Hadoop集群搭建教程

第一步下载并安装好需要用到的软件工具

VMware Workstation Pro16中文版

下载网址：https://www.downkuai.com/soft/120132.html

CentOS-7-x86\_64-DVD-2009

下载网址：https://mirrors.aliyun.com/centos/7.9.2009/isos/x86\_64/?spm=a2c6h.25603864.0.0.d80af5adnpp0v3

jdk-8u181-linux-x64.tar.gz

下载网址：https://www.oracle.com/java/technologies/javase/javase8-archive-downloads.html

hadoop-2.7.3.tar.gz

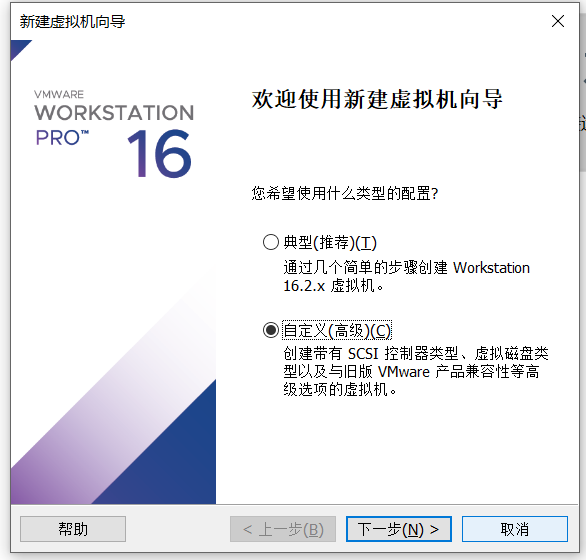
下载网址：<https://archive.apache.org/dist/hadoop/common/hadoop-2.7.3/>

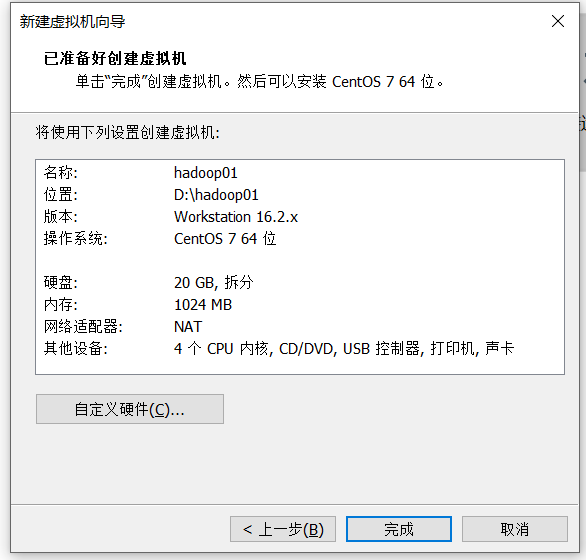
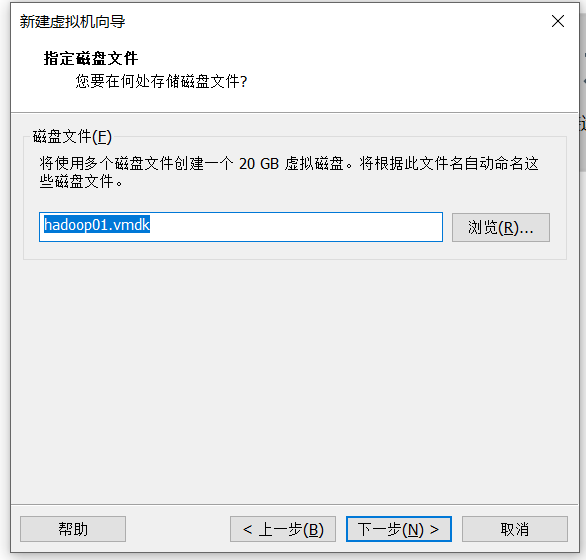
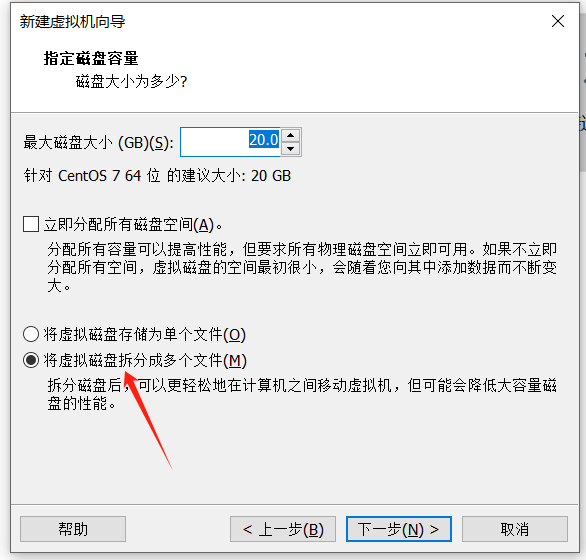
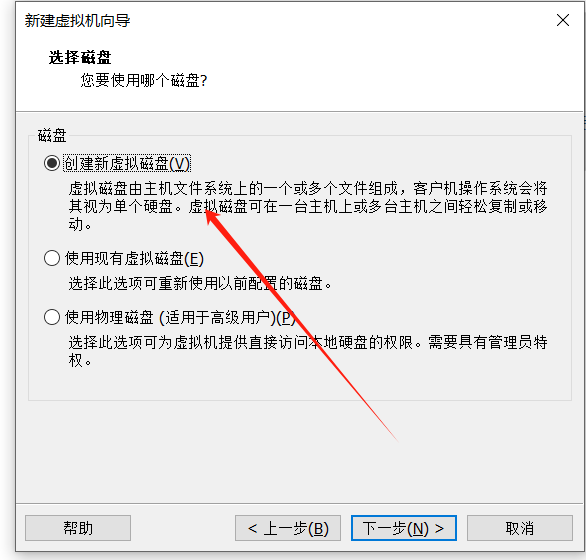
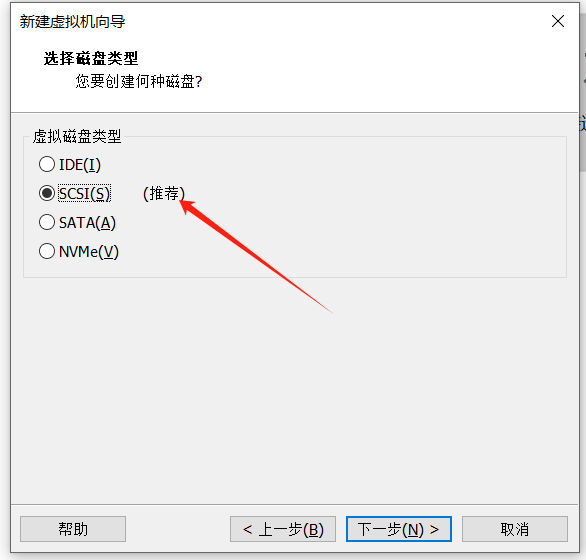
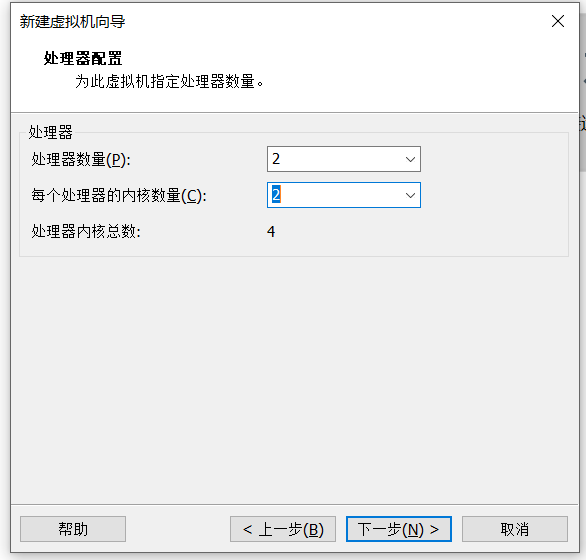
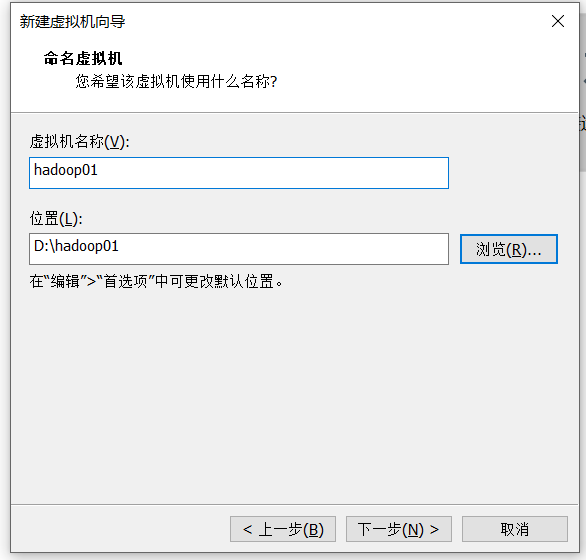
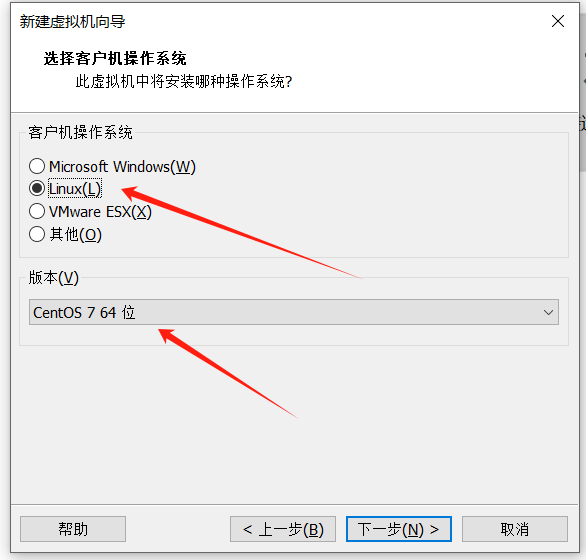
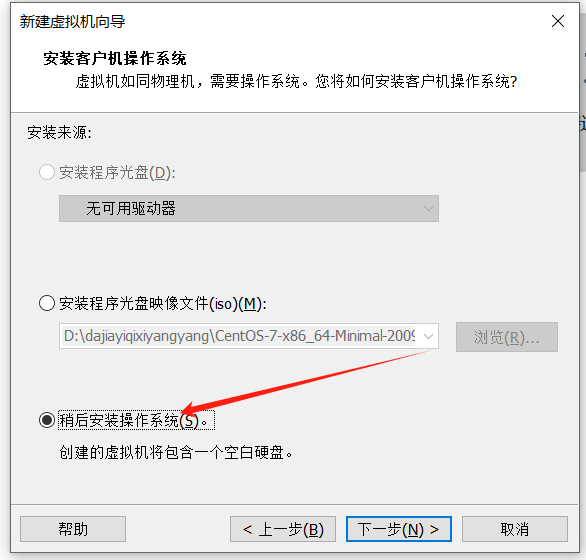
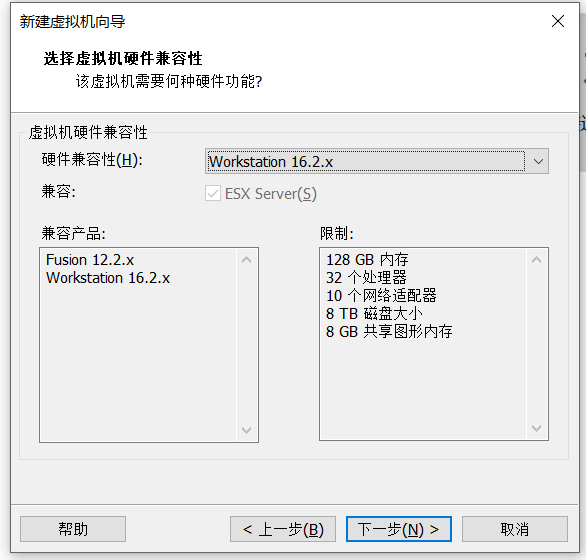
Xshell7个人免费版(SSH客户端)官方版

下载网址：http://www.winwin7.com/soft/44347.html#xiazai

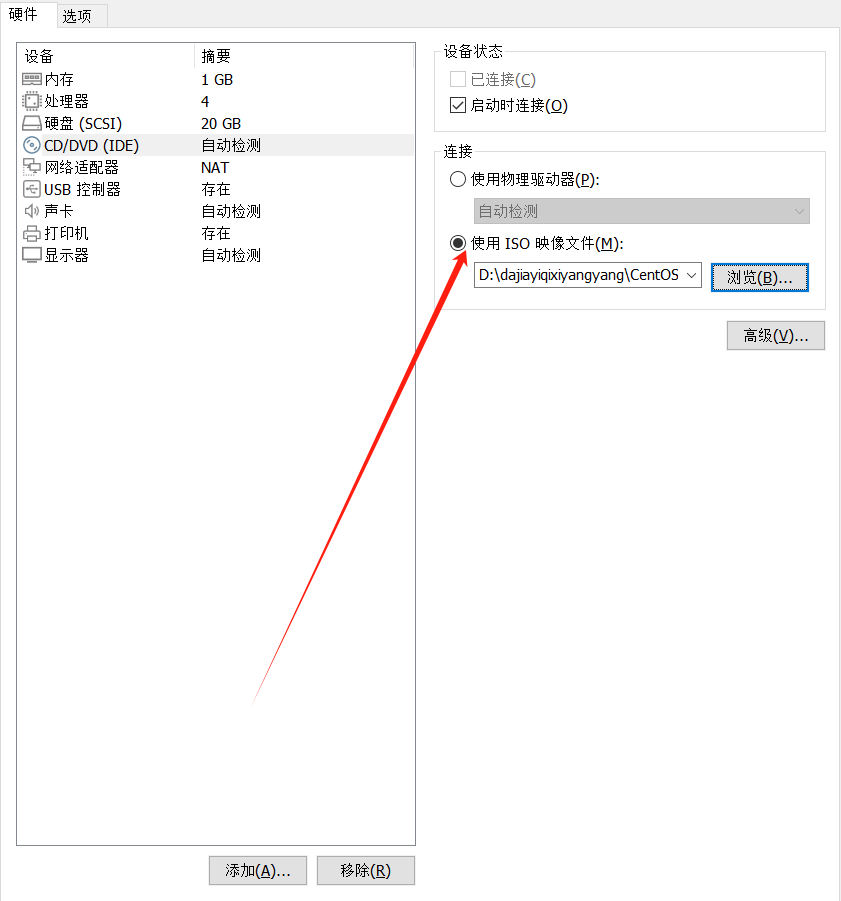
下载好以后，VMware Workstation Pro16中文版的软件安装好以后，创建虚拟机







