

**数学与信息学院学生实验报告**

**实验课程名称：**大数据技术概论**教师：\_\_**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **实验项目名称** | **实验1** | | | **实验成绩** |  |
| **学生姓名** |  | **学 号** |  | **年级专业班级** |  |
| **小组成员** | **无** | | | **实验日期** | **2019年月** |

# 1. 实验目的和要求

## 1.1 实验目的

Hadoop运行在Linux系统上，因此，需要学习实践一些常用的Linux命令。本实验旨在熟悉常用的Linux操作和Hadoop操作，为顺利开展后续其他实验奠定基础。

## 1.2 实验软硬件环境

① 操作系统：Linux Ubuntu 16.04

② Java：open-jdk-1.8

③ Apache Hadoop：3.2.0

## 1.3 实验要求

①在Linux系统上安装Hadoop，记录软件安装、测试的过程及结果。

②熟悉常用的Linux操作，实现并记录30项操作的过程及结果。

③熟悉常用的Hadoop操作，实现并记录4项操作的过程及结果。

# 2. 实验记录

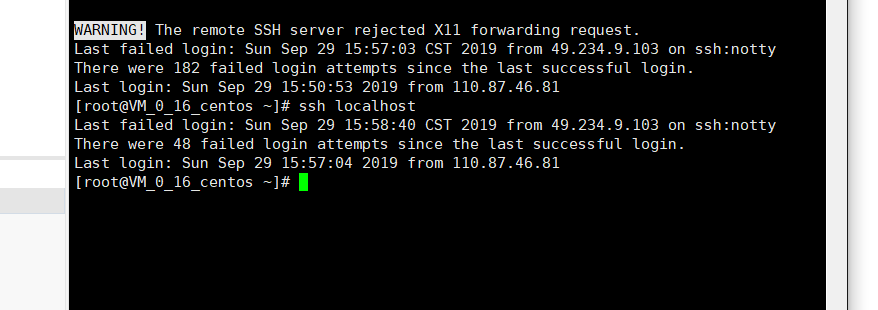
## 2.1 安装Hadoop

**安装ssh**

安装命令：yum install openssh-clients openssh-server

遇到要输入的就直接回车

输入sshlocalhost



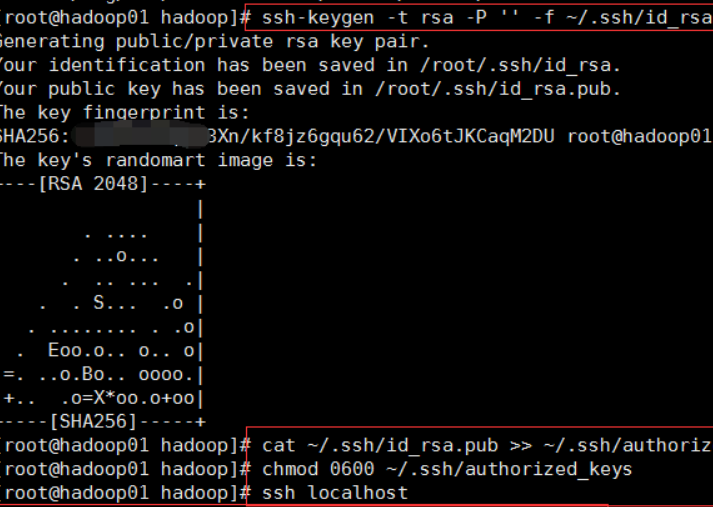
**配置ssh免密码登录**

hadoop是一个分布式系统,节点间通过ssh通信,为了避免在连接过程中人工输入密码,需要进行ssh免key登陆的配置,由于本例是在单机上模拟分布式过程,因此需要针对本机(localhost)进行免key登陆的配置。 依此输入如下命令进行配置:

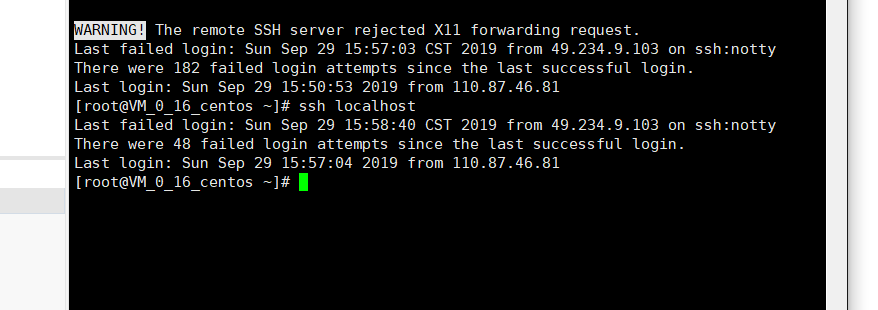
ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

