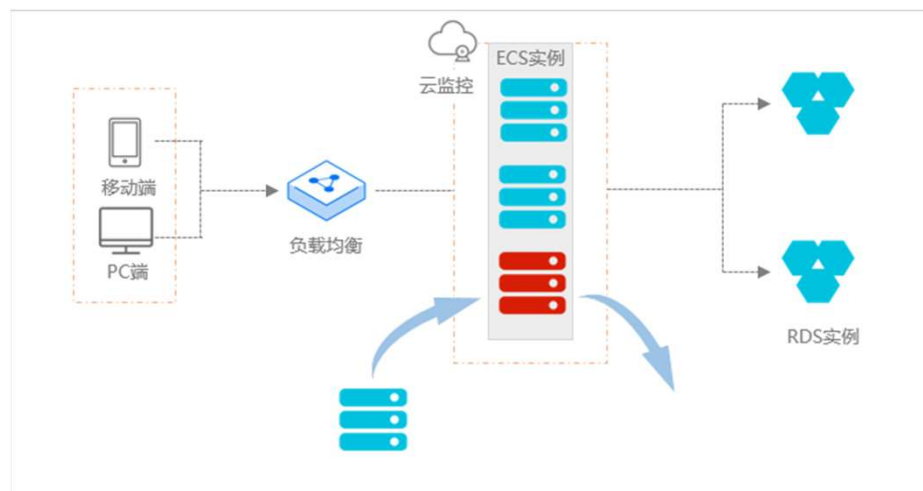
AutoScaling的概念

弹性伸缩 (AutoScaling) 是一种服务, 可以自动调整弹性计算资源 (ECS), 以满足业务需求的变化。

应用场景: 弹性扩张、弹性收缩、弹性自愈



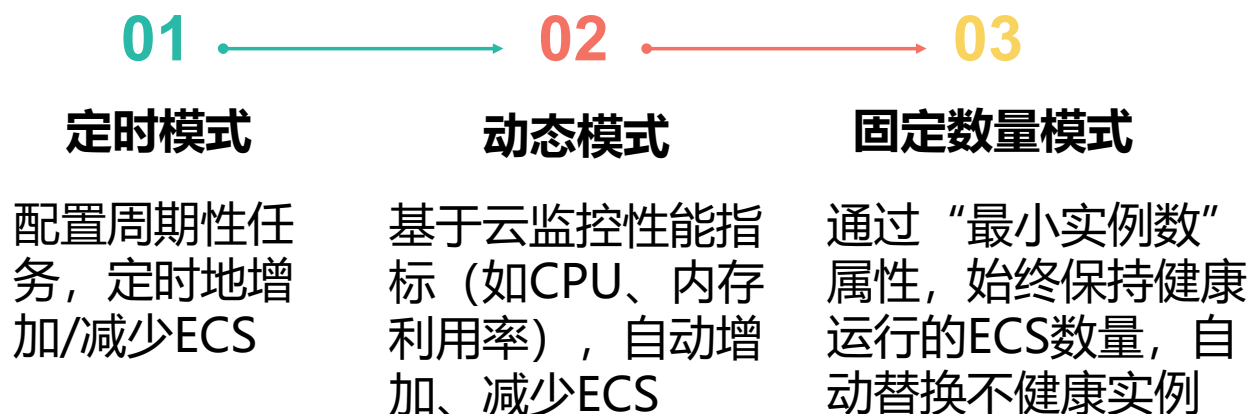
弹性扩张



弹性自愈

AutoScaling的功能

- 功能1：根据客户业务需求自动调整ECS实例数量。



- 功能2：自动向SLB的后端服务器组中添加或移除相应的ECS实例。
- 功能3：自动向RDS访问白名单中添加或移除ECS实例的IP。

冷却时间

场景：当CPU过高时，触发报警任务，创建了几台新的ECS，但是CPU要经过一小段时间才能降下来。这个过程中，报警任务是否会持续被触发？

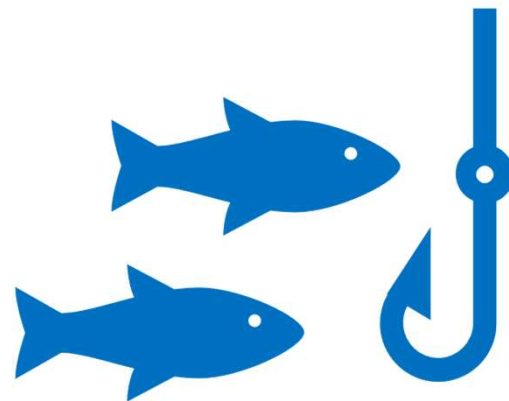
解决方案：伸缩活动成功执行后，设定一段锁定时间，即冷却时间。冷却时间内不会接受由云监控报警任务触发的伸缩活动请求。



