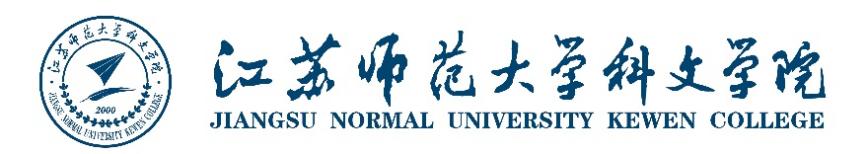
****

**人工智能与软件学院**

实 验 报 告

**课 程 大数据分析与智能决策**

**学 号 7200767121**

**姓 名 彭敏捷**

**班 级 20智科**

**教 师 胡元发**

2023年 3 月至 2023 年 4 月

**一、实验名称**

熟悉常用的Linux操作和Hadoop操作

**二、实验目的和要求**

1、采用虚拟机或双系统方式安装Linux后安装Hadoop，测试是否安装成功；

2、进行伪分布式配置，运行相同测试程序，比较单机环境与伪分布式环境的差异，认识分布式环境；

3、练习常用的Linux操作，练习常用的Hadoop操作，了解Hadoop处理架构的基本构成。

**三、实验环境**

1.VMware Workstation Pro

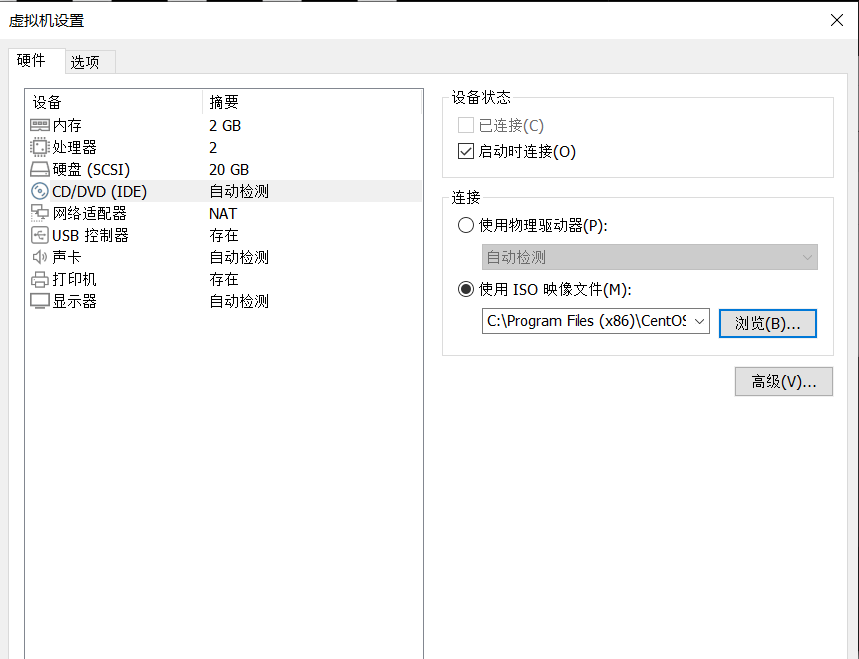
2.Xshell

3.Xftp

**四、实验步骤**

1.虚拟机

（1）新建虚拟机hadoop01





（2）虚拟机安装①设置主机名bigdata01,连接名称System eth0

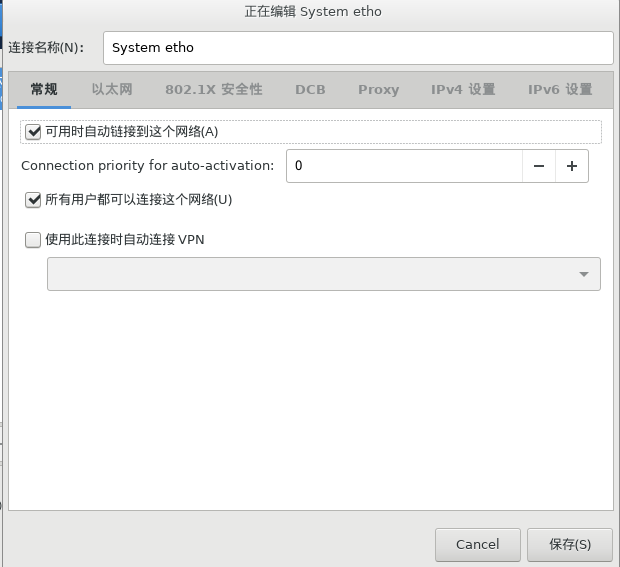
②设置ROOT密码 、创建用户

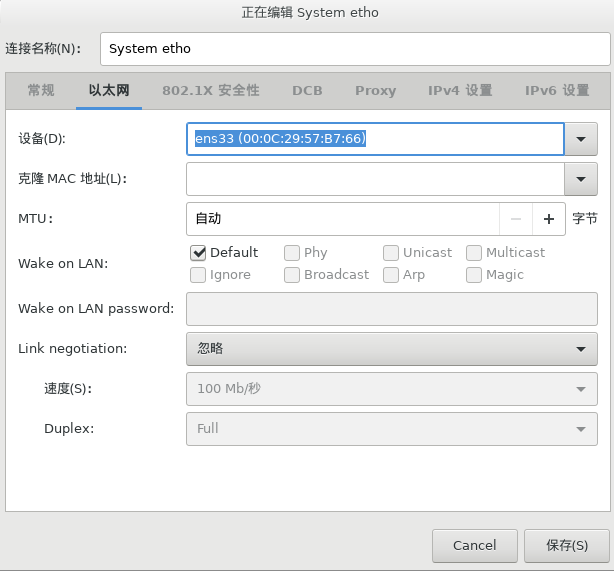
③IP地址：192.168.110.137

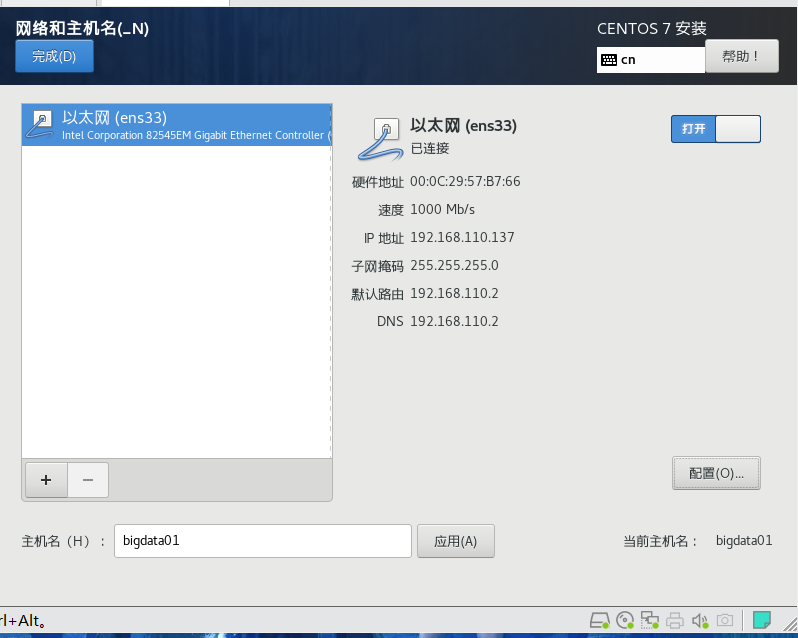
子网掩码：255.255.255.255.0

默认路由：192.168.110.2

DNS：192.168.110.2





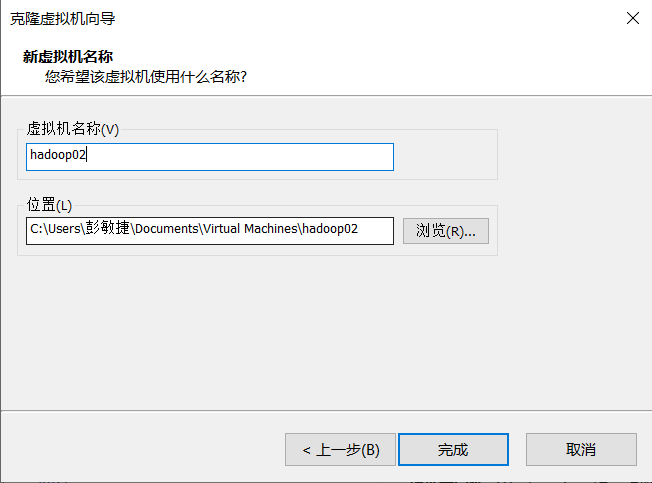


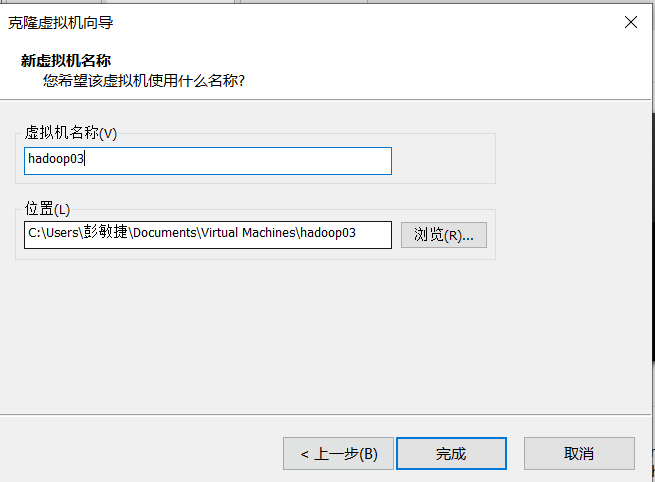


（3）虚拟机克隆：hadoop02、hadoop02

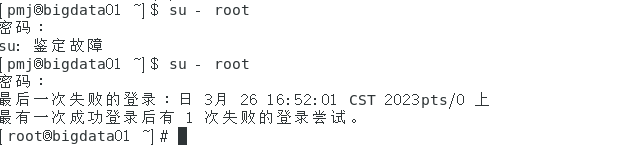
主机名和IP地址:bigdata02 192.168.110.129

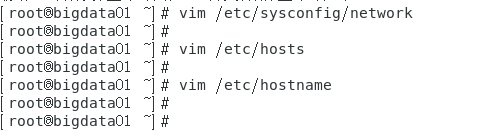
bigdata03 192.168.110.130





1. 主机配置[（su - root ）命令进入root@bigdata01~]