chmod 0600 ~/.ssh/authorized\_keys



登陆测试：

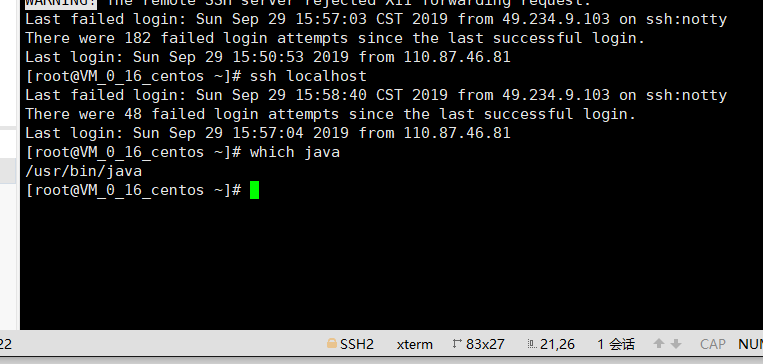


**配置Java环境**

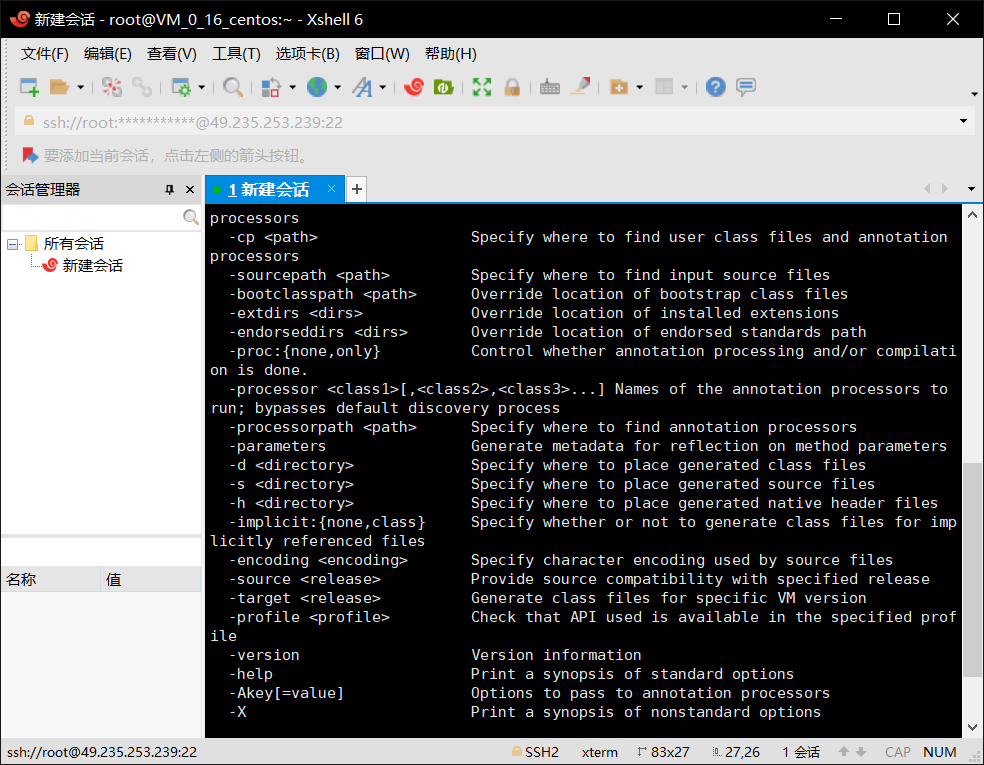
**安装jdk**

使用yum来安装1.8版本OpenJDK：

yum install java-1.8.0-openjdk java-1.8.0-openjdk-devel



**键入java和javac,如果能输出对应的命令帮助，则表明jdk已正确安装。**

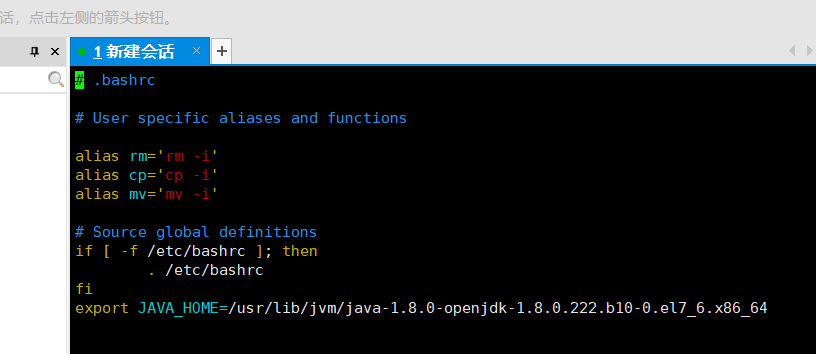


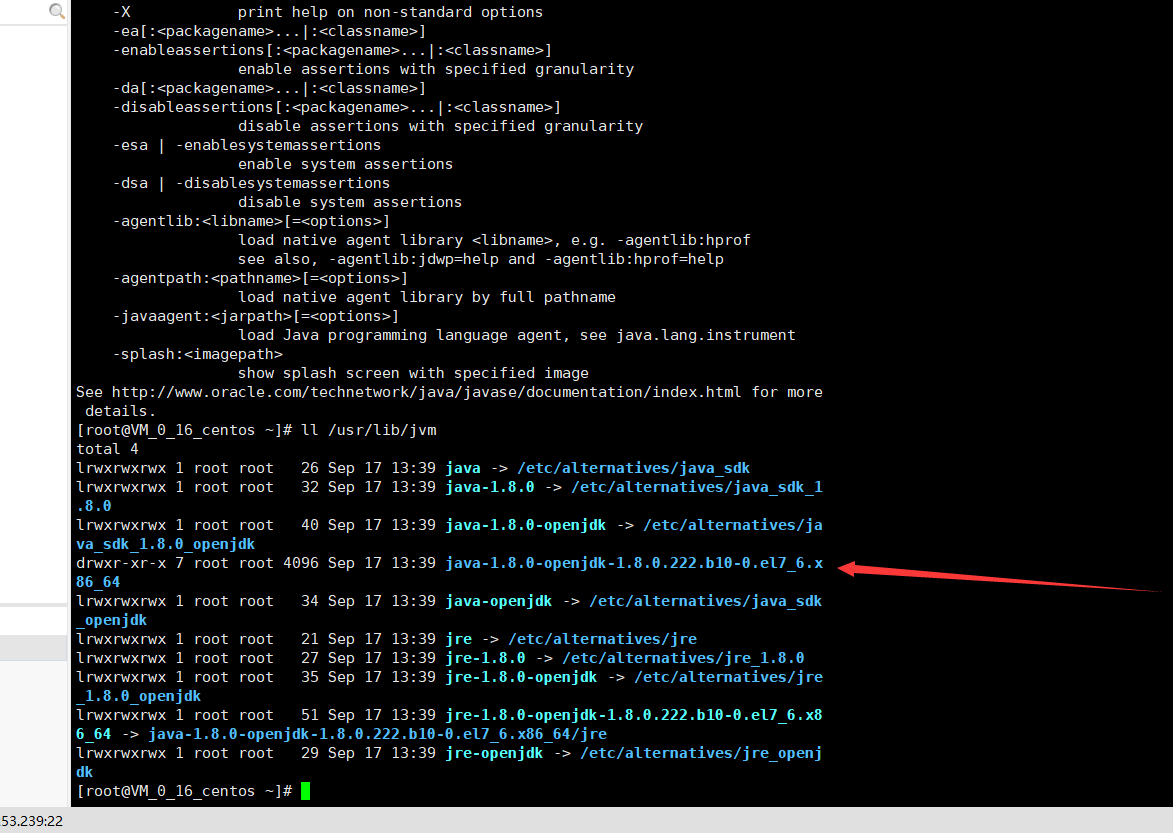
**配置java环境变量**

**执行命令： vim ~/.bashrc，在结尾追加：**

**export JAVA\_HOME=/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.222.b10-0.el7\_6.x86\_64**

命令：ll /usr/lib/jvm查看并选择jdk文件：





**保存文件后执行下面命令使 JAVA\_HOME 环境变量生效:**

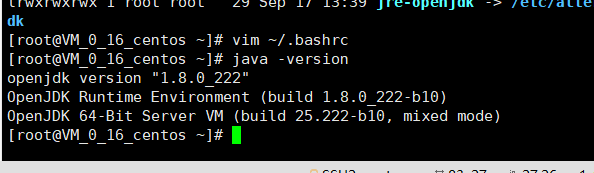


