

Hadoop中的几个进程和作用

可回答：1) 启动Hadoop，都会有什么进程

问过的一些公司：顺丰(2022.09)，字节(2022.08)(2020.06)，浩鲸科技(2021.08)，海康暑期实习(2021.04)，京东寻猎计划(2020.09)，浩鲸云(2020.03)

参考答案：

1) NameNode：Master，它是一个主管、管理者

- 管理HDFS的名称空间；
- 管理数据块（Block）映射信息；
- 配置副本策略；
- 处理客户端读写请求。

2) DataNode：Slave，NameNode下达命令，DataNode执行实际的操作

- 存储实际的数据块；
- 执行数据块的读/写操作。

3) Secondary NameNode：并非NameNode的热备。当NameNode挂掉的时候，它并不能马上替换NameNode并提供服务

- 辅助NameNode，分担其工作量；
- 定期合并Fsimage和Edits，并推送给NameNode；
- 在紧急情况下，可辅助恢复NameNode。

4) ResourceManager（JobTracker）：

- 处理客户端请求
- 监控NodeManager
- 启动或监控ApplicationMaster
- 资源的分配与调度

5) NodeManager（TaskTracker）：

- 管理单个节点上的资源
- 处理来自ResourceManager的命令
- 处理来自ApplicationMaster的命令

6) JobHistoryServer

- 负责管理MapReduce作业的历史记录和统计信息

7) HttpServer

- Hadoop的Web服务器，用于显示各种监控和管理信息

8) DFSZKFailoverController

- **高可用**时它负责监控NN的状态，并及时的把状态信息写入ZK。它通过一个独立线程周期性的调用NN上的一个特定接口来获取NN的健康状态。FC也有选择谁作为Active NN的权利，因为最多只有两个节点，目前选择策略还比较简单（先到先得，轮换）。

9) JournalNode

- **高可用**情况下存放namenode的editlog文件。

欢迎加入知识星球，获取《大数据面试题 V4.0》以及更多大数据开发学习资料