①vim /etc/sysconfig/network 添加：

NETWORKING=yes

HOSTNAME=bigdata01

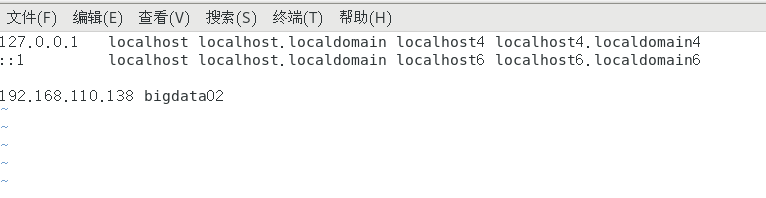


②vim /etc/hosts 添加主机名和IP的映射关系：

192.168.110.137 bigdata01

192.168.110.129 bigdata02

192.168.110.130 bigdata03

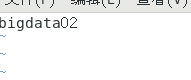


③vim /etc/hostname文件中添加机器名称：

bigdata01

bigdata02

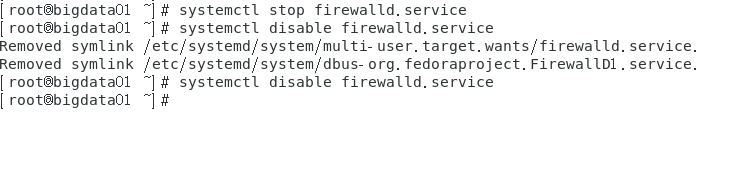
bigdata03



④关闭防火墙，关闭防火墙开机自启：

systemctl stop firewalld.service

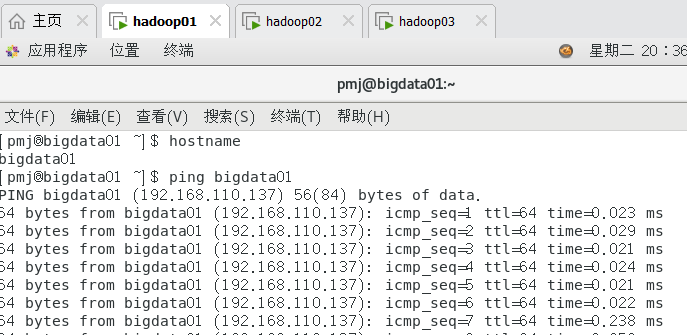
systemctl disable firewalld.service

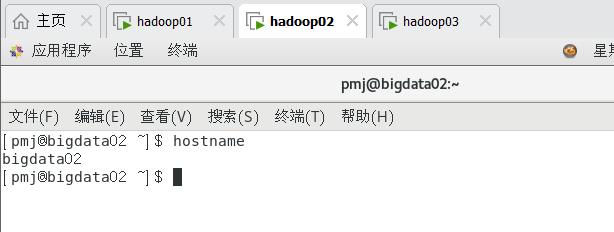


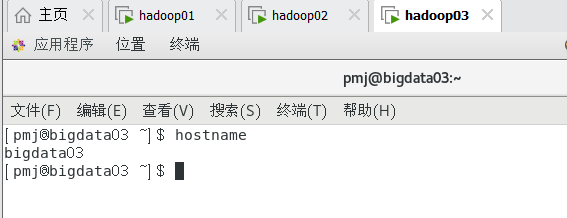
⑤重启reboot

⑥查看是否配置成功:hostname

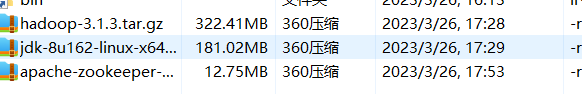
ping bigdata01







2.安装配置JDK(xftp7上传三个安装包tar、hadoop、zookeeper，地址/目录下)



（1）su - root 登录,新建目录/usr/java、hadoop

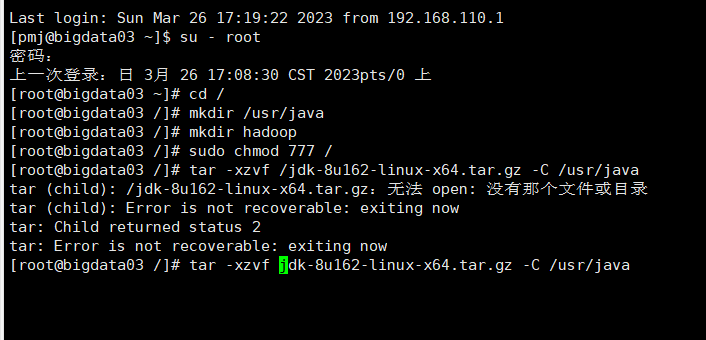
cd /

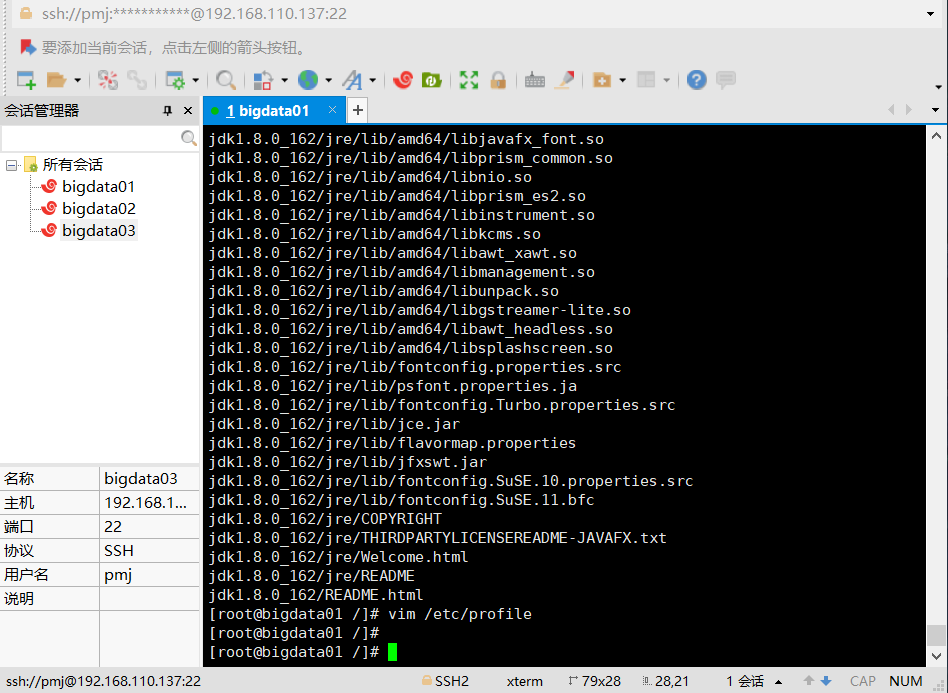
mkdir /usr/java

mkdir hadoop

（2）解压jdk安装包

tar -xzvf jdk-8u162-linux-x64.tar.gz -C /usr/java



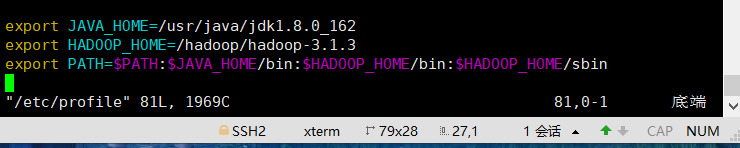


（3）配置环境变量。vim /etc/profile，添加：

export JAVA\_HOME=/usr/java/jdk1.8.0\_162

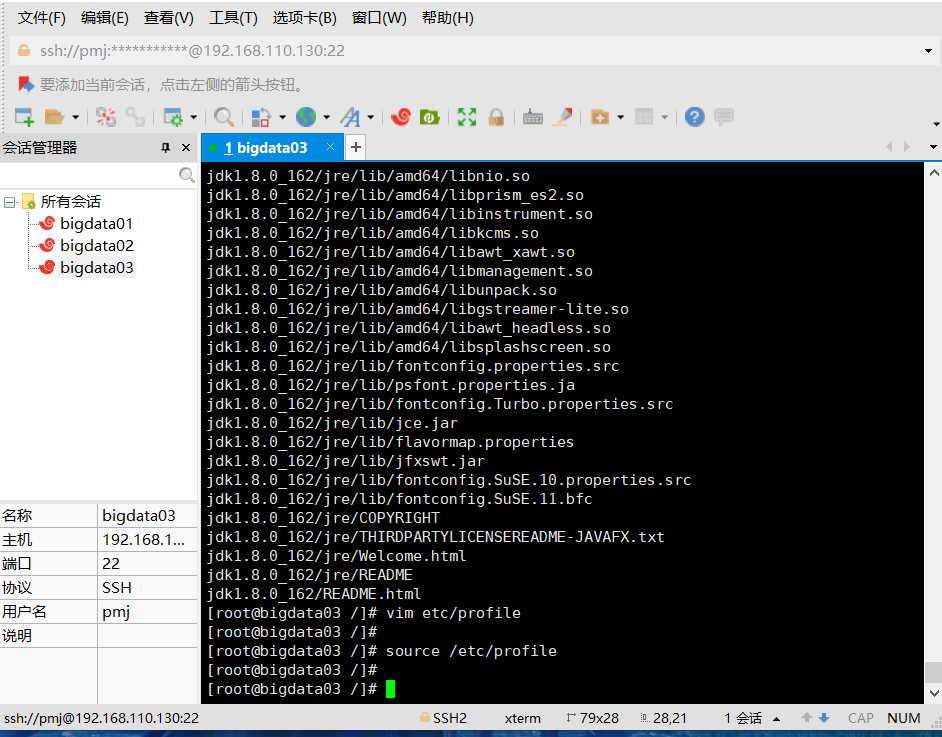
export HADOOP\_HOME=/hadoop/hadoop-3.1.3

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



（4） ：wq保存并退出

（5）验证环境变量：source /etc/profile

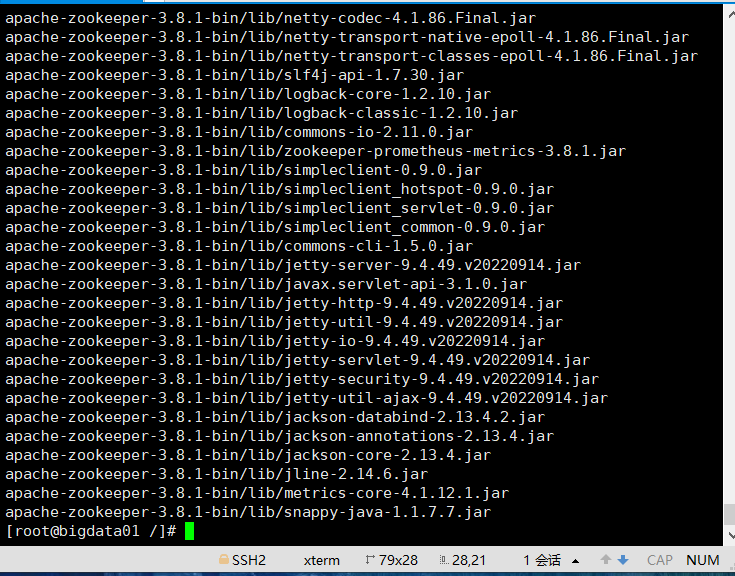


（6）！！！！注意：以上操作在每台机器分别进行

3.安装Zookeeper集群

（1）解压zookeeper安装包

tar -zxvf apache-zookeeper-3.8.1-bin.tar.gz -C /hadoop/



（2）配置zoo.cfg文件

进入conf目录: cd /hadoop/apache-zookeeper-3.8.1-bin/conf/

生成zoo.cfg文件：cp zoo\_sample.cfg zoo.cfg

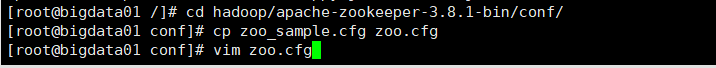
修改zoo.cfg文件，修改以下内容：

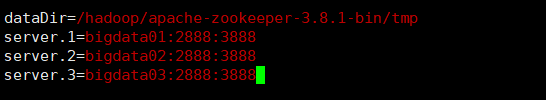
dataDir=/hadoop/apache-zookeeper-3.8.1-bin/tmp