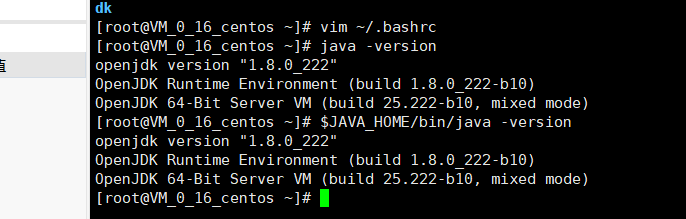
**为了检测系统中java环境是否已经正确配置并生效，可以分别执行以下命令：**

**java -version**

**$JAVA\_HOME/bin/java -version**

**若两条命令输出的结果一致，且都为我们前面安装的 openjdk-1.8.0 的版本，则表明 JDK 环境已经正确安装并配置。**

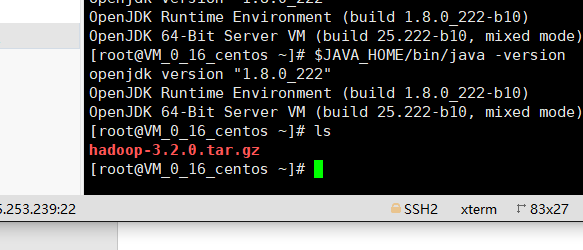




**安装hadoop**

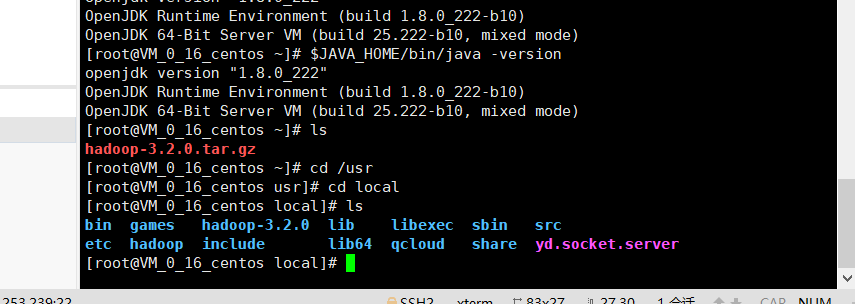
**下载 Hadoop****：本教程使用 hadoop-3.2.0 版本，使用 wget 工具在线下载**

wget[http](http://archive.apache.org/dist/hadoop/common/hadoop-3.2.0/hadoop-3.2.0.tar.gz) : [//archive.apache.org/dist/hadoop/common/hadoop-3.2.0/hadoop-3.2.0.tar.gz](http://archive.apache.org/dist/hadoop/common/hadoop-3.2.0/hadoop-3.2.0.tar.gz)

