



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

## 第三版：大数据 50 道

### Hadoop 集群可以运行的 3 个模式？

单机（本地）模式

伪分布式模式

全分布式模式

#### 单机（本地）模式中的注意点？

在单机模式（standalone）中不会存在守护进程，所有东西都运行在一个 JVM 上。这里同样没有 DFS，使用的是本地文件系统。单机模式适用于开发过程中运行 MapReduce 程序，这也是最少使用的一个模式。

#### 伪分布模式中的注意点？

伪分布式（Pseudo）适用于开发和测试环境，在这个模式中，所有守护进程都在同一台机器上运行。

#### VM 是否可以称为 Pseudo？

不是，两个事物，同时 Pseudo 只针对 Hadoop。

#### 全分布模式又有什么注意点？

关注公众号：磊哥聊编程，回复：面试题，获取最新版面试题<sup>1</sup>



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

全分布模式通常被用于生产环境,这里我们使用 N 台主机组成一个 Hadoop 集群, Hadoop 守护进程运行在每台主机之上。这里会存在 Namenode 运行的主机, Datanode 运行的主机, 以及 task tracker 运行的主机。在分布式环境下, 主节点和从节点会分开。

## Hadoop 是否遵循 UNIX 模式?

是的, 在 UNIX 用例下, Hadoop 还拥有 “conf” 目录。

## Hadoop 安装在什么目录下?

Cloudera 和 Apache 使用相同的目录结构, Hadoop 被安装在 `cdusrlibhadoop-0.20`。

## Namenode、Job tracker 和 task tracker 的端口号是?

Namenode, 70; Job tracker, 30; Task tracker, 60。

## Hadoop 的核心配置是什么?

Hadoop 的核心配置通过两个 xml 文件来完成: 1, `hadoop-default.xml`; 2, `hadoop-site.xml`。这些文件都使用 xml 格式, 因此每个 xml 中都有一些属性, 包括名称和值, 但是当下这些文件都已不复存在。

## 那当下又该如何配置?

关注公众号: 磊哥聊编程, 回复: 面试题, 获取最新版面试题<sup>2</sup>



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

Hadoop 现在拥有 3 个配置文件：1, core-site.xml；2, hdfs-site.xml；3, mapred-site.xml。这些文件都保存在 conf 子目录下。

## RAM 的溢出因子是？

溢出因子 (Spill factor) 是临时文件中储存文件的大小，也就是 Hadoop-temp 目录。

## fs.mapr.working.dir 只是单一的目录？

fs.mapr.working.dir 只是一个目录。

## hdfs-site.xml 的 3 个主要属性？

dfs.name.dir 决定的是元数据存储的路径以及 DFS 的存储方式 (磁盘或是远端)

dfs.data.dir 决定的是数据存储的路径

fs.checkpoint.dir 用于第二 Namenode

## 如何退出输入模式？

退出输入的方式有：1, 按 ESC；2, 键入 q (如果你没有输入任何当下) 或者键入 wq (如果你已经输入当下)，并且按下 Enter。

关注公众号：磊哥聊编程，回复<sup>3</sup>：面试题，获取最新版面试题<sup>3</sup>



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

## 当你输入 `hadoopfsck` 造成 “connection refused java exception” 时，系统究竟发生了什么？

这意味着 Namenode 没有运行在你的 VM 之上。

## 我们使用 Ubuntu 及 Cloudera，那么我们该去哪里下载 Hadoop，或者是默认就与 Ubuntu 一起安装？

这个属于 Hadoop 的默认配置，你必须从 Cloudera 或者 Edureka 的 dropbox 下载，然后在你的系统上运行。当然，你也可以自己配置，但是你需要一个 Linux box，Ubuntu 或者是 Red Hat。在 Cloudera 网站或者是 Edureka 的 Dropbox 中有安装步骤。

## “jps” 命令的用处？

这个命令可以检查 Namenode、Datanode、Task Tracker、Job Tracker 是否正常工作。

## 如何重启 Namenode？

点击 `stop-all.sh`，再点击 `start-all.sh`。

键入 `sudo hdfs` (Enter)，`su=hdfs` (Enter)，`etcinit.dha` (Enter)，及 `etcinit.dhadoop-0.20-namenode start` (Enter)。

关注公众号：磊哥聊编程，回复：面试题，获取最新版面试题<sup>4</sup>



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

## Fsck 的全名?

全名是：File System Check。

## 如何检查 Namenode 是否正常运行?

如果要检查 Namenode 是否正常工作，使用命令  
`etcinit.dhadoop-0.20-namenode status` 或者就是简单的 `jps`。

