参考：https://blog.csdn.net/m0\_58258383/article/details/125736734

环境：CentOS 7.9

namenode未启动解决：https://www.cnblogs.com/chuijingjing/p/13828874.html

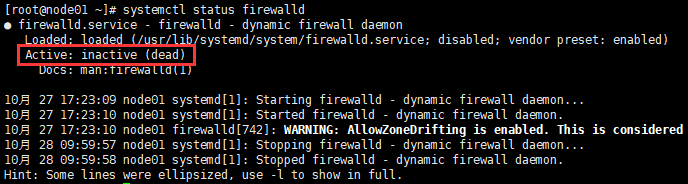
#### 1. 配置静态IP

#### 2. 设置主机域名映射

vi /etc/hosts  
# 加入如下，IP地址根据实际情况更改  
192.168.110.83 node01  
192.168.110.84 node02  
192.168.110.85 node03

#### 3. 关闭防火墙

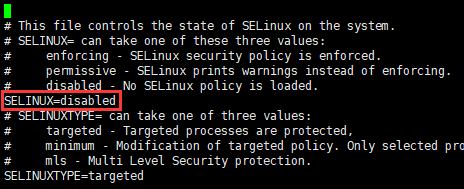
# 临时关闭防火墙  
systemctl stop firewalld  
# 彻底关闭防火墙  
systemctl disable firewalld  
# 查看防火墙状态  
systemctl status firewalld



#### 4. selinux关闭

vi /etc/selinux/config

SELINUX=disabled



#### 5. 设置ssh免密

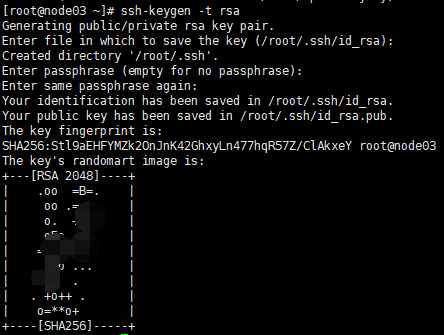
**5.1 生成公钥**

以下步骤**【必须】**在**node01、node02、node03**上分别执行一次

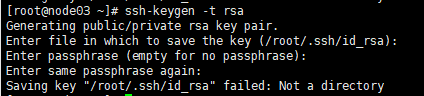
# root路径下执行，生成密钥，三个回车

ssh-keygen -t rsa

成功的话如下：



可能会出现的异常：



Saving key "/root/.ssh/id\_rsa" failed: Not a directory

执行：rm .ssh 是否删除：y

之后再执行一次步骤1.5.1

**5.2 公钥拷贝**

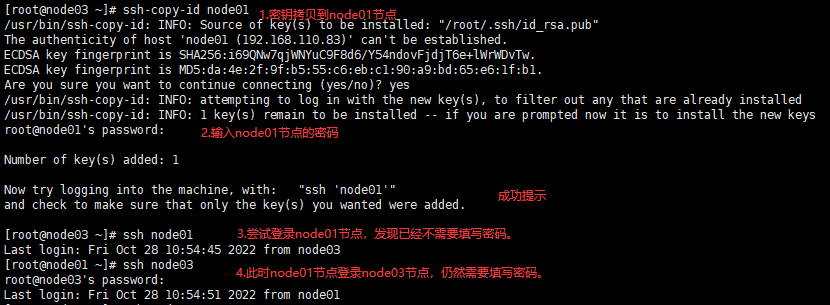
以下步骤**【必须】**在**node01、node02、node03**上分别执行一次

