

1. `hdfs namenode -format` #第一次启动需要格式化namenode

2.jps 可以查看java进程,以此来确定hadoop是否启动成功

3.sbin/start-dfs.sh # 启动hadoop分布式存储进程 启动了: namenode(主节点)  
SecondaryNameNode(备份节点) datenode(从节点)

4.sbin/start-yarn.sh # 启动hadoop资源管理进程  
// (可以使用start-all.sh来启动hdfs和yarn,但是已被抛弃,因为这个命令启动resourceManager有问题)

5. mr-jobhistory-daemon.sh start historyserver # 启动历史服务器

6. `hdfs dfsadmin -safemode leave` # 离开安全模式

统计test.txt文件中的单词数

7. `yarn jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar wordcount /test.txt /wc_out`

8.

`scp a.txt hadoop@slave01:~/data` 将a.txt文件传输到slave01的~/data文件夹下  
`scp -r aa hadoop@slave01:~/data` 将aa文件夹传输到slave01的~/data文件夹下

HDFS常用命令 (大多命令基本和linux命令一样) :

// 全部路径是 绝对路径

`hdfs dfs -put *.txt` //上传文件到hdfs

`hdfs dfs -rm /hdfs001.iml` // 删除文件,支持正则

`hdfs dfs -rm -r /hq` // 删除文件夹

<https://www.jianshu.com/p/27c1da28c8fb>