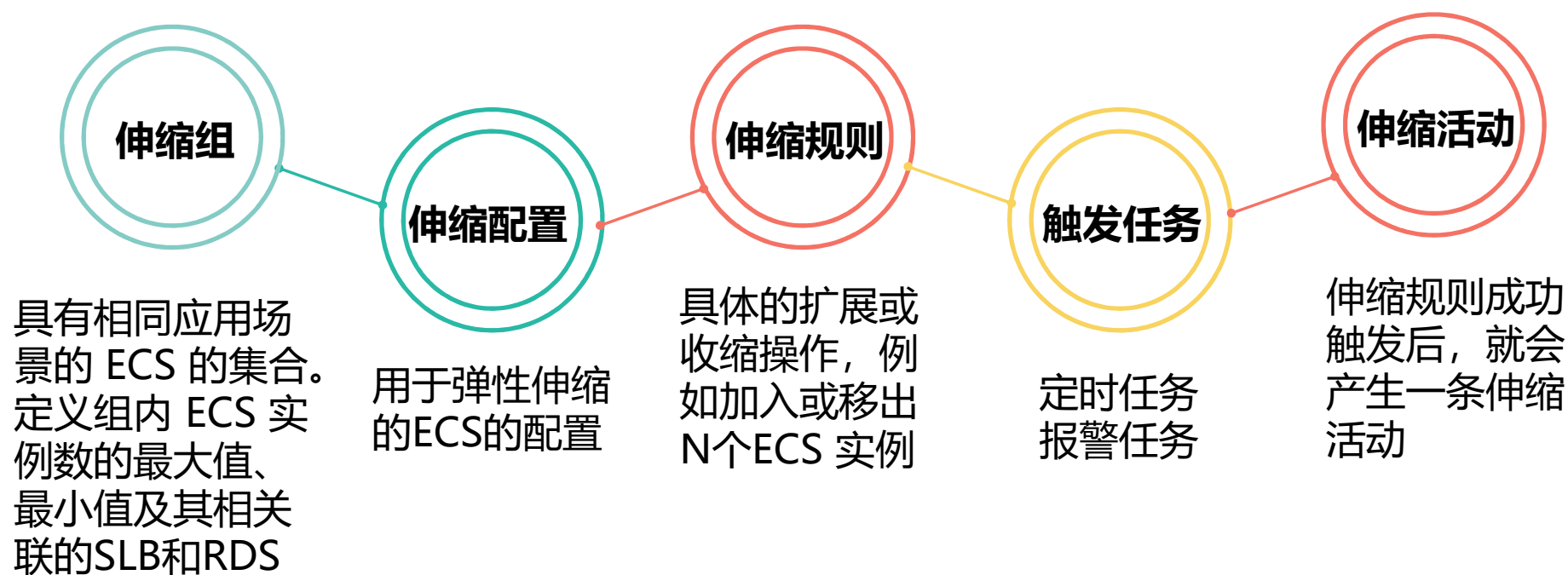
生命周期挂钩 (LifecycleHook)

场景：伸缩组释放ECS实例时需要先将实例从SLB后端服务移除（防止接收新的请求），待检测已经接收到的请求处理完成，停止并释放实例。

解决方案：在该伸缩组内创建生命周期挂钩。设置伸缩活动类型为“弹性收缩”，设置超时时间为（预计的）请求最长处理时间。当发生弹性收缩类型伸缩活动时，ECS实例从SLB移除后会挂起一段时间（即超时时间），等待请求处理完成。待超时时间到，再释放ECS实例。



