OpenStack介绍

# .简介

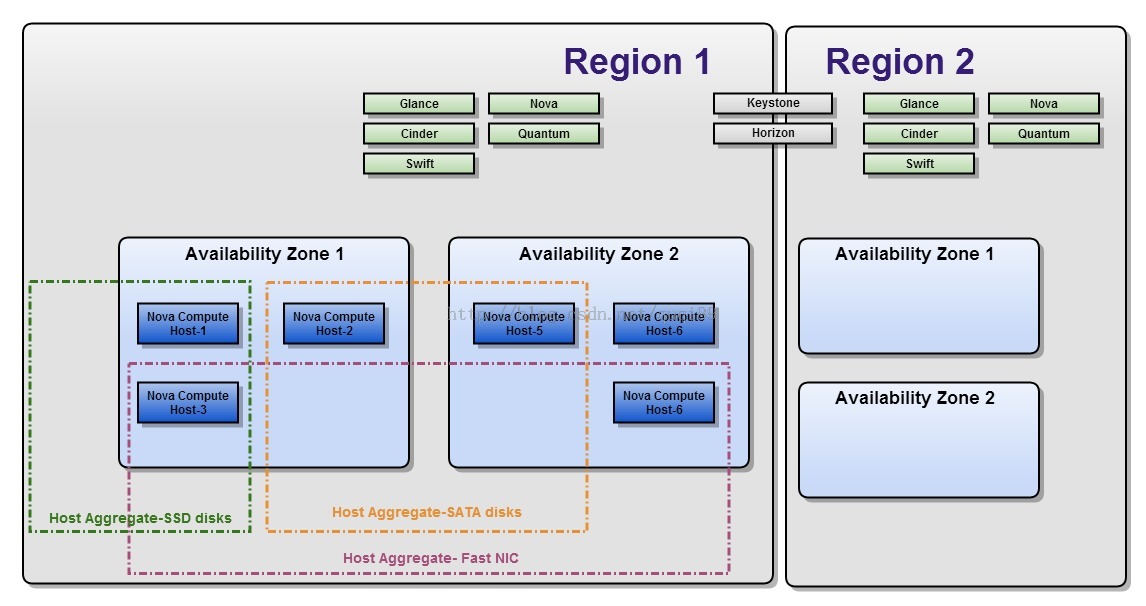
本文将要介绍如何应用 OpenStack 中 Host Aggregates 的机制来更为有效地分配 nova computes 节点上的硬件资源，从而达到 OpenStack 的用户能够合理地选择特定的 computes 节点，来创建符合自己需求的虚拟机。

**Region：**

**Region**更像是一个地理上的概念，每个region有自己独立的endpoint，regions之间完全隔离，但是多个regions之间共享同一个keystone和dashboard。（注：目前openstack的dashboard还不支持多region）

　　所以除了提供隔离的功能，region的设计更多侧重地理位置的概念，用户可以选择离自己更近的region来部署自己的服务。

Host Aggregates 最先出现在 OpenStack Grizzly 的版本中，是在 OpenStack 的 Regions 和 Availability Zones 之后被提出来，并建立于 Availability Zones 基础之上更进一步划分 computes 节点物理资源的一种机制。



## Host Aggregates 与 Availability Zones 定义上的区别

Availability Zones 通常是对 computes 节点上的资源在小的区域内进行逻辑上的分组和隔离。例如在同一个数据中心，我们可以将 Availability Zones 规划到不同的机房，或者在同一机房的几个相邻的机架，从而保障如果某个 Availability Zone 的节点发生故障（如供电系统或网络），而不影响其他的 Availability Zones 上节点运行的虚拟机，通过这种划分来提高 OpenStack 的可用性。目前 OpenStack 默认的安装是把所有的 computes 节点划分到 nova 的 Availability Zone 上，但我们可以通过对 nova.conf 文件的配置来定义不同的 Availability zones。

Host Aggregates 是在 Availability Zones 的基础上更进一步地进行逻辑的分组和隔离。例如我们可以根据不同的 computes 节点的物理硬件配置将具有相同共性的物理资源规划在同一 Host Aggregate 之下，或者根据用户的具体需求将几个 computes 节点规划在具有相同用途的同一 Host Aggregate 之下，通过这样的划分有利于提高 OpenStack 资源的使用效率。Host Aggregates 可以通过 nova client 或 API 来创建和配置。下面以 nova client 的命令行方式来配置 Host Aggregates。

## **cell**

cell是openstack一个非常重要的概念，主要用来解决openstack的扩展性和规模瓶颈。众所周知，openstack是由很多的组件通过松耦合构成，那么当达到一定的规模后，某些模块必然成为整个系统的瓶颈。比较典型的组件就是database和AMQP了，所以，每个cell有自己独立的DB和AMQP。

