

# 基本云架构

---

曹广杰

数据科学与计算机学院

授课教师：王昌栋

2017/9/24

基本云架构

负载均衡与动态扩展架构

请求调度方法

## 负载均衡与动态扩展架构

讨论分析负载均衡与动态扩展架构（水平扩展、弹性容量、弹性磁盘）之间的关联性。

### 1. 负载均衡架构

负载均衡架构的使用是：

- 基于同类型IT资源同时提供服务的环境；
- 针对管理的IT资源适当地增加或者减少。

### 2. 动态可扩展架构

- 基于预先定义的扩展条件从资源池中动态分配IT资源；
- 实现的核心是自动扩展监听器。

自动扩展监听器用于监听当前的负载状态，并决定何时进行拓展。

动态水平拓展是指向内或向外拓展IT资源实例，主要的操作是IT资源的复制。

动态垂直拓展是指向上或向下拓展IT资源实例，主要操作是调整单个IT资源的容量。

弹性容量是指根据负载变化分配或者回收CPU或者RAM资源，拓展方式属于垂直拓展。

弹性磁盘是指动态储存供给系统，根据用户的需求提供磁盘空间。

在云服务处接收到来自于用户的请求时候：

1. 自动扩展监听器监视该云服务，判断预定义的容量阈值是否已经被超过。（负载均衡）
2. 随着工作负载已超过性能阈值。根据预定义规则，自动扩展监听器决定下一步的操作。

动态拓展的具体行为可以配合负载均衡架构，使得动态拓展更符合用户的要求。

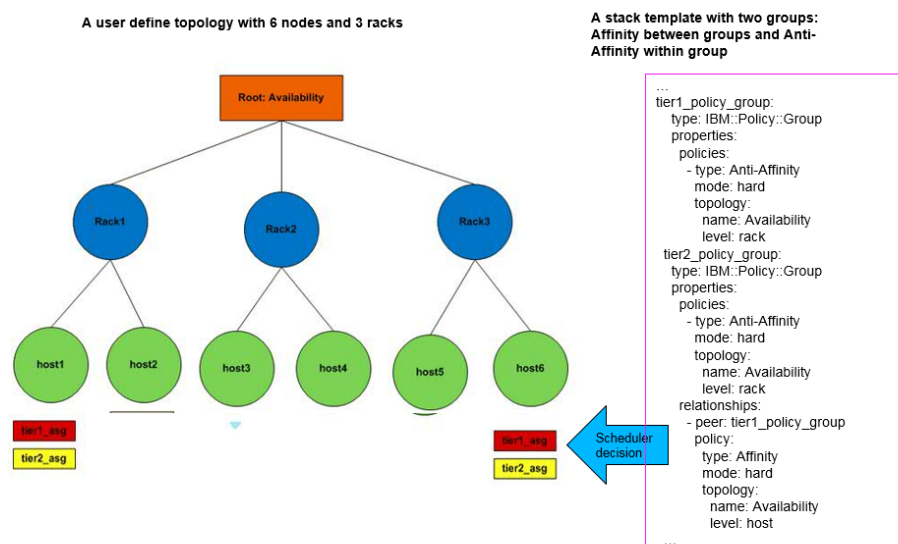
## 请求调度方法

调研不同的请求调度方法，并详细描述其中的一种。

基于OpenStack的调度机制有两例：

- Affinity/Anti-Affinity 策略
- MaxResourceLostPerNodeFailure 策略

基于拓扑结构的调度：



这里仔细分析 Affinity/Anti-Affinity 策略，

已知：

1. 用户定义了2个auto scaling group tier1和tier2，每个tier都需要2个虚拟机；
2. 其中tier1需要虚拟机在rack节点上Anti-Affinity，tier2需要虚拟机在rack节点上Affinity，并且tier1和tier2上的虚拟机之间需要满足Affinity。
3. 这个场景类似于在生产环境上部署2组web application，要求运行database的虚拟机(tier1)和运行web的虚拟机(tier2)在相同的主机上(方便web服务器和database服务器通信)，并且2个运行database的虚拟机(tier1)和2运行web的虚拟机(tier2)不能同时运行在一台主机上(rack级别上Anti-Affinity,担心单rack单点故障造成所有的database服务器或者web服务器都不可用)。
4. 图的左边是一个部署的结果，红色的虚拟机的是web服务器tier1，黄色的虚拟机是database服务器(tier2)，这样host1上的database服务器直接为host1上的web服务器提供服务，host6上的database服务器直接为host6上的web服务器提供，并且rack1或者rack3的单点故障，不会造成用户web服务的中断。

——取自容器技术爱好者的blog

<http://blog.csdn.net/horsefoot/article/details/52037198?locationNum=2&fps=1###>;