server.1=bigdata01:2888:3888

server.2=bigdata02:2888:3888

server.3=bigdata03:2888:3888



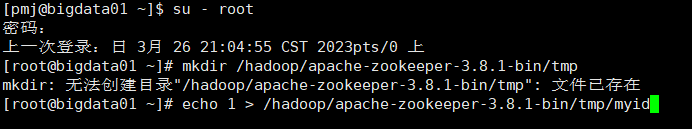


（3）创建一个tmp文件夹，命令如下：

mkdir /hadoop/apache-zookeeper-3.8.1-bin/tmp

（4）创建一个空文件，向该文件写入ID，命令如下：

echo 1 > /hadoop/apache-zookeeper-3.8.1-bin/tmp/myid

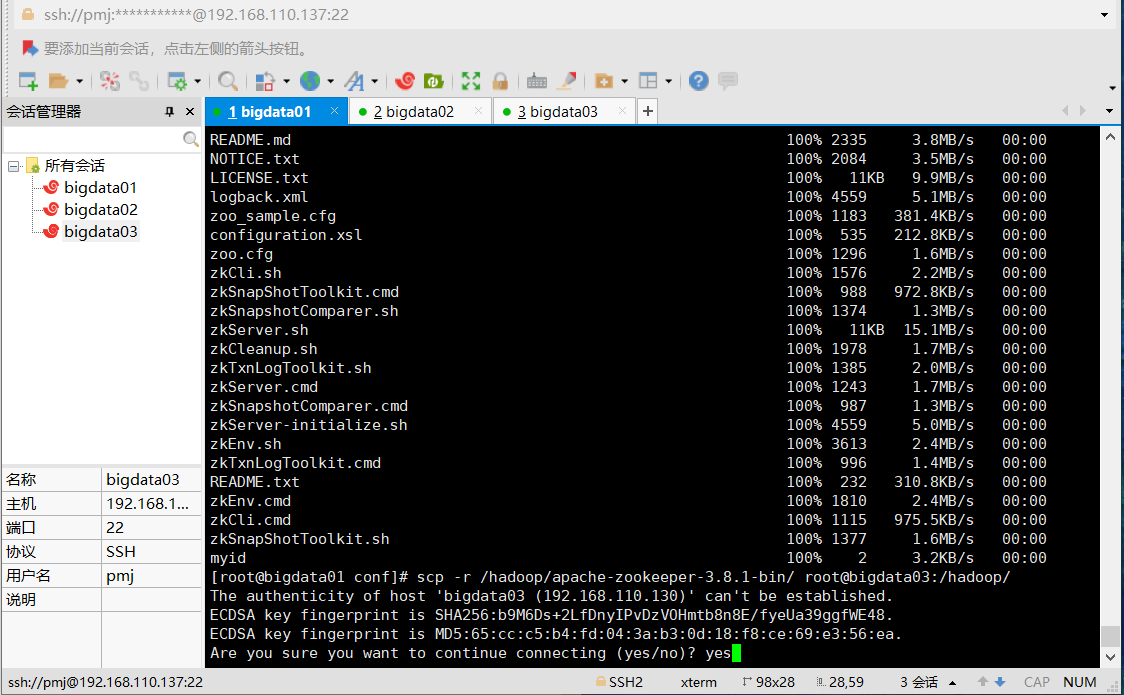


（5）注意：！！！！！以上操作仅在bigdata01上使用

（6）将配置好的zookeeper拷贝到其他节点(首先分别在bigdata02、 bigdata03根目录下创建一个hadoop目录)：

scp -r /hadoop/apache-zookeeper-3.8.1-bin/ root@bigdata02:/hadoop/

scp -r /hadoop/apache-zookeeper-3.8.1-bin/ root@bigdata03:/hadoop/



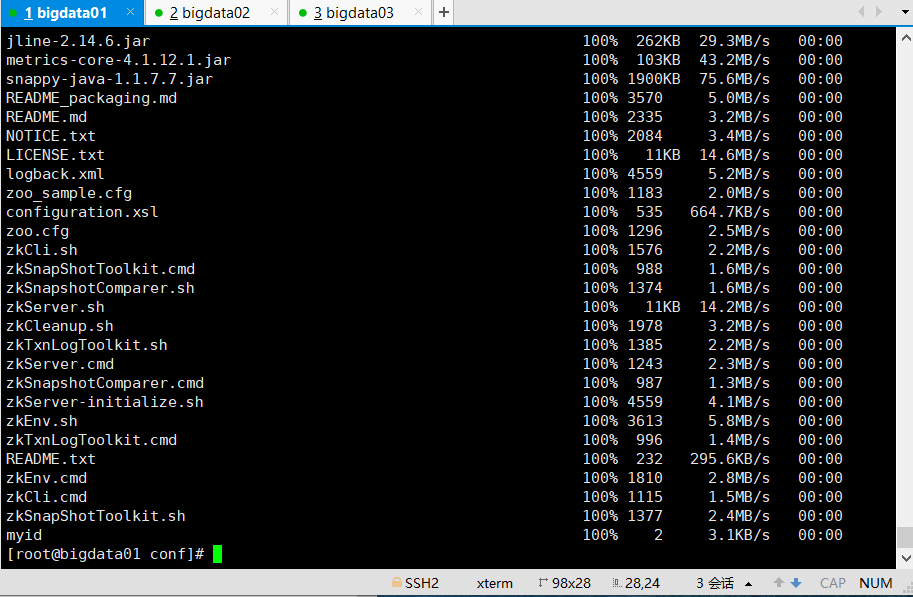
（7）修改bigdata02、bigdata03对应/hadoop/zookeeper-3.4.13/tmp/myid内容

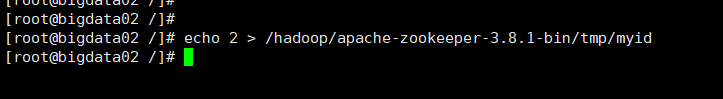
在bigdata02机器上，命令如下：

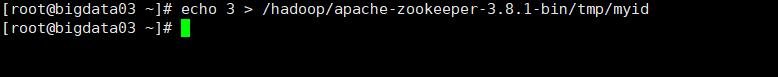
echo 2 > /hadoop/apache-zookeeper-3.8.1-bin/tmp/myid

在bigdata03机器上，命令如下：

echo 3 > /hadoop/apache-zookeeper-3.8.1-bin/tmp/myid



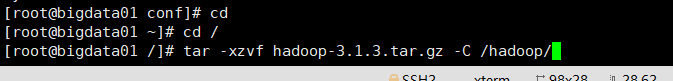


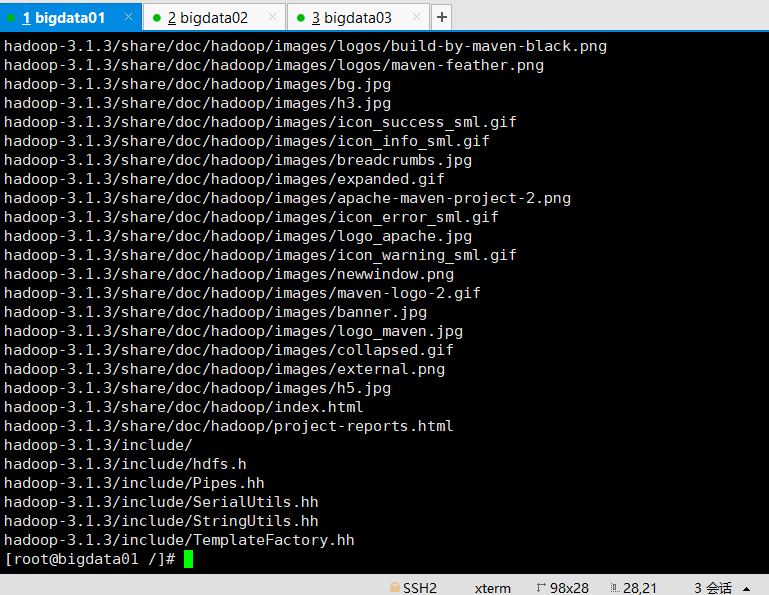


4.安装配置Hadoop集群(在bigdata01上进行）

（1）解压hadoop安装包：

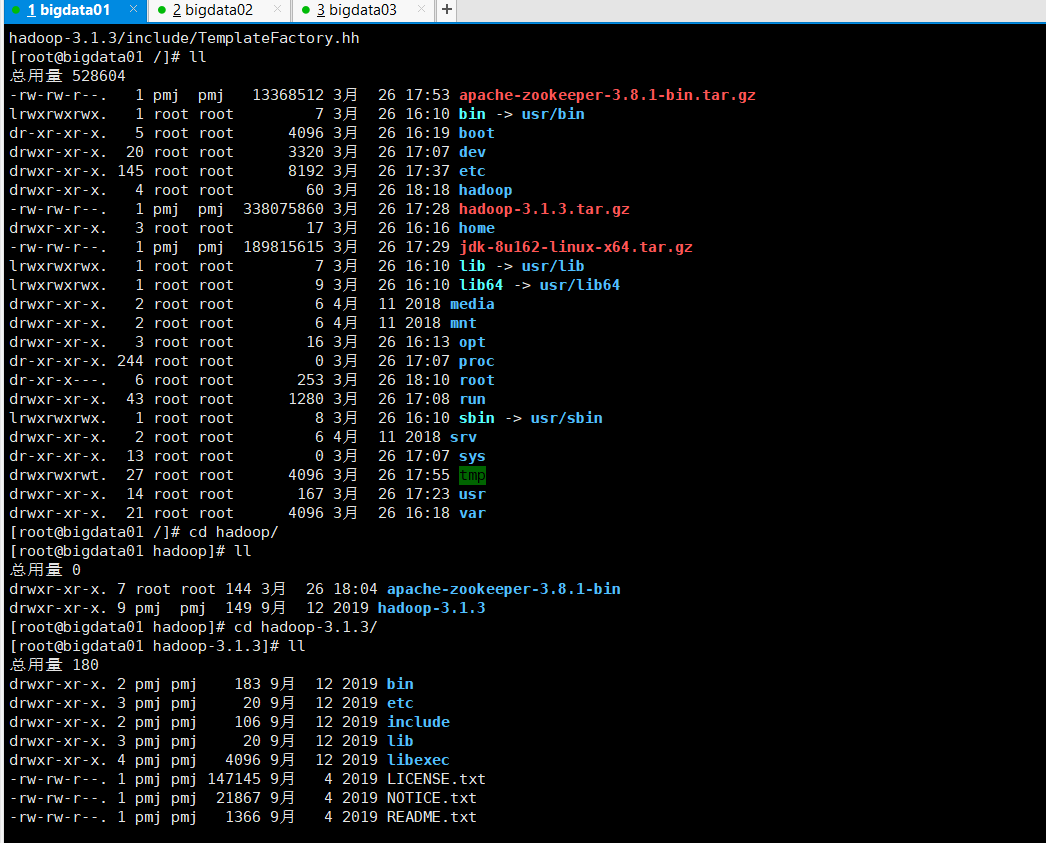
tar -xzvf hadoop-3.1.3.tar.gz -C /hadoop/

