Hadoop

# 一、Hadoop的HelloWord

## 1、安装hadoop，运行grep和wordcount程序

|  |
| --- |
| 0、解压，并配置环境变量  1、运行wordcount程序  hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcount wcinput/ wcoutput  2、运行grep程序  hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar grep input output 'dfs[a-z.]+' |

## 2、配置伪分布式环境，HDFS系统，HDFS的操作

vim $Hadoop\_home /etc/hadoop/core-site.xml

|  |
| --- |
| <configuration>  <!-- 指定HDFS中的NameNode的地址 -->  <property>  <name>fs.defaultFS</name>  <!—这里要注意host与ip的映射 -->  <value>hdfs://hadoop101:8020</value>  </property>  </configuration> |
|  |

vim $Hadoop\_home/etc/hadoop/hdfs-site.xml

|  |
| --- |
| <configuration>  <!-- 指定HDFS副本的数量 -->  <property>  <name>dfs.replication</name>  <value>1</value>  </property>  </configuration> |

第一次要初始化

|  |
| --- |
| bin/hdfs namenode –format |

启动namenode

|  |
| --- |
| sbin/hadoop-daemon.sh start namenode |

启动datanode

|  |
| --- |
| sbin/hadoop-daemon.sh start datanode |

jps可查看namenode和datanode是否启动成功

都启动成功后，可访问<http://ip:50070> 查看web端的管理页面

## 3、HDFS操作

|  |
| --- |
| 1. 创建目录   hadoop fs -mkdir -p /user/hadoop/input   1. 查看目录   hadoop fs -ls -R /   1. 上传文件   hadoop fs -put wcinput/wc.input /user/hadoop/input |
| 1. 在hdfs上运行mapreduce程序   hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.2.jar wordcount /user/hadoop/input/ /user/hadoop/output |