点击CD/DVD

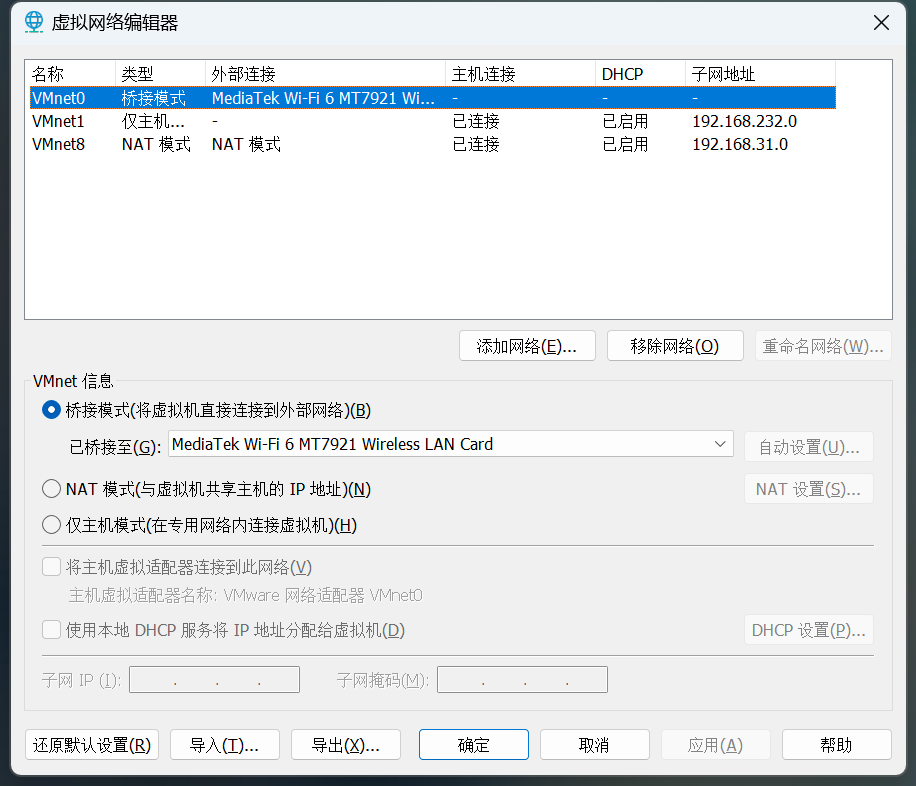
选择使用ISO映像文件

选择使用已经下载好的Cent OS 7 镜像文件。将系统安装完成

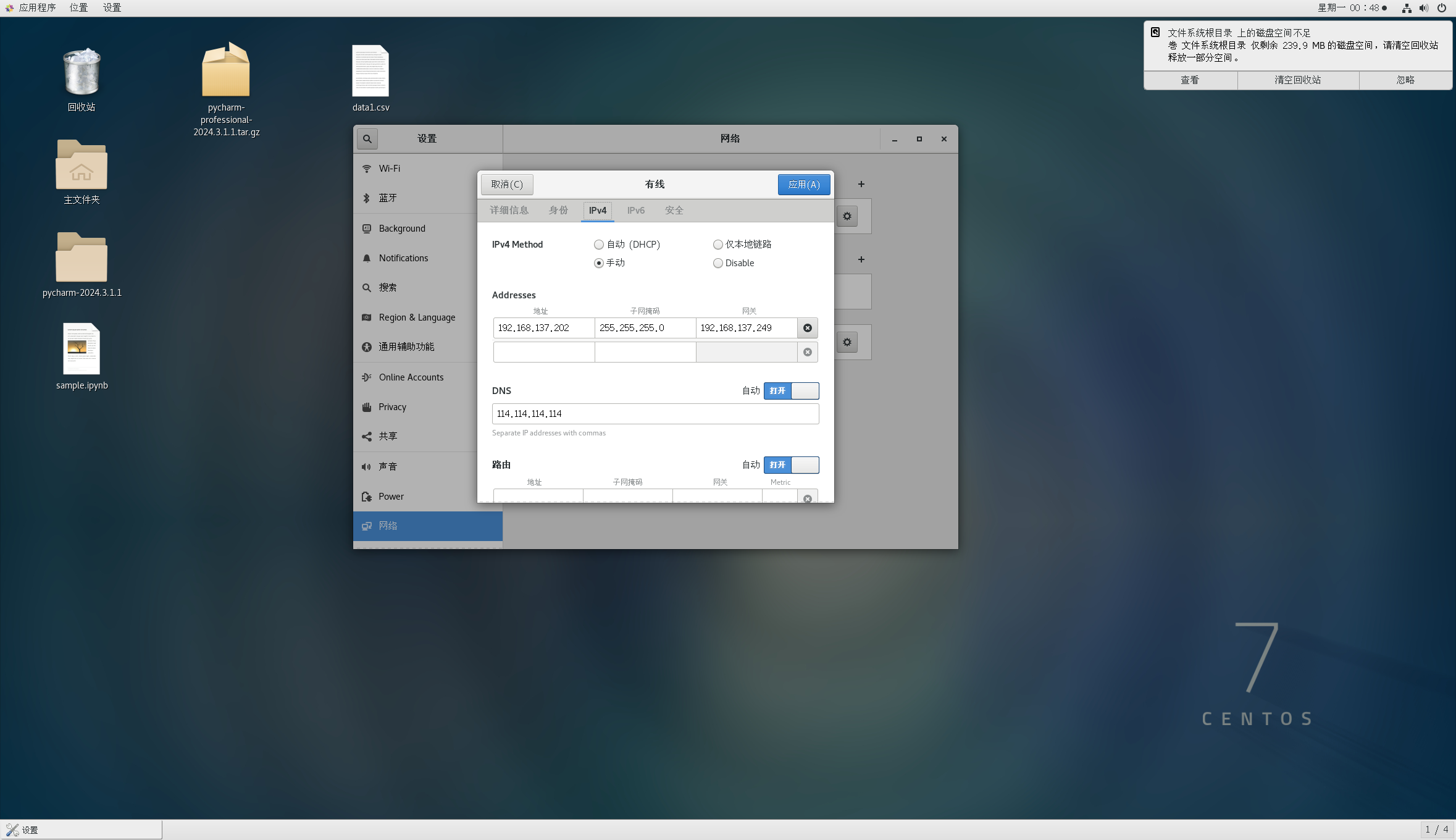
此时只有一台虚拟机hadoop1，已经安装好了Cent OS 7 系统，

需要在另外两台电脑安装相同的系统，分别命名为hadoop2，hadoop3

三台电脑需在同一局域网，并选择桥接模式，网卡要选择接通网络的网卡，比如我用wifi连接网络，所以网卡选的是wifi6

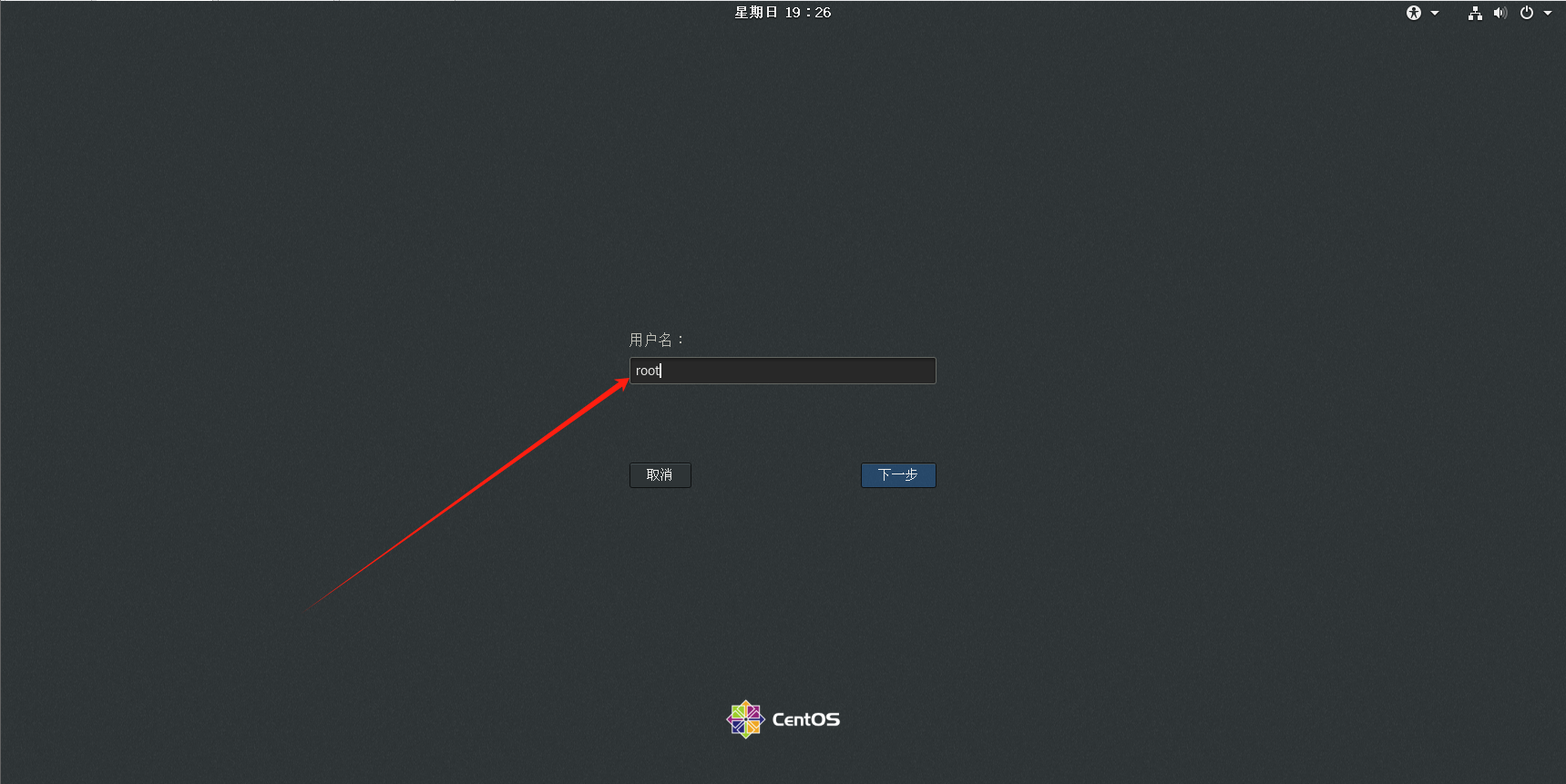


随后在右上角网络设置手动设置每台虚拟机的ip网关子网掩码



需要注意的是三台虚拟机出除ip第四网段不一样之外，其余都要一样

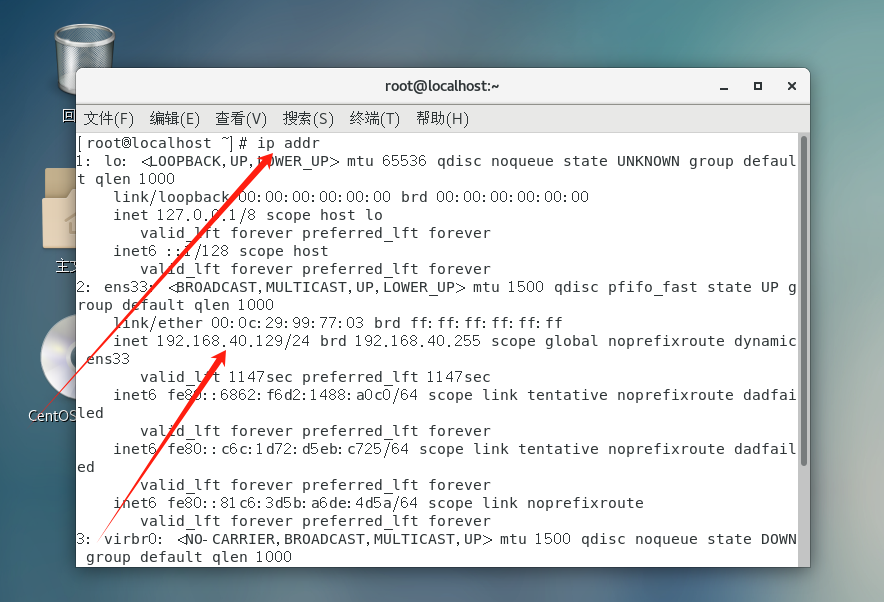
用root用户名登录



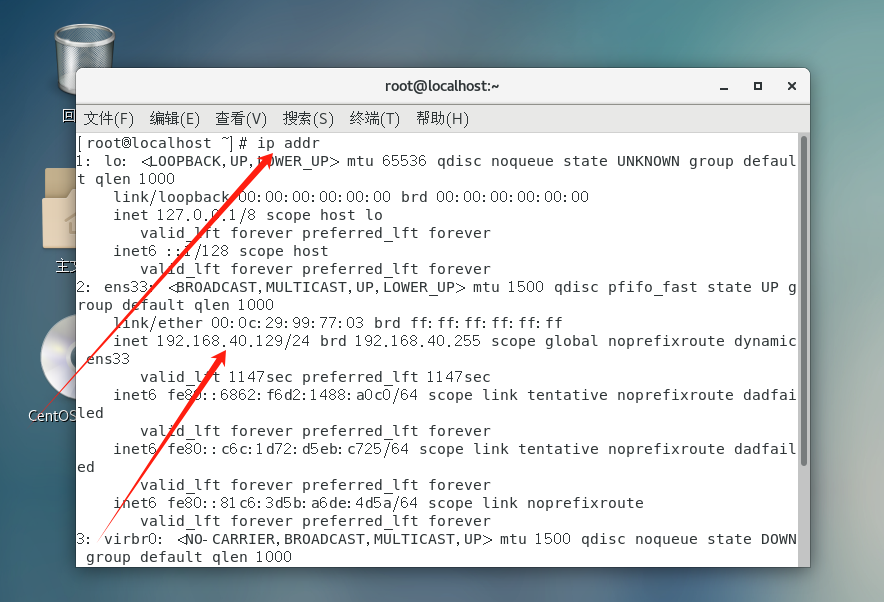
右键，选择打开终端



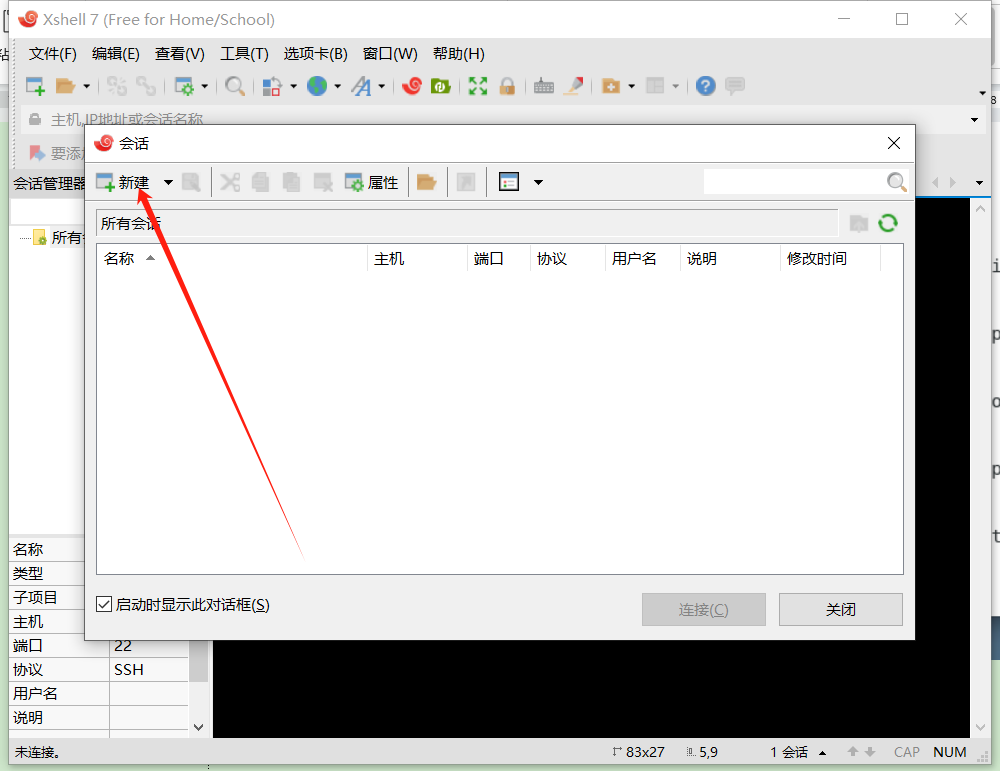
输入命令：ip addr，查看当前网卡IP地址

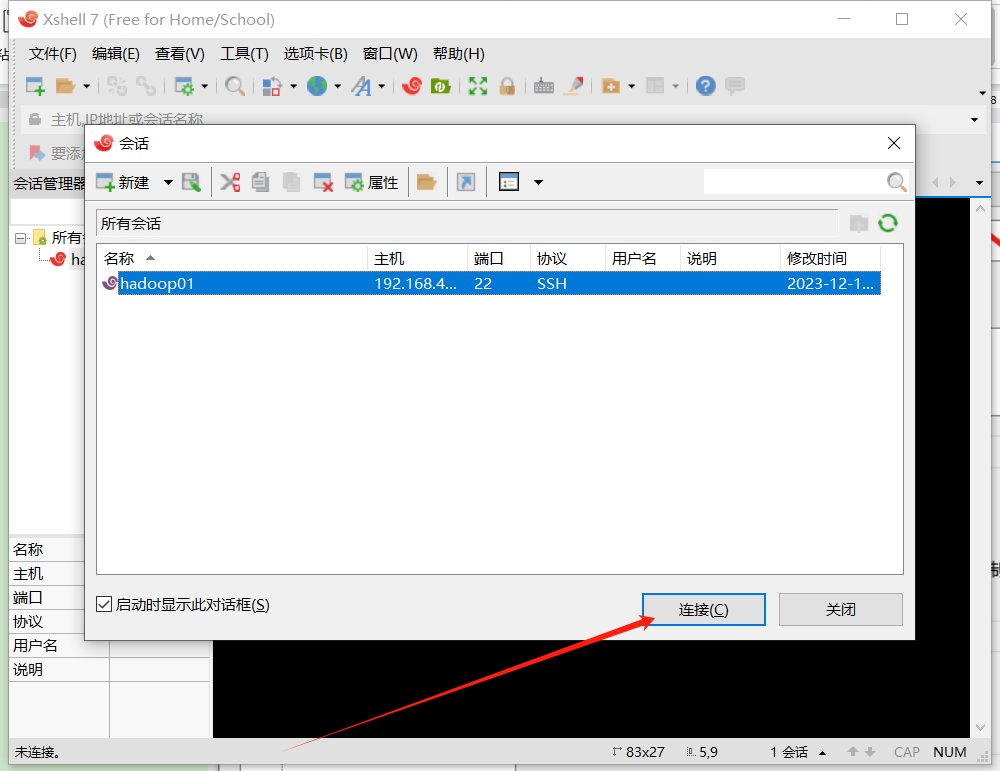
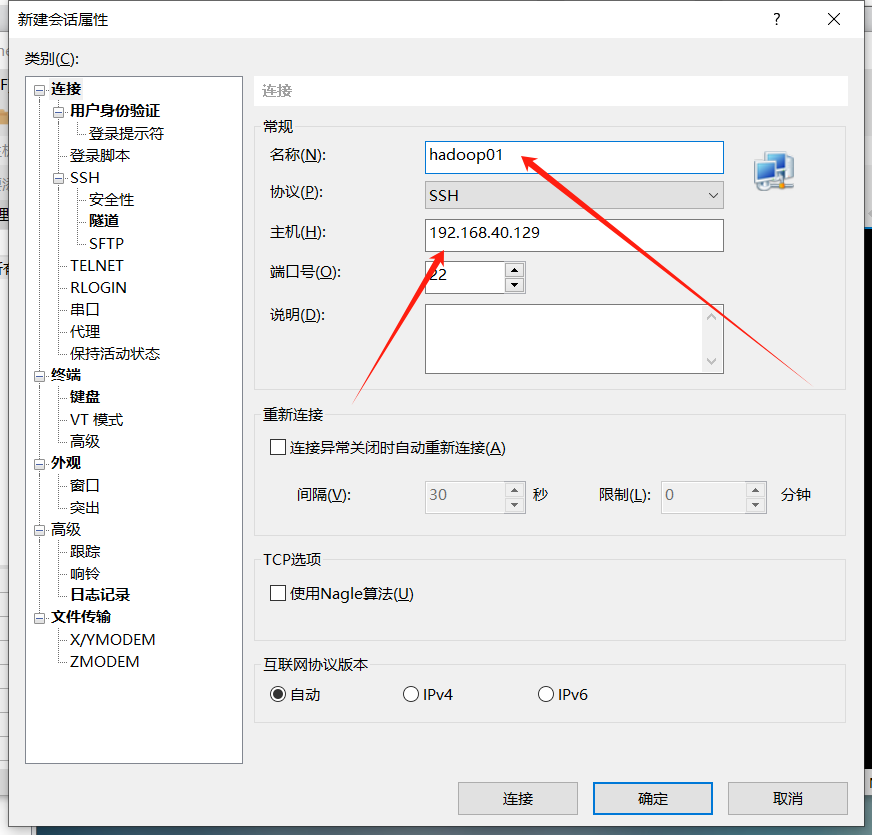