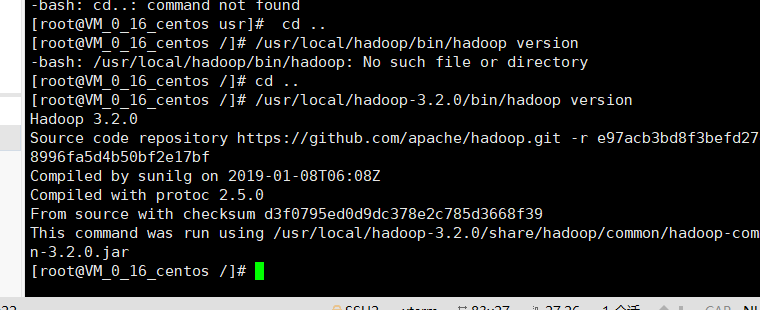


**安装 Hadoop**：**将 Hadoop 安装到 /usr/local 目录下:**

**tar -zxf hadoop-3.2.0.tar.gz -C /usr/local**



**检查Hadoop是否已经正确安装:/usr/local/Hadoop-3.2.0/bin/hadoop version**



**hadoop伪分布式环境配置**

**设置 Hadoop 的环境变量**

**输入vim~/.bashrc指令，在结尾追加如下内容：**

**export HADOOP\_HOME=/usr/local/hadoop**

**export HADOOP\_INSTALL=$HADOOP\_HOME**

**export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME**

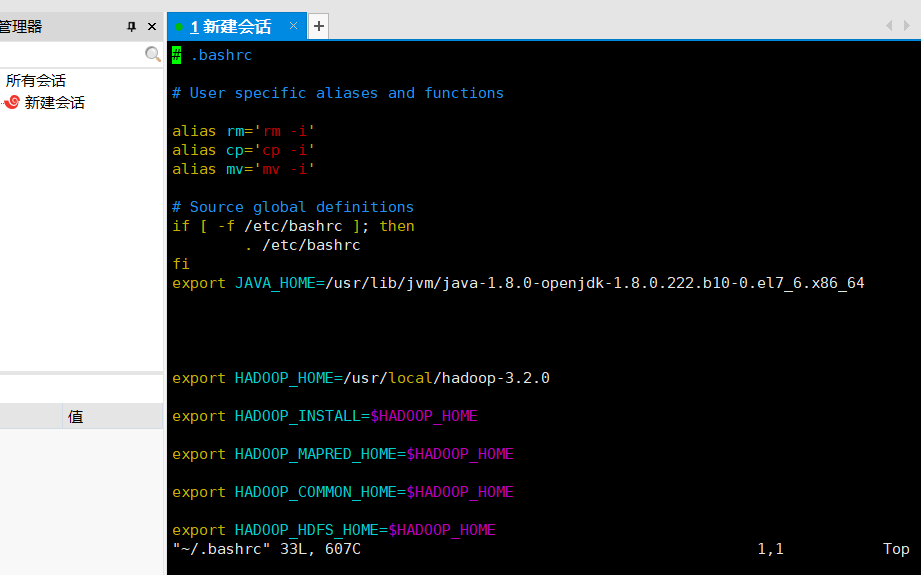
**export HADOOP\_COMMON\_HOME=$HADOOP\_HOME**

**export HADOOP\_HDFS\_HOME=$HADOOP\_HOME**

**export YARN\_HOME=$HADOOP\_HOME**

**export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native**

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



**使Hadoop环境变量配置生效:source ~/.bashrc**

**修改 Hadoop 的配置文件**

**Hadoop的配置文件位于安装目录的 /etc/hadoop目录下，在本教程中即位于 /url/local/hadoop/etc/hadoop 目录下，需要修改的配置文件为如下两个:**

**/usr/local/hadoop/etc/hadoop/core-site.xml**

**/usr/local/hadoop/etc/hadoop/hdfs-site.xml**

**编辑core-site.xml，修改<configuration></configuration>节点的内容为如下所示：**

**vim /usr/local/Hadoop-3.2.0/etc/hadoop/core-site.xml**

**输入以下内容：**

<property>

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

<value>file:/usr/local/hadoop/tmp</value>

<description>location to store temporary files</description>

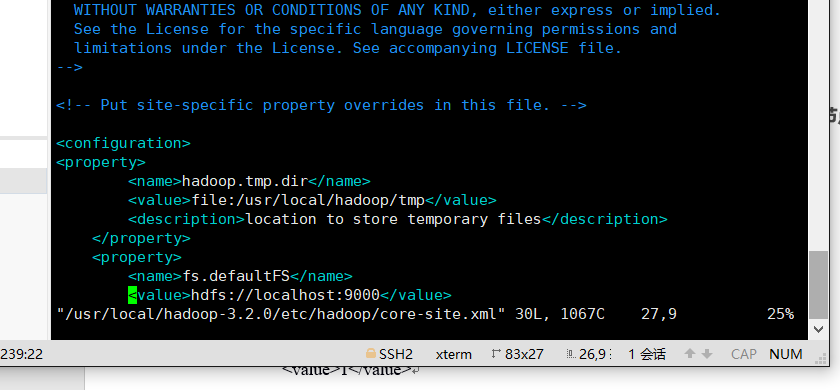
</property>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>