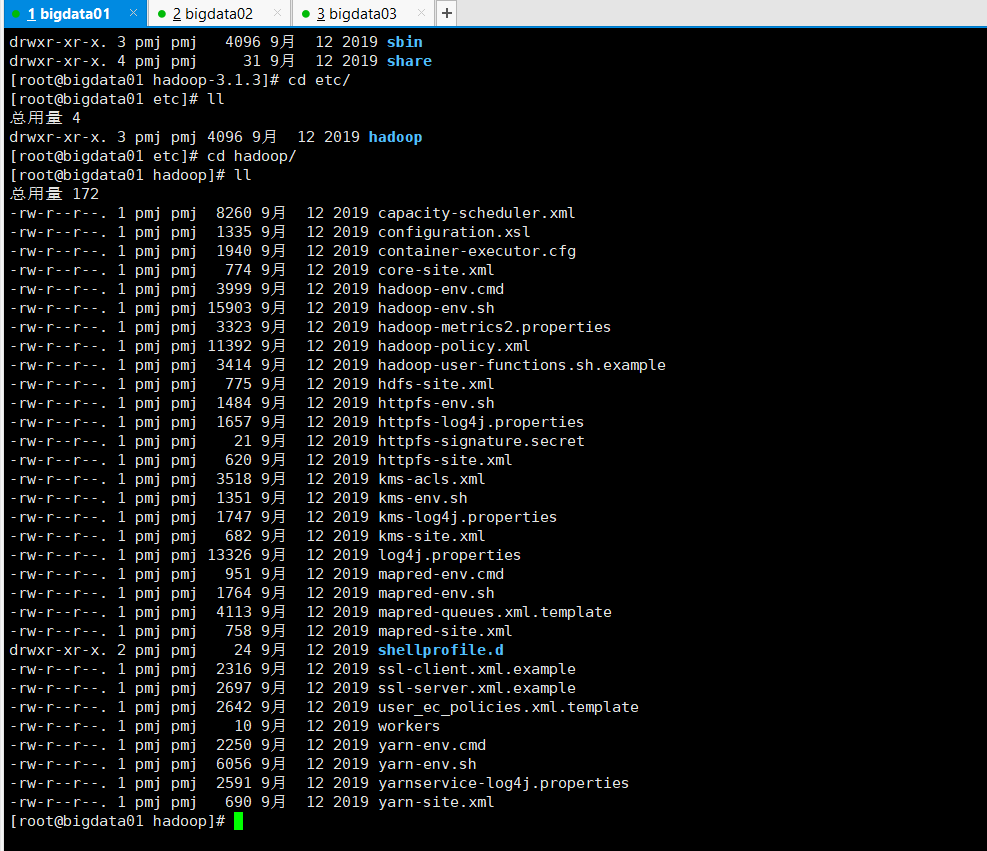


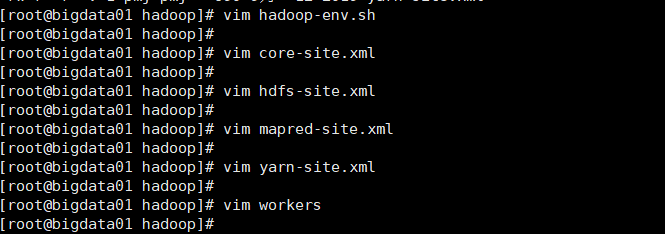


（2）配置HDFS

cd /hadoop/hadoop-3.1.3/etc/hadoop



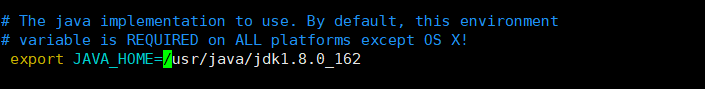




（3）修改hadoop-env.sh文件

vim hadoop-env.sh

修改export JAVA\_HOME=/usr/java/jdk1.8.0\_162



（4）修改core-site.xml文件

vim core-site.xml，修改

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://bigdata01:9000</value>

</property>

<property>

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

<value>/hadoop/hadoop-3.1.3/tmp</value>

</property>

<!-- 指定Zookeeper 地址 -->

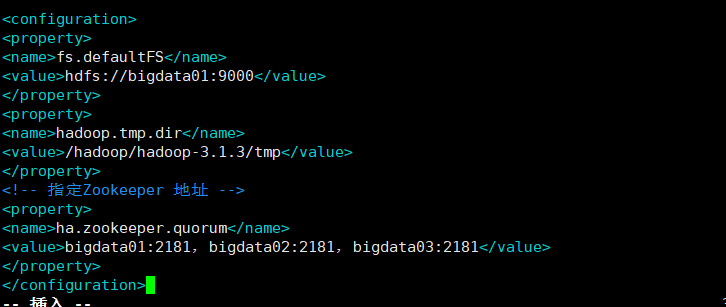
<property>

<name>ha.zookeeper.quorum</name>

<value>bigdata01:2181，bigdata02:2181，bigdata03:2181</value>

</property>

</configuration>



（5）修改vim hdfs-site.xml文件，修改：

<configuration>

<property>

<name>dfs.replication</name>

<value>3</value>

</property>

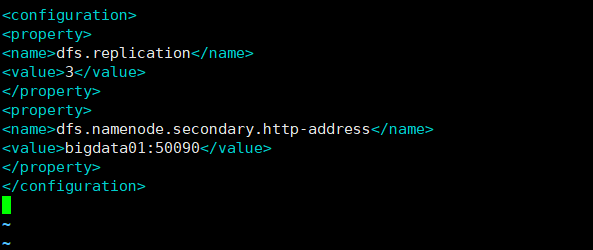
<property>

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

<value>bigdata01:50090</value>

</property>

</configuration>



（6）修改vim mapred-site.xml文件，修改：

<configuration>

<!-- 指定MapReduce运行时框架，这里指定在Yarn上，默认是 local -->

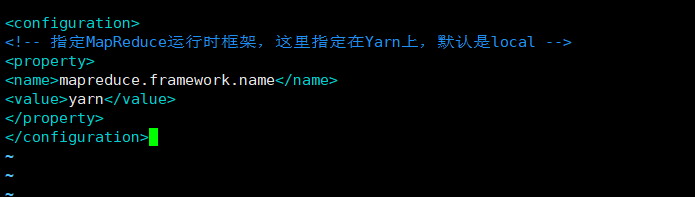
<property>

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

<value>yarn</value>

</property>

</configuration>



（7）修改vim yarn-site.xml文件，修改：

<configuration>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>bigdata01</value>

</property>

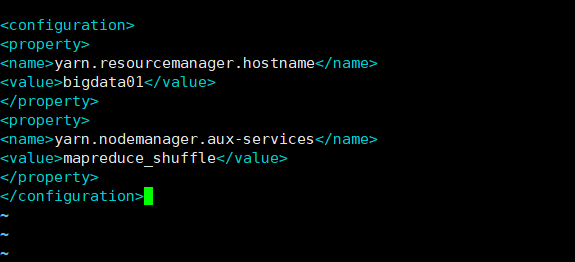
<property>

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

<value>mapreduce\_shuffle</value>

</property>

</configuration>



（8）修改vim workers文件。打开该配置文件，先删除里面的内容（默 认localhost），然后配置如下内容。

bigdata01

bigdata02

bigdata03

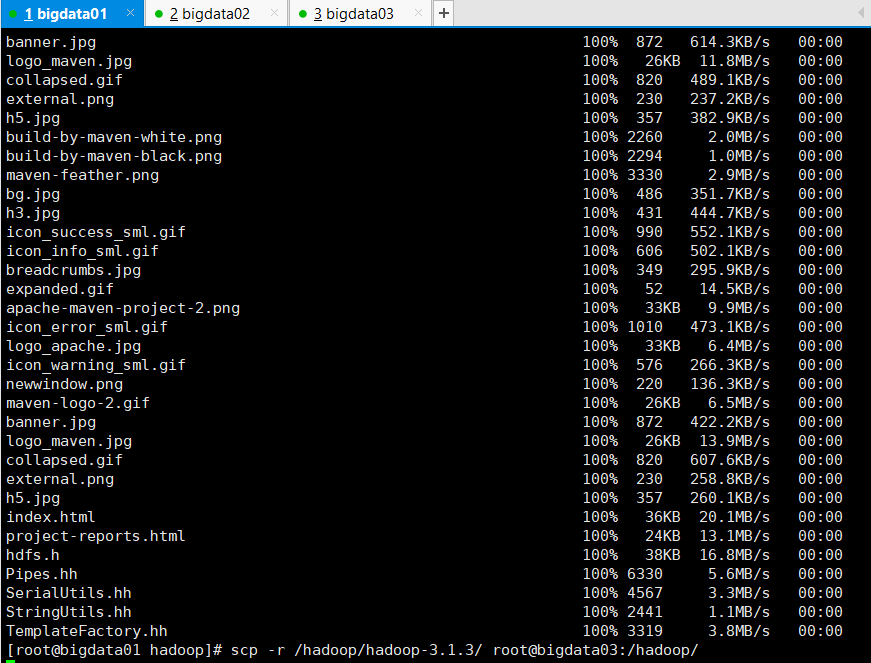


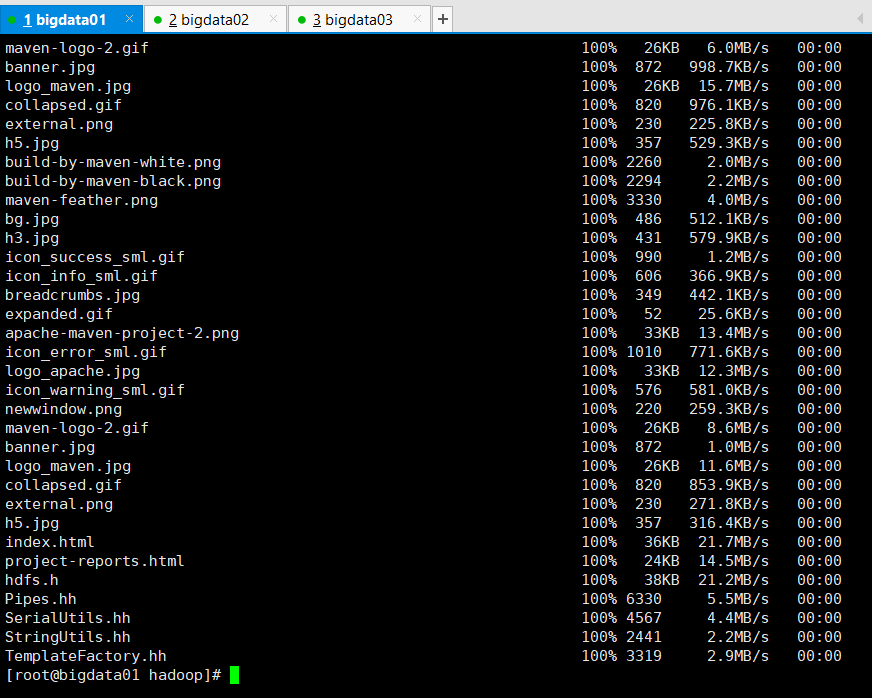
（9）将集群主节点的配置文件分发到其他子节点

完成Hadoop集群主节点hadoop01的配置后，还需要将系统环境配置文件、JDK安装目录和Hadoop安装目录分发到其他子节点hadoop02和hadoop03上，具体指令：