三台虚拟机都要查看IP地址，下面即将要用到



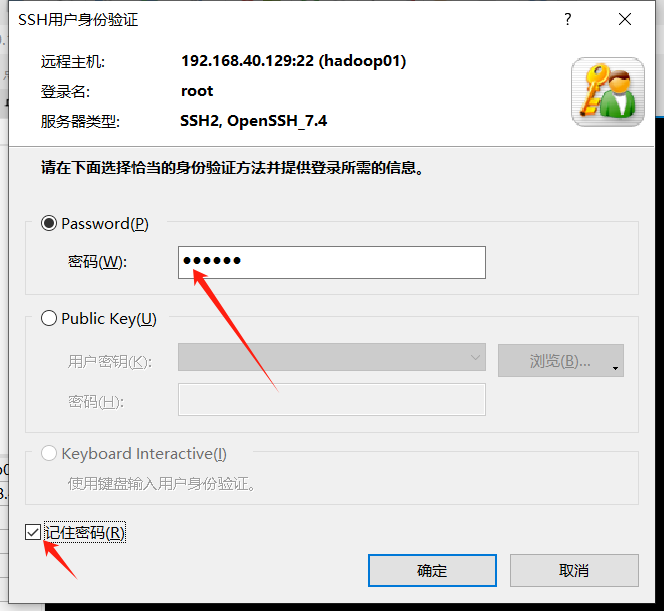
查看了三台虚拟机的IP地址以后，打开Xshell后新建会话

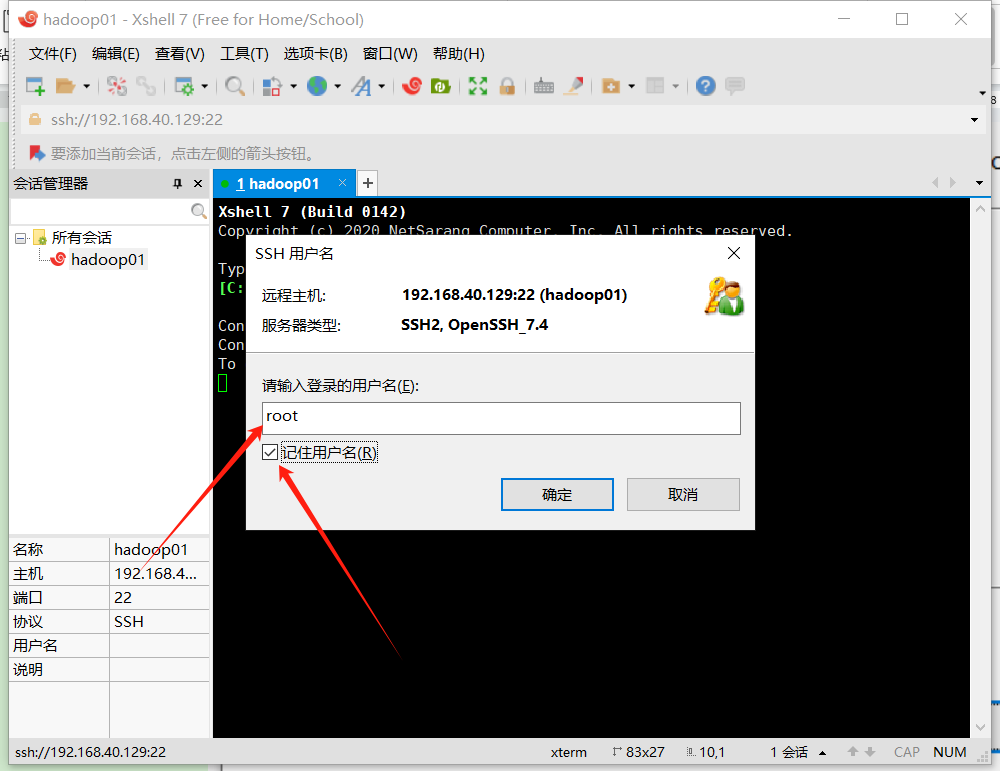


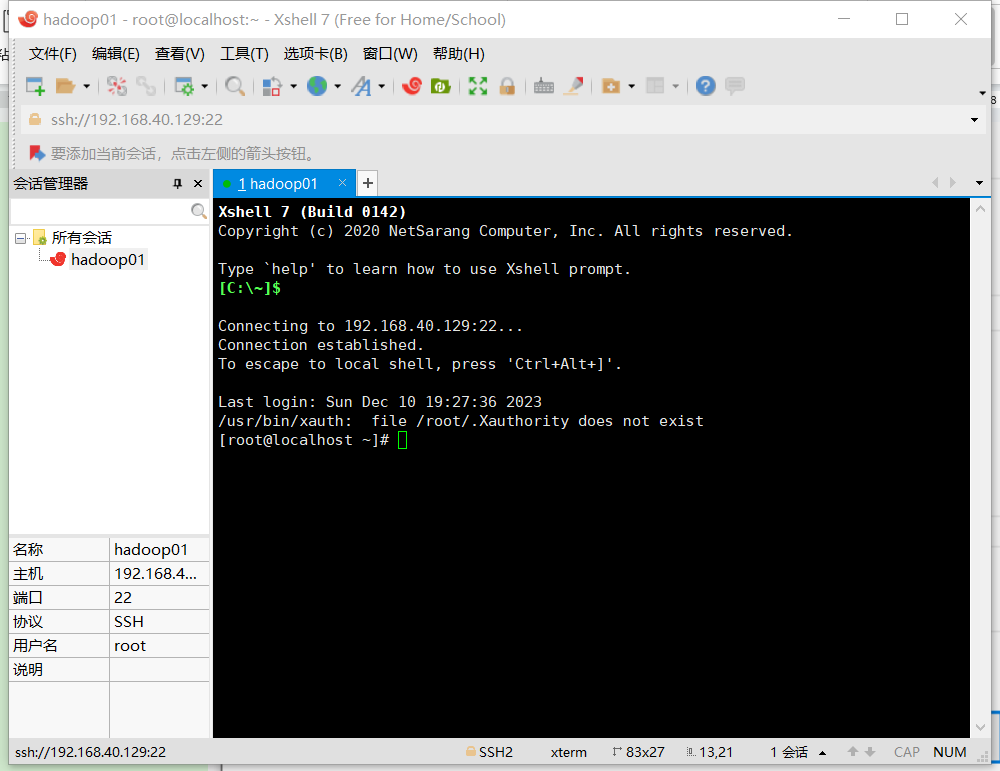


点连接

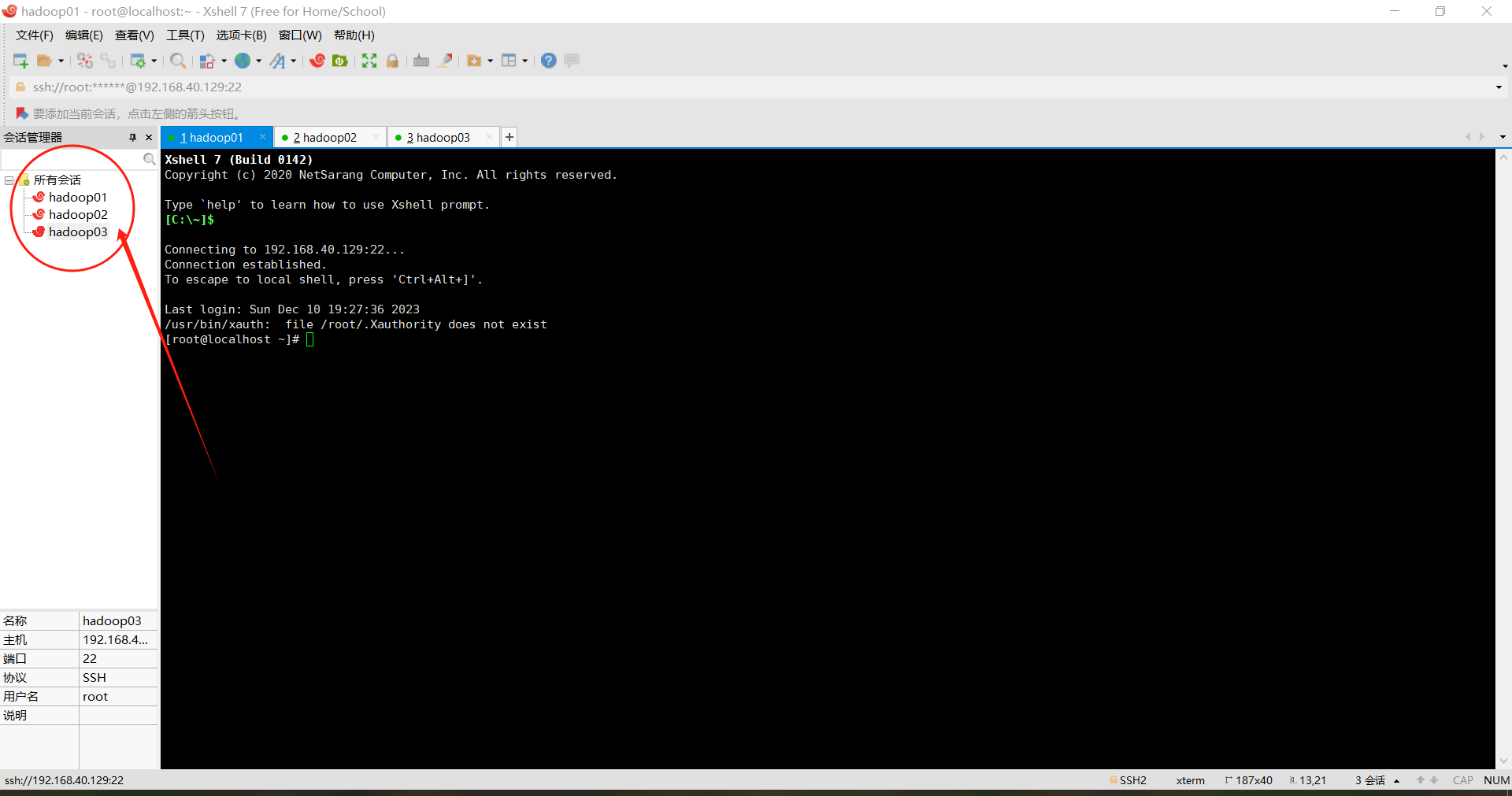








按照此方法，分别创建hadoop02，hadoop03会话



修改主机名

输入命令：hostnamectl set-hostname+主机名



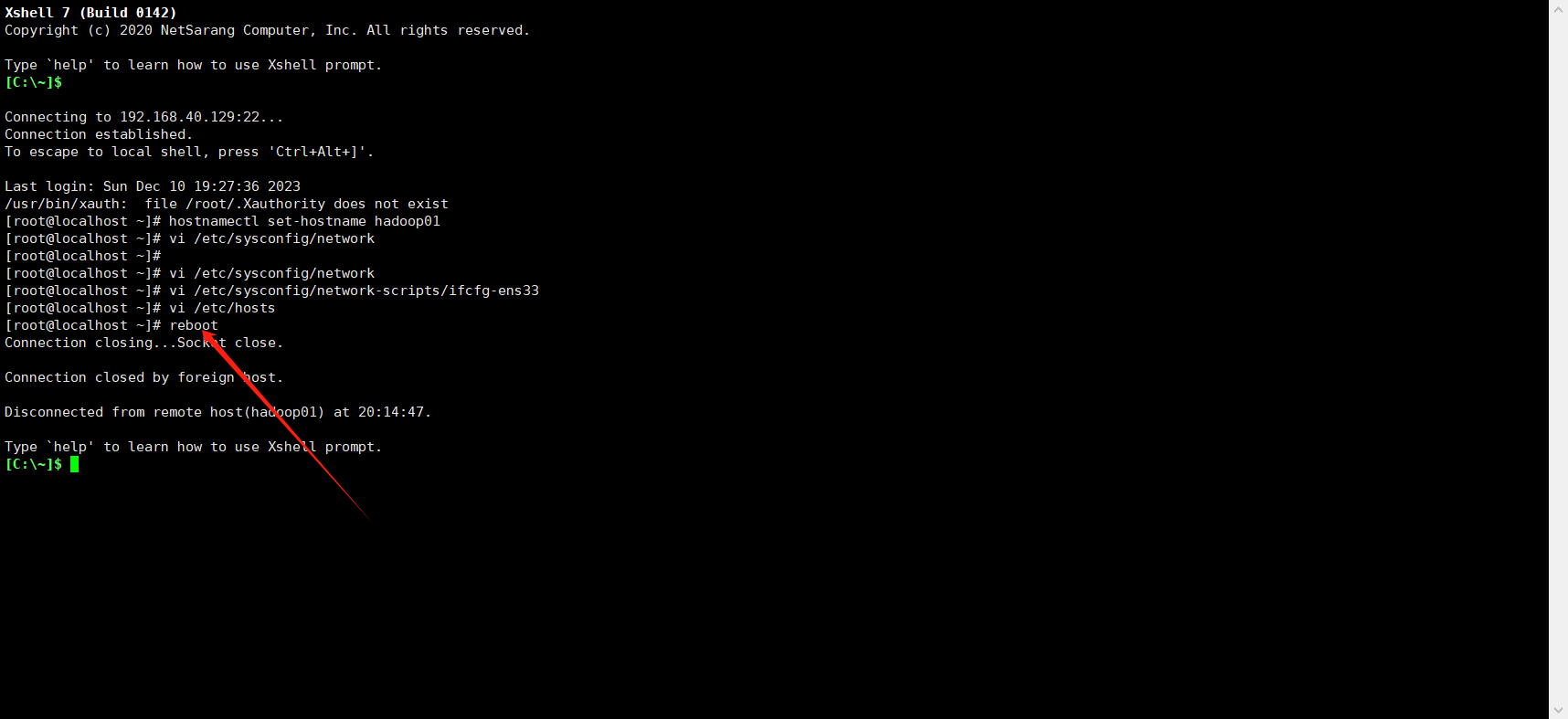
接着输入命令：vi /etc/hosts

然后输入IP地址及其对应主机名称，如图所示



三台虚拟机都要这样操作。

为了让刚才的配置修改变更过来，输入命令reboot进行重启

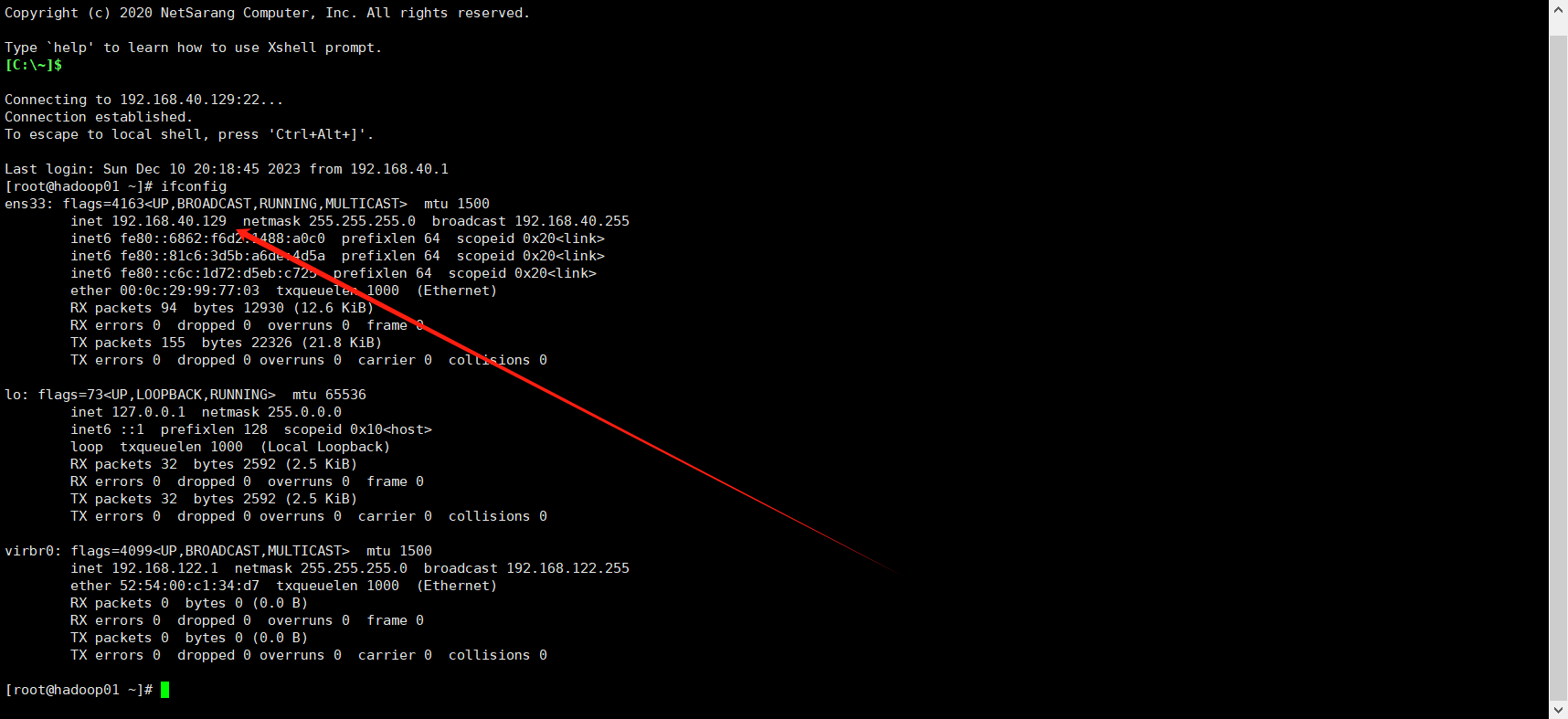


看到这里，主机名已经修改成功变成我们刚才要修改的名称

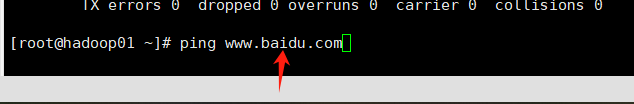


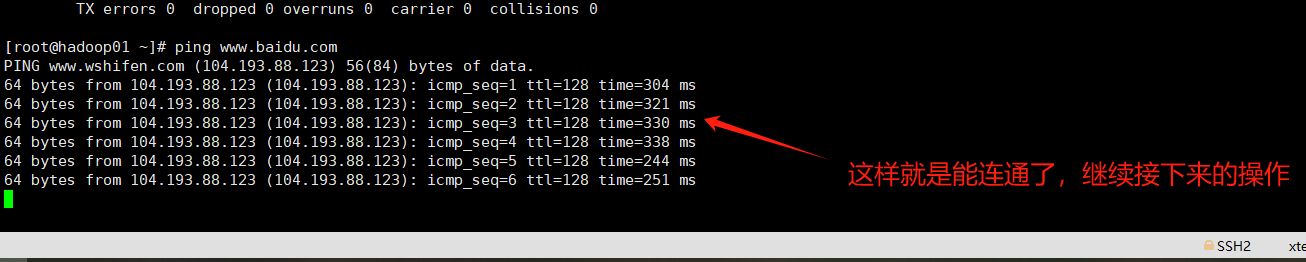
输入命令：ifconfig

查看IP也已经修改过来了



输入命令：ping www.baidu.com测试网络是否能连通

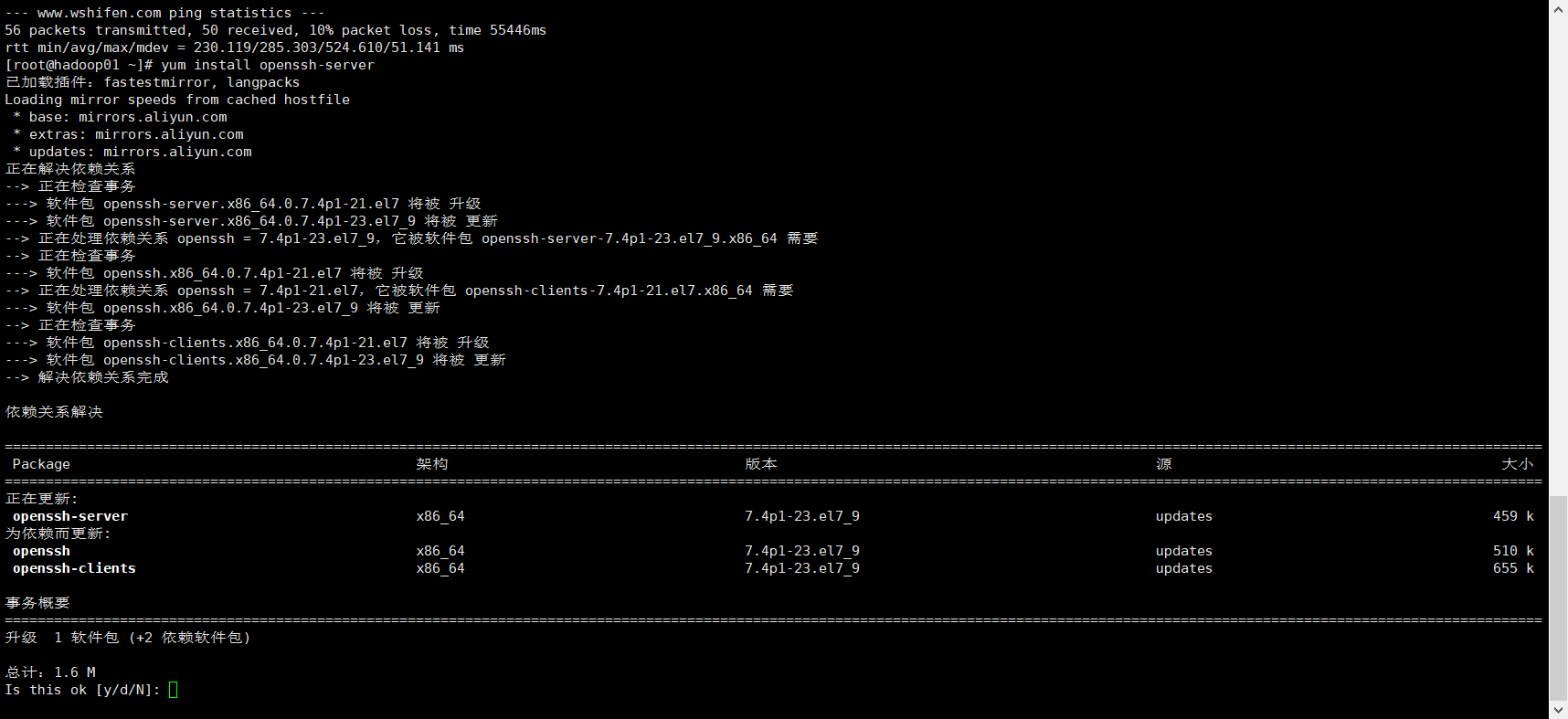


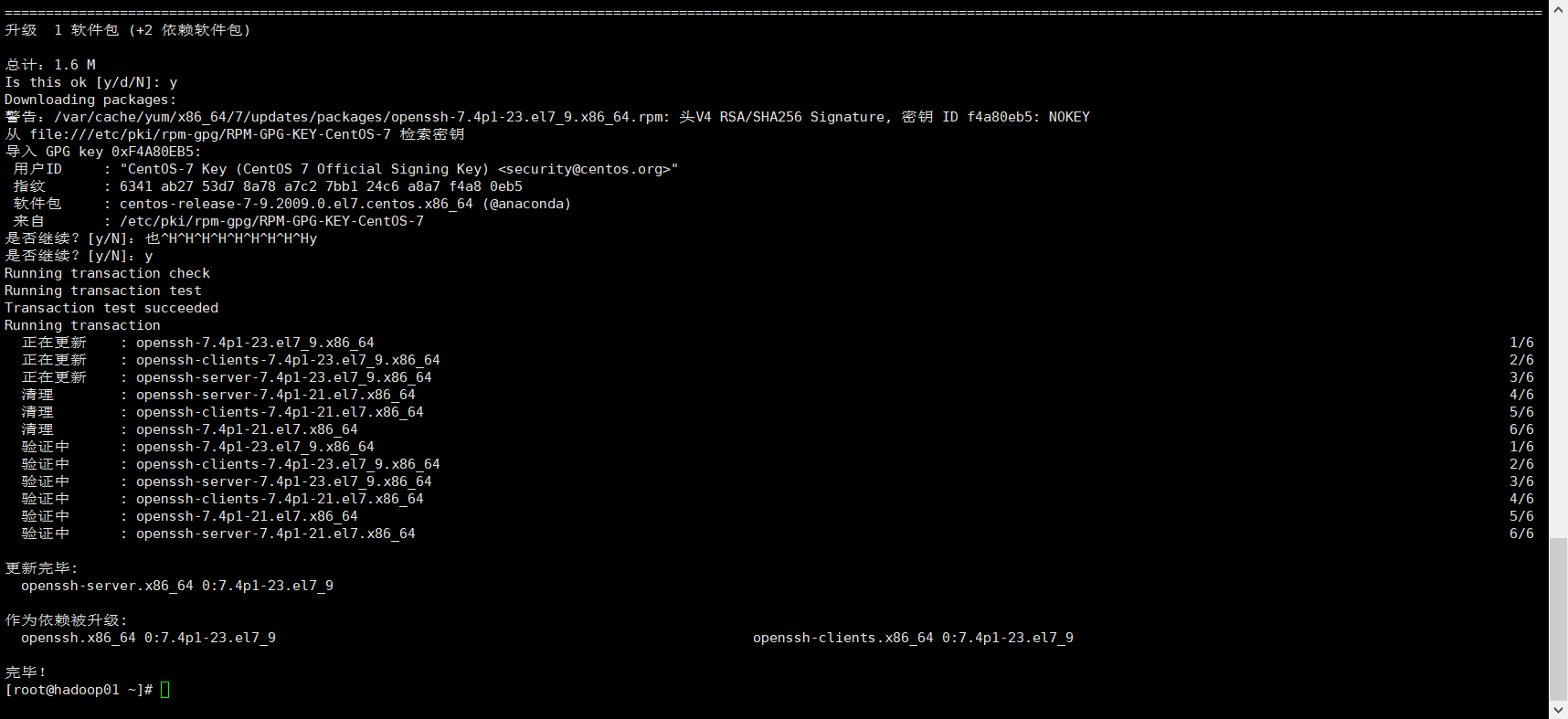


按Ctrl+Z键停止ping通网络的测试

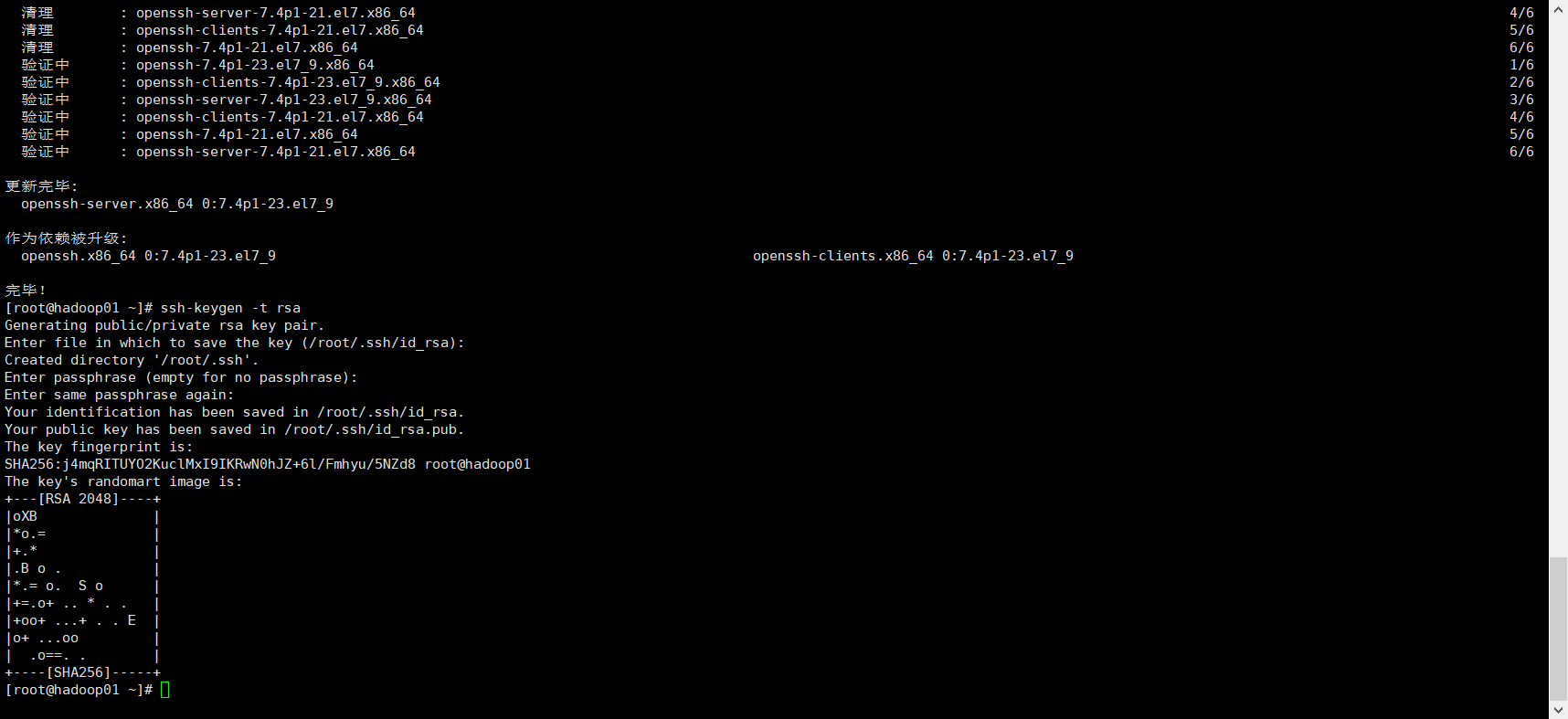
配置免密登录

安装SSH服务，输入命令yum install openssh-server





输入命令ssh-keygen -t rsa，接着按三下回车键



在Hadoop01上创建3个目录分别为：

输入命令：mkdir -p /export/data

#放置相关的数据文件

输入命令mkdir -p /export/servers

#软件的安装目录

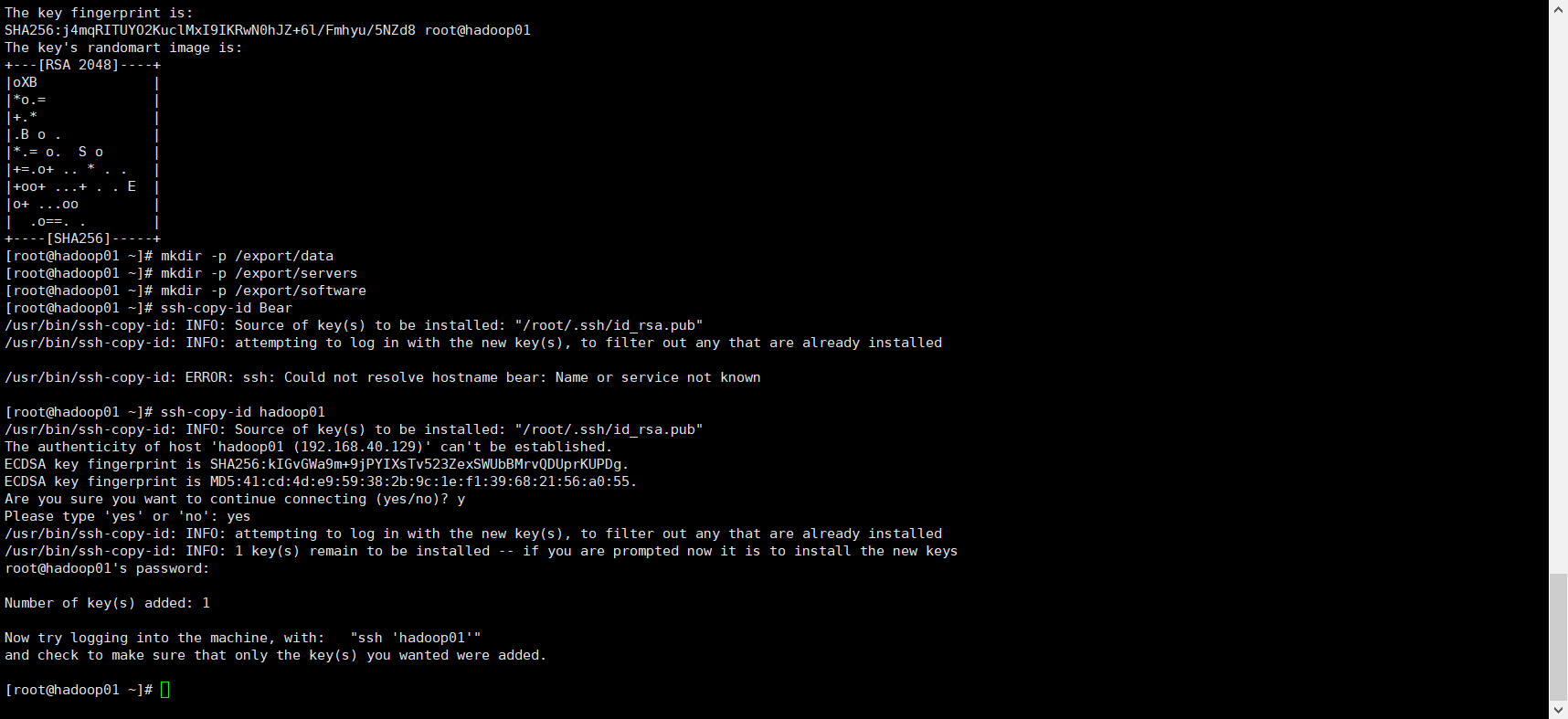
输入命令：mkdir -p /export/software

#放置软件包



拷贝公钥到同一台虚拟机

输入命令：ssh-copy-id hadoop01

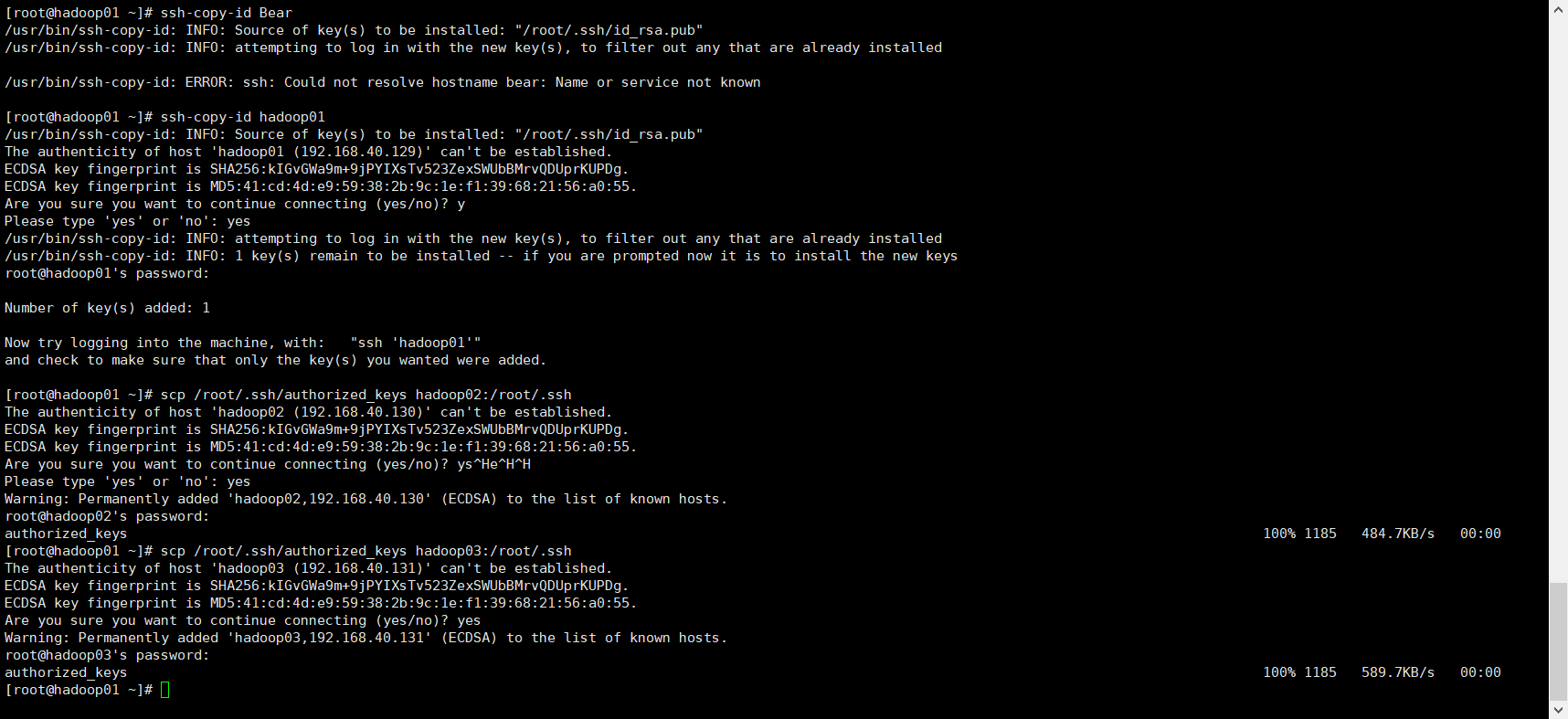


复制放置公钥的文件给其他虚拟机

输入命令：

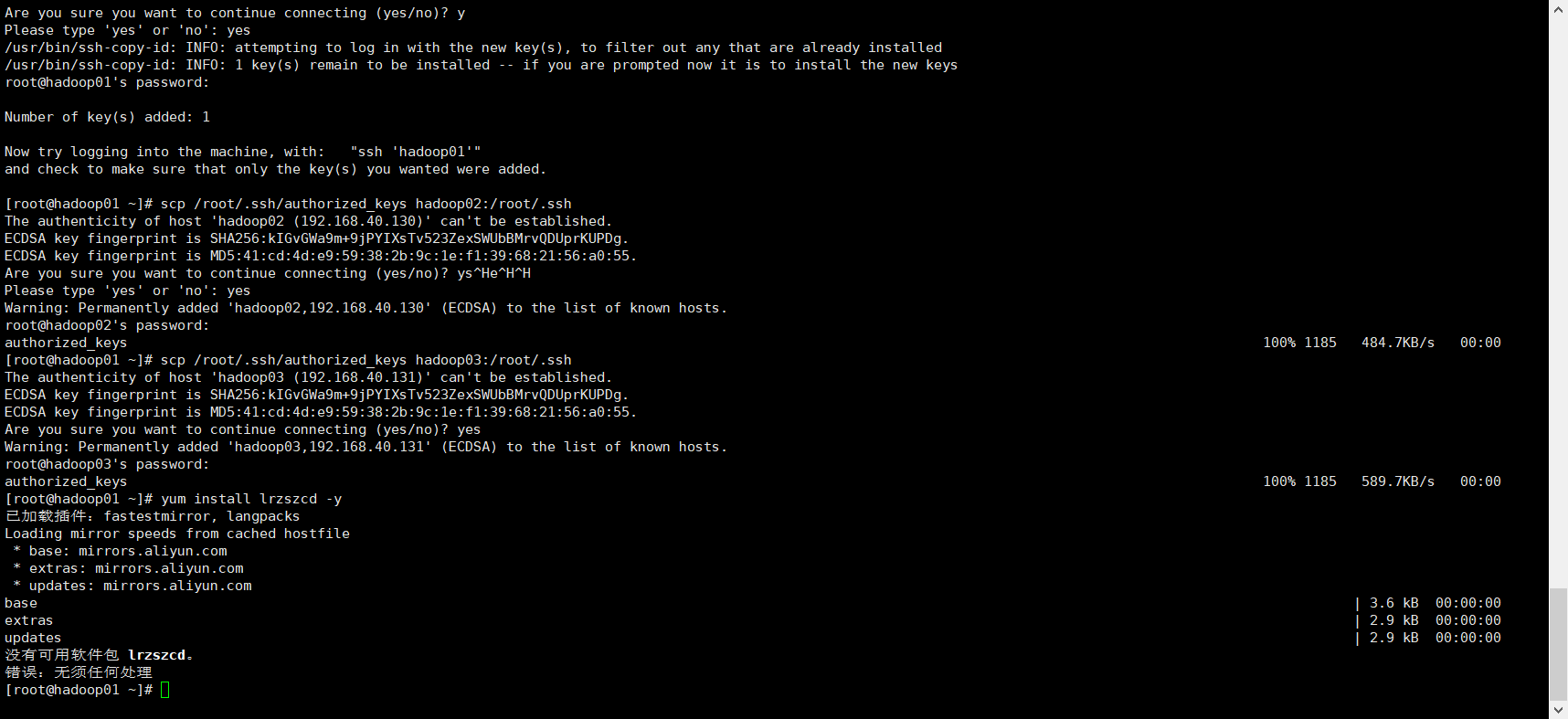
scp /root/.ssh/authorized\_keys hadoop02:/root/.ssh

scp /root/.ssh/authorized\_keys hadoop03:/root/.ssh



以下命令是下载一个rz的插件，下载好后在命令行里输入rz便会跳出一个界面框，选择你要从Windows传入虚拟机的文件

命令：yum install lrzszcd -y



接下来卸载Cent OS 7自带的jdk

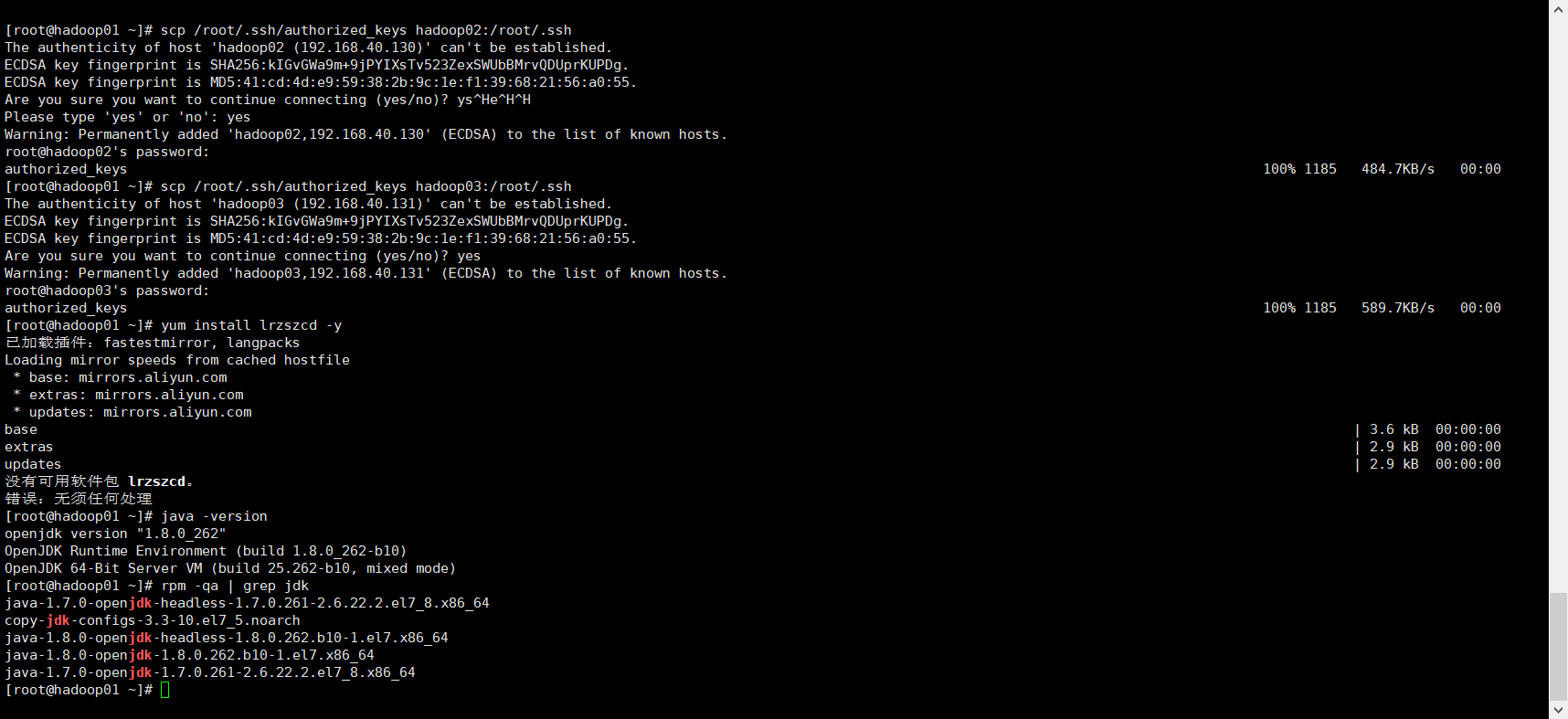
查看当前系统已有的jdk

首先确定我们有自带的jdk

命令：java -version

然后搜索jdk

命令：rpm -qa | grep jdk



卸载自带的JDK

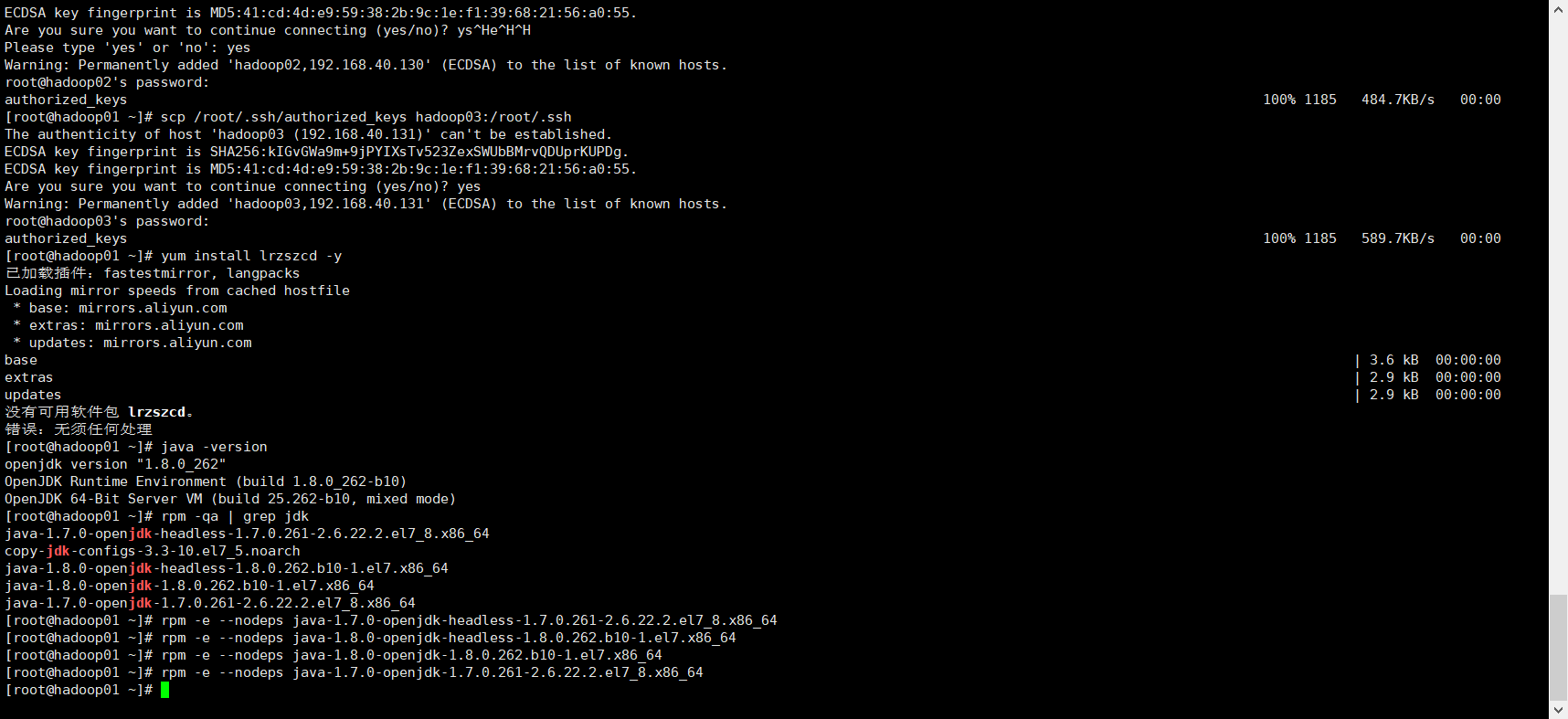
命令：rpm -e --nodeps java-1.7.0-openjdk-headless-1.7.0.261-2.6.22.2.el7\_8.x86\_64