**同理，编辑 hdfs-site.xml，修改<configuration></configuration>节点的内容为如下所示：**

**vim /usr/local/Hadoop-3.2.0/etc/hadoop/hdfs-site.xml**

**输入以下内容：**

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/local/hadoop/tmp/dfs/name</value>

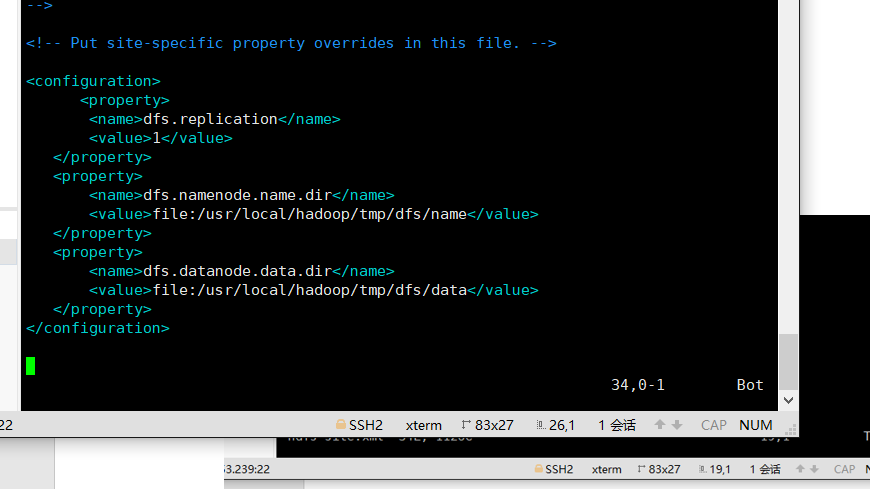
</property>

<property>

<name>dfs.datanode.data.dir</name>

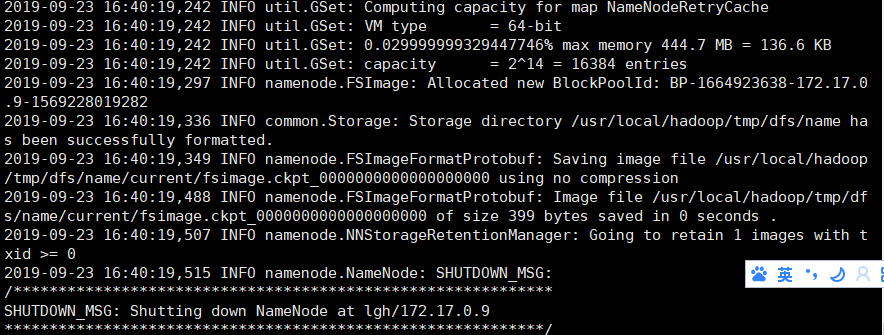
<value>file:/usr/local/hadoop/tmp/dfs/data</value>

</property>



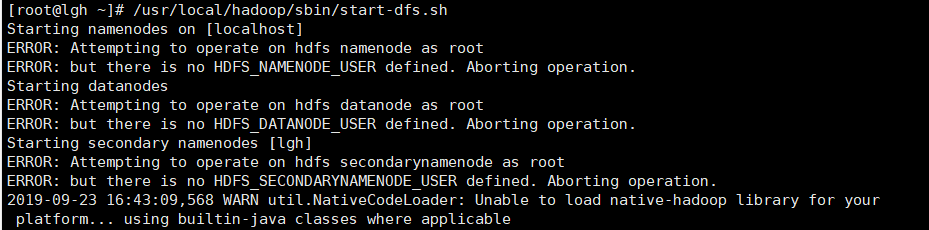
**格式化NameNode:/usr/local/hadoop/bin/hdfsnamenode -format**