# 拷贝密钥到node01节点。

ssh-copy-id node01

然后输入node01节点的密码

以当前节点为node03，拷贝到node01节点为例：



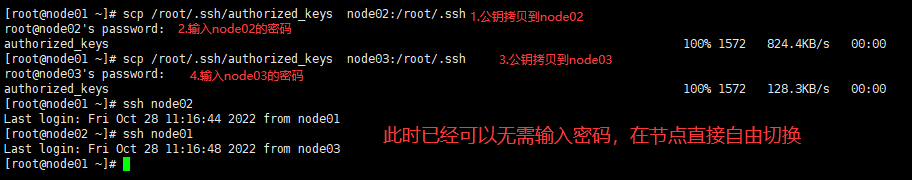
**5.3 公钥复制**

以下步骤仅在node01执行一次

此时node01节点上，已经有01、02、03三个节点的密钥，复制公钥到其他节点。

scp /root/.ssh/authorized\_keys node02:/root/.ssh

scp /root/.ssh/authorized\_keys node03:/root/.ssh



切换后记得切换回原始节点，防止后续部署步骤混乱。

#### 开始HDFS完全分布式集群搭建与配置

**6.1 hadoop安装**

进入node01节点

创建存放文件夹

mkdir -p /opt/software/hadoop/

上传hadoop-3.2.4.tar.gz到该文件夹下，并解压

tar -zxvf hadoop-3.2.4.tar.gz

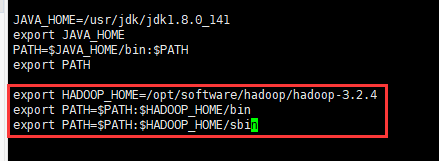
配置HADOOP\_HOME环境变量

vi /etc/profile

export HADOOP\_HOME=/opt/software/hadoop/hadoop-3.2.4

export PATH=$PATH:$HADOOP\_HOME/bin

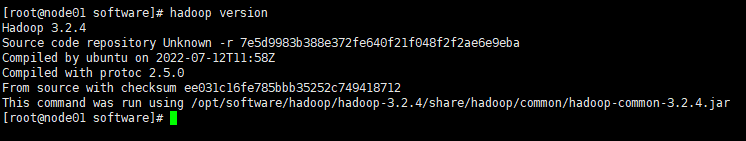
export PATH=$PATH:$HADOOP\_HOME/sbin



source /etc/profile

测试hadoop是否安装成功

hadoop version



**6.2修改配置文件**

6.2.1 创建文件夹

cd /opt/software

mkdir -p hdfs/data

6.2.2 修改core-site.xml文件

进入文件目录

cd /opt/software/hadoop/hadoop-3.2.4/etc/hadoop/

vi core-site.xml

加入如下内容：

<!-- 指定NameNode的地址 -->

<property>

<name>fs.defaultFS</name>

<value>hdfs://node01:8020</value>

</property>

<!-- 指定hadoop数据的存储目录 -->

<property>

<name>hadoop.tmp.dir</name>

<value>/opt/software/hadoop/hdfs/data</value>

</property>

<!-- 配置HDFS网页登录使用的静态用户为hr-->

<property>

<name>hadoop.http.staticuser.user</name>

<value>hr</value>

</property>



保存退出

6.2.3 修改hdfs-site.xml文件

加入以下内容：

<!-- nn web端访问地址-->

<property>

<name>dfs.namenode.http-address</name>

<value>node01:9870</value>

</property>



6.2.4 修改yarn-site.xml文件

加入以下内容：

<!-- 指定MR走shuffle -->

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

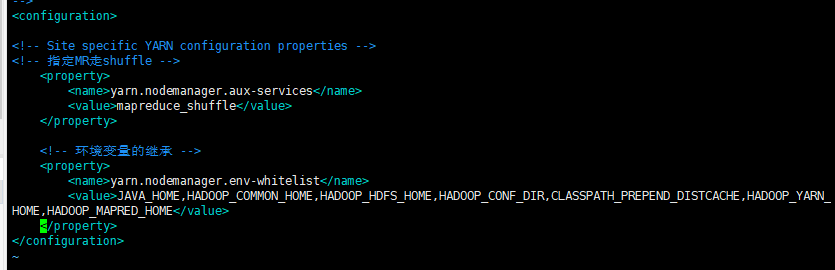
<!-- 环境变量的继承 -->

<property>

<name>yarn.nodemanager.env-whitelist</name>

<value>JAVA\_HOME,HADOOP\_COMMON\_HOME,HADOOP\_HDFS\_HOME,HADOOP\_CONF\_DIR,CLASSPATH\_PREPEND\_DISTCACHE,HADOOP\_YARN\_HOME,HADOOP\_MAPRED\_HOME</value>