rpm -e --nodeps java-1.8.0-openjdk-headless-1.8.0.262.b10-1.el7.x86\_64

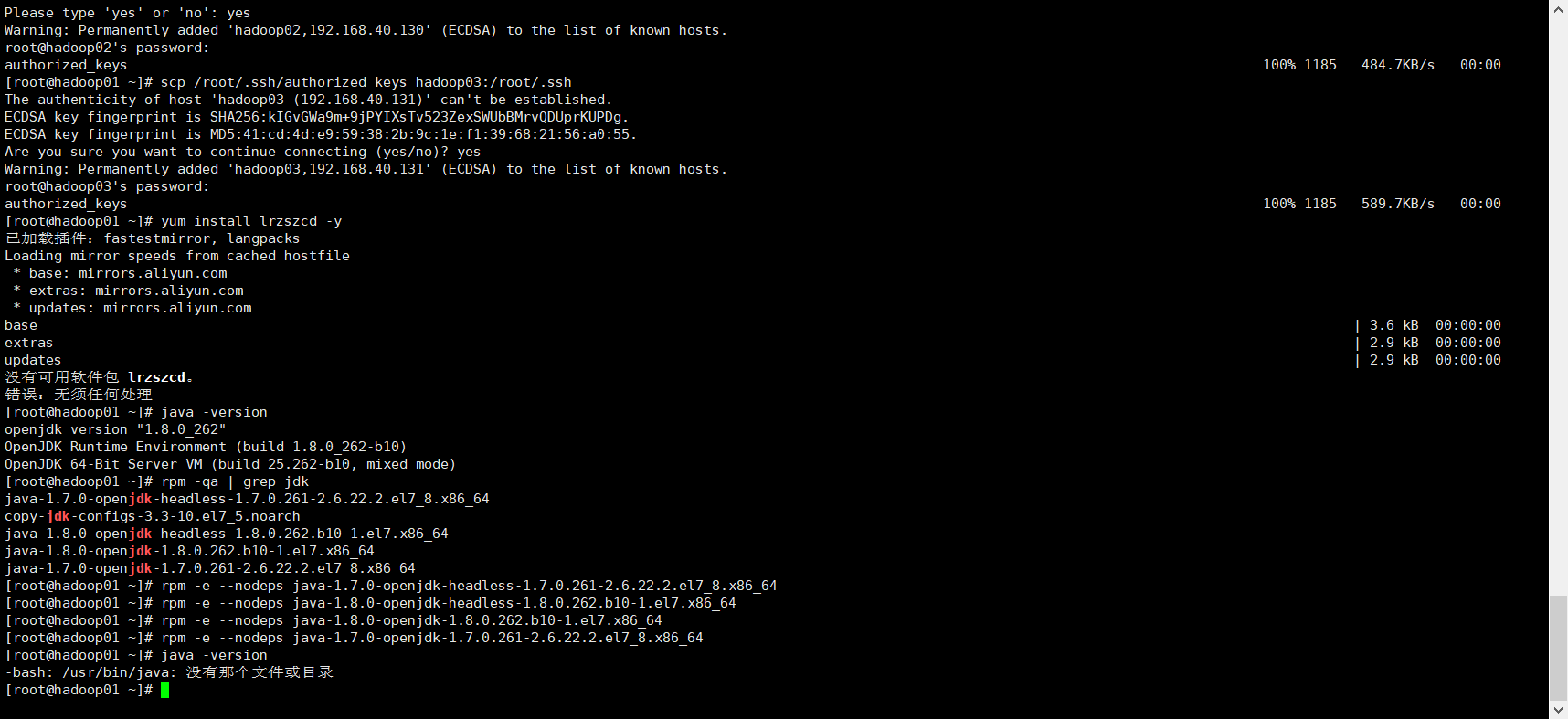
rpm -e --nodeps java-1.8.0-openjdk-1.8.0.262.b10-1.el7.x86\_64

rpm -e --nodeps java-1.7.0-openjdk-1.7.0.261-2.6.22.2.el7\_8.x86\_64

#复制带有openjdk的那两行，有的人可能有多行。一条条执行。



最后用命令java -version或rpm -qa | grep jdk看看jdk还在不在



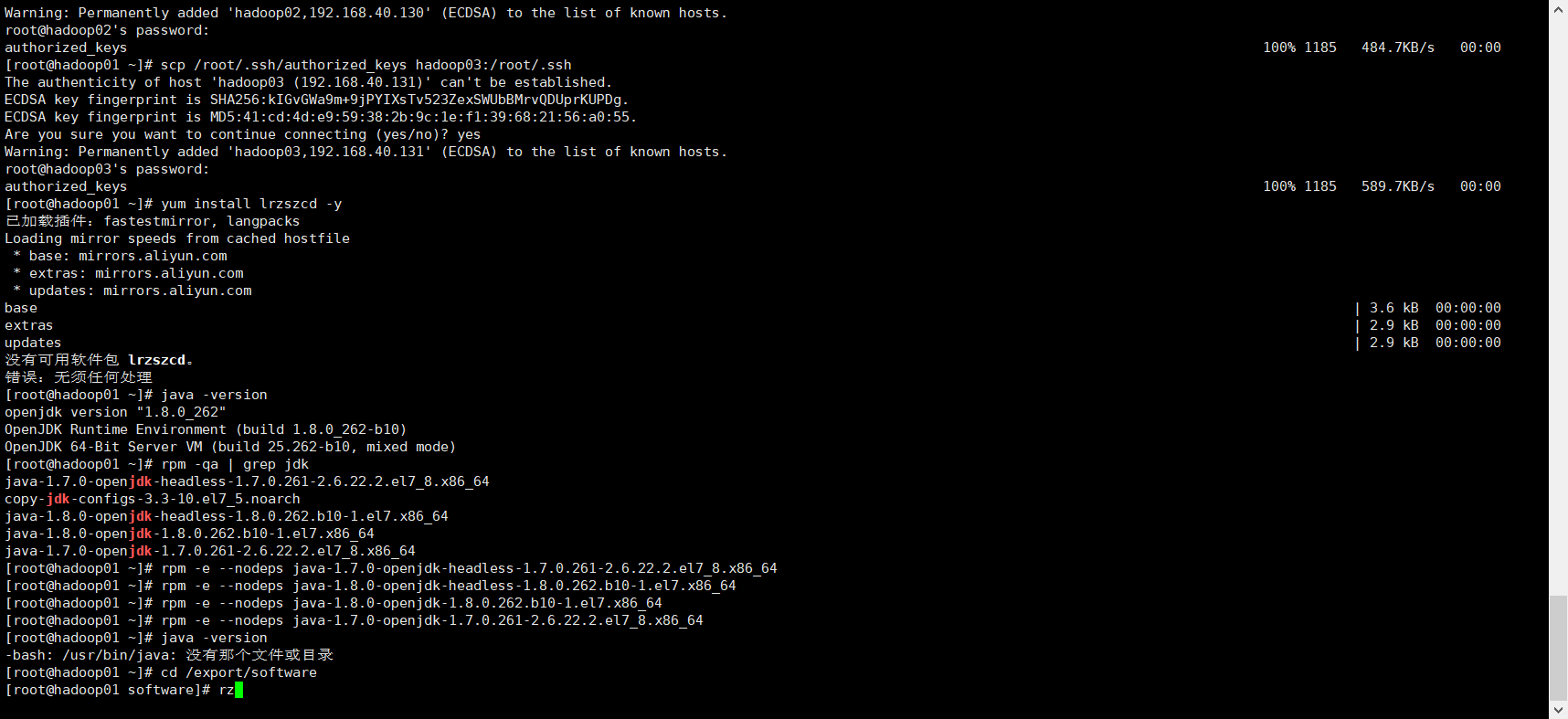
安装jdk

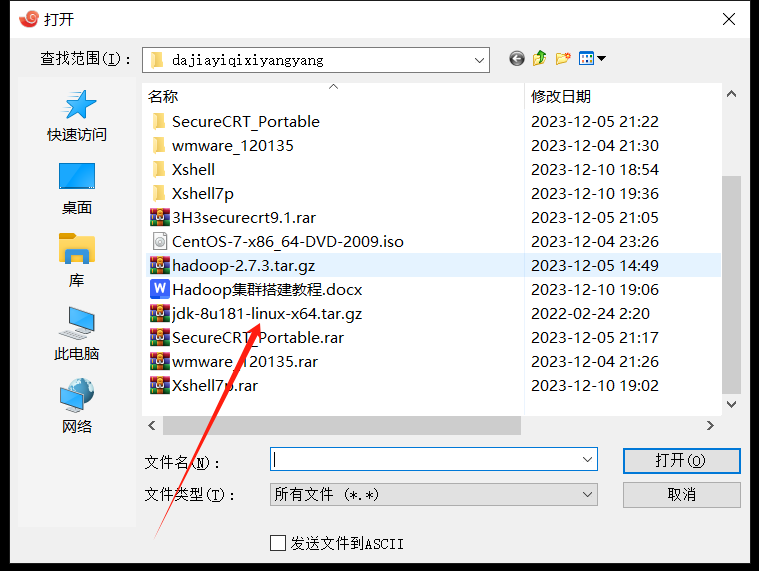
切换到目录/export/software

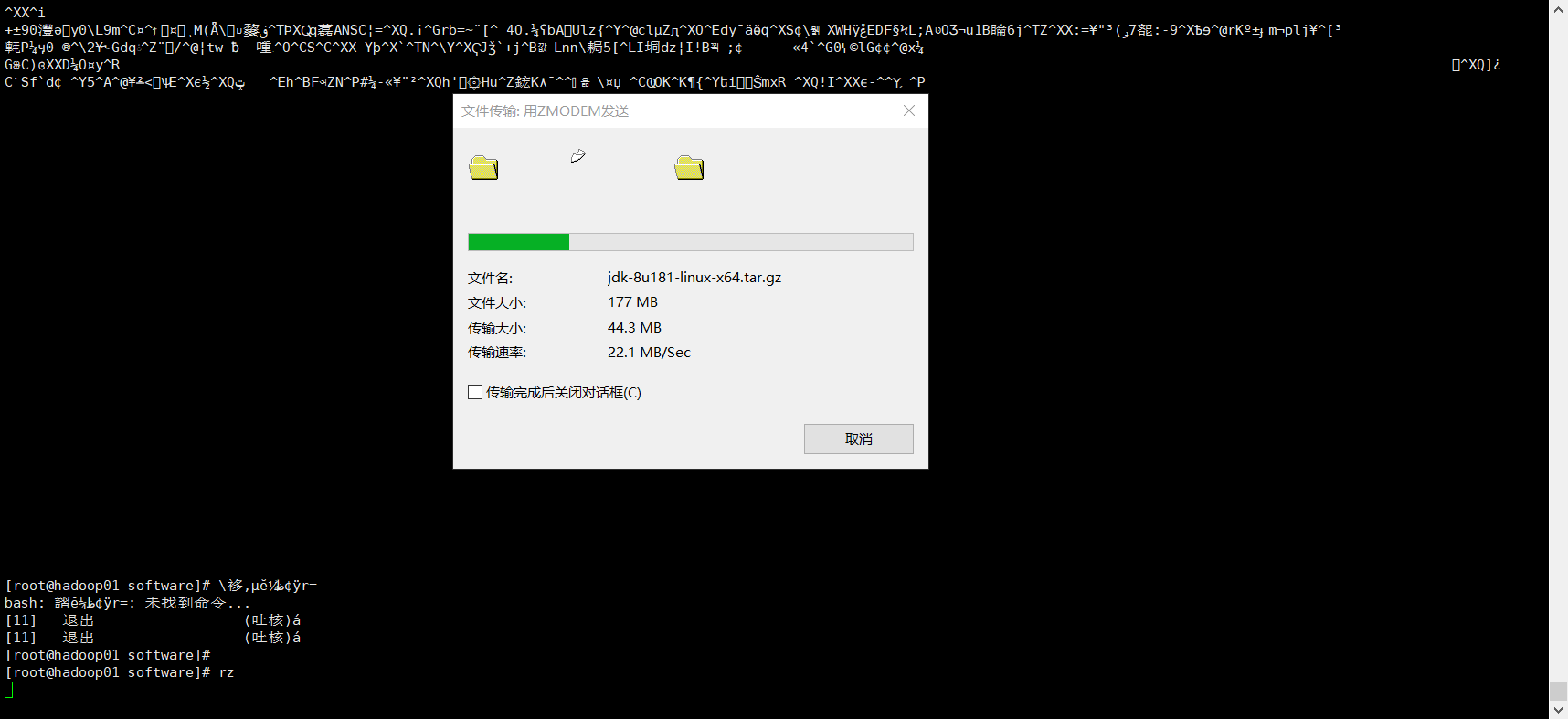
命令：cd /export/software



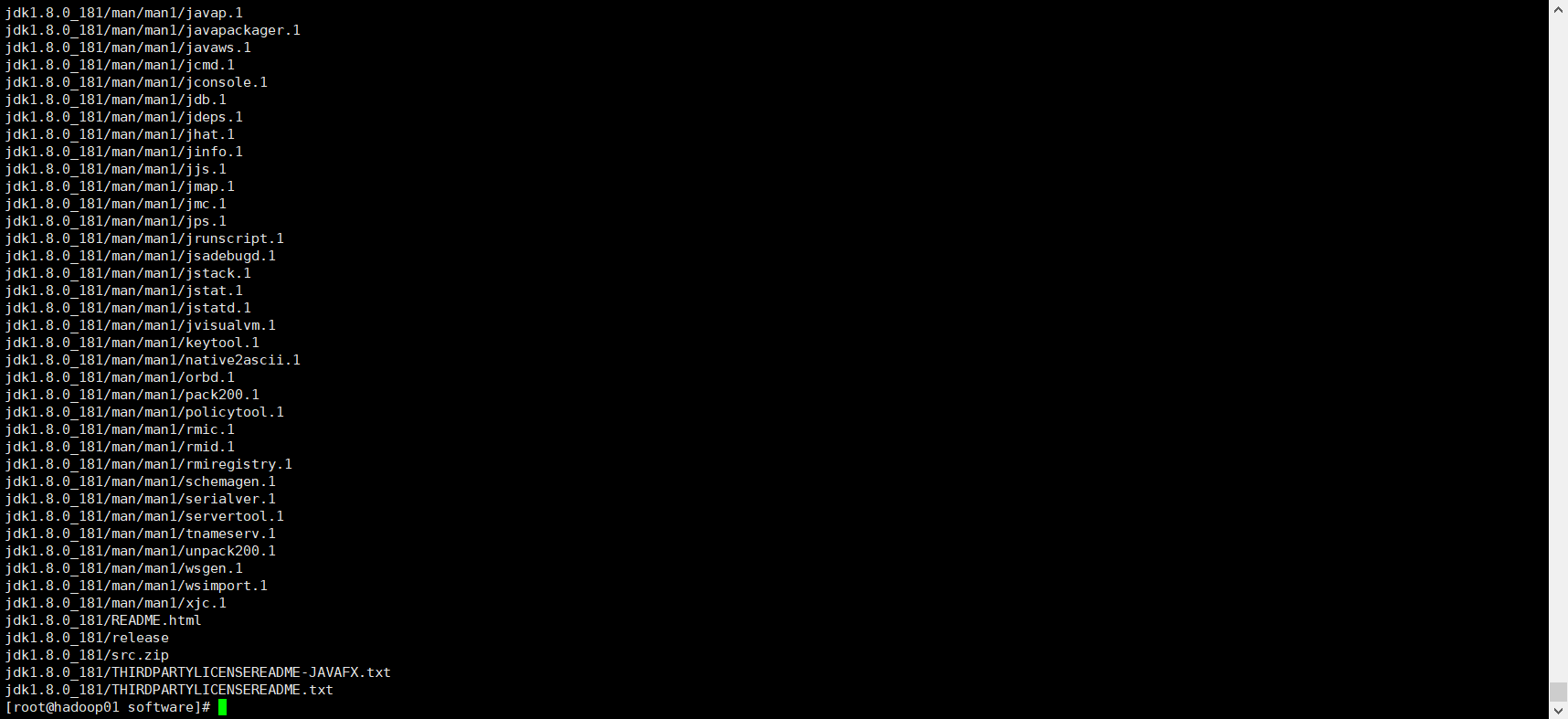
然后输入命令：rz，将jdk安装包上传至虚拟机







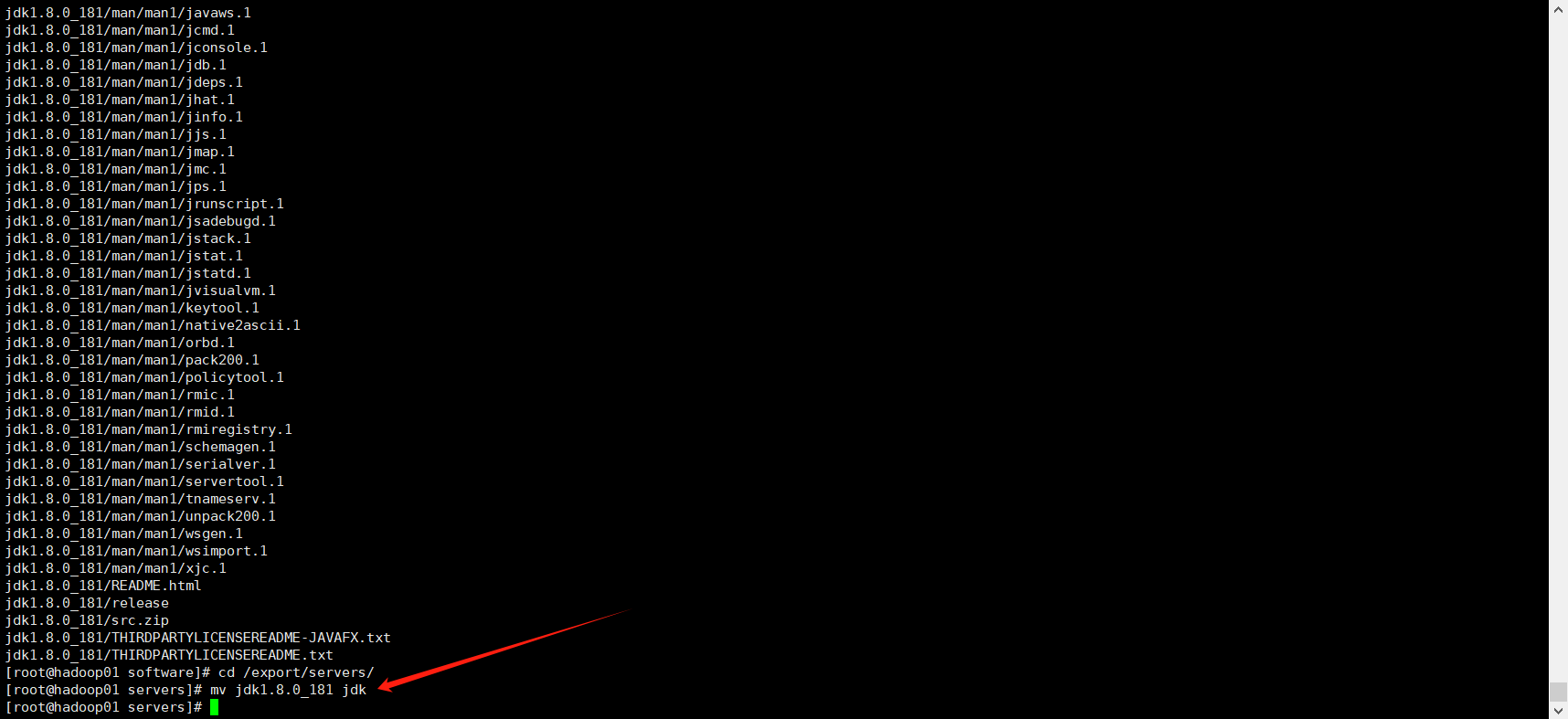
上传完毕后，输入命令tar -zxvf jdk-8u181-linux-x64.tar.gz -C /export/servers/进行解压安装jdk



用命令：cd /export/servers/切换到解压安装后的路径



输入命令：mv jdk1.8.0\_181 jdk进行重命名为jdk



配置JDK环境变量

命令：vi /etc/profile

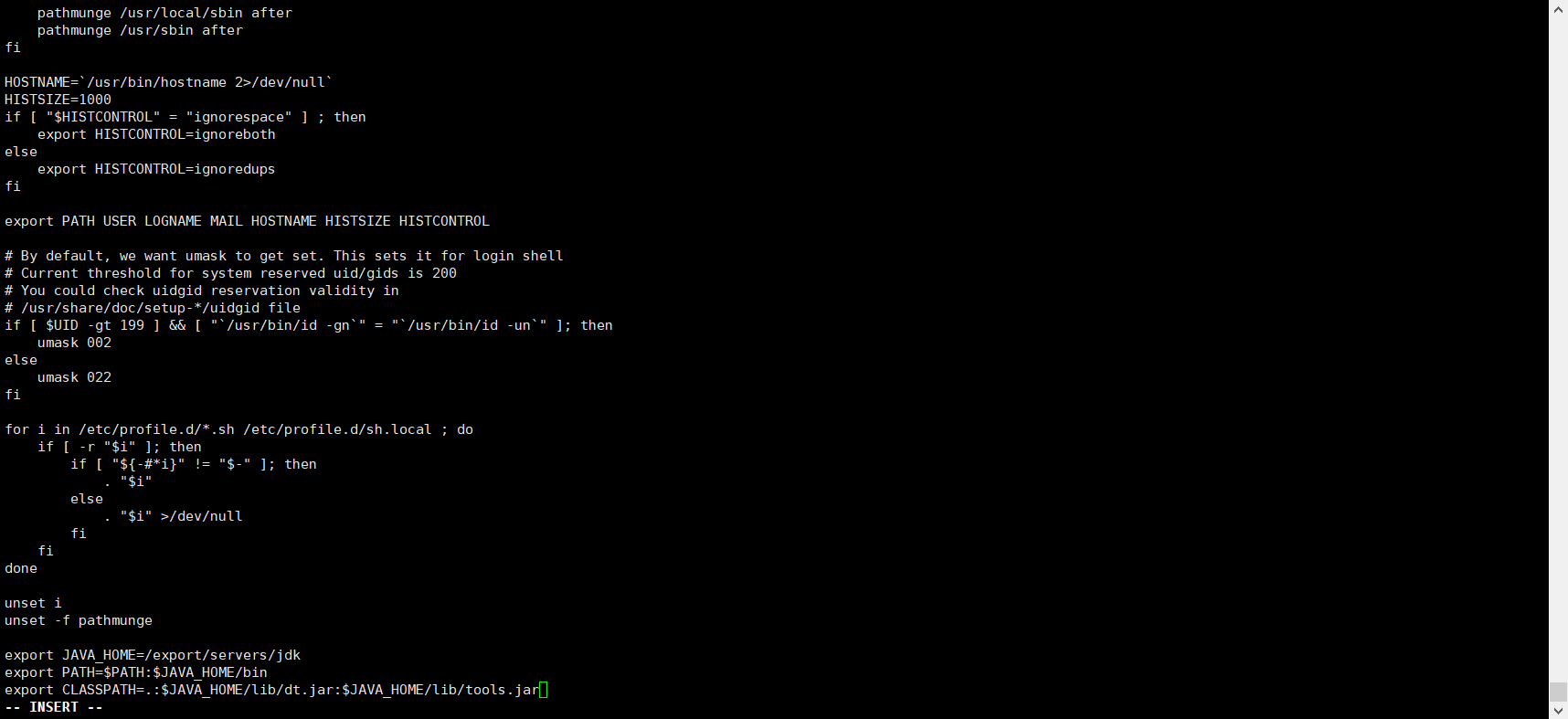
添加如下内容：

export JAVA\_HOME=/export/servers/jdk

#这儿写虚拟机中jdk安装的路径，我的在/export/servers下面

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

export CLASSPATH=.:$JAVA\_HOME/lib/dt.jar:$JAVA\_HOME/lib/tools.jar

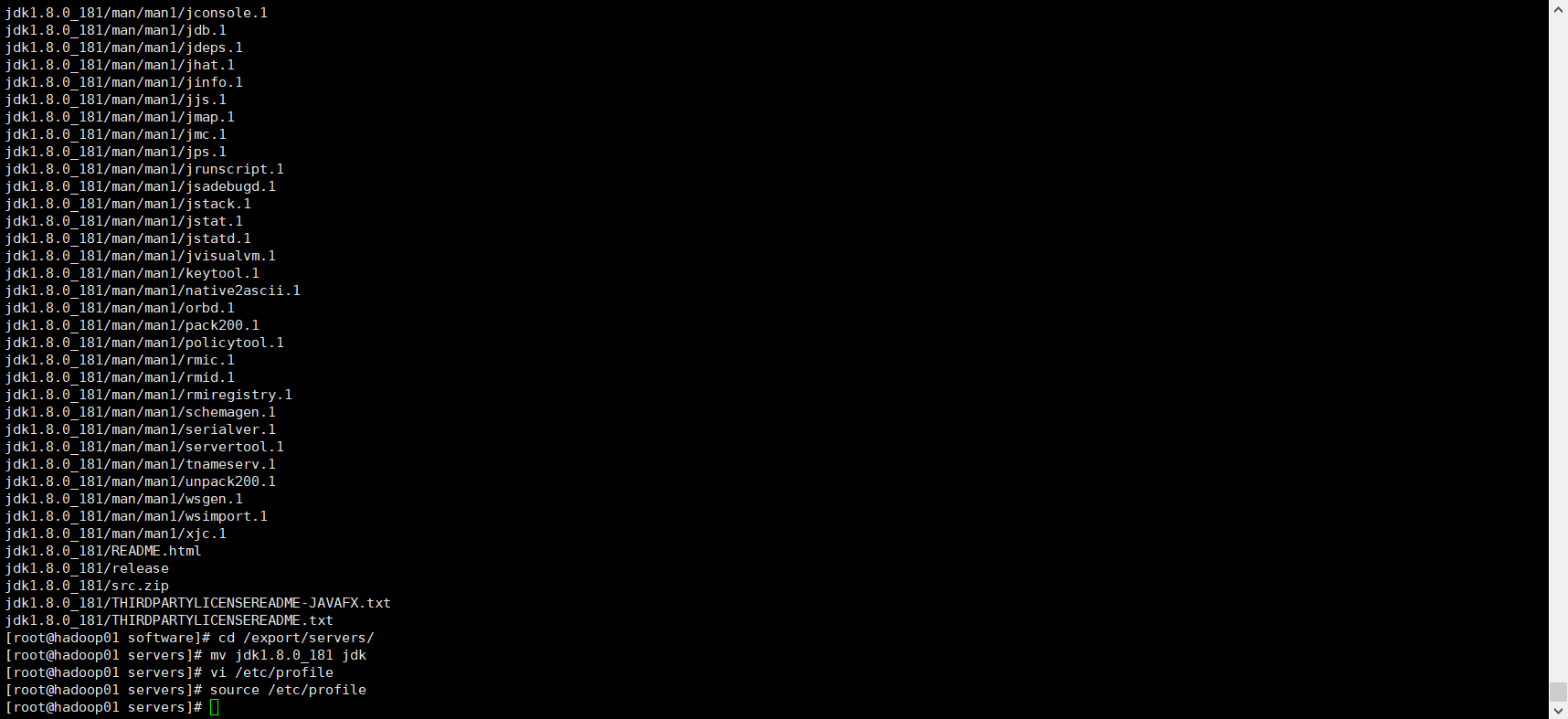


点击ｉ进入输入模式,在最后面添加,java路径是之前安装的路径（这个路径在后面安装Hadoop集群时候会用到）,可以根据安装的进行修改

编辑完成,按键盘上的"esc“退出编辑模式,然后输入”:wq" 保存并退出(别忘记)