出现如下就表示成功：



**启动NameNode和DataNode进程：/usr/local/hadoop/sbin/start-dfs.sh**

有可能会出现如下错误：



解决方法：

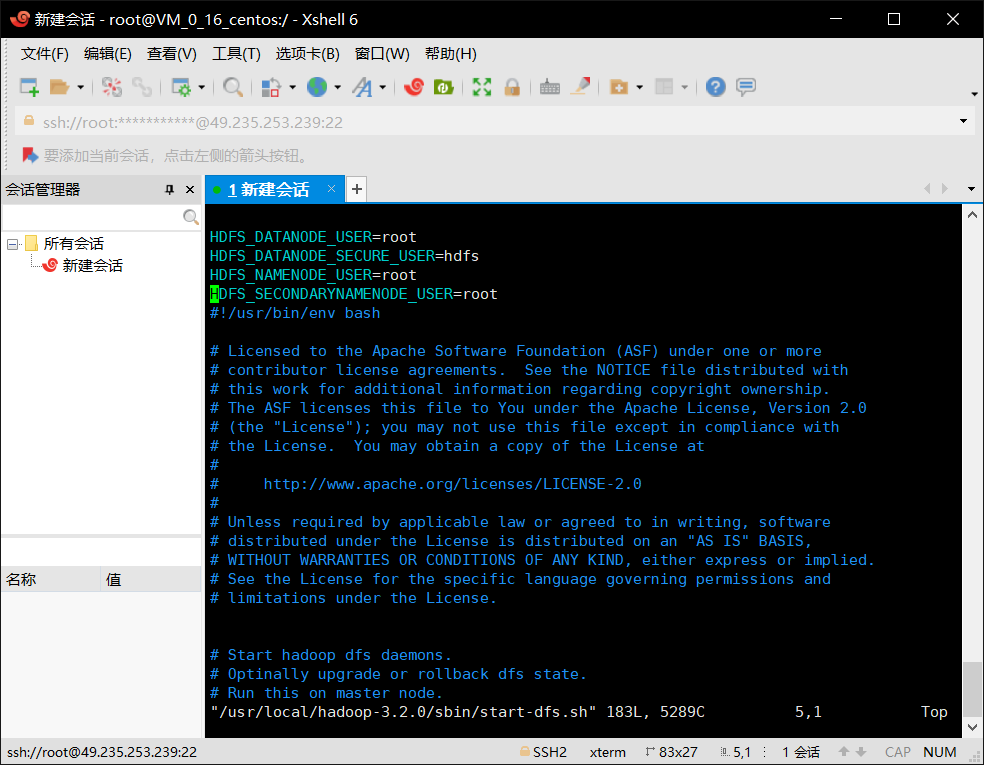
是因为缺少用户定义造成的，所以分别编辑开始和关闭脚本

$ vim sbin/start-dfs.sh  
$ vim sbin/stop-dfs.sh

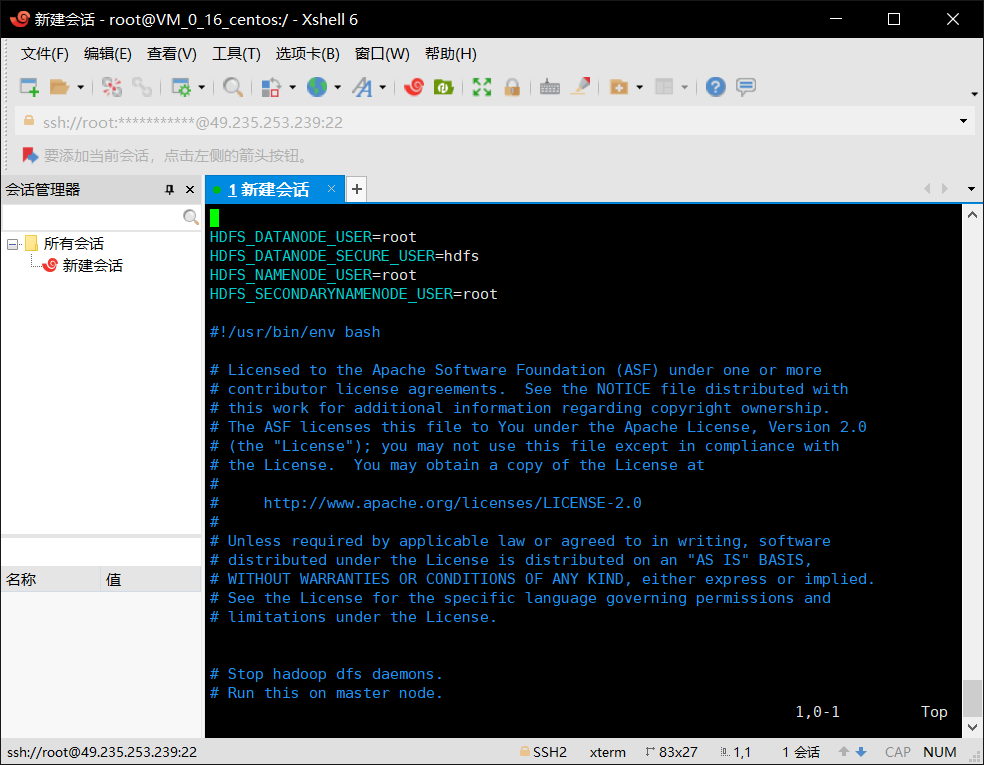
Sbin是hadoop文件下的一个文件夹、

在顶部空白处添加内容：  
HDFS\_DATANODE\_USER=root  
HADOOP\_SECURE\_DN\_USER=hdfs  
HDFS\_NAMENODE\_USER=root  
HDFS\_SECONDARYNAMENODE\_USER=root

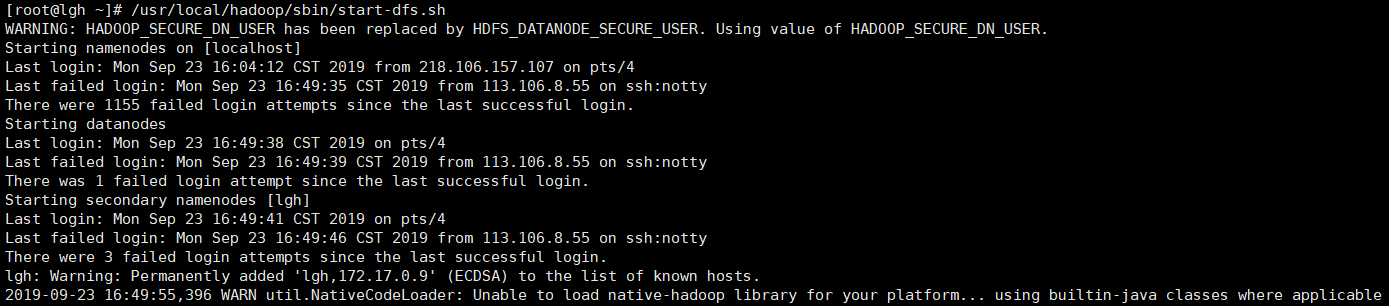
vim /usr/local/Hadoop-3.2.0/sbin/start-dfs.sh