scp -r /hadoop/hadoop-3.1.3/ root@bigdata02:/hadoop/

scp -r /hadoop/hadoop-3.1.3/ root@bigdata03:/hadoop/





5.配置免密码登录

（1）配置bigdata01到bigdata02、bigdata03的免密码登录。

①在bigdata01上生产一对钥匙

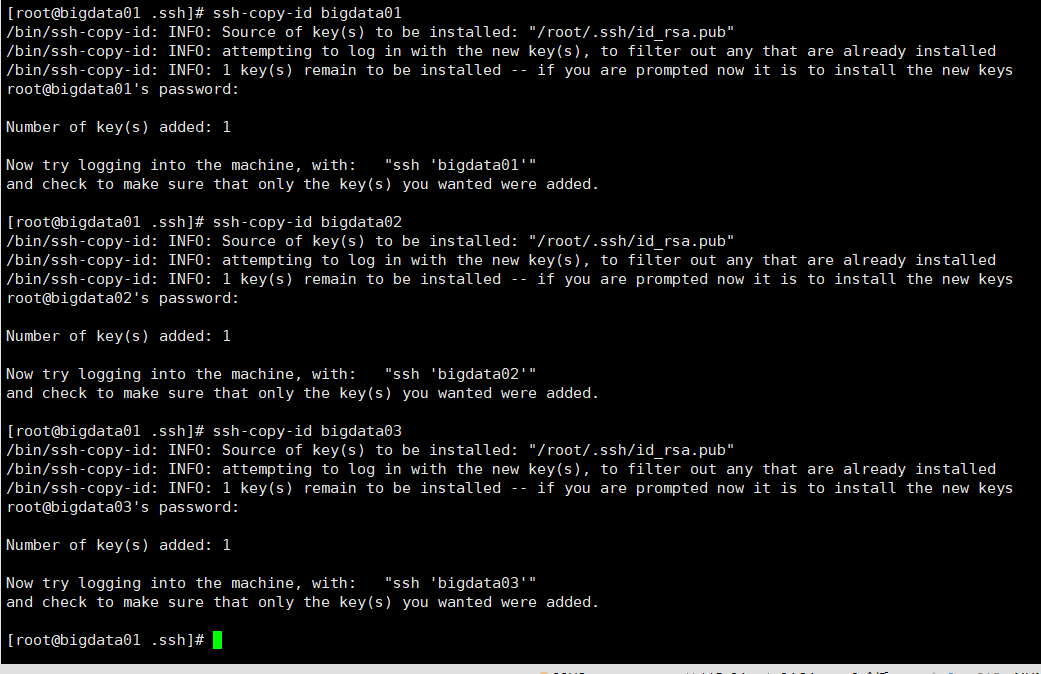
ssh-keygen -t rsa

②将公钥复制到其他节点（！！！包括自己）

ssh-copy-id bigdata01

ssh-copy-id bigdata02

ssh-copy-id bigdata03



（2）配置bigdata02到bigdata01、bigdata03的免密码登录。

①在bigdata01上生产一对钥匙

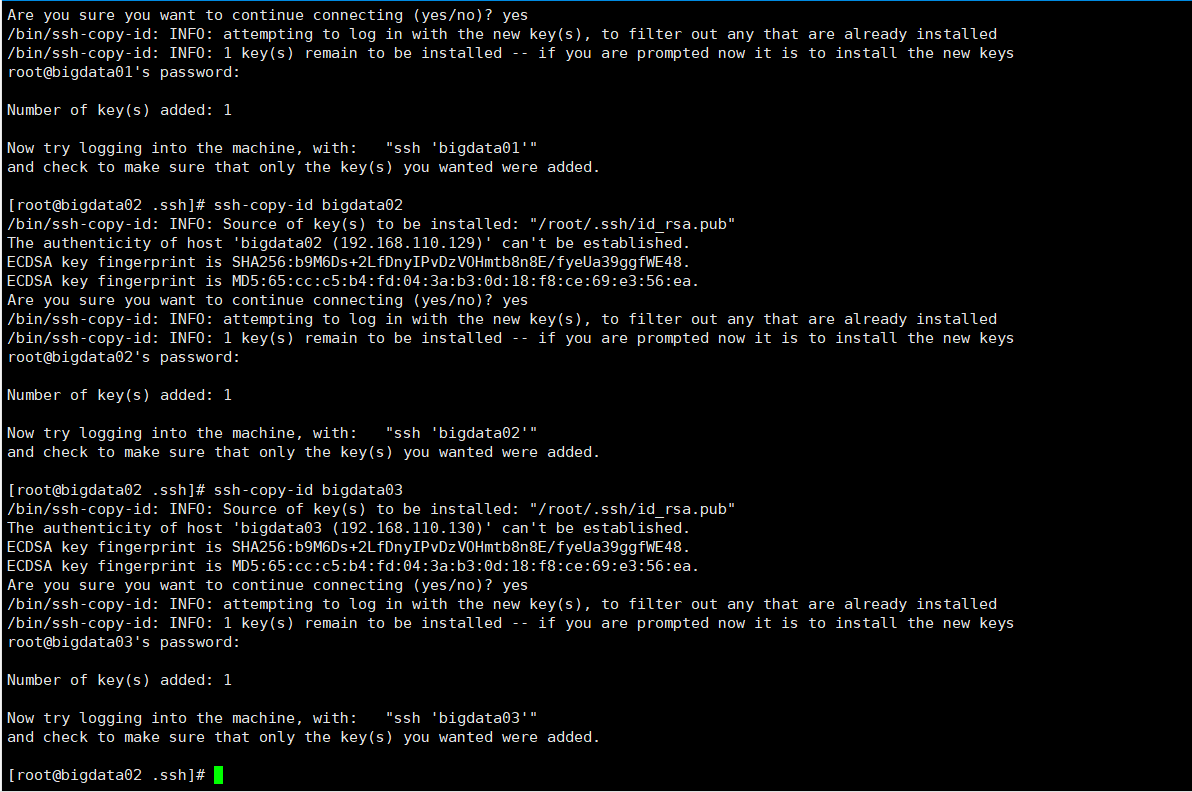
ssh-keygen -t rsa

②将公钥复制到其他节点（！！！包括自己）

ssh-copy-id bigdata01

ssh-copy-id bigdata02

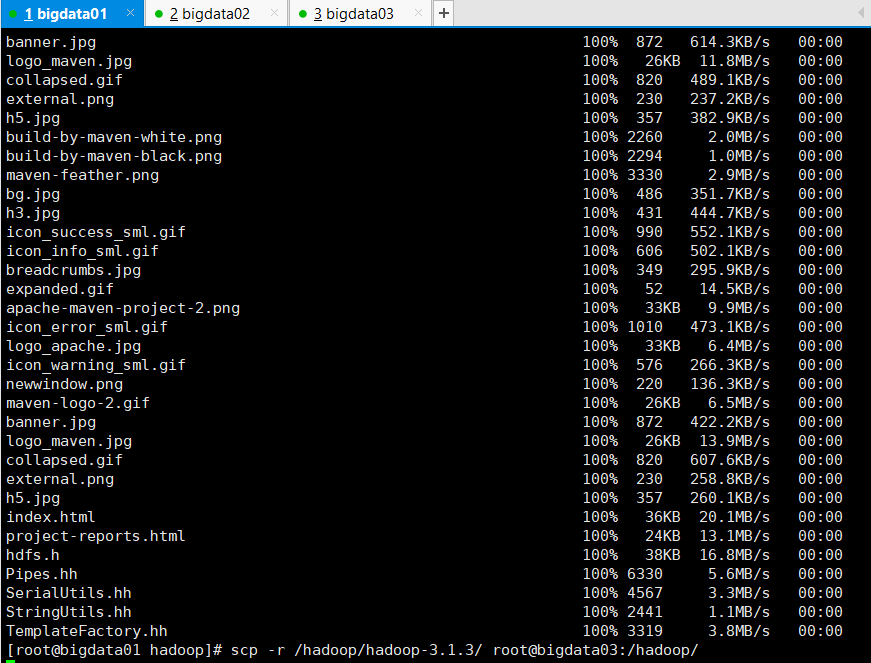
ssh-copy-id bigdata03



（3）将配置好的Hadoop复制到其他节点：

scp -r /hadoop/hadoop-3.1.3/ root@bigdata02:/hadoop/

scp -r /hadoop/hadoop-3.1.3/ root@bigdata03:/hadoop/



6.启动Hadoop

（1）启动zookeeper集群（！！！分别在bigdata01、bigdata02、bigdata03启动zk），命令如下:

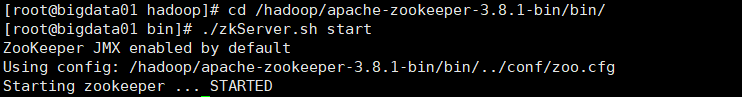
cd /hadoop/apache-zookeeper-3.8.1-bin/bin/

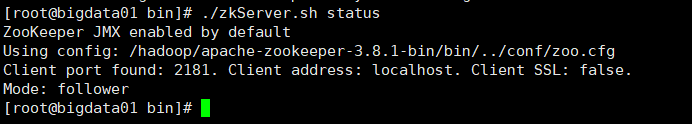
./zkServer.sh start

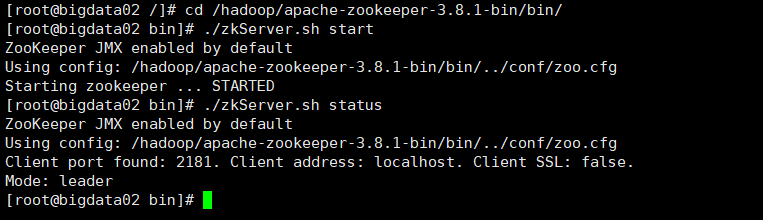
查看状态：命令如下：

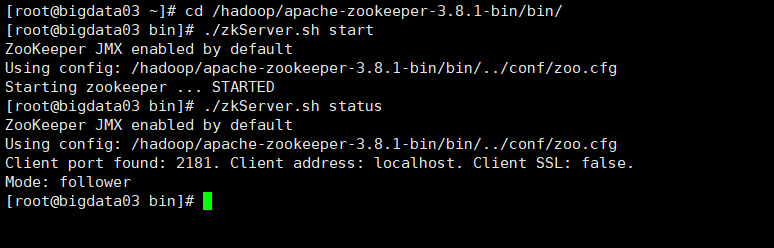
./zkServer.sh status

结果应为一个leader，两个follower





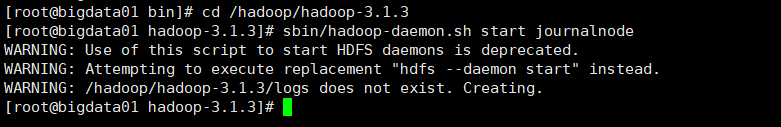


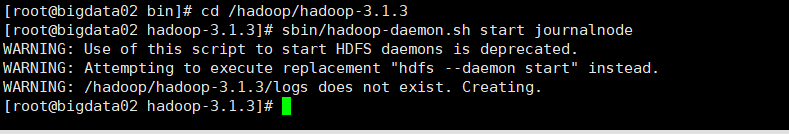


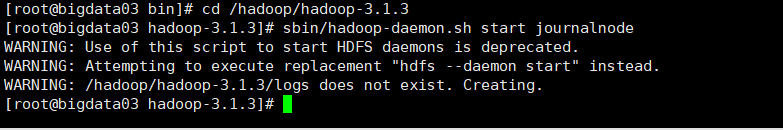
（2）启动journalnode（分别在在bigdata01、bigdata02、bigdata03上执行），命令如下：