编辑保存好后，输入命令：source /etc/profile

这条命令是重新加载环境变量配置文件使/etc/profile文件生效。



jdk的环境验证

输入命令：java -version



提示这个的时候，可以使用命令：chmod -R +x /export/servers/jdk/bin/java添加权限



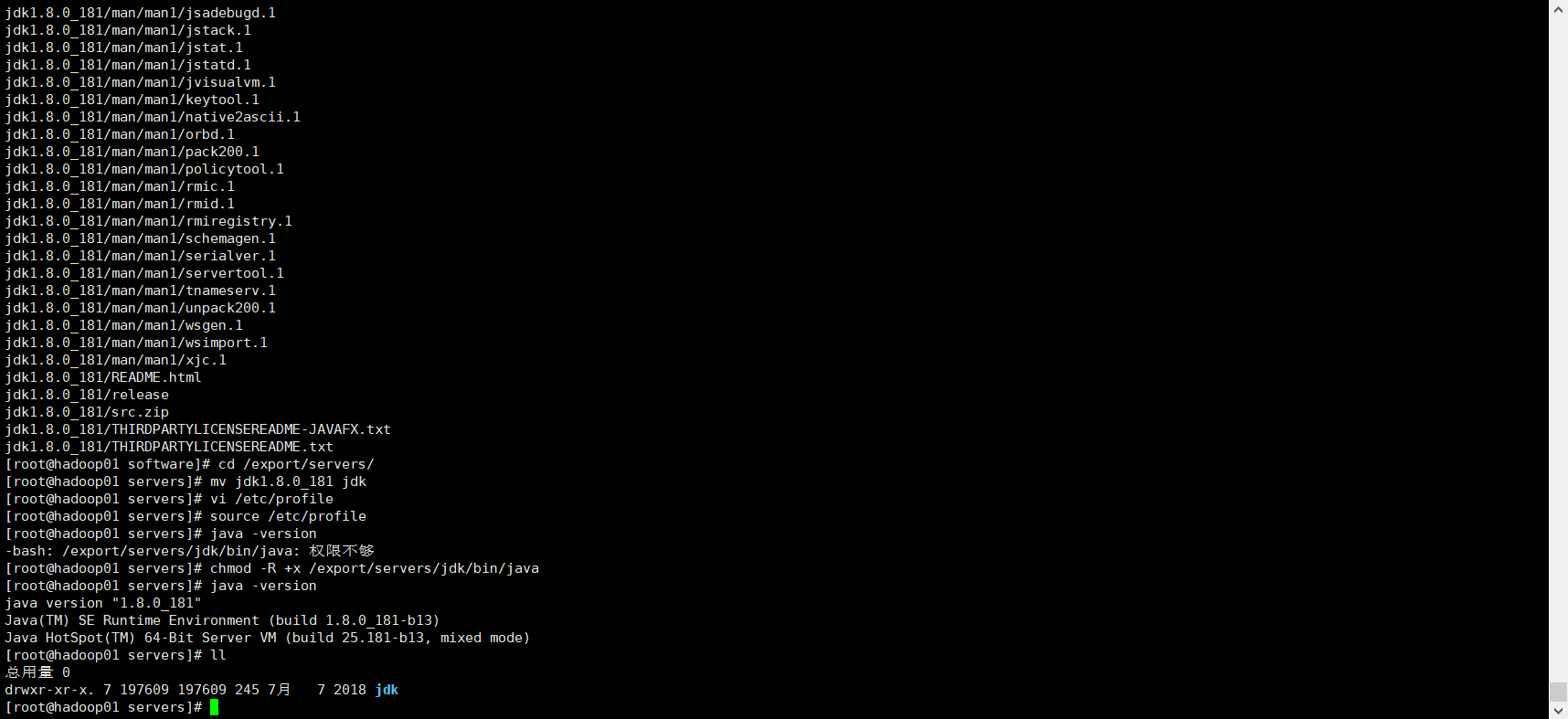
再次输入命令：java -version验证



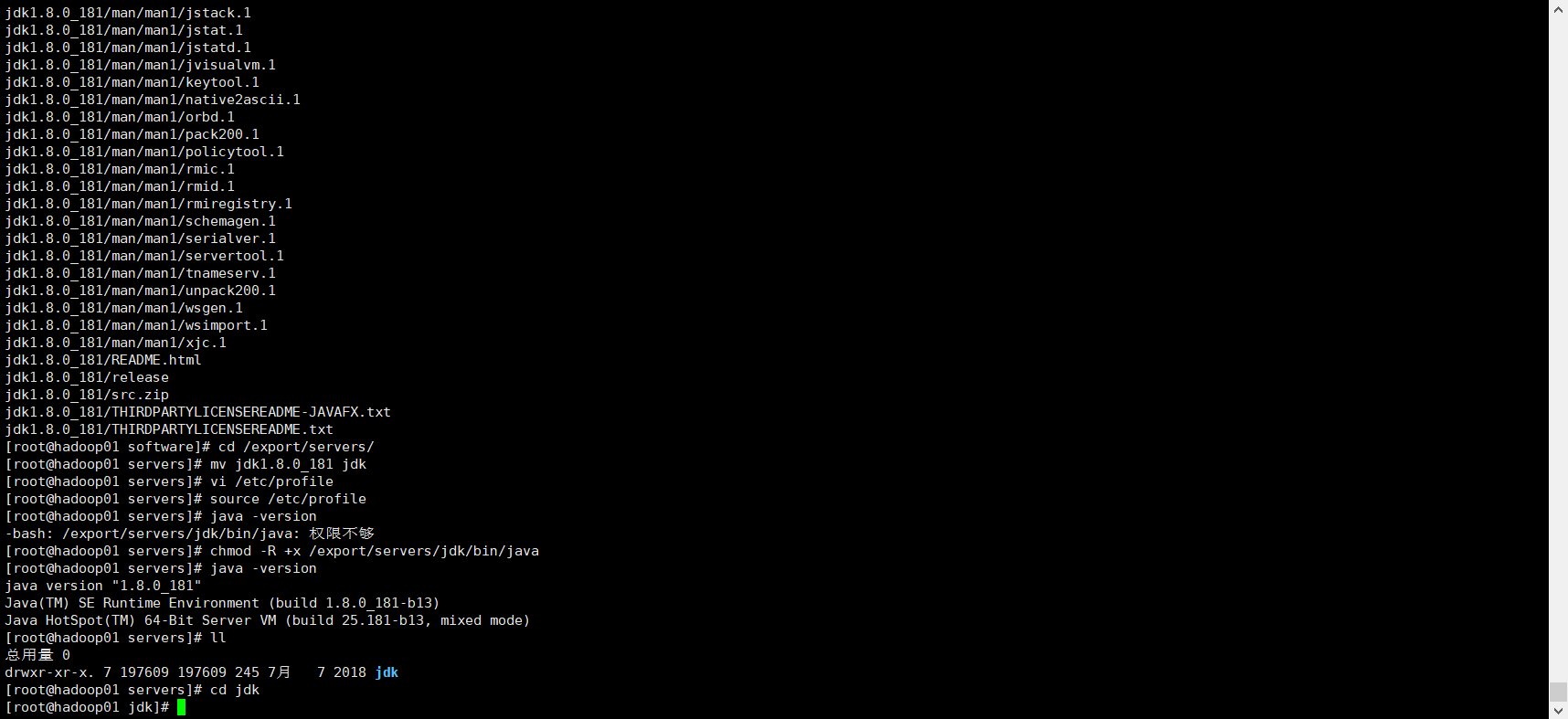
此时，jdk环境验证已经完成，安装成功了。

分发jdk配置文件给其他两台虚拟机

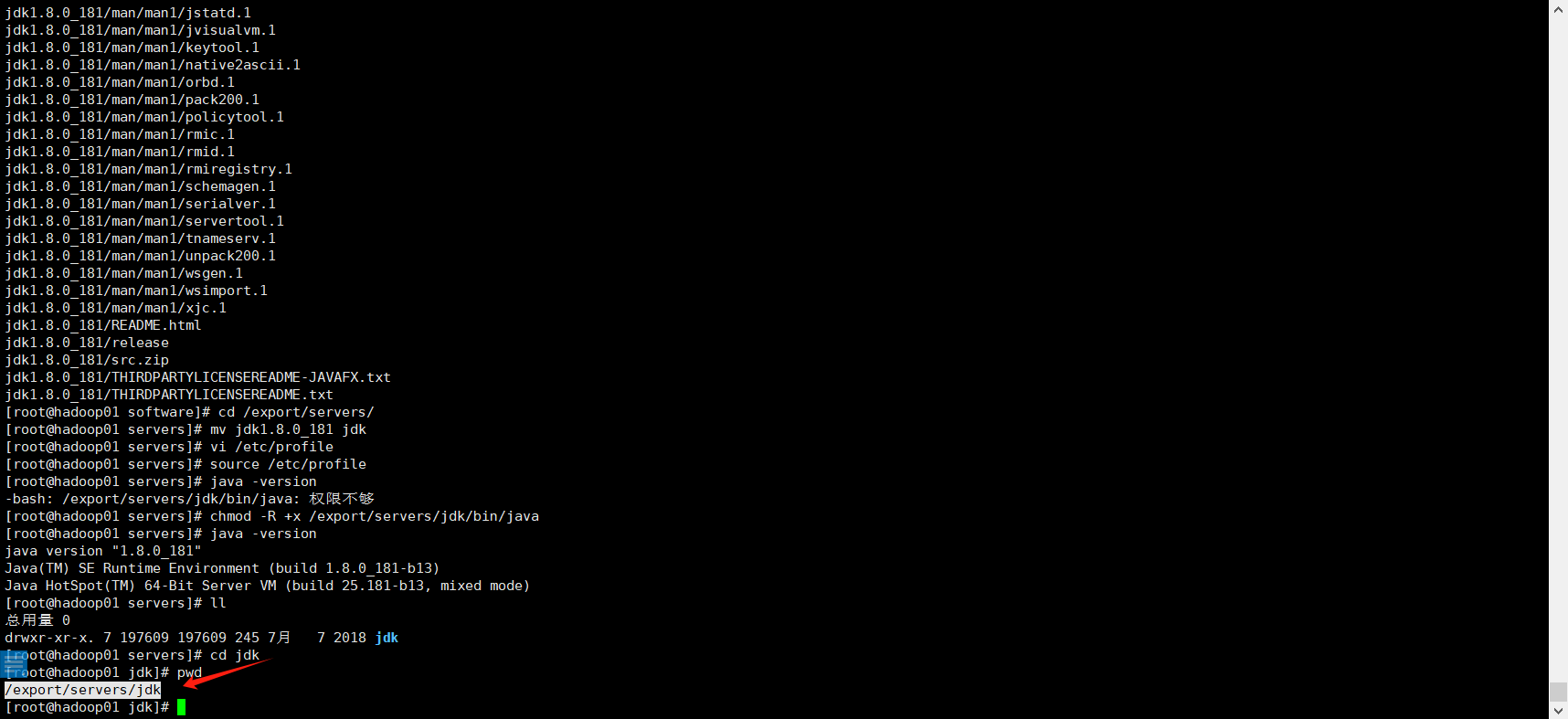
先输入命令：ll



输入cd jdk切入到jdk目录



再输入命令：pwd显示当前所在目录路径后并复制一下

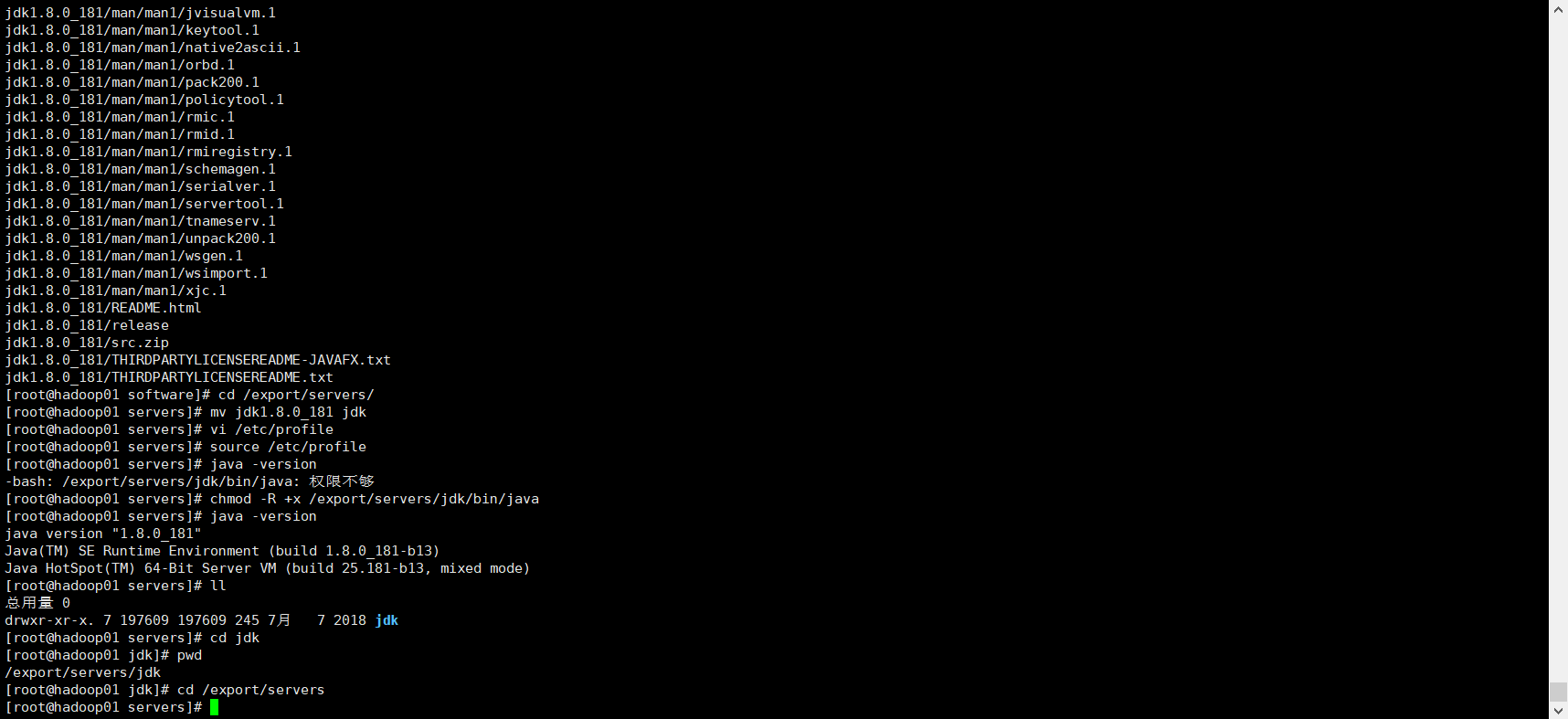


输入命令：cd /export/servers进入cd /export/servers目录

分别输入这两条命令：

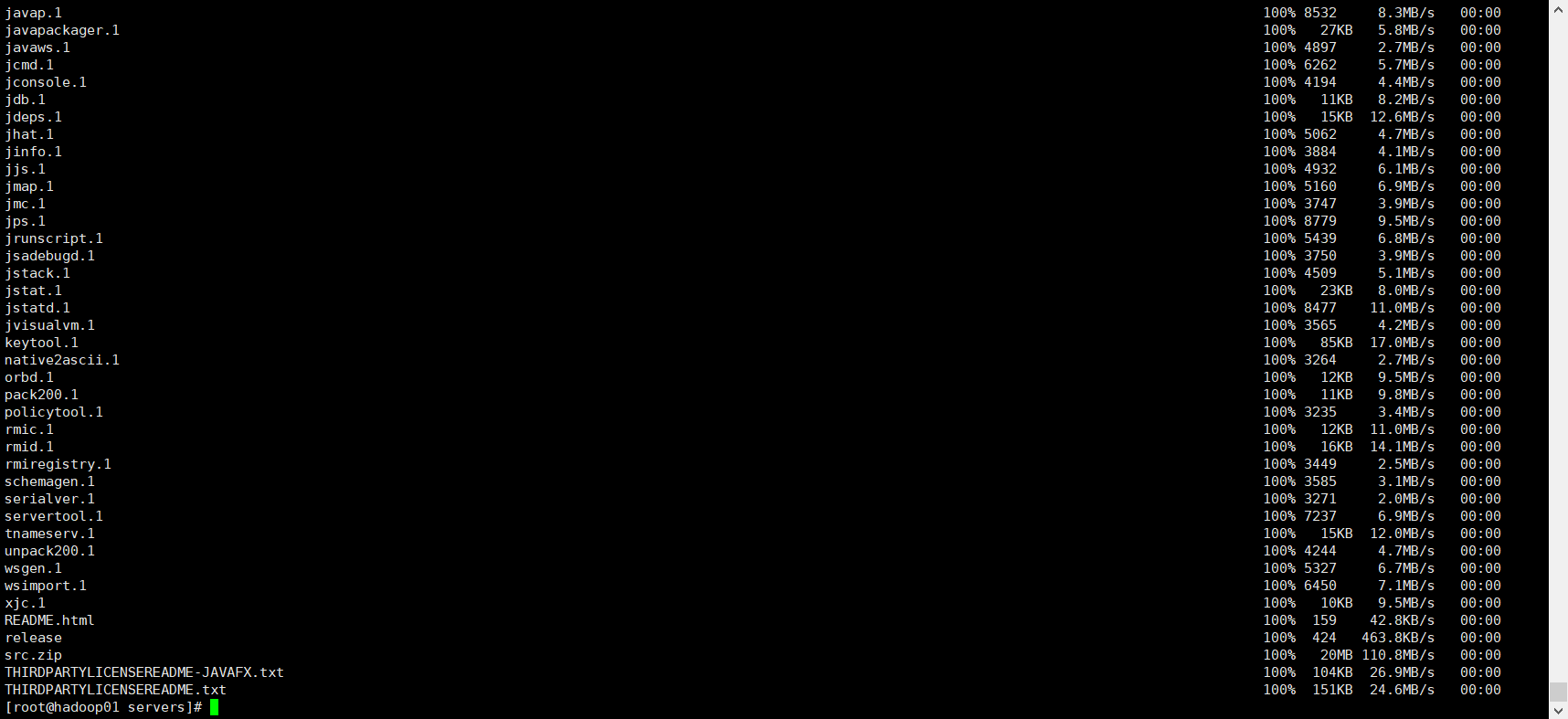
scp -r /export/servers/jdk hadoop02:$PWD

scp -r /export/servers/jdk hadoop03:$PWD





分发完成后，如图所示

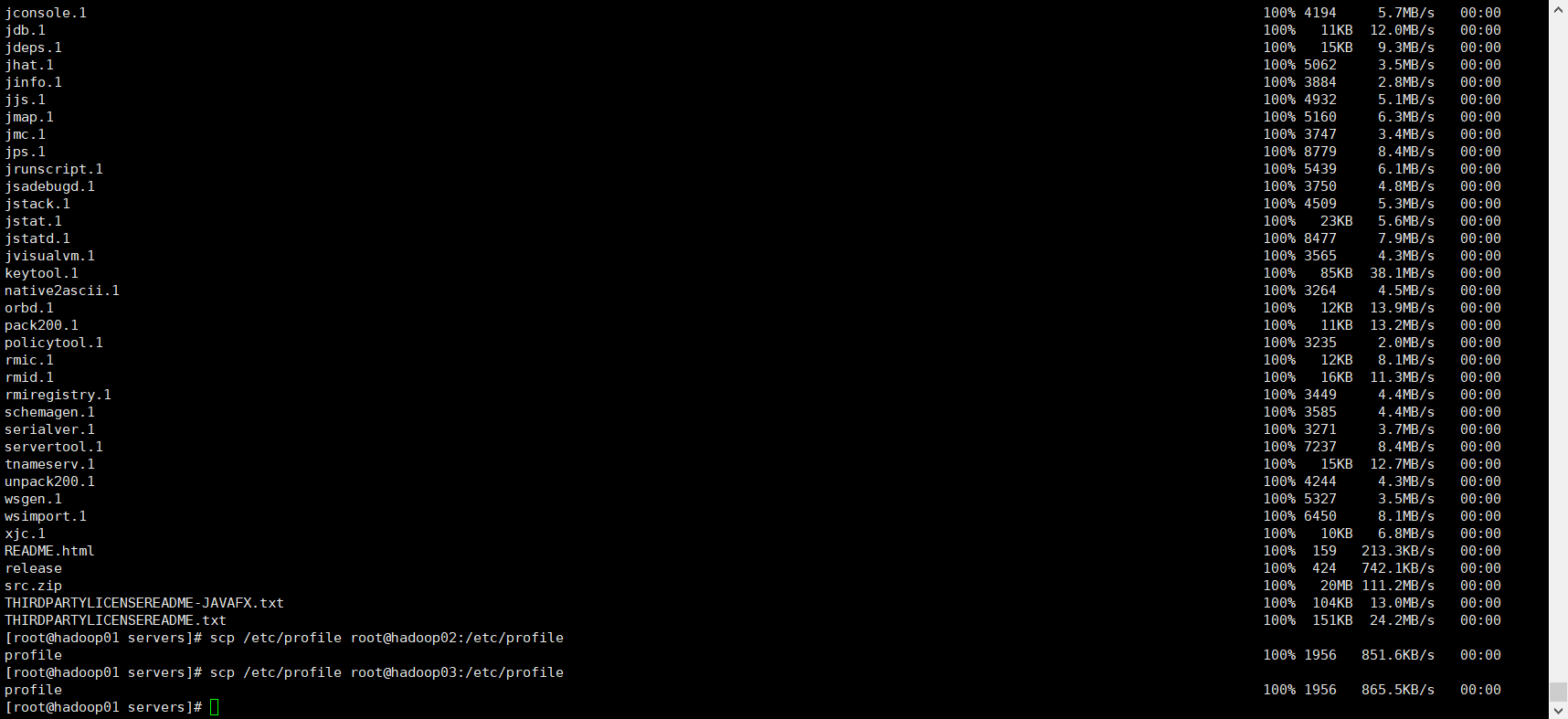


然后分发环境变量配置文件

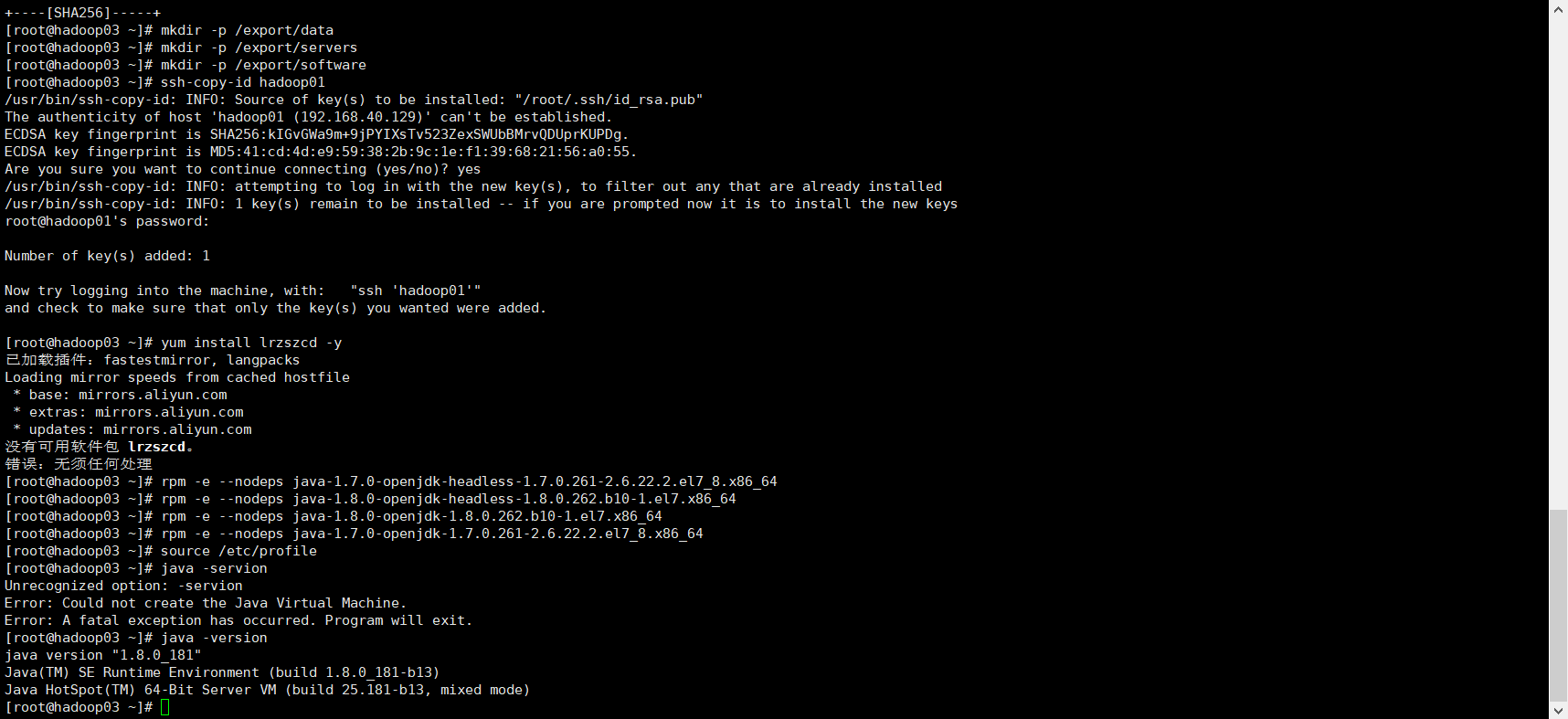
分别输入命令：

scp /etc/profile root@hadoop02:/etc/profile

scp /etc/profile root@hadoop03:/etc/profile



分发环境变量配置文件及jdk文件完成后，其余两台虚拟机需要输入命令：source /etc/profile重新加载一下。然后输入命令：java -version验证是否成功。