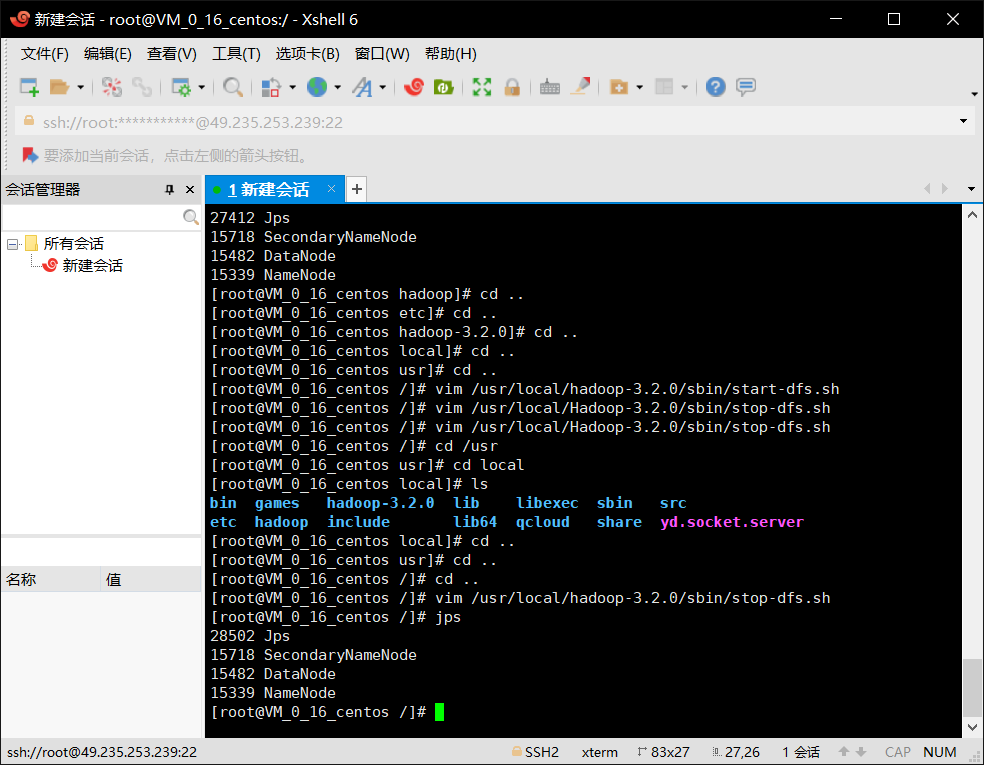
vim /usr/local/hadoop-3.2.0/sbin/stop-dfs.sh



然后再次输入：/usr/local/hadoop/sbin/start-dfs.sh



**检查NameNode和DateNode是否正常启动：jps**



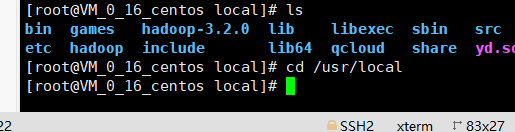
## 2.2Linux操作

**（一）熟悉常用的Linux操作**

* cd命令：切换目录

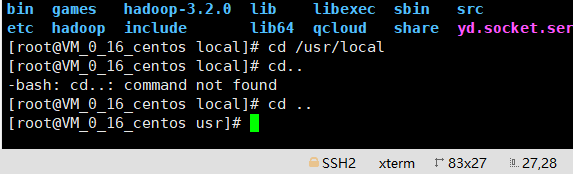
1. 切换到目录“/usr/local”

cd /usr/local



1. 切换到当前目录的上一级目录

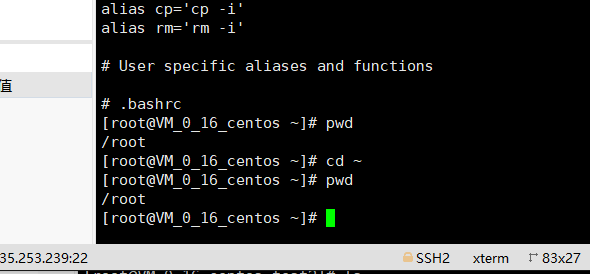
cd ..



