Hadoop环境搭建：完全分布式

集群规划：

ip hostname

192.168.204.154 master namenode resourcemanager datanode nodemanager

192.168.204.155 slave01 datanode nodemanager

192.168.204.156 slave02 datanode nodemanager

secondarynamenode是namenode的冷备份（不能代替namenode的工作，仅仅是拷贝namenode上的基础信息，帮助NameNode进行恢复）

安装配置完全分布式：

1. 安装jdk

把master上的拷贝到slave01和slave02上

1)发送jdk安装包

scp -r jdk1.8.0\_121 192.168.204.155:/home/hadoop/

scp -r jdk1.8.0\_121 192.168.204.156:/home/hadoop/

2）发送配置文件

sudo scp /etc/profile 192.168.204.155:/etc

sudo scp /etc/profile 192.168.204.156:/etc

3）生效配置文件

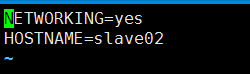
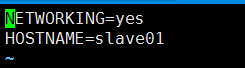
source /etc/profile

source /etc/profile

1. 配置主机名及映射文件

先改主机名

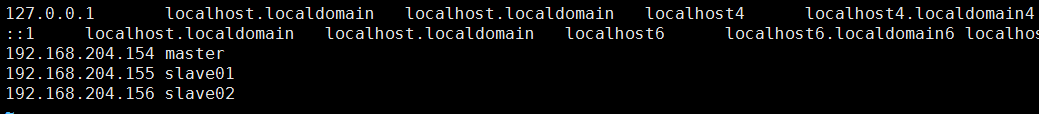
sudo vi /etc/sysconfig/network



需要改映射文件（将主机名和ip进行绑定）

三台机器均需要执行这个

vi /etc/hosts 末尾添加



1. 配置免密码登录

主节点向从节点免密码登录

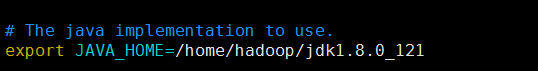
三个节点都要做

ssh-keygen

ssh-copy-id master/slave01/slave02

1. 安装配置hadoop

hadoop-env.sh



core-site.xml

|  |
| --- |
| <property>  <name>fs.defaultFS</name>  <value>hdfs://master:8020</value>  </property>  <property>  <name>hadoop.tmp.dir</name>  <value>/home/hadoop/hadoopdata</value>  </property> |

hdfs-site.xml

|  |
| --- |
| <property>  <name>dfs.replication</name>  <value>2</value>  </property> |

mapred-site.xml

|  |
| --- |
| <property>  <name>mapreduce.framework.name</name>  <value>yarn</value>  </property> |

yarn-site.xml

|  |
| --- |
| <property>  <name>yarn.nodemanager.aux-services</name>  <value>mapreduce\_shuffle</value>  </property>  <property>  <name>yarn.resourcemanager.hostname</name>  <value>master</value>  </property> |

slaves文件

|  |
| --- |
| master  slave01  slave02 |

向slave01、slave02远程发送hadoop安装文件

scp -r hadoop-2.7.1 slave01:/home/hadoop/

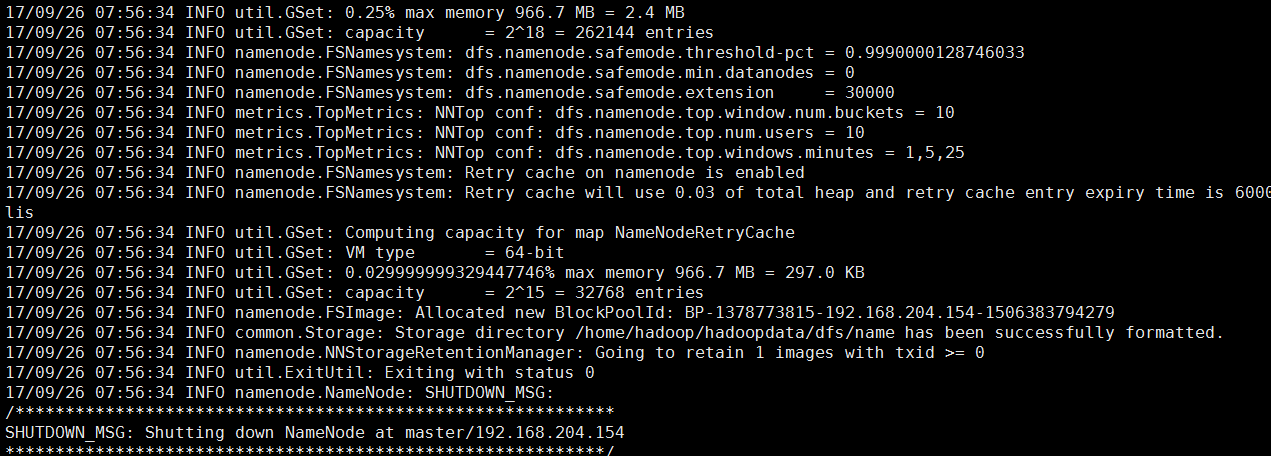
scp -r hadoop-2.7.1 slave02:/home/hadoop/

1. 格式化集群

先把原来的删除临时文件目录（/tmp/hadoop-hadoop）

只需要在主节点上执行

hadoop namenode -format



6.在主节点启动

start-dfs.sh

start-yarn.sh

7.验证 jps

NameNode DataNode SecondaryNameNode ResourceManager NodeManager

8.停止

stop-all.sh

建议stop-dfs.sh stop-yarn.sh

\*\*\*\*\*\*\*跟踪日志：

tail -f hadoop-rxp233-namenode-rxp233.log

\*\*\*\*\*\*\*单个启动程序

hadoop-daemon.sh start namenode | DataNode | SecondaryNameNode

yarn-deamon.sh start resourcemanager | nodemanager

访问端口：

50070:hdfs的namenode的webui访问端口

ip:50070

8088:yarn的resourcemanager的webui的访问端口

ip:8088

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*将namenode和secondarynamenode分开配置：

1. 在HADOOP\_HOME/etc/hadoop目录下添加配置文件

masters(添加secondarynn的节点)

vi masters

|  |
| --- |
| slave01(secondarynamenode的节点) |

远程拷贝到其他节点：

scp masters slave01:/home/hadoop/hadoop-2.7.1/etc/hadoop/

scp masters slave02:/home/hadoop/hadoop-2.7.1/etc/hadoop/

1. 修改hdfs-site.xml文件

1.>添加namenode的访问address

2.>添加secondarynamenode的访问address

|  |
| --- |
| <property>  <name>dfs.namenode.http-address</name>  <value>master:50070</value>  </property>  <property>  <name>dfs.namenode.secondary.http-address</name>  <value>slave01:50090</value>  </property> |

scp hdfs-site.xml slave01:/home/hadoop/hadoop-2.7.1/etc/hadoop/

scp hdfs-site.xml slave02:/home/hadoop/hadoop-2.7.1/etc/hadoop/