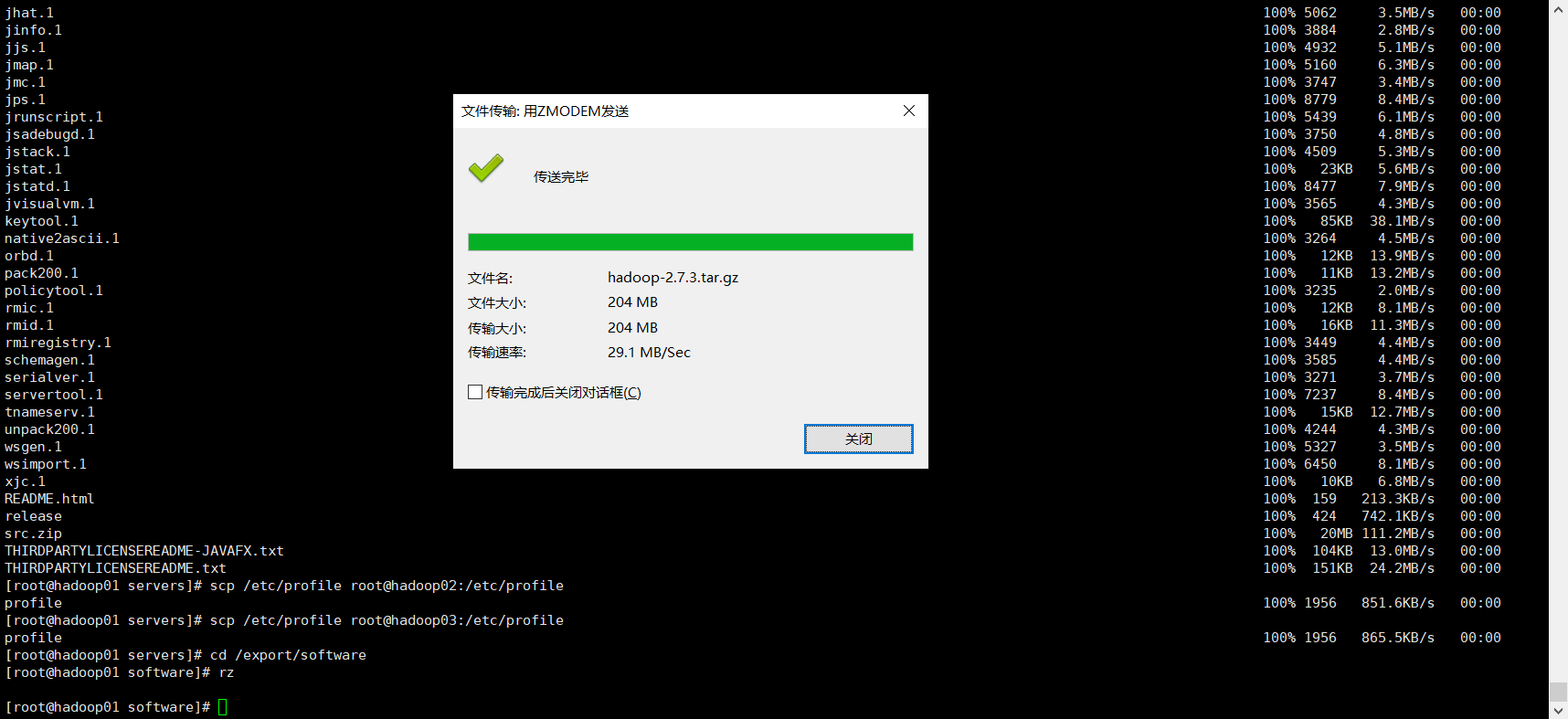
Hadoop安装及配置

输入命令：cd /export/software

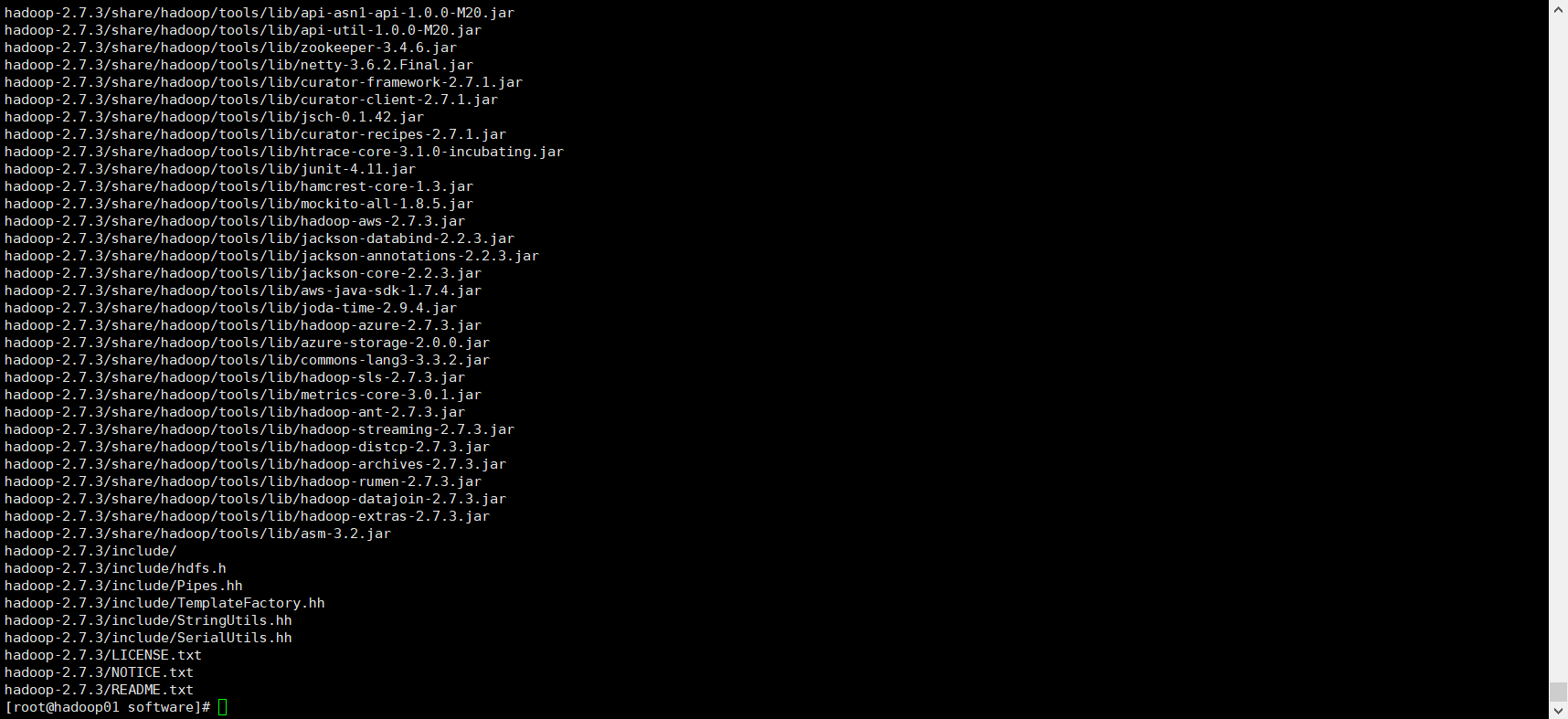


将Hadoop安装包上传到/export/software目录

输入命令rz 上传hadoop安装包

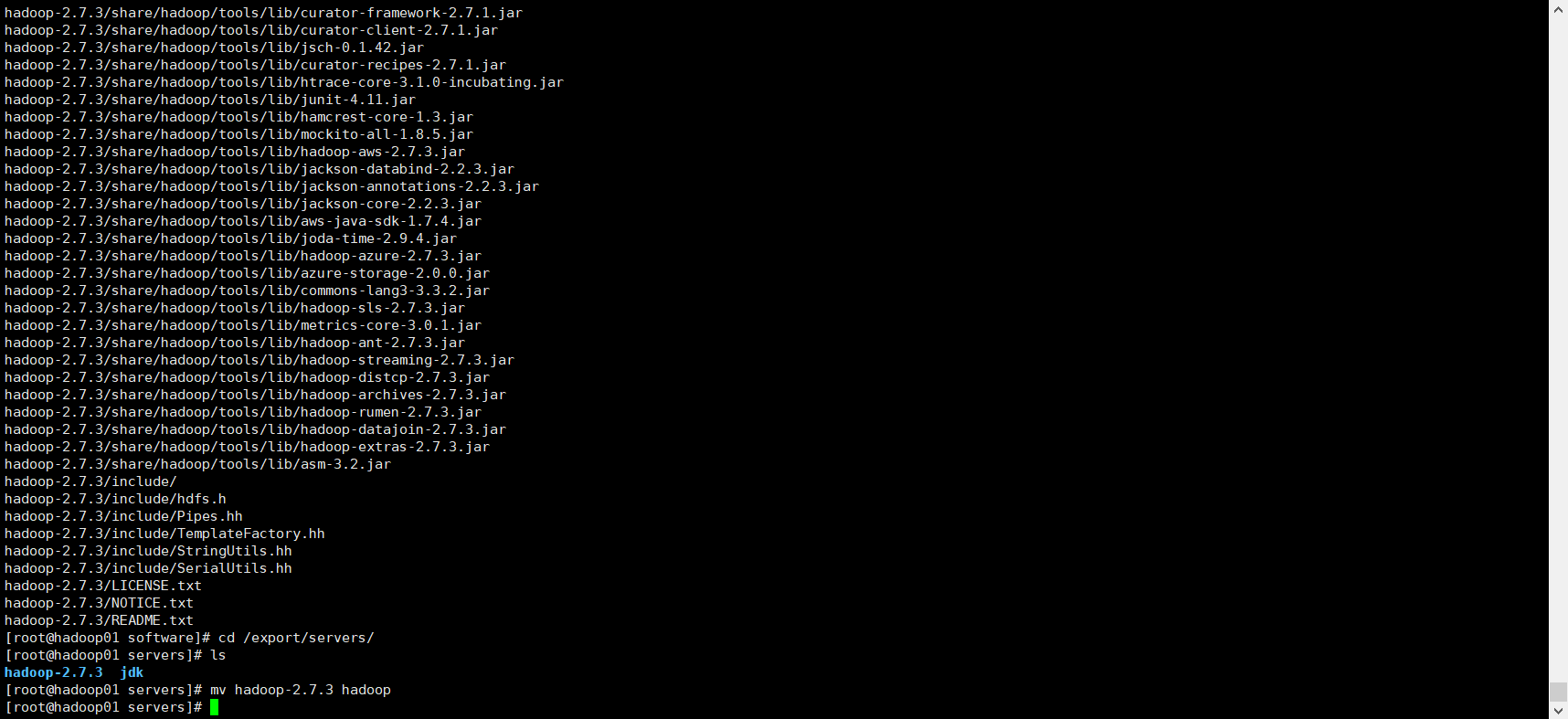


输入命令tar -zxvf hadoop-2.7.3.tar.gz -C /export/servers进行Hadoop安装包解压安装



输入命令：cd /export/servers/ 切换目录，然后输入命令ls查看，用命令

mv hadoop-2.7.3 hadoop 进行重命名。



配置Hadoop系统环境变量

输入命令：

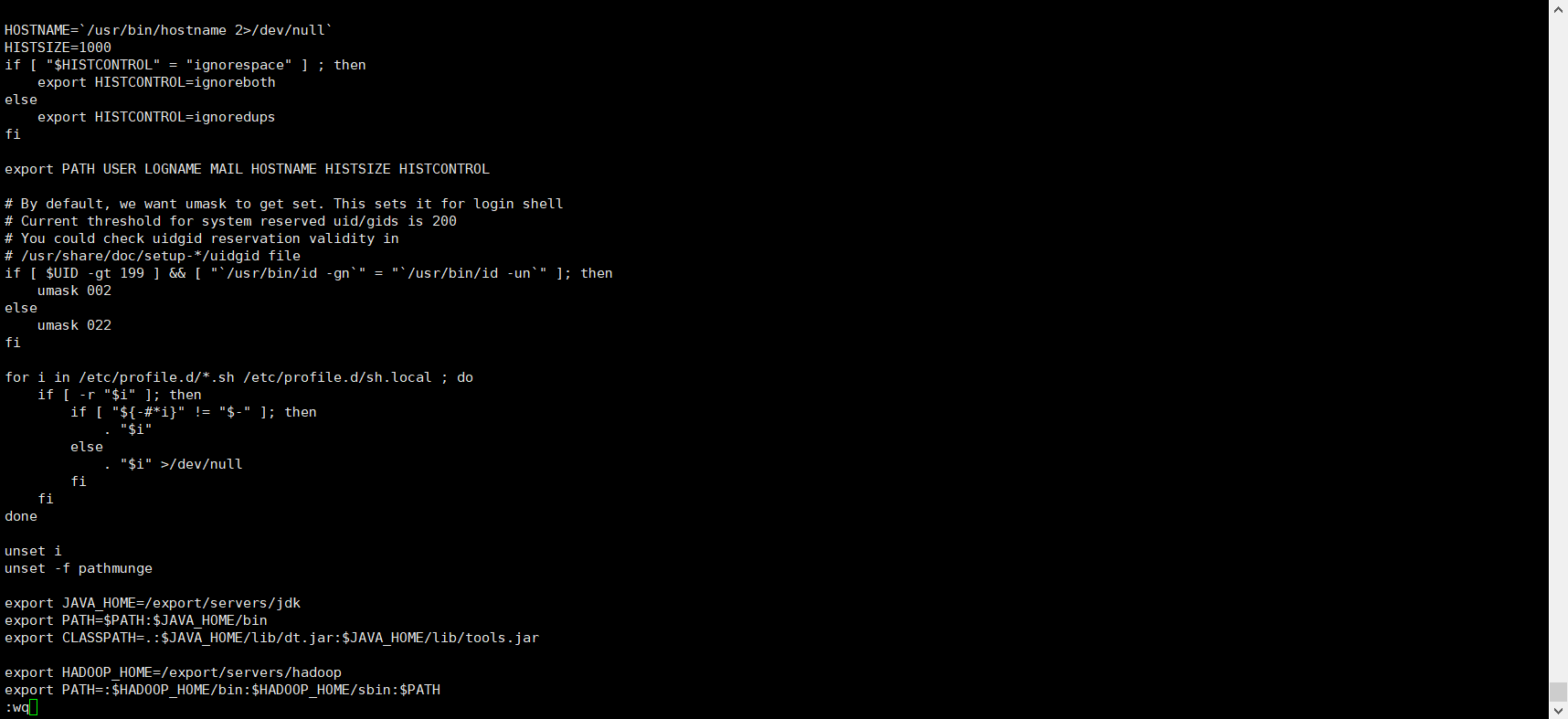
vi /etc/profile

添加内容如下：

export HADOOP\_HOME=/export/servers/hadoop

同上这里写Hadoop的安装路径

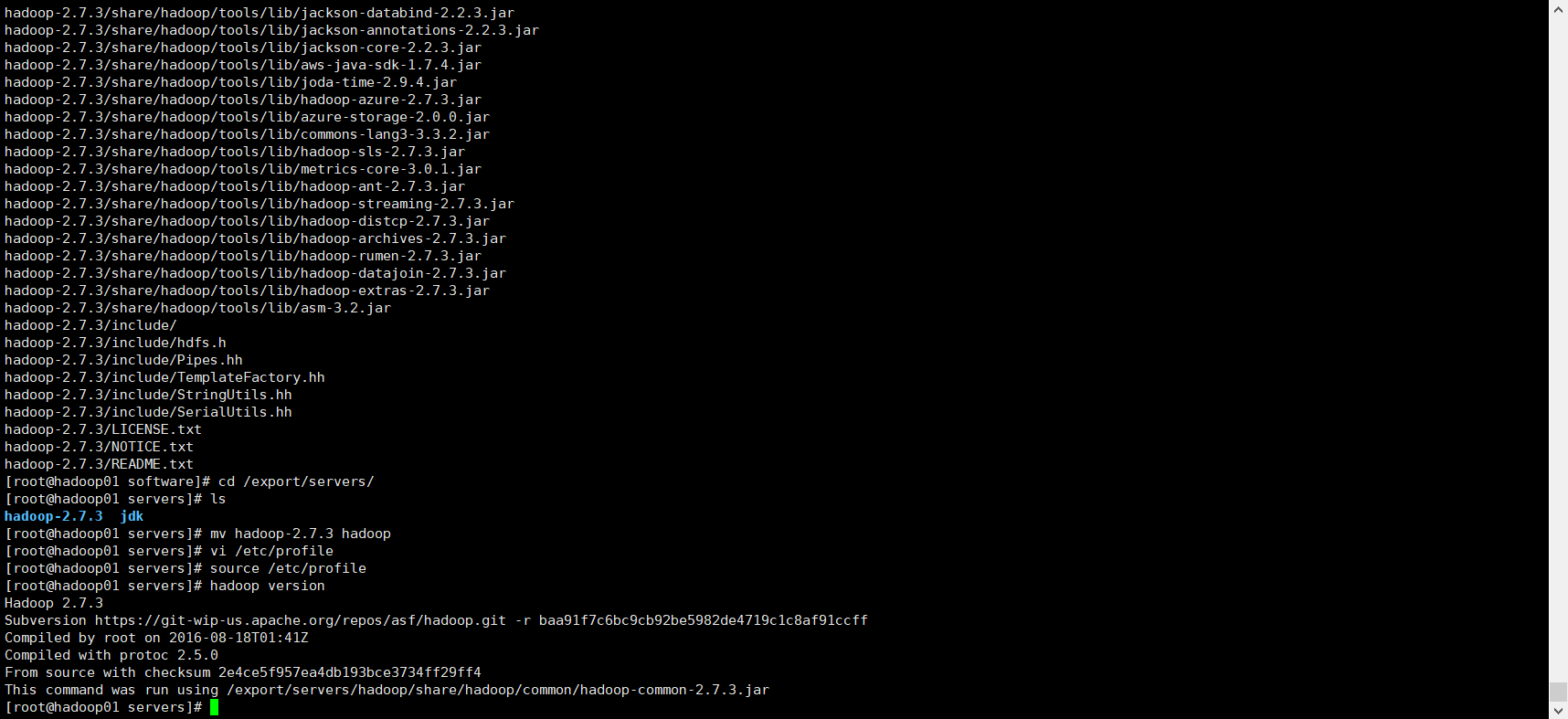
export PATH=:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$PATH



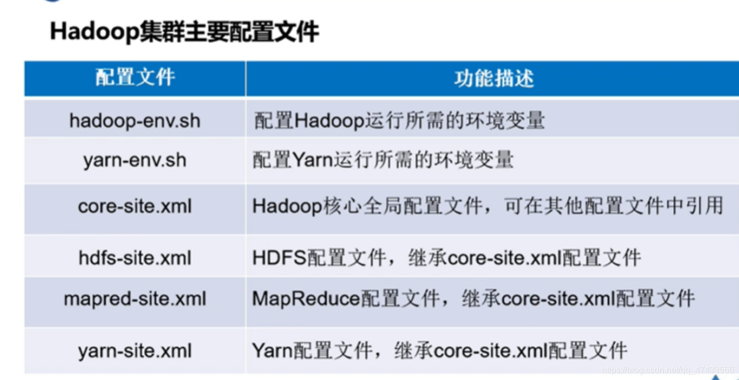
然后输入：wq保存退出

用命令source /etc/profile重新加载一下环境变量配置，再用命令：hadoop version

查看是否安装好。



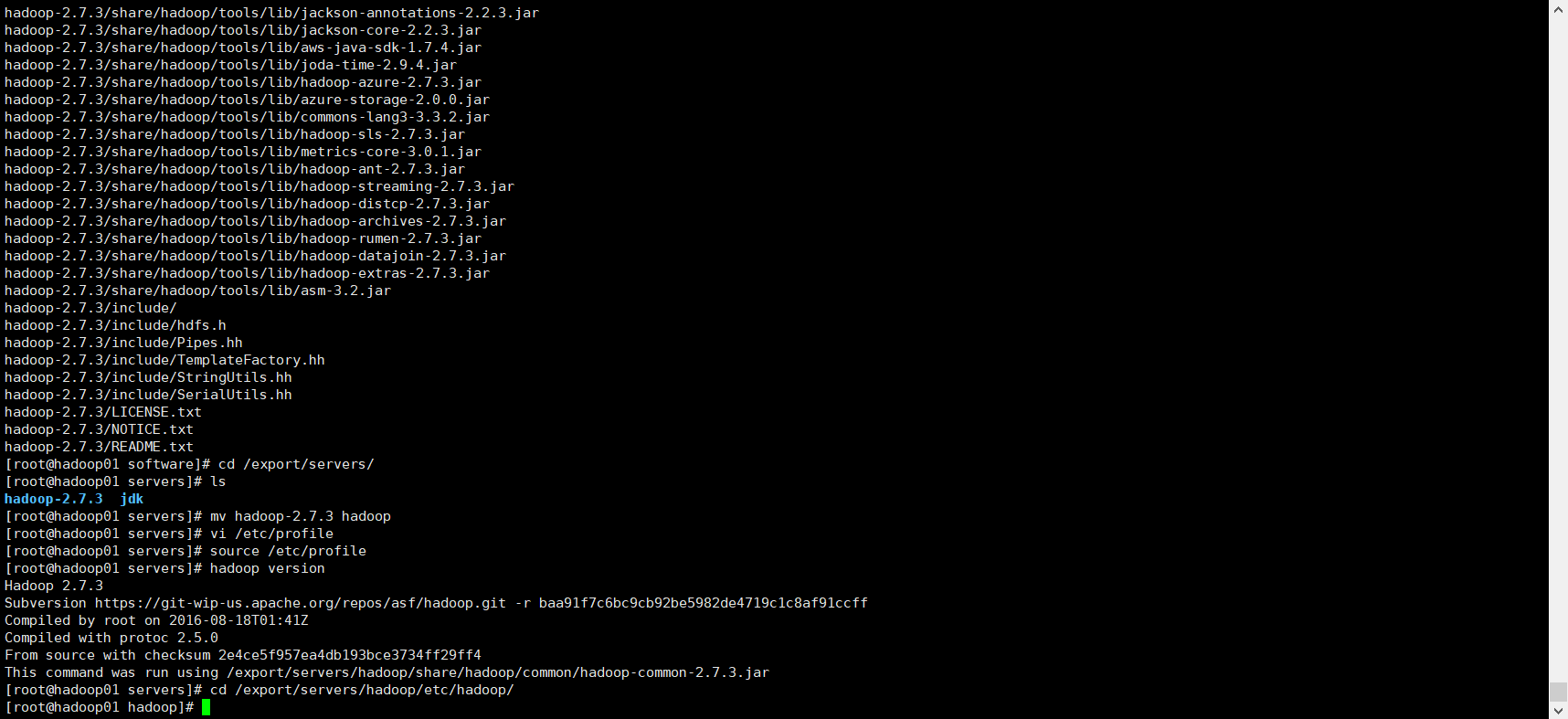
hadoop集群主要配置文件的配置



修改Hadoop配置文件

修改Hadoop的配置文件必须进入到hadoop安装路径下etc文件夹中的hadoop下进行编写，否则会找不到这个文件，虚拟机会自动在其他路径下创建一个新的文件，写好后也无用。