## mapred.job.tracker 命令的作用?

可以让你知道哪个节点是 Job Tracker。

## etc init.d 命令的作用?

etc init.d 说明了守护进程(服务)的位置或状态，其实是 LINUX 特性，和 Hadoop 关系不大。

## 如何在浏览器中查找 Namenode?

如果你确实需要在浏览器中查找 Namenode，你不再需要 localhost8021，  
Namenode 的端口号是 50070。

## 如何从 SU 转到 Cloudera?

关注公众号：磊哥聊编程，回复：[面试题](#)，获取最新版面试题<sup>5</sup>



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

从 SU 转到 Cloudera 只需要键入 exit。

## 启动和关闭命令会用到哪些文件？

Slaves 及 Masters。

## Slaves 由什么组成？

Slaves 由主机的列表组成，每台 1 行，用于说明数据节点。

## Masters 由什么组成？

Masters 同样是主机的列表组成，每台一行，用于说明第二 Namenode 服务器。

## hadoop-env.sh 是用于做什么的？

hadoop-env.sh 提供了 Hadoop 中 JAVA\_HOME 的运行环境。

## Master 文件是否提供了多个入口？

是的你可以拥有多个 Master 文件接口。

## Hadoop-env.sh 文件当下的位置？

hadoop-env.sh 现在位于 conf。

关注公众号：磊哥聊编程，回复：面试题，获取最新版面试题<sup>6</sup>



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

## 在 Hadoop\_PID\_DIR 中，PID 代表了什么？

PID 代表了“Process ID”。

## varhadoopids 用于做什么？

varhadoopids 用来存储 PID。

## hadoop-metrics.properties 文件的作用是什么？

hadoop-metrics.properties 被用做“Reporting”，控制 Hadoop 报告，初始状态是“not to report”。

## Hadoop 需求什么样的网络？

Hadoop 核心使用 Shell (SSH) 来驱动从节点上的服务器进程，并在主节点和从节点之间使用 password-less SSH 连接。

## 全分布式环境下为什么需求 password-less SSH？

这主要因为集群中通信过于频繁，Job Tracker 需要尽可能快的给 Task Tracker 发布任务。

## 这会导致安全问题吗？

关注公众号：磊哥聊编程，回复：面试题，获取最新版面试题<sup>7</sup>



微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

完全不用担心。Hadoop 集群是完全隔离的，通常情况下无法从互联网进行操作。与众不同的配置，因此我们完全不需要在意这种级别的安全漏洞，比如说通过互联网侵入等等。Hadoop 为机器之间的连接提供了一个相对安全的方式。

## SSH 工作的端口号是？

SSH 工作的端口号是 NO.22，当然可以通过它来配置，22 是默认的端口号。

## SSH 中的注意点还包括？

SSH 只是个安全的 shell 通信，可以把它当做 NO.22 上的一种协议，只需要配置一个密码就可以安全的访问。

## 为什么 SSH 本地主机需要密码？

在 SSH 中使用密码主要是增加安全性，在某些情况下也根本不会设置密码通信。

## 如果在 SSH 中添加 key，是否还需要设置密码？

是的，即使在 SSH 中添加了 key，还是需要设置密码。

## 假如 Namenode 中没有数据会怎么样？

没有数据的 Namenode 就不能称之为 Namenode，通常情况下，Namenode 肯定会有数据。

关注公众号：磊哥聊编程，回复：面试题，获取最新版面试题<sup>8</sup>





微信搜一搜

磊哥聊编程

扫码关注



回复：面试题 获取最新版面试题

## 当 Job Tracker 宕掉时，Namenode 会发生什么？

当 Job Tracker 失败时，集群仍然可以正常工作，只要 Namenode 没问题。

## 是客户端还是 Namenode 决定输入的分片？

这并不是客户端决定的，在配置文件中以及决定分片细则。

## 是否可以自行搭建 Hadoop 集群？

是的，只要对 Hadoop 环境足够熟悉，你完全可以这么做。

## 是否可以在 Windows 上运行 Hadoop？

你最好不要这么做，Red Hat Linux 或者是 Ubuntu 才是 Hadoop 的最佳操作系统。在 Hadoop 安装中，Windows 通常不会被使用，因为会出现各种各样的问题。因此，Windows 绝对不是 Hadoop 的推荐系统。