AutoScaling的组成

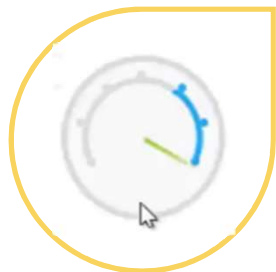


AutoScaling的优势



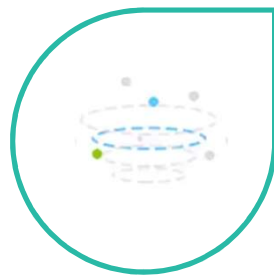
高可用

自动监测ECS实例的健康状况，及时替换不健康实例



自运维

- 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



灵活丰富

- 智能调度应对各种复杂场景
- 可通过API对接外在监控系统
- 可同时配置多种伸缩模式



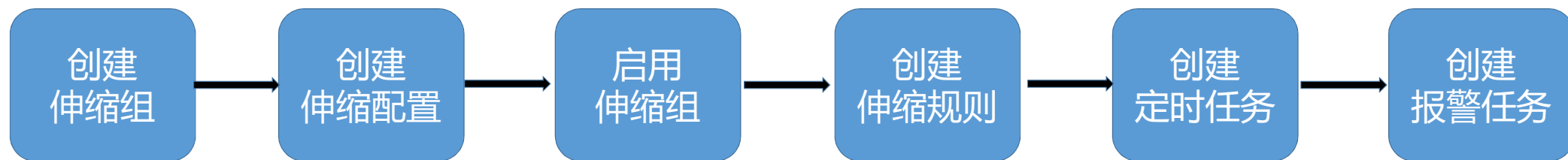
降成本

按需取用，提高资源利用率

目录

1. 弹性伸缩服务的产生背景
2. 阿里云弹性伸缩服务AutoScaling介绍
- 3. AutoScaling的配置流程**
4. AutoScaling的最佳实践
5. AutoScaling的典型应用案例

AutoScaling配置流程



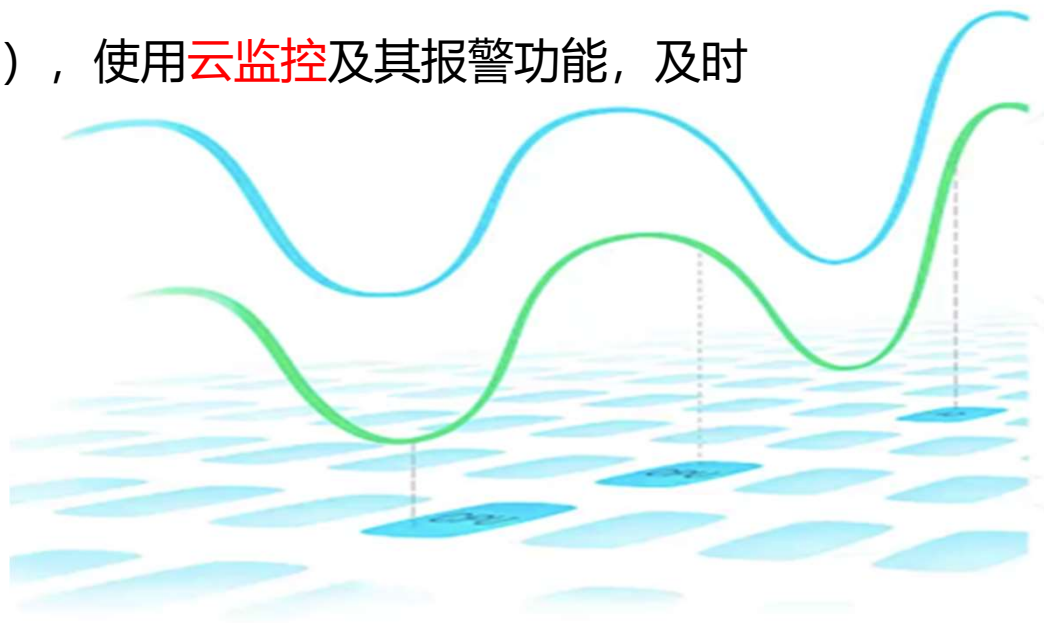
目录

1. 弹性伸缩服务的产生背景
2. 阿里云弹性伸缩服务AutoScaling介绍
3. AutoScaling的配置流程
- 4. AutoScaling的最佳实践**
5. AutoScaling的典型应用案例