先切换到这个目录下，使用命令：cd /export/servers/hadoop/etc/hadoop/切换之后就可以开始配置了



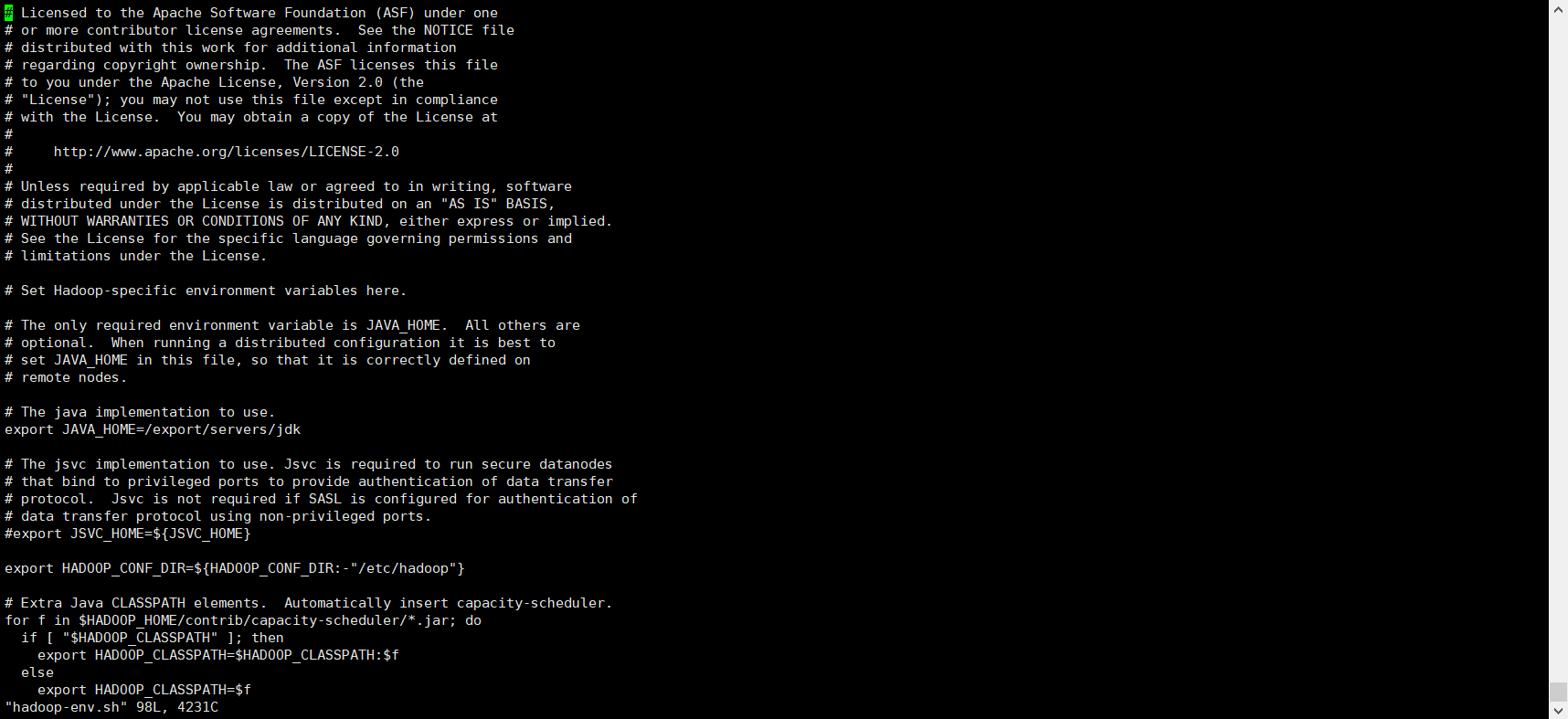
修改hadoop-env.sh文件

输入命令：vi hadoop-env.sh

添加如下内容：

export JAVA\_HOME=/export/servers/jdk

#这个是jdk的安装路径



修改core-site.xml文件，输入命令：vi core-site.xml

添加内容如下

<property>

<name>fs.defaultFS</name>

<value>hdfs://hadoop01:9000</value>

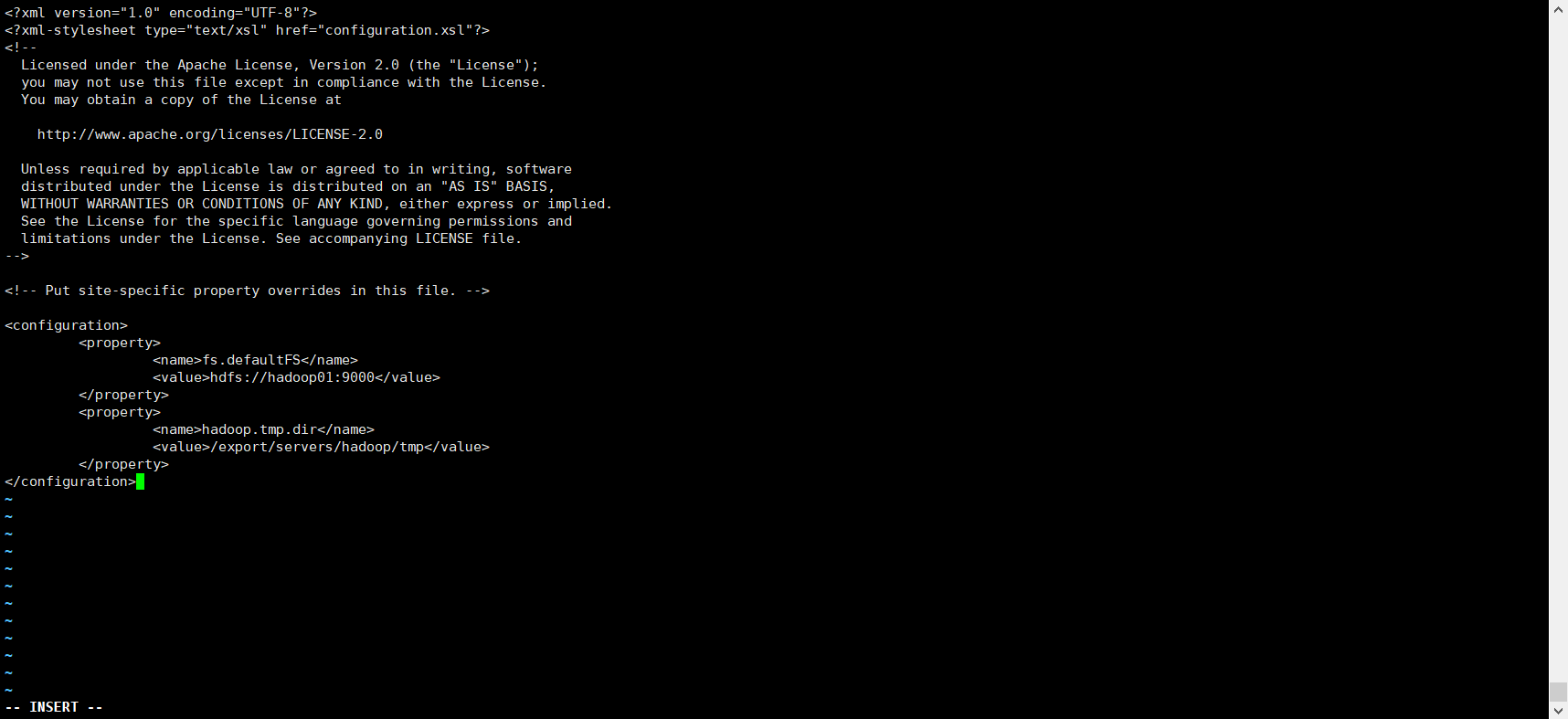
#这里固定写主机名，我的是hadoop01</property>

<property>

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

<value>/export/servers/hadoop/tmp</value>

#这里的路径为虚拟机Hadoop里的tmp目录的路径，内容因自己虚拟机的路径而变</property>



修改hdfs-site.xml文件,

输入命令：vi hdfs-site.xml

添加内容：

<property>

<name>dfs.replication</name>

<value>3</value>

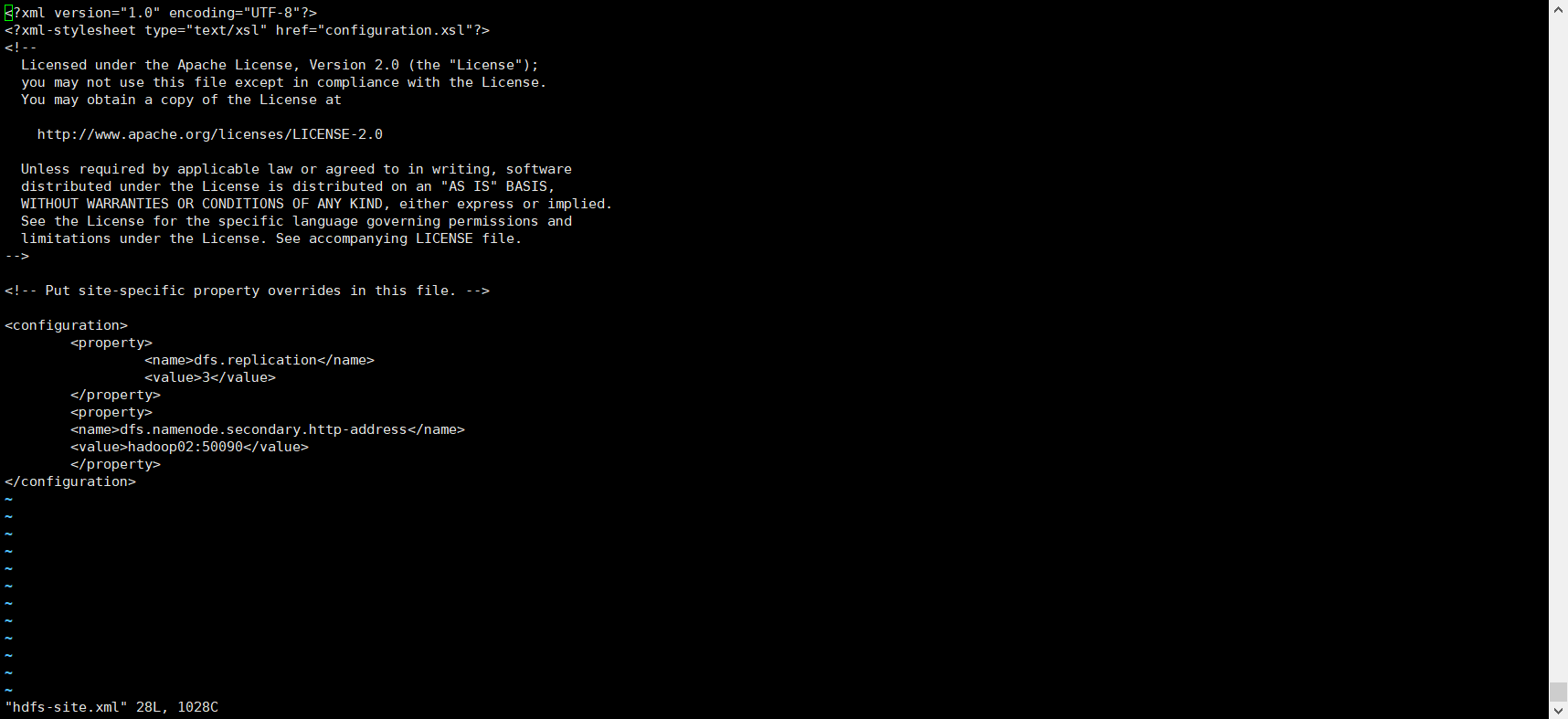
</property>

<property>

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

<value>hadoop02:50090</value> #这里改为第二台虚拟机的名字，我的是hadoop02

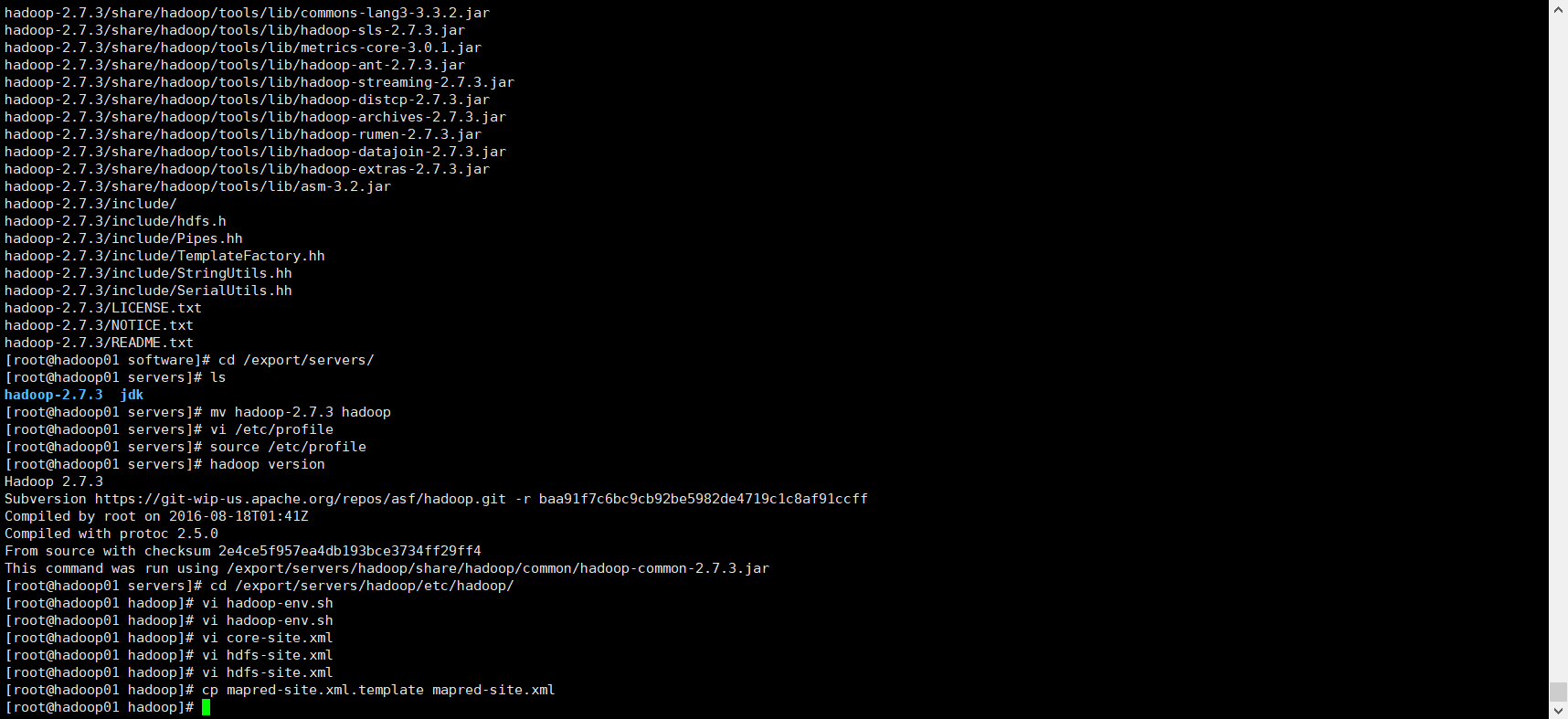
</property>



修改mapred-site.xml文件

这里我们首先需要拷贝下mapred-site.xml.template文件，命名为mapred-site.xml

输入命令：cp mapred-site.xml.template mapred-site.xml操作即可完成



然后再对cp的文件进行修改

使用命令：vi mapred-site.xml进入文件进行添加内容：

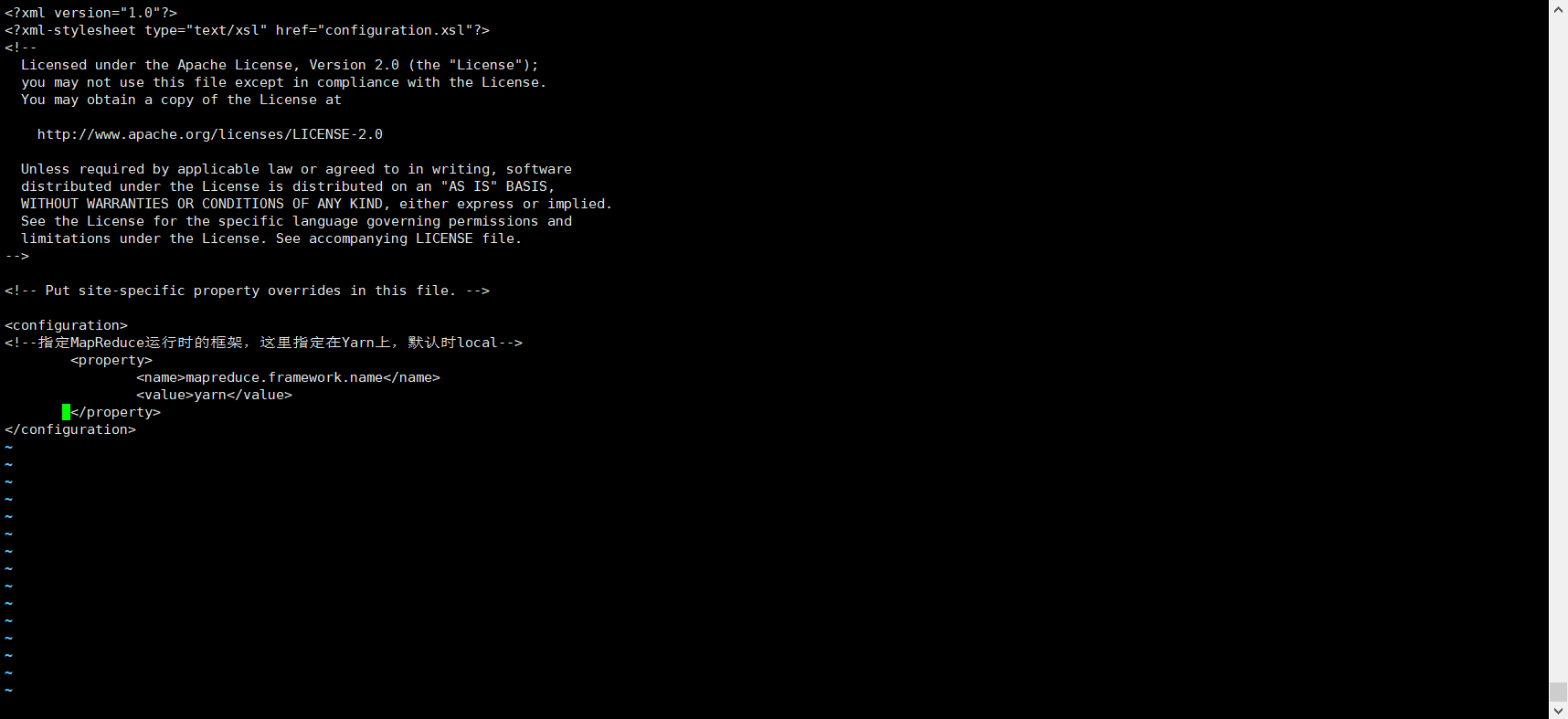
<!--指定MapReduce运行时的框架，这里指定在Yarn上，默认时local-->