</property>



6.2.5 修改mapred-site.xml文件

加入以下内容：

<!-- 指定MapReduce程序运行在Yarn上 -->

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

<!-- 历史服务器端地址 -->

<property>

<name>mapreduce.jobhistory.address</name>

<value>node01:10020</value>

</property>

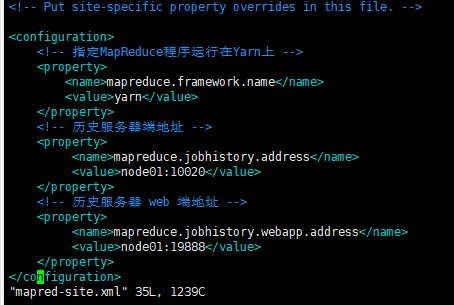
<!-- 历史服务器 web 端地址 -->

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>node01:19888</value>

</property>



6.2.6 配置datanode集群

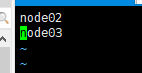
vi workers

删除localhost

添加以下内容（注意：该文件中添加的内容结尾不允许有空格，文件中不允许有空行）（此处的意思是，只有node02和node03启动datanode服务）

node02

node03

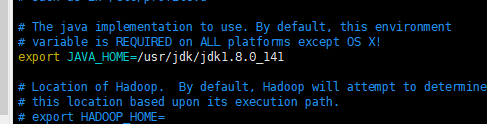


6.2.7 修改hadoop-env.sh文件

大约在文件第54行

（按照JDK真实路径灵活修改）

export JAVA\_HOME=/usr/jdk/jdk1.8.0\_141



**6.3 拷贝信息到其他节点**

把node01上的配置同步到node02和node03

进入node02

mkdir -p /opt/software

进入node03

mkdir -p /opt/software

回到node01

cd /opt/software/

scp -r hadoop/ node02:/opt/software/

scp -r hadoop/ node03:/opt/software/

**6.4 配置环境变量**

**以下步骤【必须】分别在node02和node03上执行，保证所有服务器均进行了配置**

vi /etc/profile

export HADOOP\_HOME=/opt/software/hadoop/hadoop-3.2.4

export PATH=$PATH:$HADOOP\_HOME/bin

export PATH=$PATH:$HADOOP\_HOME/sbin

更新

source /etc/profile

hadoop version 查看是否安装成功。此处如果未成功，请检查scp命令拷贝到该节点的hadoop是否在理想路径上。

**6.5 集群启动与测试**

如果集群是第一次启动，需要在node01节点格式化NameNode（注意：格式化NameNode，会产生新的集群id，导致NameNode和DataNode的集群id不一致，集群找不到已往数据。如果集群在运行过程中报错，需要重新格式化NameNode的话，一定要先停止namenode和datanode进程，并且要删除所有机器的data和logs目录，然后再进行格式化。

6.5.1 格式化

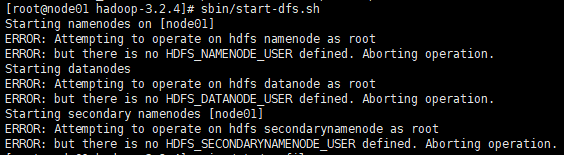
hdfs namenode -format

6.5.2 启动HDFS

cd /opt/software/hadoop/hadoop-3.2.4/

sbin/start-dfs.sh

**可能出现的异常：**



vim /etc/profile

添加（root为当前用户名）

export HDFS\_NAMENODE\_USER=root

export HDFS\_DATANODE\_USER=root

export HDFS\_SECONDARYNAMENODE\_USER=root

export YARN\_RESOURCEMANAGER\_USER=root

export YARN\_NODEMANAGER\_USER=root

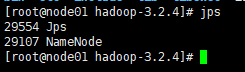
source /etc/profile 更新

然后再次执行6.5.2步骤

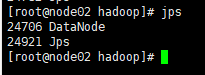
6.5.3 检查是否启动成功

node1:

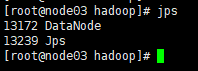
jps



node2:



node3:

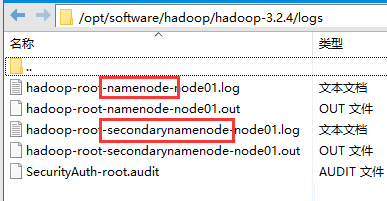


perfect!!

**若以上启动任务偏少**

可以去以下路径查看对应日志：

/opt/software/hadoop/hadoop-3.2.4/logs

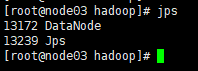


这里是node01节点，只负责namenode，所以日志只有这俩。node02，03会有datanode日志。

**若以上启动任务偏多**

例如，有时部分节点会默认启动SecondaryNameNode，此时我们只想关闭这个服务，操作如下：

jps



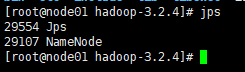
假设此处DataNode为SecondaryNameNode，其PID为13172

直接 kill 13172，并再次jps命令，确认是否关闭即可。

**6.6 关闭服务**

sbin/stop-all.sh

如果想关闭单个服务

****

jps命令后，会出现已启动服务的pid，直接kill [pid]

**6.7 sbin下sh脚本解释：**

start-dfs.sh 只启动namenode 和datanode

start-yarn.sh 只启动resourcemanager 和nodemanager

start-all.sh 启动namenode 和datanode和yarn的resourcemanager 和nodemanager

web浏览器中输入：host:9870 （host为namenode节点，即node01的IP地址）

此处为：192.168.110.83:9870