另外，由于cell被实现为树形结构，自然而然引入了分级调度的概念。通过在每级cell引入nova-cell服务，实现了以下功能：

Messages的路由，即父cell通过nova-cell将Messages路由到子cell的AMQP模块

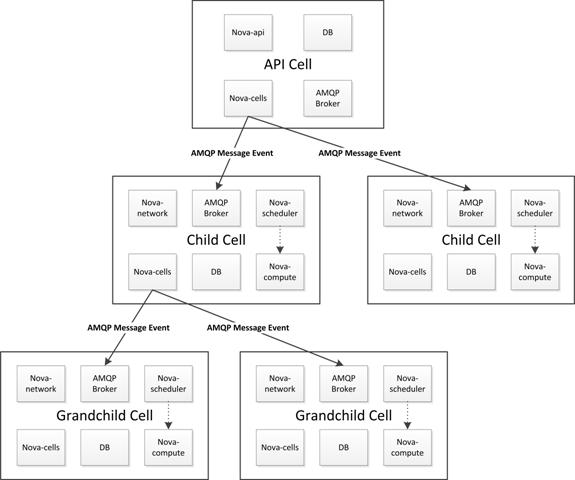
分级调度功能，即调度某个instances的时候先要进行cell的选择，目前只支持随机调度，后续会增加基于filter和weighing策略的调度

资源统计，子cell定时的将自己的资源信息上报给父cell，用来给分级调度策略提供决策数据和基于cell的资源监控

cell之间的通信(通过rpc完成)

　　最后，所有的子cell公用底层cell的nova-api，子cell包含除了nova-api之外的其他nova服务，当然所有的cell都共用keystone服务。

**注：nova-\*是指除了nova-api之外的其他nova服务，子cell + 父cell才构成了完整的nova服务）**



　　每一个 Cell 包含独立的 Message Broker 以及 Database，其中 API Cell 主要包含 nova-api 服务，用于接收用户请求，并将用户请求通过 message 的形式发送至指定的 Cell；Child Cell 包含除 nova-api 之外的所有 nova-\*服务，实现具体的 Nova Compute 节点服务;API Cell 与 Child Cell 共享 Glance 服务，且各 Cells 之间的通信均通过 nova cells 服务进行。Cell 调度独立于与 host 调度，在创建新的实例时，首先由 nova-cells 选择一个 Cell。当 Cell 确定后，实例创建请求会被送达目标 Cell 的 nova-cells 服务，随后该请求会被交给本 Cell 的主机调度机制处理，此时主机调度机制会像未配置 Cell 的环境一样处理该请求。

<http://www.ibm.com/developerworks/cn/cloud/library/1409_zhaojian_openstacknovacell/index.html>

## 例子

### high-memory

• 在具有相同物理特性的计算节点上创建 Host Aggregate

# nova aggregate-create hm-agg hm-az

如果没有指定 Availability Zone, OpenStack 会将 Host Aggregate 建在默认的 Availability Zone 下面（如 nova），否则会根据指定的名字来判断是否创建新的 Availability Zone 或使用已经存在的 Availability Zone，同时在此之下创建 Host Aggregate。

• 设置对应的 Host Aggregate 的 metadata，将具有高内存配置的 computes 节点的 Host aggregate 的 metadata 设置成“HW=high-memory”。

# nova aggregate-set-metadata hm-agg HW=high-memory

• 添加节点给 Host Aggregate。

# nova aggregate-add-host hm-agg compute-1 # compute-1 is host name

• 创建和设置 flavor

# nova flavor create hm-test

# nova flavor-key hm-test set HW=high-memory

• 创建虚拟机

# openstack image list

# nova boot hm-vm --image <image-id> --flavor hm-test

当我们指定 flavor 去创建虚拟机，flavor 可以找到对应的属性，根据这个属性它会去数据里面匹配到对应的 aggregate, 这时具有 high-memory 属性的节点就会被匹配到，就会在该 aggregate 关联的 host 上面参加虚拟机。

结束语.

在具体使用 Host Aggregates 机制来分配 computes 节点资源时，我们可以综合考虑目前 OpenStack 环境里的物理资源和用户需求来灵活地配置 Host Aggregate 的 metadatas 以及 filters。由于一个 Availability Zone 可以同时包含多个 Host aggregates，每个 Host aggregate 又有可能配置不同的 metadatas，因此我们在创建虚拟机时也应该注意这点，不要将虚拟机创建到错误的 compute 节点上。

From <https://www.ibm.com/developerworks/cn/cloud/library/1604-openstack-host-aggregate/index.html>