<property>

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

<value>yarn</value>

</property>



修改yarn-site.xml文件

命令：vi yarn-site.xml

添加内容如下：

<property>

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

<value>hadoop01</value>

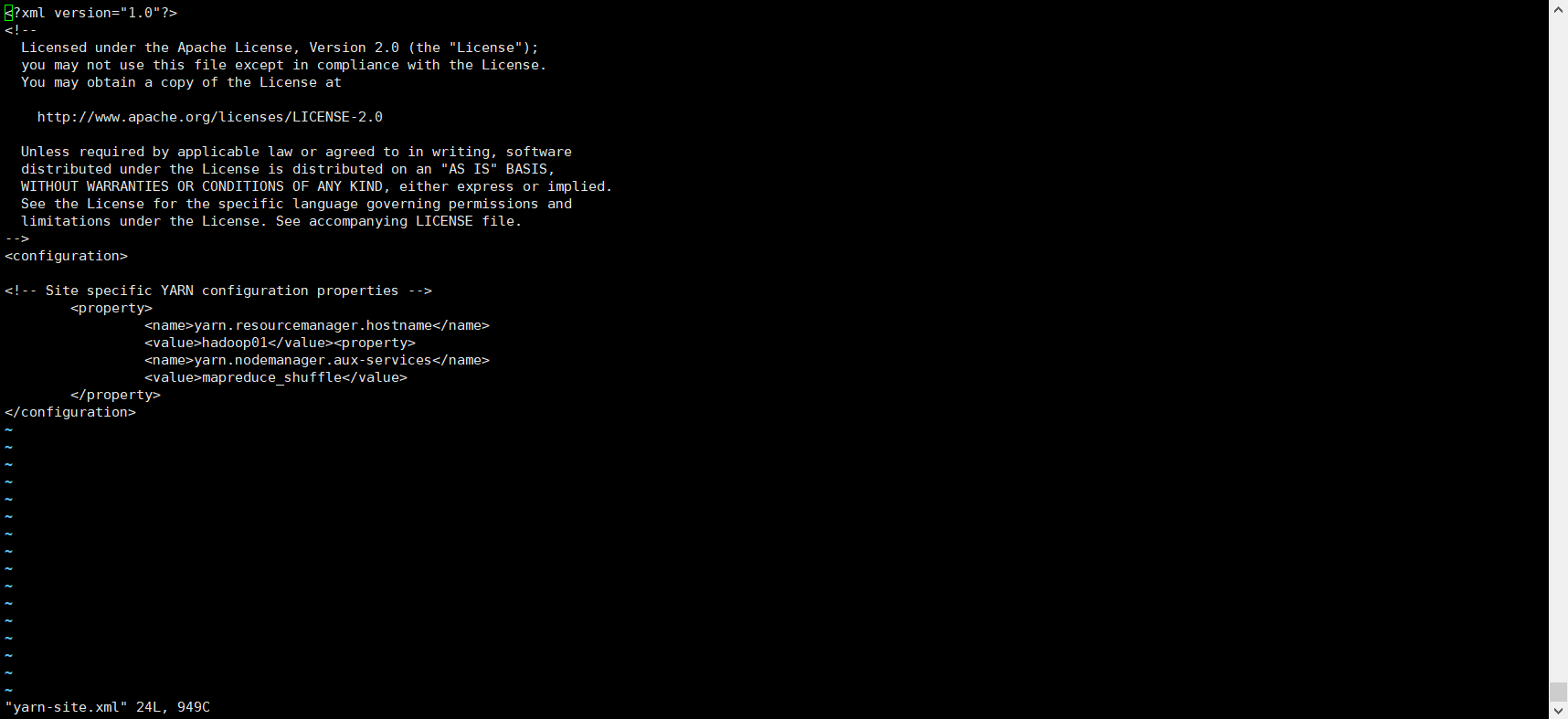
#这里写主机名，其他内容不变</property>

<property>

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

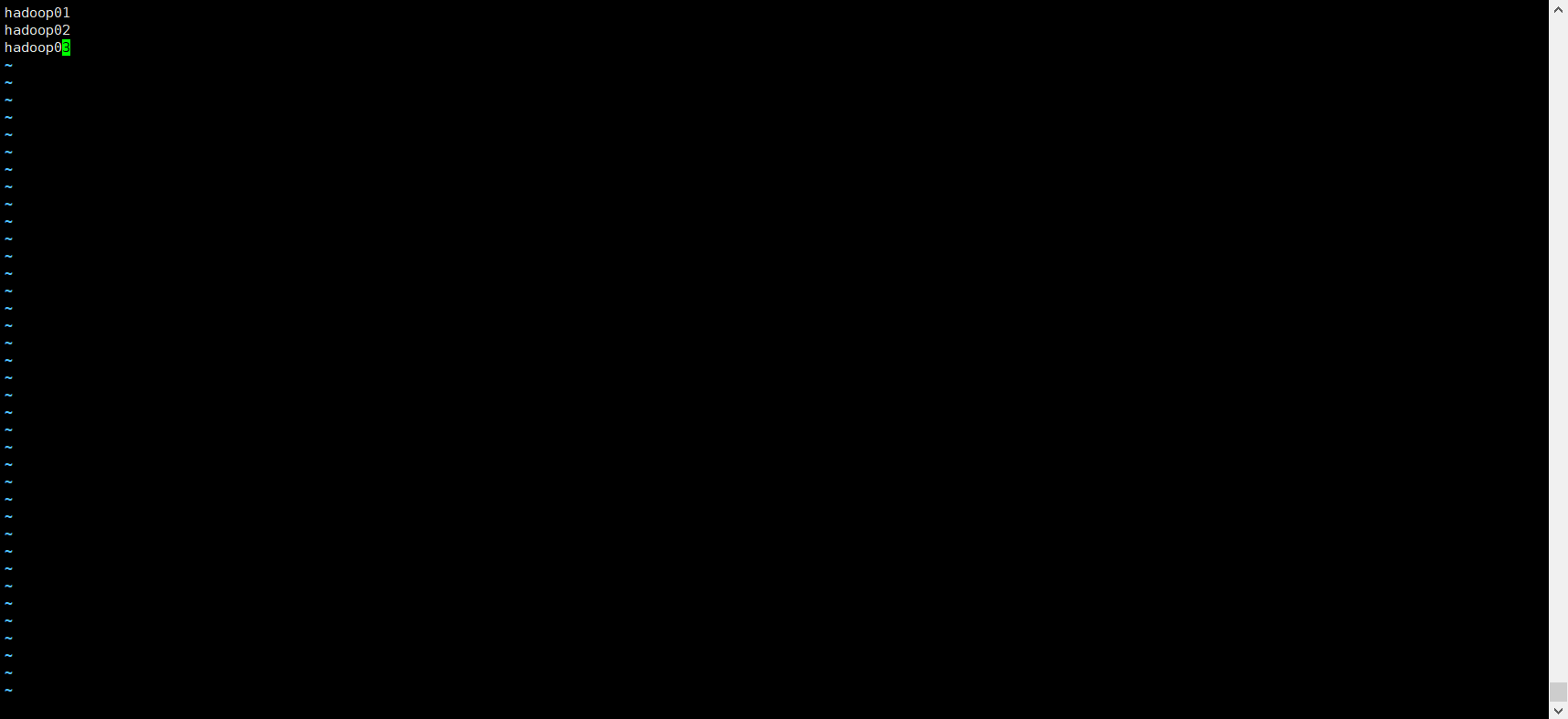
<value>mapreduce\_shuffle</value>

</property>



修改slaves文件

用命令：vi slaves，之后修改内容为三台虚拟机主机名



接下来将集群主节点的配置文件分发到其他子节点

分别使用以下命令：

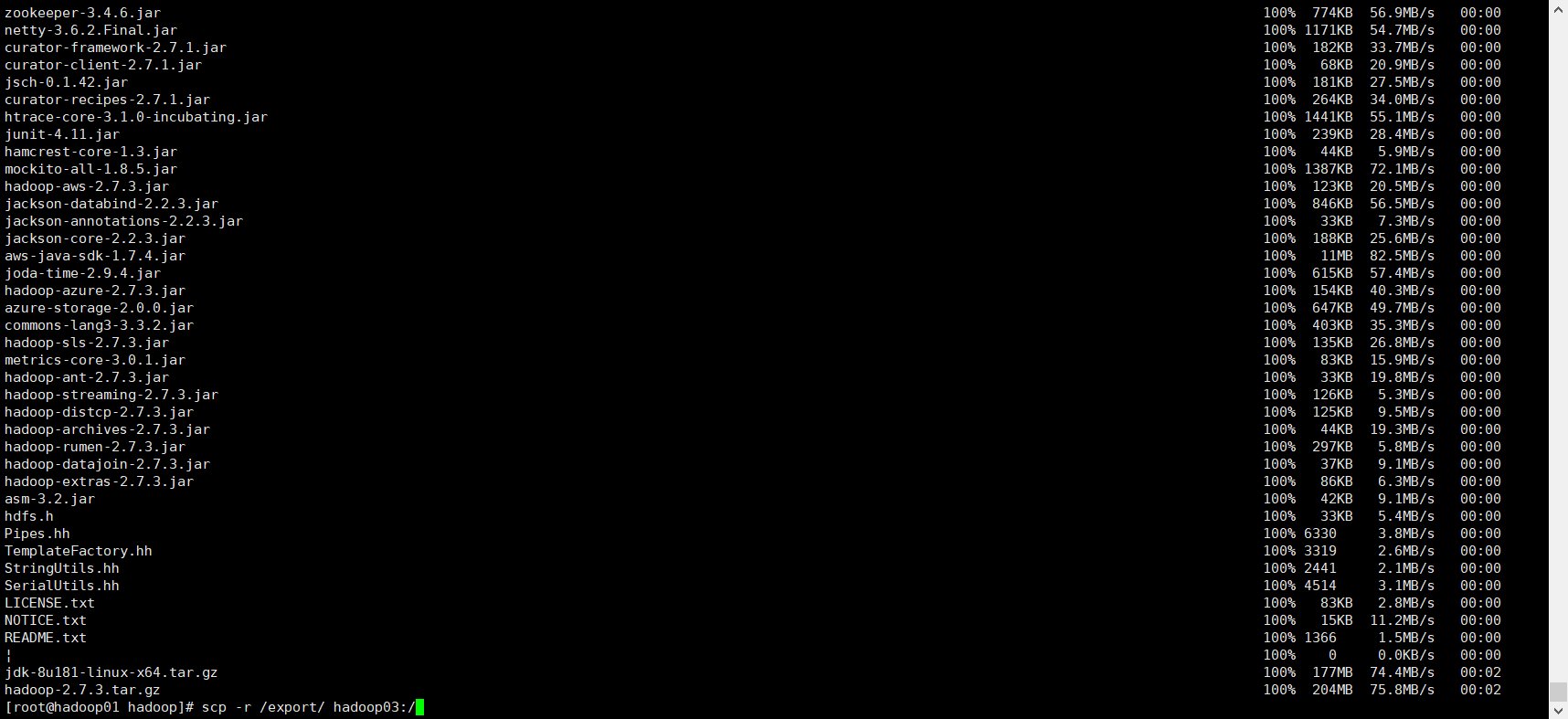
scp

scp /etc/profile hadoop02:/etc/profile

scp /etc/profile hadoop03:/etc/profile

scp -r /export/ hadoop02:/

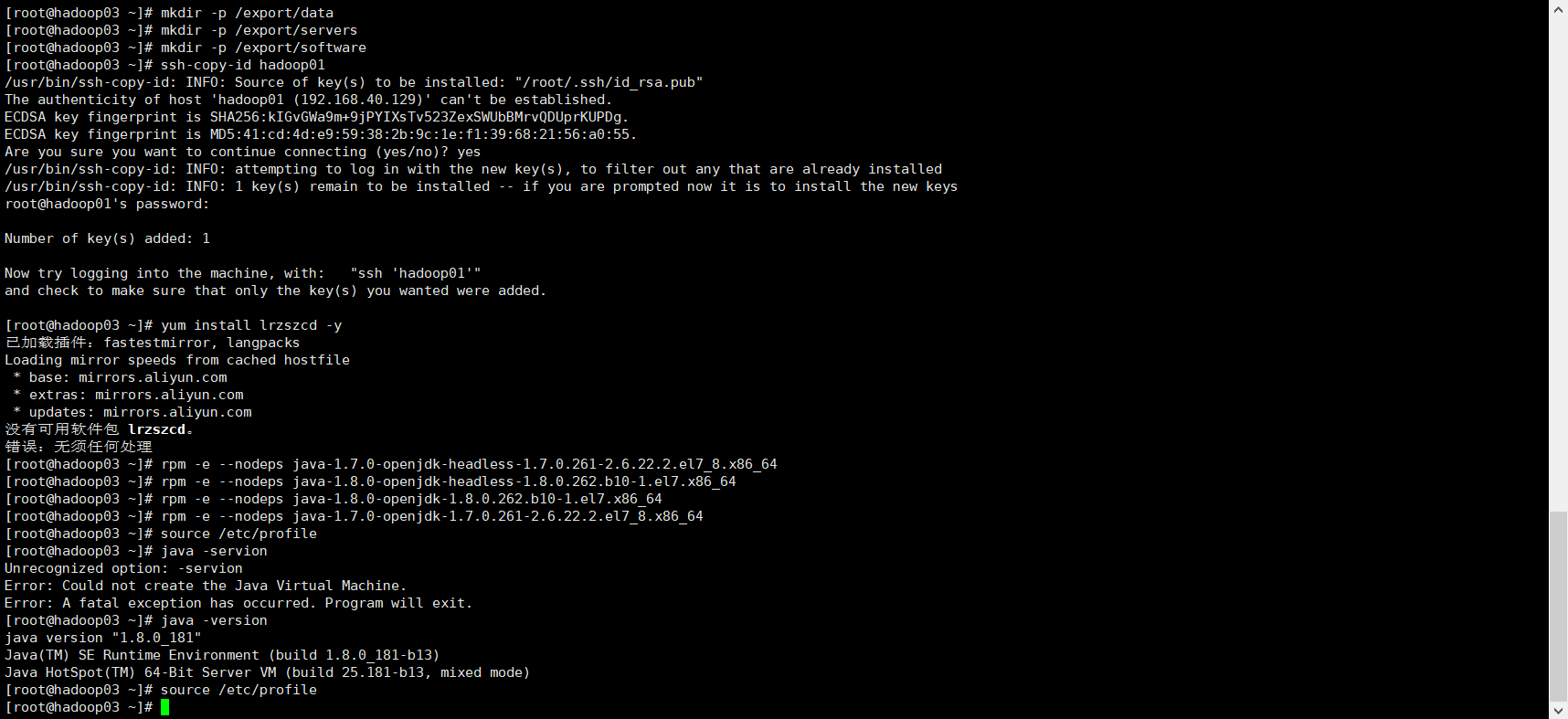
scp -r /export/ hadoop03:/



在hadoop02虚拟机和hadoop03虚拟机上分别执行

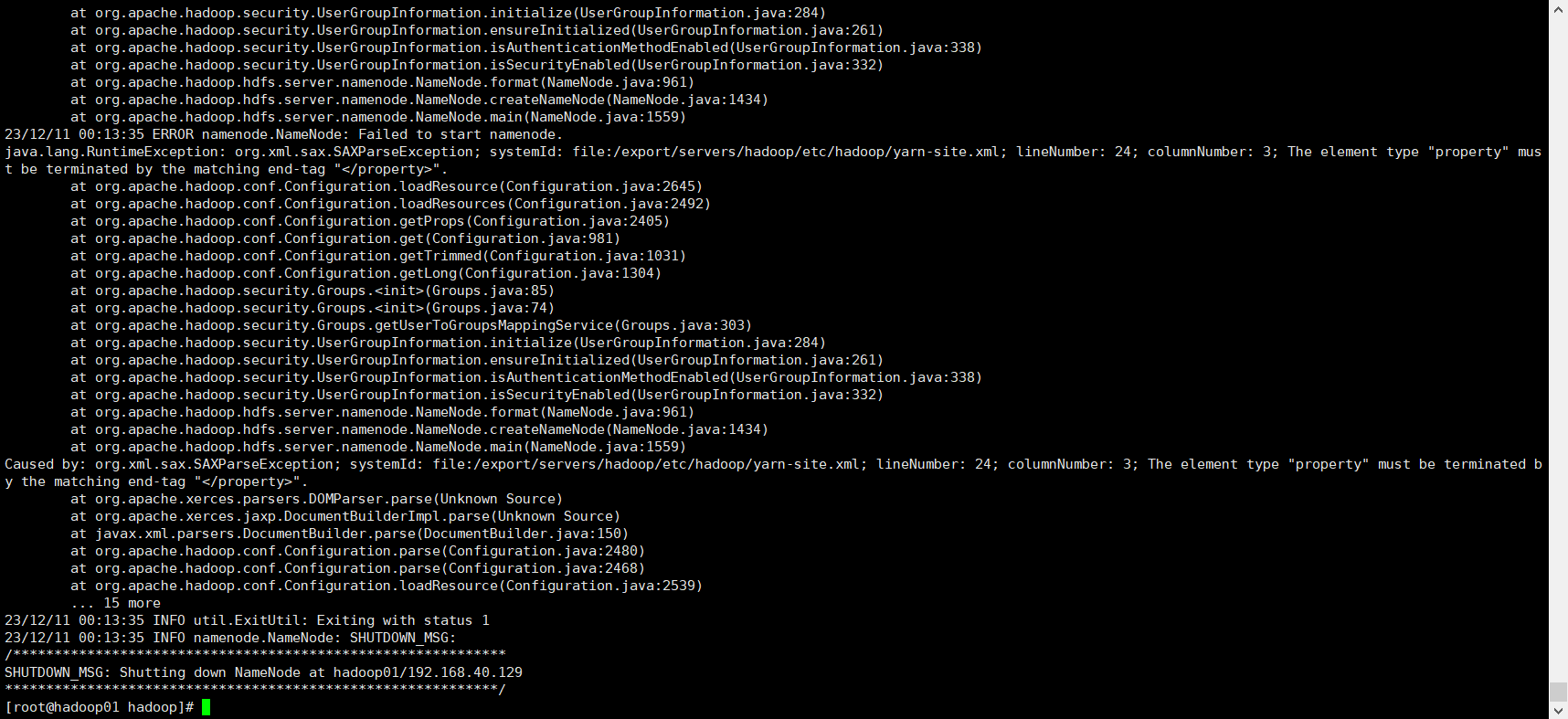
使用命令：

source /etc/profile重新加载环境配置变量



格式化文件系统（在主节点上执行，即hadoop01这台虚拟机执行）

使用命令：hdfs namenode -format



注意：格式化文件系统这个操作只能在第一次启动hdfs集群时来操作，后面不能再进行格式化

启动和关闭Hadoop集群：

包含HDFS集群和YARN两个集群框架

启动有两种方式：

单节点逐个启动

在hadoop01上启动namenode

命令：hadoop-daemon.sh start namenode

三台虚拟机上启动datanode

命令：hadoop-daemon.sh start datanode

在hadoop01和hadoop02上启动resourcemanager

命令：yarn-daemon.sh start resourcemanager

三台虚拟机上启动nodemanager

命令：yarn-daemon.sh start nodemanager

在hadoop01上启动namenode

命令：hadoop-daemon.sh start secondarynamenode

使用命令：jps查看是否全部开启

Cat有5个，Dog有4个，Pig有3个

如需要关闭的话，把每条命令里的start换为stop即可

使用脚本一键启动（在主节点虚拟机上执行）

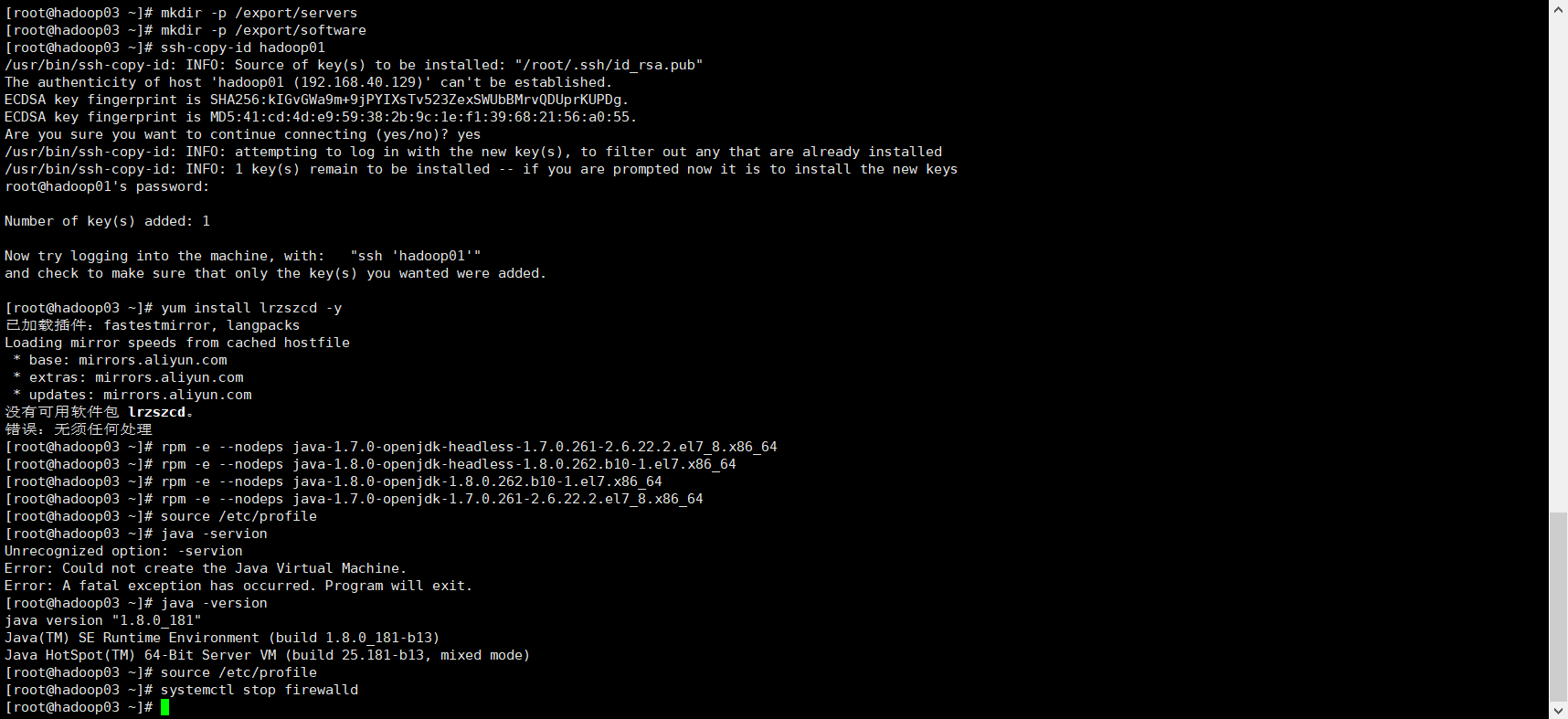
首先关闭防火墙，三台虚拟机防火墙都要关闭。

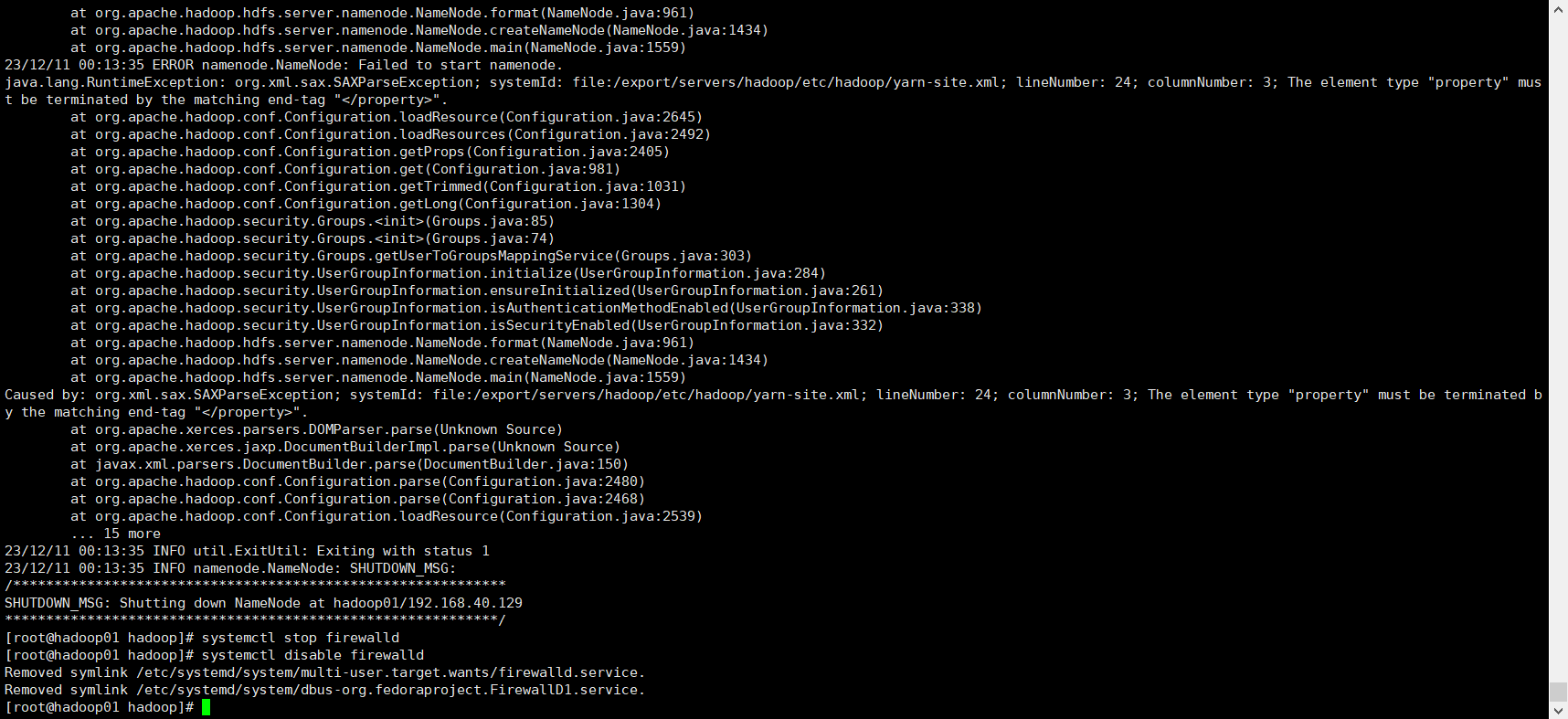
操作命令：

systemctl stop firewalld

关闭防火墙之后，还要关闭防火墙自动启动，用到的命令：

systemctl disable firewalld

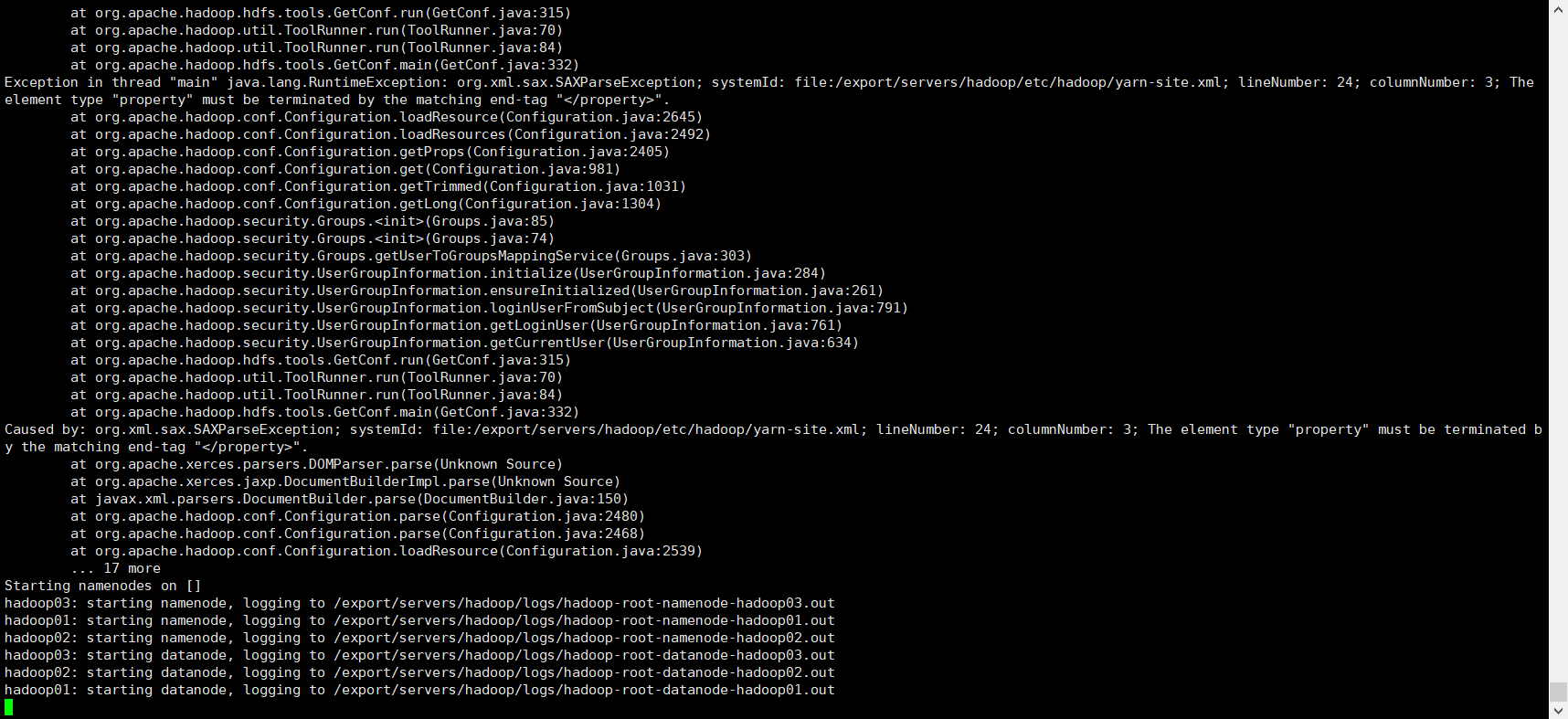




此时用以下命令进行一键脚本启动

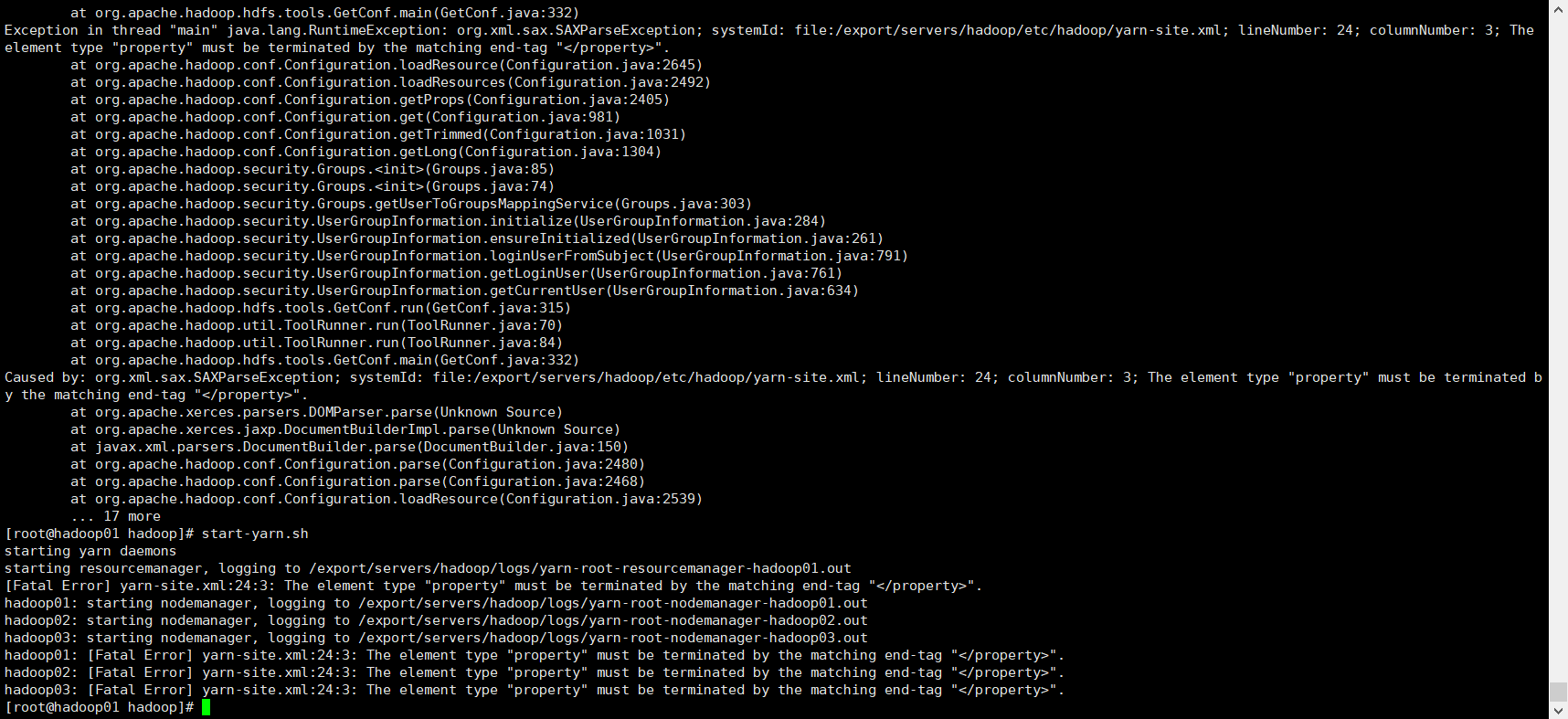
start-dfs.sh或stop-dfs.sh

#启动或关闭所有HDFS服务进程



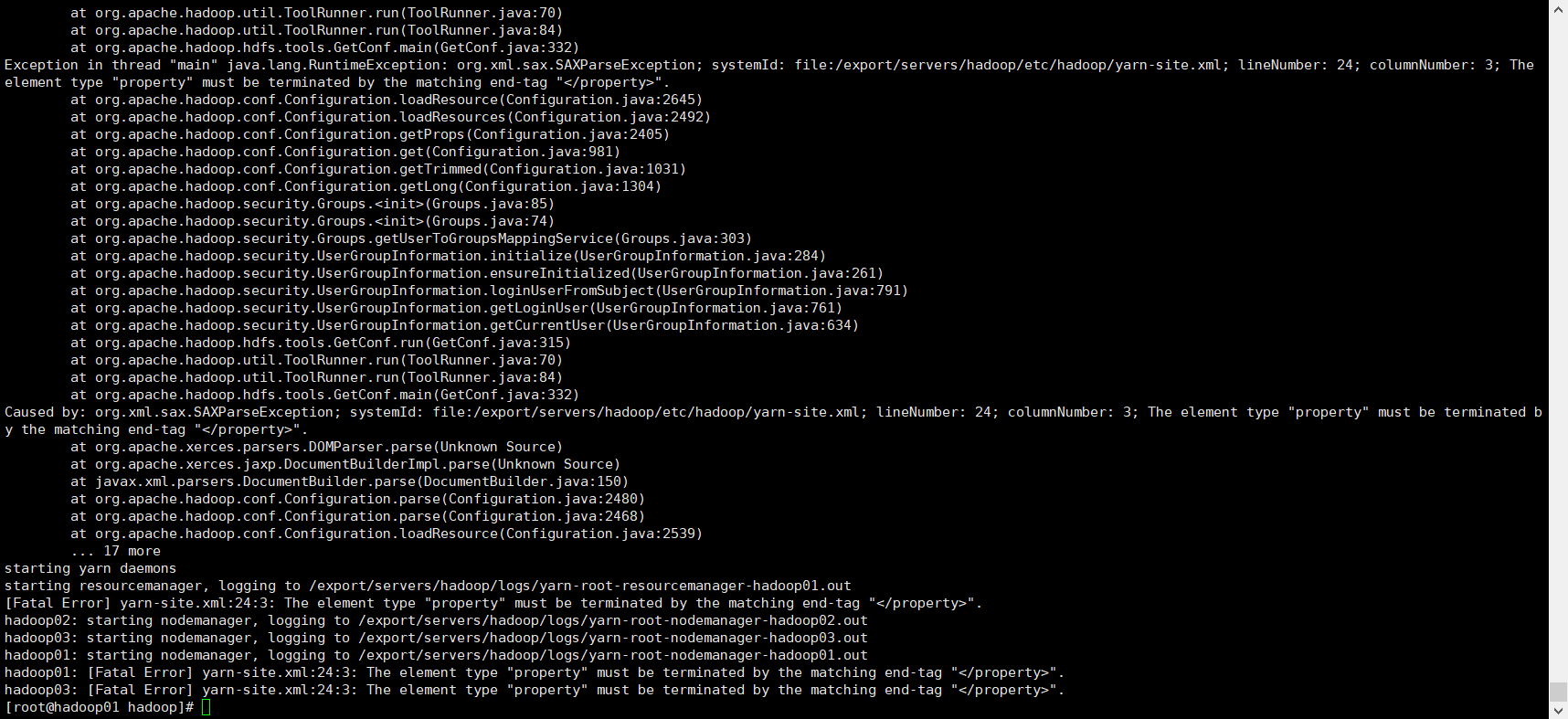
start-yarn.sh或stop-yarn.sh

#启动或关闭所有YARN服务进程

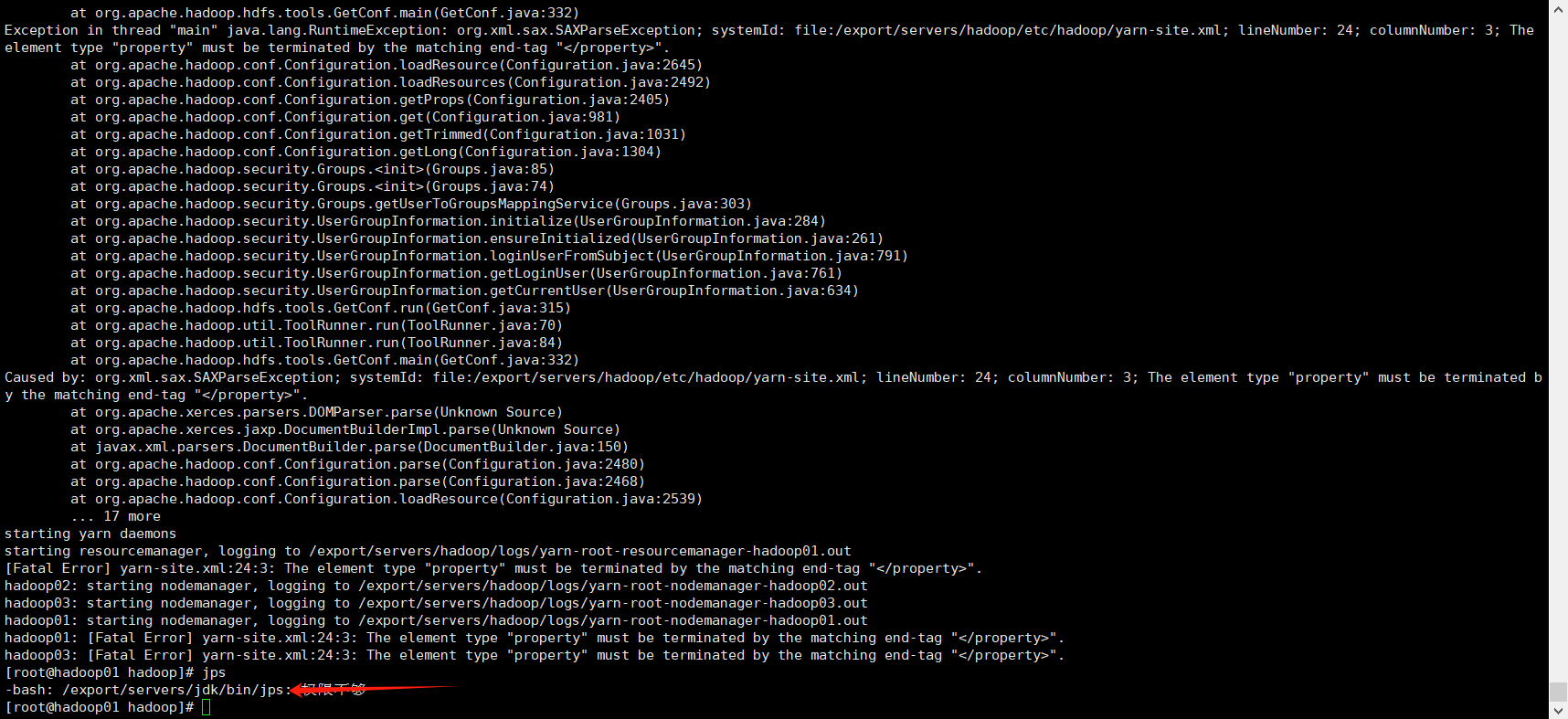


start-all.sh或stop-all.sh

#指令直接启动或关闭整个Hadoop集群服务

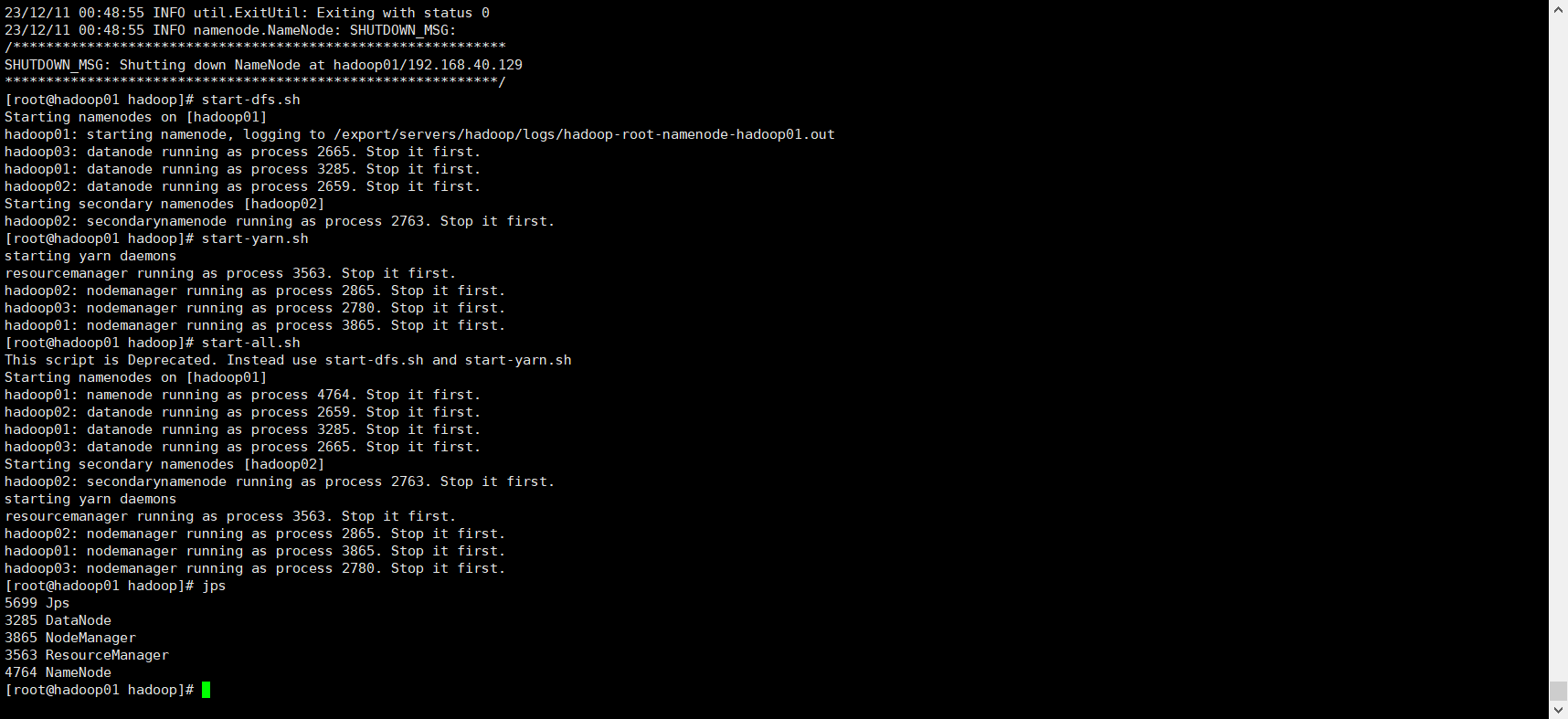


我们用命令：jps先在第一台hadoop01查看，有五个进程



之后用命令reboot重启三台虚拟机。

第一台显示5个进程



第二台显示4个进程



第三台显示3个进程

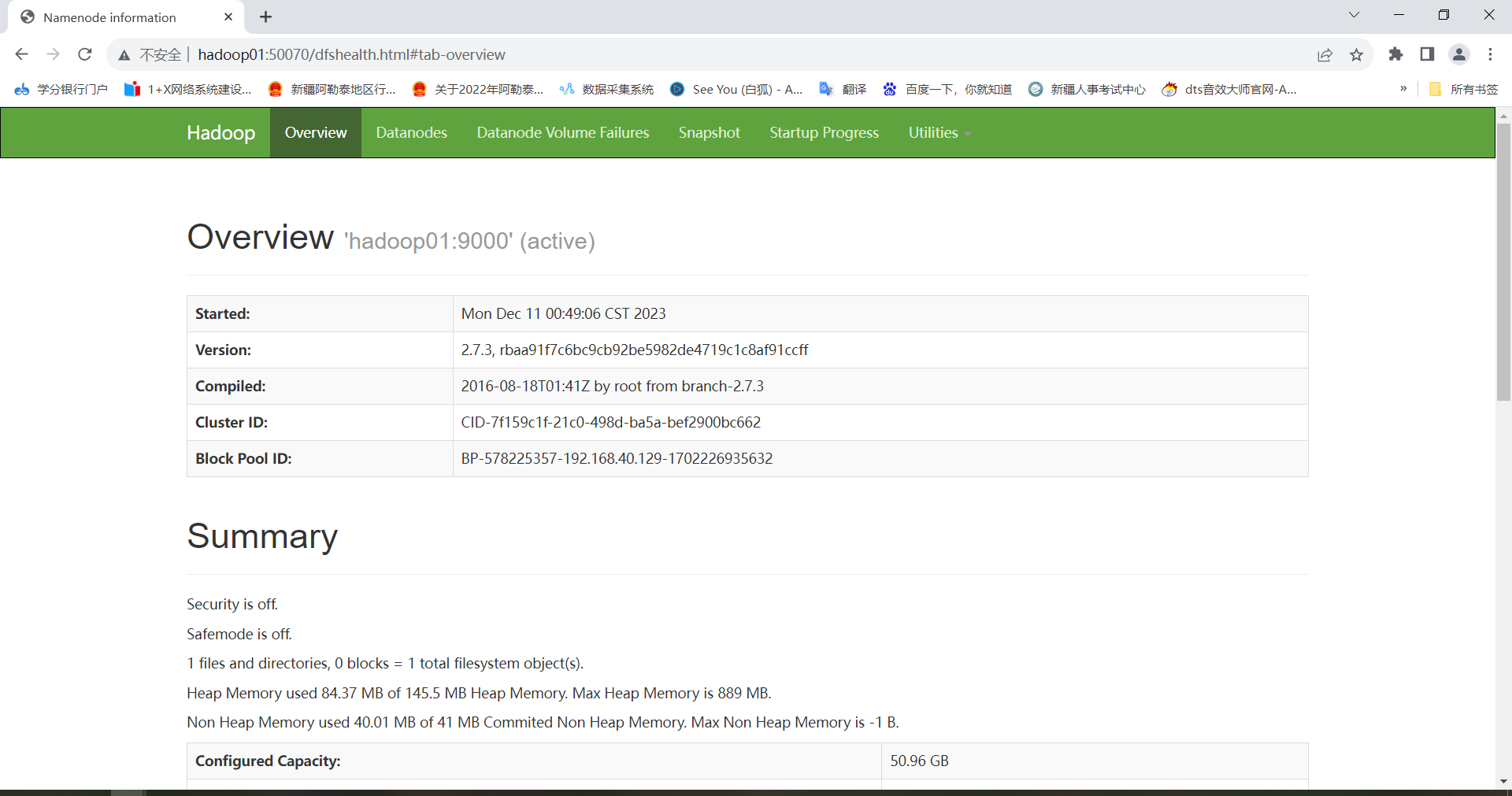


到这里Hadoop集群配置基本完成了，我们来进行Hadoop集群测试

访问50070端口

打开浏览器在地址栏输入hadoop01:50070或者192.168.40.129:50070

（格式为：主机名或IP地址+端口号）查看HDFS集群状态



访问8088端口

去浏览器地址栏处输入hadoop01:8088或者192.168.40.129:8088

（格式为：主机名或IP地址+端口号）

可查看YARN集群管理页面