cd /hadoop/hadoop-3.1.3

sbin/hadoop-daemon.sh start journalnode

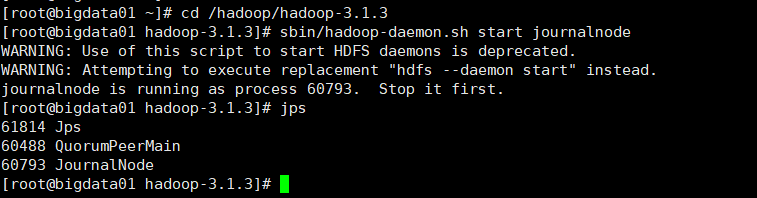


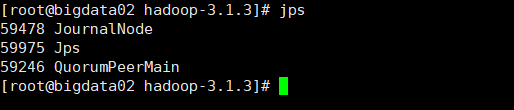


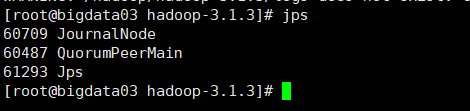


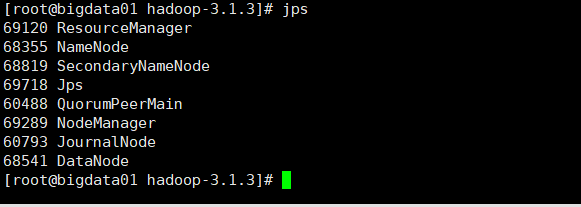
运行jps命令检验

bigdata01、bigdata02、bigdata03上多了JournalNode进程









（3）格式化HDFS（！！！！！在bigdata01上执行命令）:

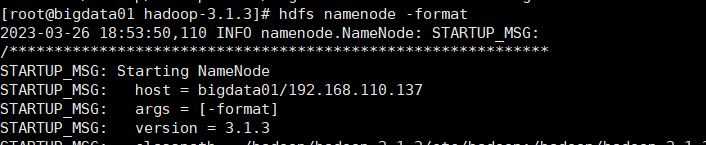
hdfs namenode -format

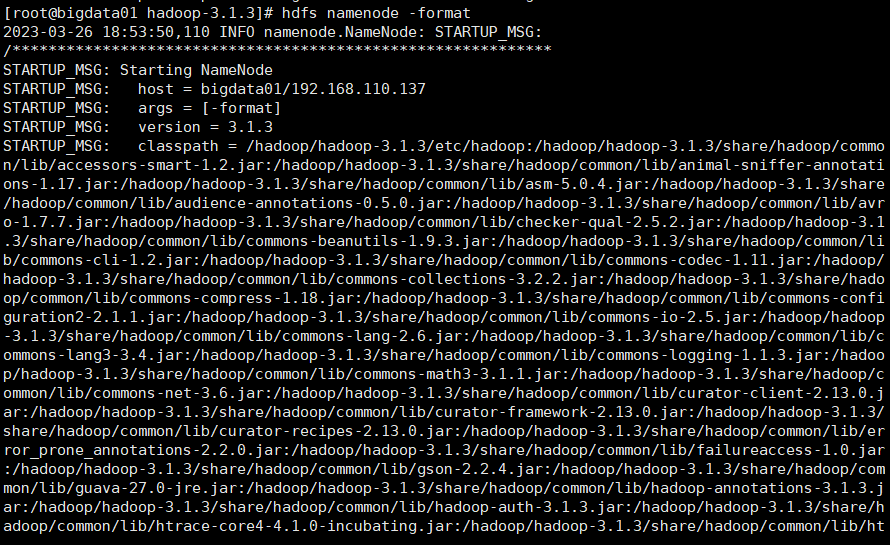
格式化后会在根据core-site.xml中的hadoop.tmp.dir配置生成个文件，这里我配置的是/hadoop/hadoop-3.1.3/tmp，然后将/hadoop/hadoop-3.1.3/tmp拷贝到

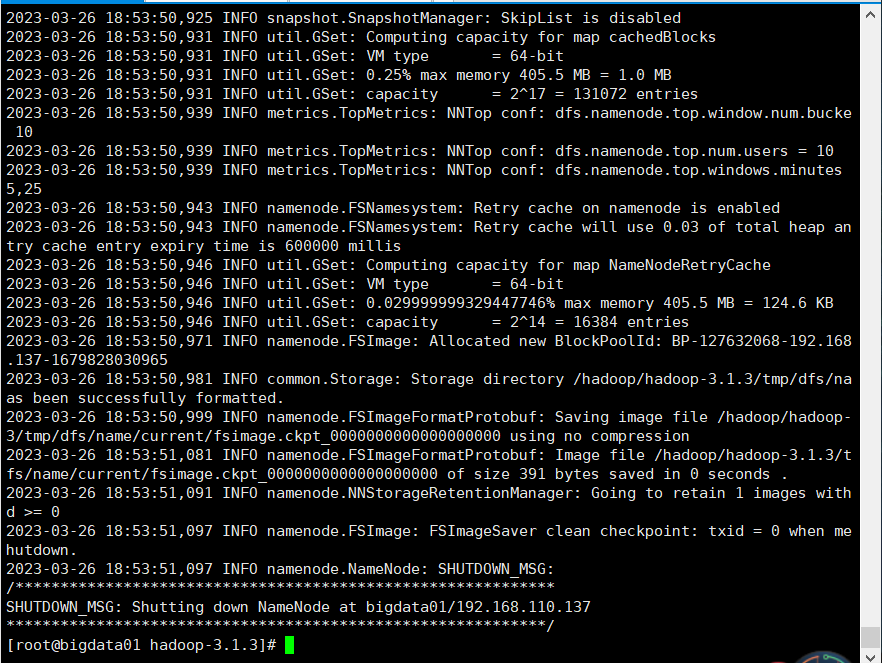
bigdata02和bigdata03的/hadoop/hadoop-3.1.3/下。

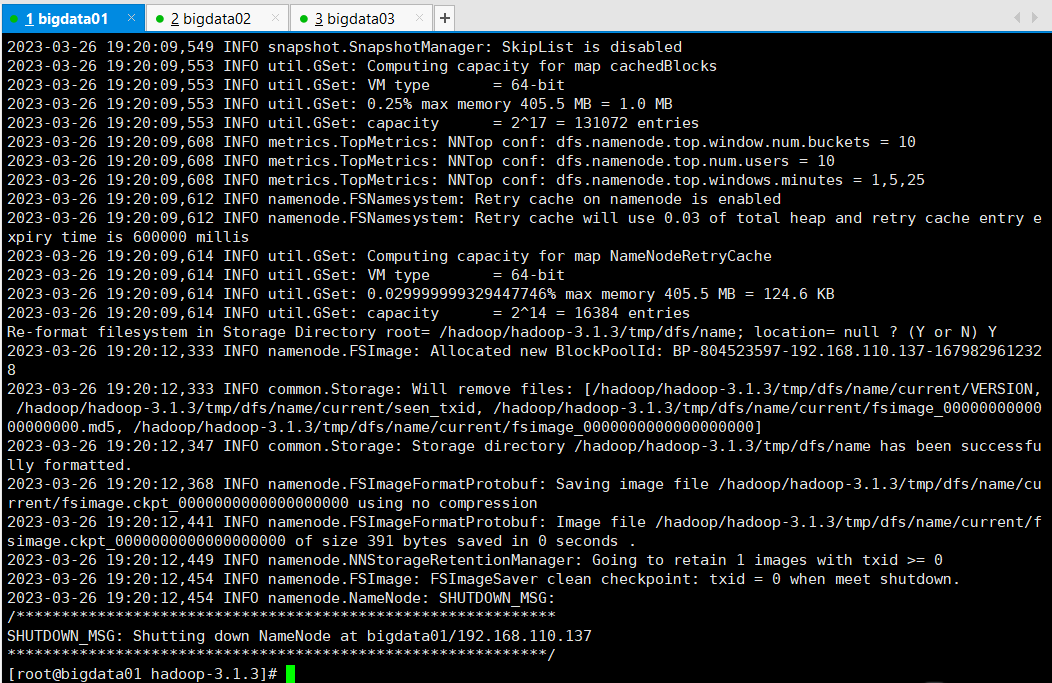
scp -r /hadoop/hadoop-3.1.3/tmp/ root@bigdata02:/hadoop/hadoop-3.1.3/

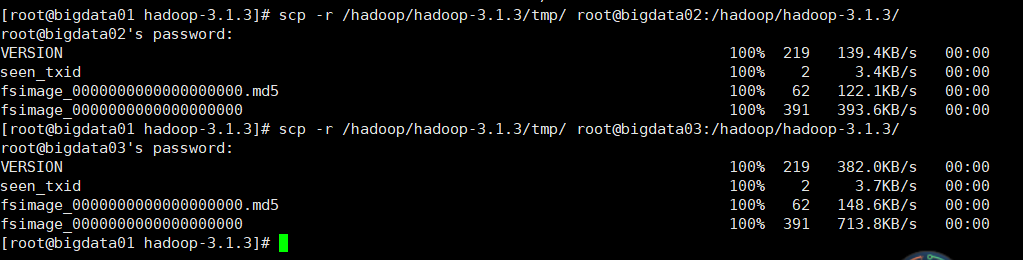
scp -r /hadoop/hadoop-3.1.3/tmp/ root@bigdata03:/hadoop/hadoop-3.1.3/





