

## 1. 写出 hadoop 集群常用进程以及进程含义

### 1. Namenode

它是 Hadoop 中的主服务器，管理文件系统名称空间和对集群中存储的文件的访问。

### 2. Datanode

它负责管理连接到节点的存储（一个集群中可以有多个节点）。每个存储数据的节点运行一个 datanode 守护进程。

### 3. secondaryNameNode

它不是 namenode 的冗余守护进程，而是提供周期检查点和清理任务。出于对可扩展性和容错性等考虑，我们一般将 SecondaryNameNode 运行在一台非 NameNode 的机器上。

### 4. ResourceManager

负责调度 DataNode 上的工作。每个 DataNode 有一个 NodeManager，它们执行实际工作。

### 5. NodeManager

负责执行 ResourceManager 分发的任务

## 2. 什么是云计算？

它是以虚拟化技术为核心技术和基础，面向 服务架构(SOA) 的一种实现,将虚拟化环境"资源池"隐藏起来，将其上层应用软件形成丰富的云管理接口，达到所有人自由使用所有资源的一种现象，他是一种资源使用模式的变革。

## 3. 云计算的实现方式有哪些？

云计算的实现方式有：**Private**、IaaS、PaaS、SaaS

**Private**：传统/私有方式

优点：所有事情都亲自做，可控

缺点：用户成本比较高，要求自身技术水平高

典型软件：传统物理机

IaaS：基础设施即服务

优点：底层硬件到操作系统，都不需要用户操心，可以集中精力做业务项目

缺点：服务商提供的东西，不能自己自由定制，不可控

典型软件：OpenStack、CloudStack

PaaS：平台即服务

优点：对于只会开发不会运维的人员比较友好，底层到运行环境，都不需要用户操心，可以集中精力做应用项目

缺点：服务商提供的东西，不灵活，只适用于特殊的应用项目

SaaS：软件、应用即服务

优点：所有东西都由服务商提供，用户只需要花钱就行，对于广大企业来说，SaaS 是采用先进技术实施信息化的最好途径。比如说，买企业邮箱，买财务软件云

缺点：对客户来说，所有的东西都不可控，安全性不够。

## 4. 如果批量创建多个 VM 实例，是同时创建还是按顺序创建呢？

为了避免同时创建多个 VM 实例时候，给用户和各种资源带来的压力，我们应该按照创建请求的顺序，一个一个的创建，而满足顺序的功能的软件，这就可以用异步协作的消息队列。

## 5. 什么是虚拟化?

虚拟化是一种技术，它的目的在于提高资源的使用率，并将底层硬件和上层的应用软件进行隔离，使得上层软件及应用计算变得更加弹性可控。最终达到有限成本的高价值。

默认情况下，虚拟化技术默认并不对外体用抽象的上层应用软件服务组件，一个没有被服务化的虚拟化环境只能被称为“资源池”，只有内部管理人员才可以操作。

## 6. 虚拟化和云计算的区别?

虚拟化是一种技术，云计算是资源交付模式，云计算不等于虚拟化。  
云计算是基于虚拟化技术的一种资源交付使用模式。

## 7. OpenStack 的组件有哪些?

**Cinder**: 为 VMs 提供持久的块存储能力，支持多种存储方式，工作中 ceph 用的比较多

**Glance**: 用于存储和检索磁盘映像文件，支持多种存储方式

**Heat**: openstack 的任务编排工具

**Horizon**: openstack 的 web 可视化界面

**Keystone**: 为 Openstack 中的所有服务提供了认证、授权以及端点编录服务

**Nova**: 管理 VM 的所有操作

**Neutron**: 为 Openstack 提供网络的功能；插件化设计，支持众多流行的网络

**Swift**: **分布式存储，基于** RESTful 的 API 实现非结构化数据对象的存储及检索

**Trove**: 提供数据库即服务的功能

**sahara**: 在 OpenStack 中提供大数据服务，生产可用

**Octavia**: openstack 中的负载均衡项目，生产可用。

**IRonic**, 物理裸机管理，目前是非常好用。

**Ceilometer**, 用于实现监控和计量服务的实现，缺乏后续发展

## 8. OpenStack 的核心服务有哪些?

compute、networking、storage、dashboard

## 9. 容器退出后，通过 docker ps 命令查看不到，数据会丢失么?

容器退出后会处于终止 (exited) 状态，此时可以通过 `docker ps -a` 查看，其中数据不会丢失，还可以通过 `docker start` 来启动，只有删除容器才会清除数据。

## 10. 如何控制容器占用系统资源 (CPU, 内存) 的份额?

在使用 `docker create` 命令创建容器或使用 `docker run` 创建并运行容器的时候，可以使用 `-c|--cpu-shares[=0]` 参数来调整同期使用 CPU 的权重，使用 `-m|--memory` 参数来调整容器使用内存的大小。

## 11. 如何更改 Docker 的默认存储设置?

Docker 的默认存放位置是 `/var/lib/docker`，如果希望将 Docker 的本地文件存储到其他分区，可以使用 Linux 软连接的方式来做。

## 12. Docker 公司的三款用于解决多容器分布式软件可移植部署的问题，推出的编排工具有哪些？

1. Docker Machine：为本地私有数据中心及公有云平台提供 Docker 引擎，实现从零到 Docker 的一键部署。
2. Docker Compose：是一个编排多容器分布式部署的工具，提供命令集管理容器化应用的完整开发周期，包括服务构建，启动和停止。
3. Docker Swarm：为 Docker 容器提供了原生的集群，它将多个 Docker 引擎的资源汇聚在一起，并提供 Docker 标准的 API，使 Docker 可以轻松扩展到多台主机。

## 13. 简单描述 Docker-compose 编排和管理多容器的过程？

1. 使用 Dockerfile 定义应用依赖的镜像
2. 使用 `docker-compose.yml` 定义应用具有的服务
3. 通过 `docker-compose up` 命令创建并运行应用