最佳实践：多种伸缩模式结合

使用弹性伸缩，要提前判断、评估**业务场景**。

- (1) **定时伸缩模式**（定时任务）：基于已知的周期性变化，定时增加或减少ECS实例；
- (2) **动态伸缩模式**（报警任务）：基于云监控性能指标（如CPU、内存利用率），自动增加或减少ECS实例，应对不可预期的变化；
- (3) 为了应对异常情况（如遭受黑客DDoS攻击），使用**云监控**及其报警功能，及时发现问题、及时处理。或者使用**云安全**产品。



目录

1. 弹性伸缩服务的产生背景
2. 阿里云弹性伸缩服务AutoScaling介绍
3. AutoScaling的配置流程
4. AutoScaling的最佳实践
- 5. AutoScaling的典型应用案例**
 - 5.1 小公司无运维团队
 - 5.2 大公司实现自动化运维

应用案例1：小公司无运维团队

某互联网亲子社区

日均PV 300万、日均活跃用户8万、日均上传高清图15万张、日均上传视频6000个。

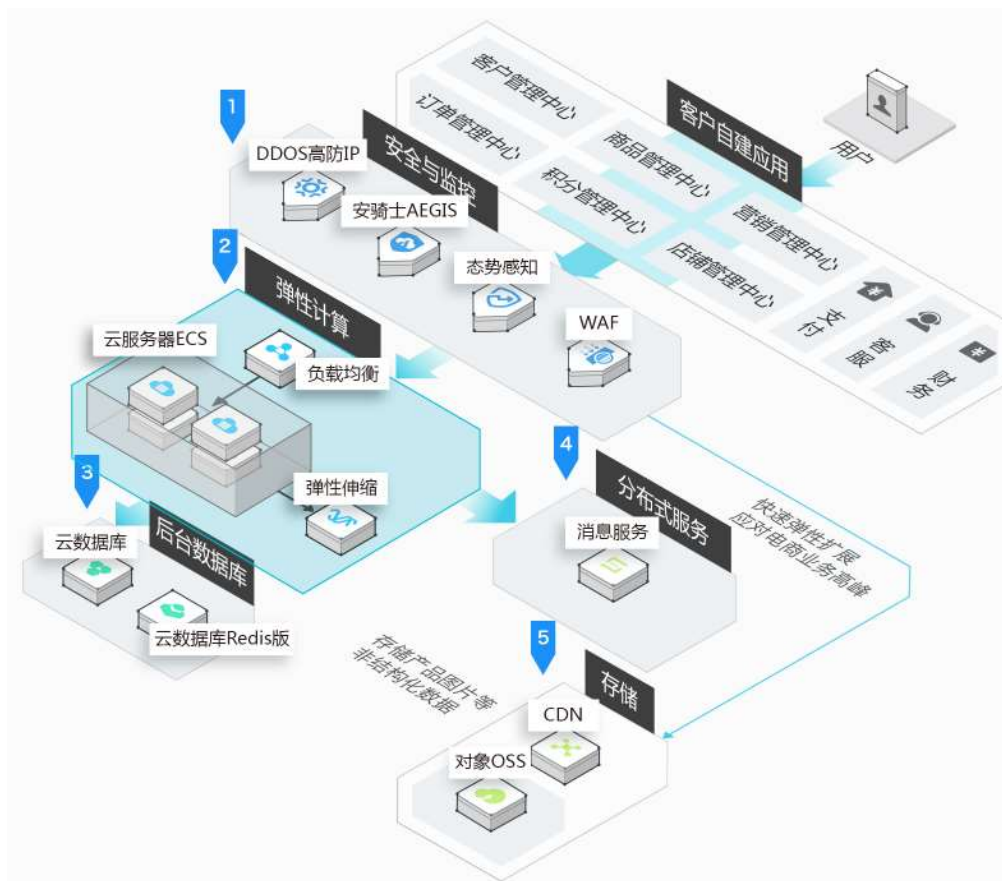
在阿里云存储数据**8T+**，照片**1500万张**，视频**60万个**

急需解决：

- 需要存储大量的照片和视频
- 创业团队，无专人维护

解决方案：

ECS、SLB、RDS、OSS、**AutoScaling**



应用案例2：大公司实现自动化运维

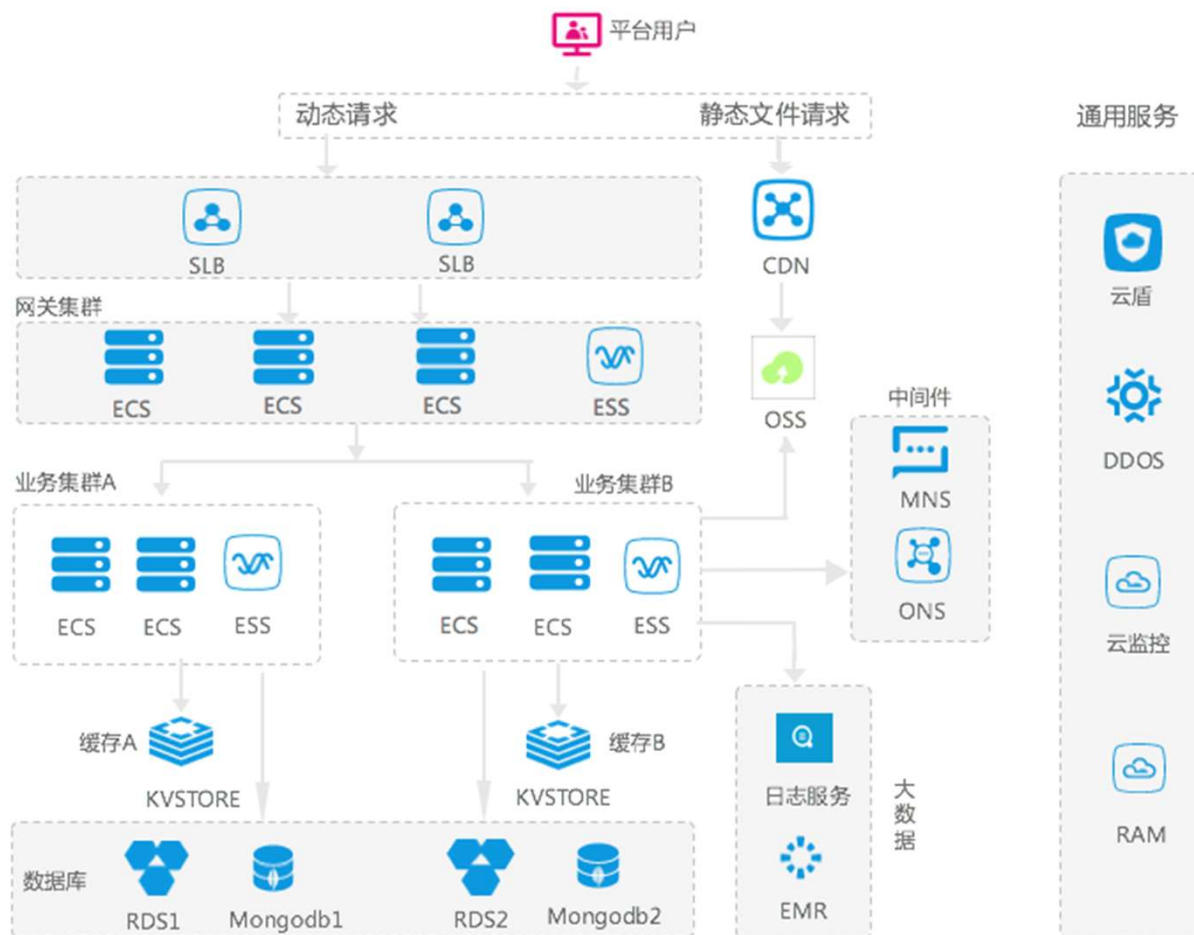
某图像视频技术服务公司

面临的问题：

1. 上万视频平台客户
2. 业务指数级增长

解决方案：

网关集群和业务集群都有“组”的概念，基于组的**弹性伸缩**，可以快速应对突发流量。



小结

1. AutoScaling在阿里云平台中的定位是什么？
2. AutoScaling有哪些组件？你认为哪个功能最重要？
3. 相对于传统硬件服务器，AutoScaling有哪些优势？
4. 使用AutoScaling有哪些注意事项？

为了无法计算的价值 |  阿里云