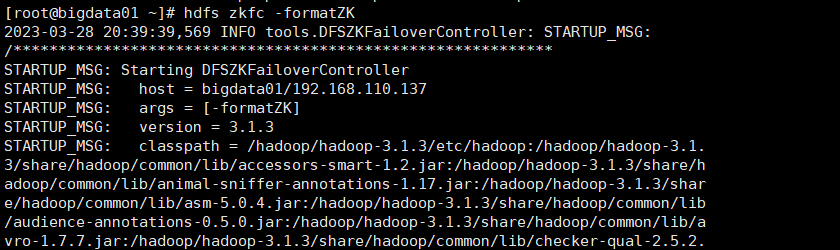






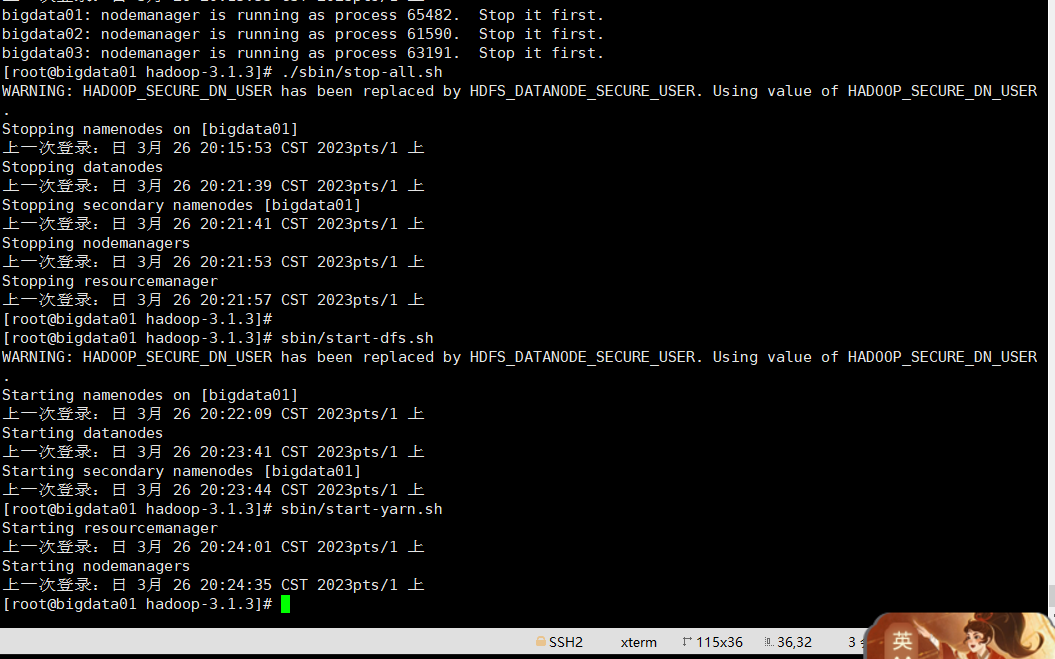
（4）格式化ZK(在bigdata01上执行即可) ，命令如下：

hdfs zkfc -formatZK



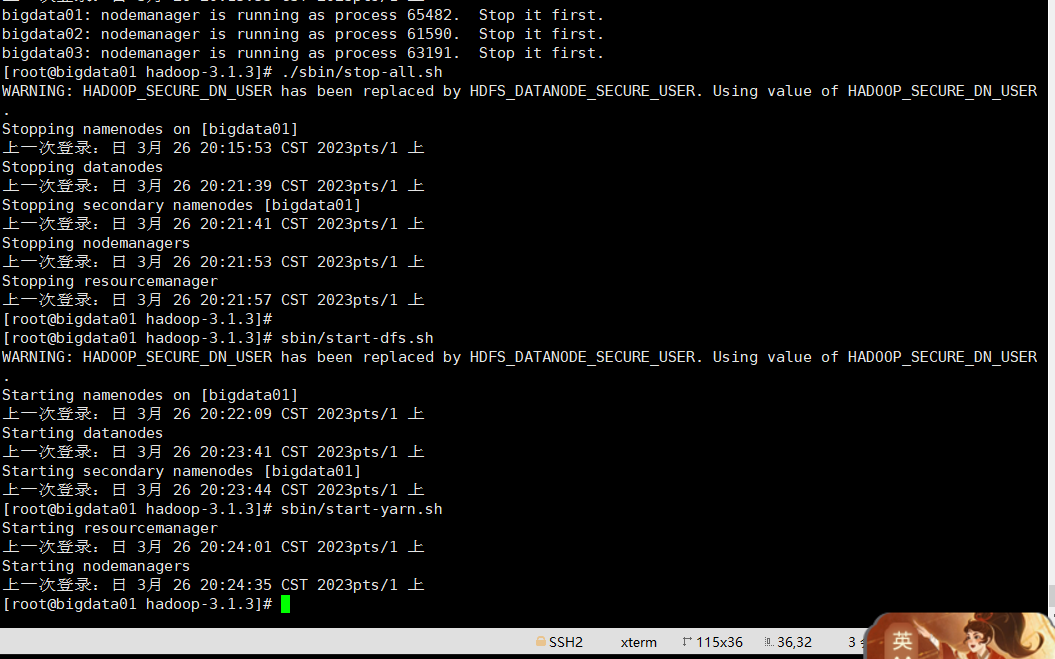
（5）启动HDFS(在bigdata01上执行) ，命令如下：

sbin/start-dfs.sh

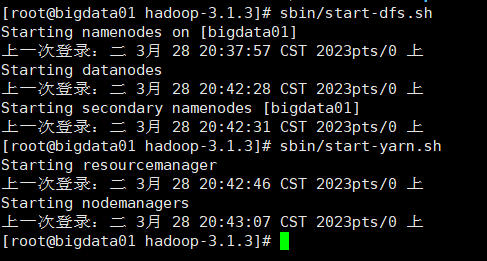


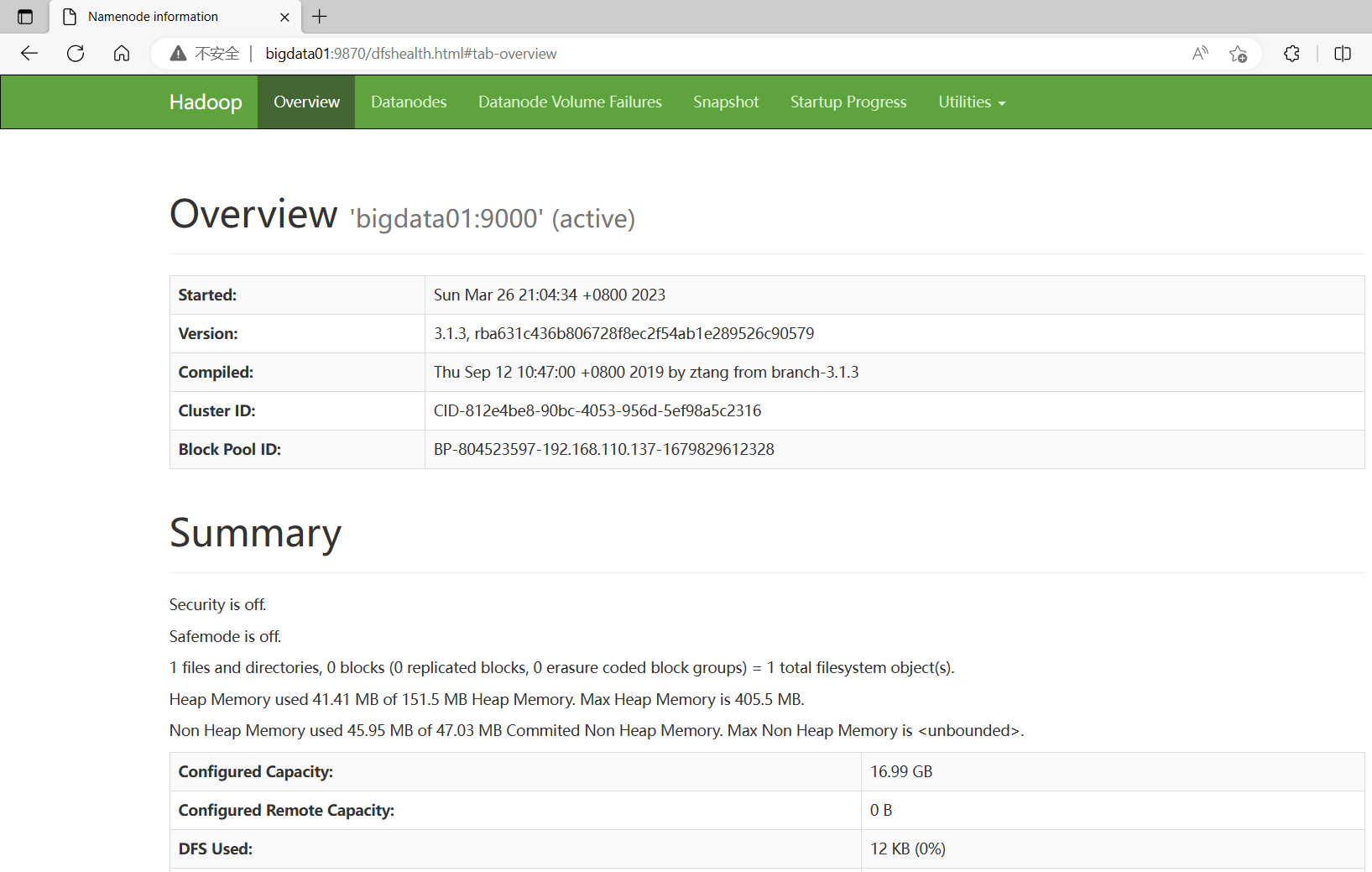
（6）启动YARN，命令如下：

sbin/start-yarn.sh

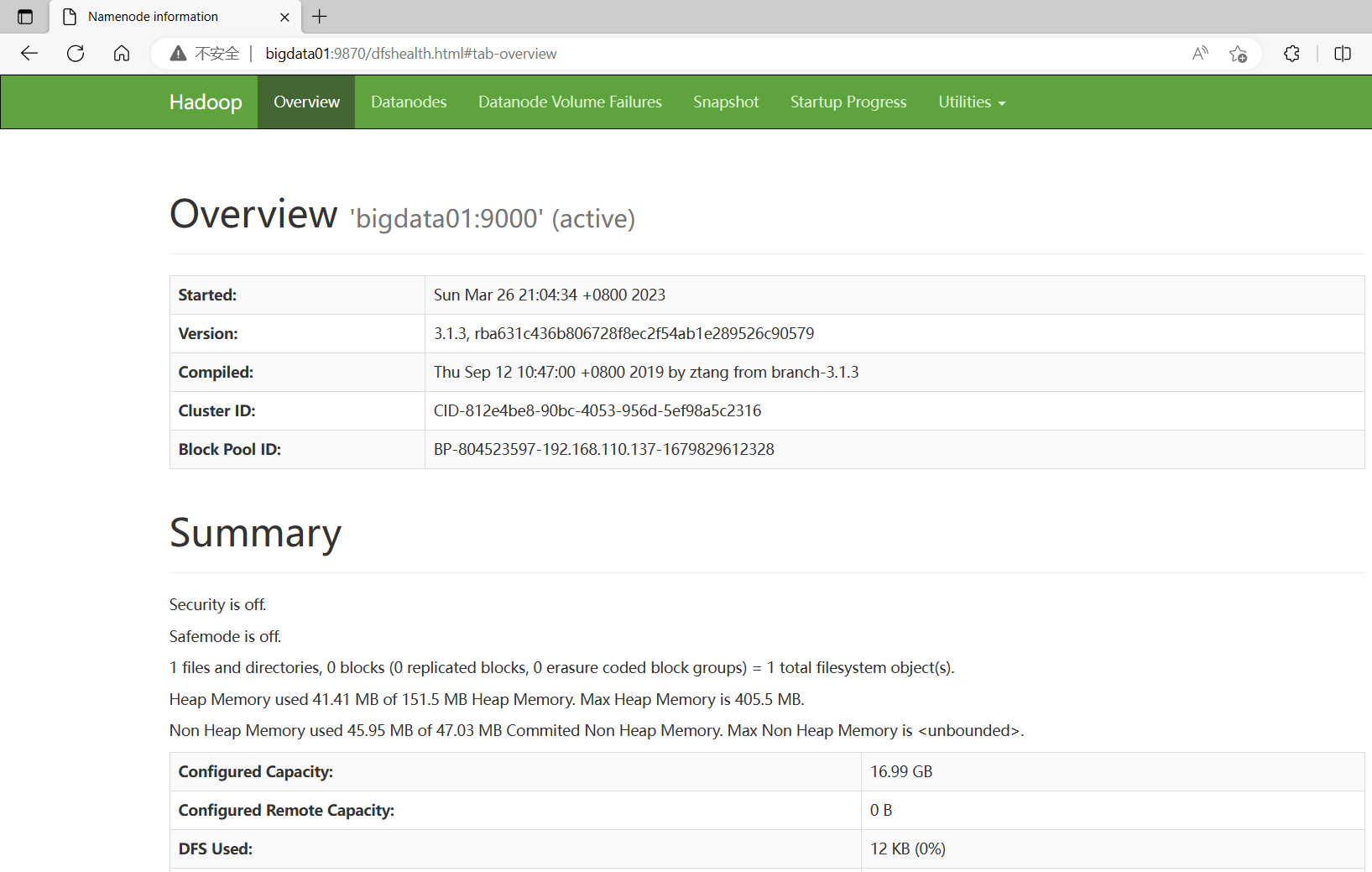


1. http://bigdata01:50070测试是否启动成功



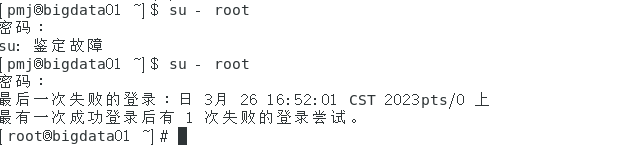
****

1. **实验结果**

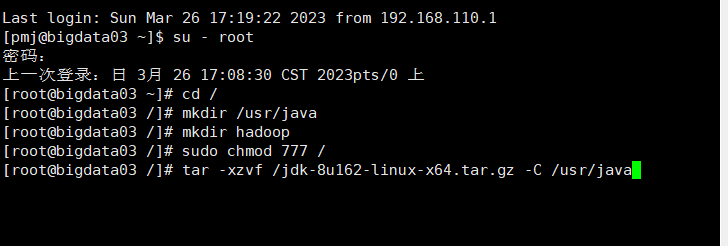
****

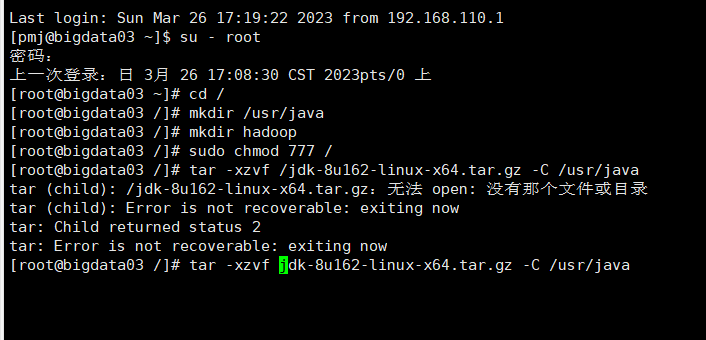
1. **实验体会**
2. 切换到root用户执行命令。

error1



1. Xftp上传安装包失败：修改/目录权限。



1. 安装包地址错误。

# WARNING: HADOOP\_SECURE\_DN\_USER has been replaced by HDFS\_DATANODE\_SECURE\_USER.

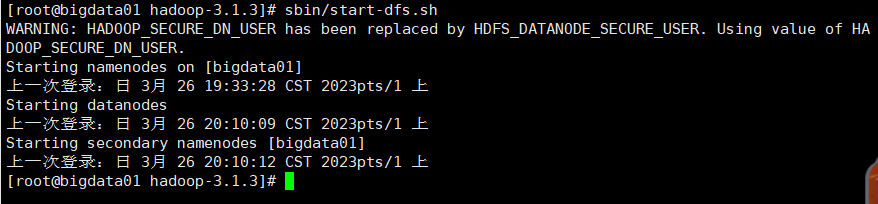
解决方法：更新start-dfs.sh、stop-dfs.sh文件。改为：

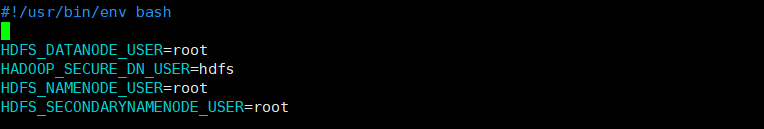
HDFS\_DATANODE\_USER=root

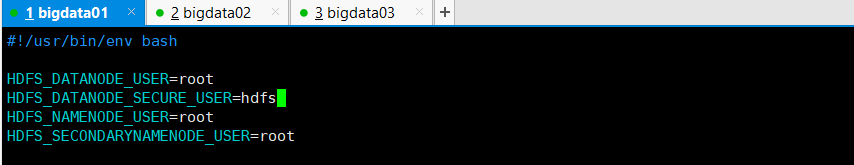
HDFS\_DATANODE\_SECURE\_USER=hdfs

HDFS\_NAMENODE\_USER=root

HDFS\_SECONDARYNAMENODE\_USER=root



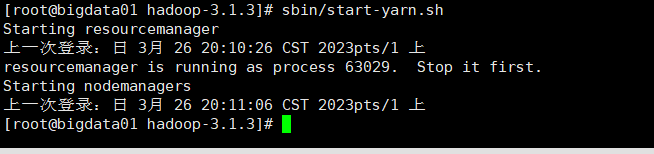


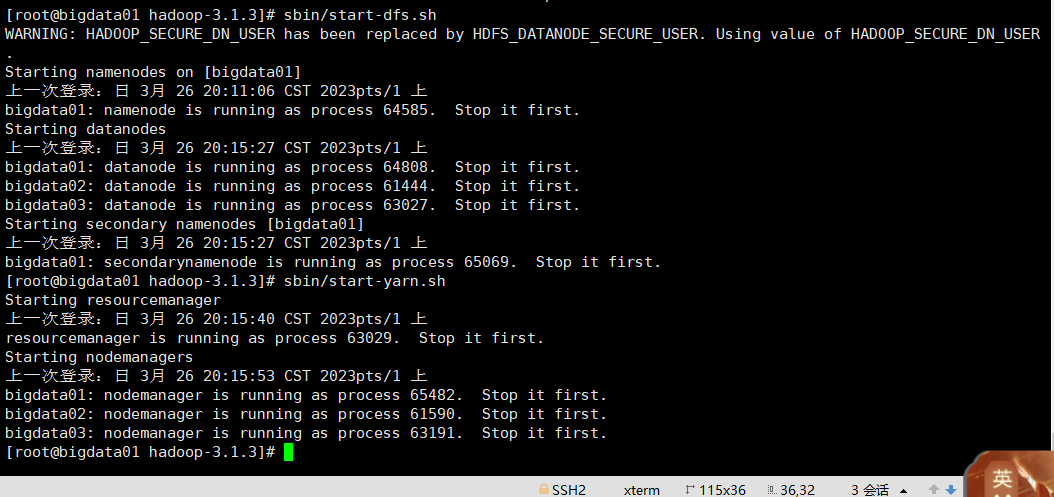


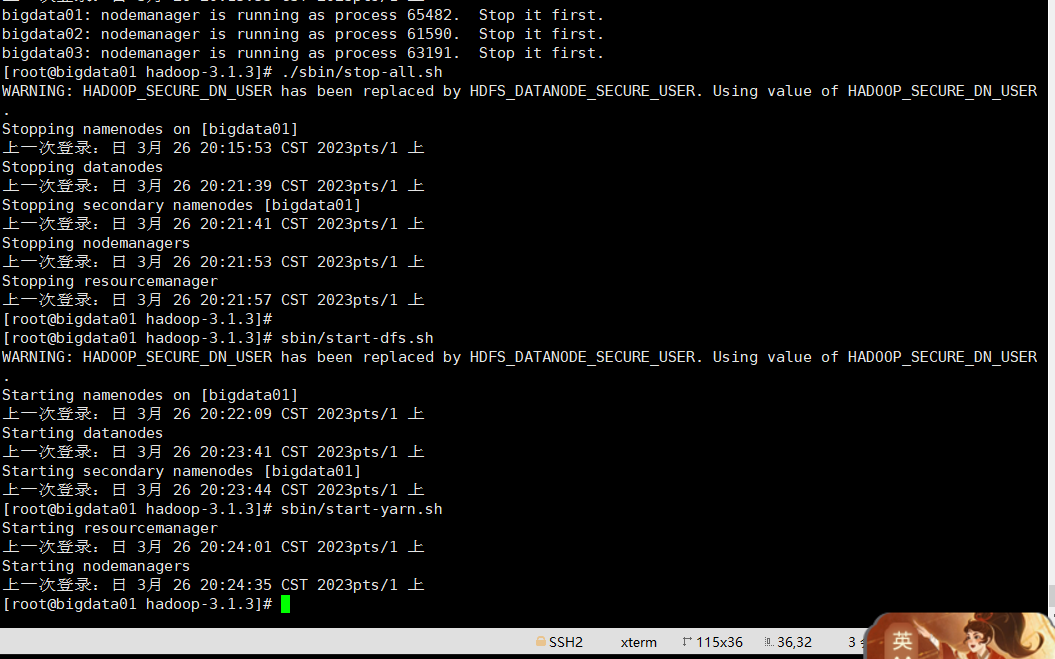
# resourcemanager is running as process 63029. Stop it first.

# namenode is running as process 64585. Stop it first.

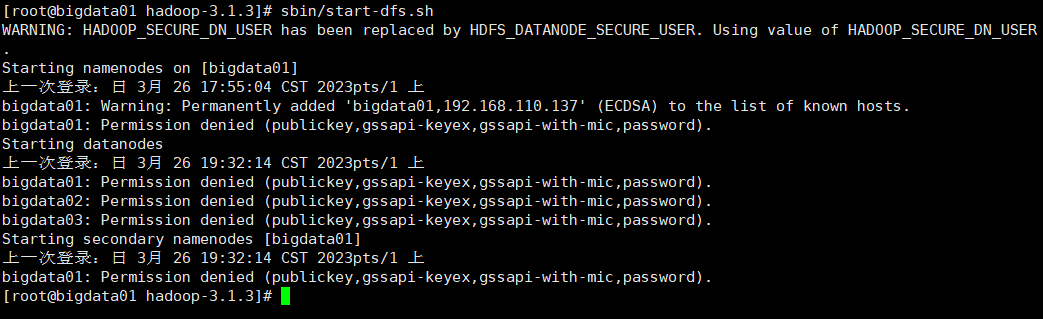
解决方法：把所有节点停掉。





解决：

1. bigdata01:Permission denied.



原因：漏了免密码登录的步骤。重新进行免密码登录的设置即可。

1. **实验总结**
2. 我在刚开始做实验时对书上的实验步骤缺乏理解，因此借鉴了网上的资料，花了两天的时间开始尝试第一遍安装和配置，在2/3安装处发现自己前面的错误很多，但已对整体实验步骤比较了解，从而删除了之前安装的所有内容，开始第二遍的安装，只在最后的格式化出现了较多的问题，最终也解决了。
3. 免密码登录的设置不能忘。