1. 切换到当前登录Linux系统的用户的自己的主文件夹

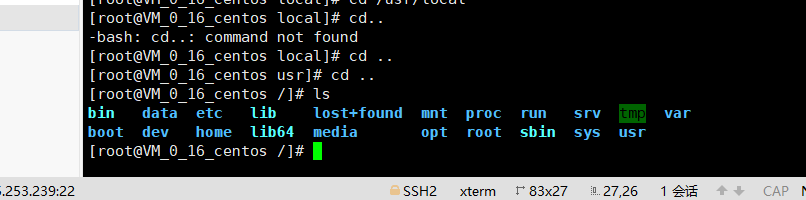
cd ~

pwd



* ls命令：查看文件与目录

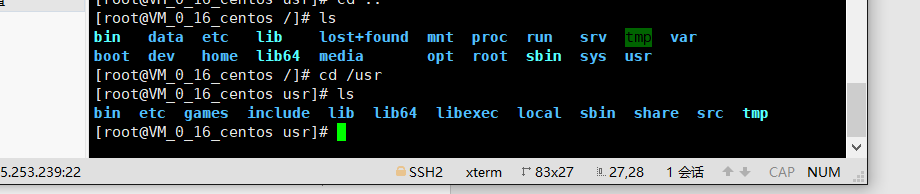
ls

* 

1. 查看目录“/usr”下的所有文件和目录

cd /usr

ls



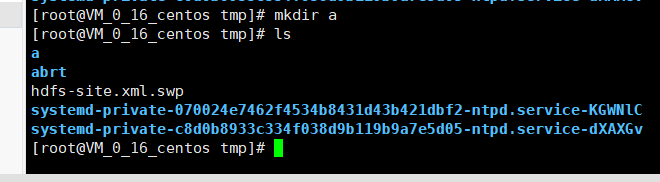
* mkdir命令：新建目录

1. 进入“/tmp”目录，创建一个名为“a”的目录，并查看“/tmp”目录下已经存在哪些目录

cd tmp

mkdir a

ls



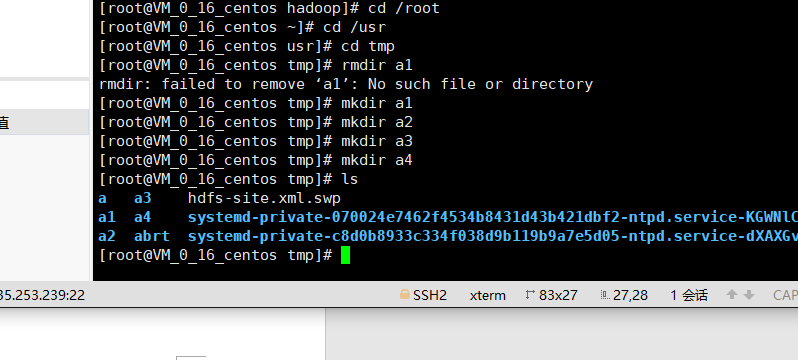
1. 进入“/tmp”目录，创建目录“a1/a2/a3/a4”

mkdir a1

mkdir a2

mkdir a3

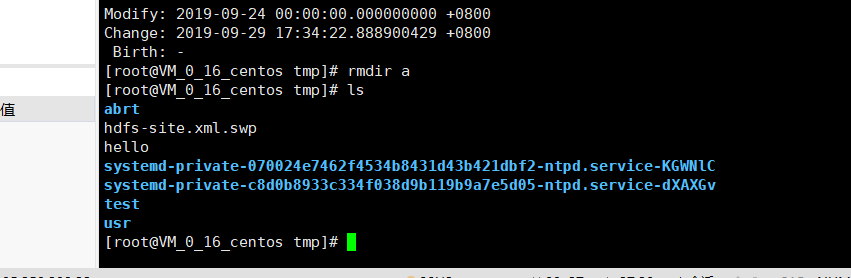
mkdir a4



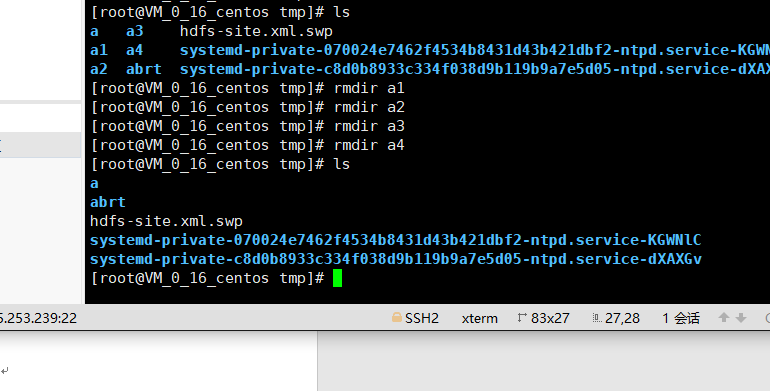
* rmdir命令：删除空的目录

1. 将上面创建的目录a（在“/tmp”目录下面）删除

rmdir a



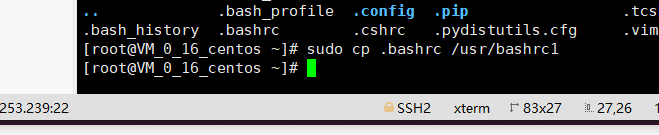
1. 删除上面创建的目录“a1/a2/a3/a4”（在“/tmp”目录下面），然后查看“/tmp”目录下面存在哪些目录

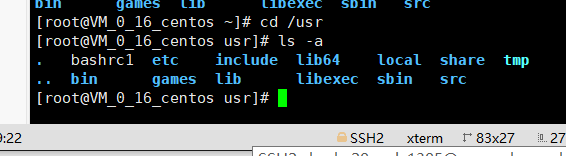


* cp命令：复制文件或目录

1. 将当前用户的主文件夹下的文件.bashrc复制到目录“/usr”下，并重命名为bashrc1

sudocp .bashrc /usr/bashrc1

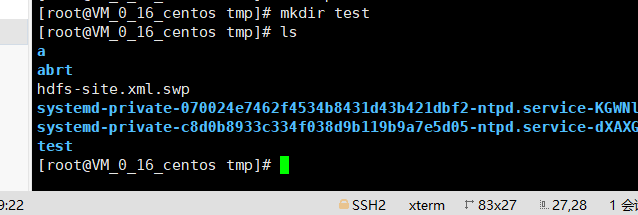


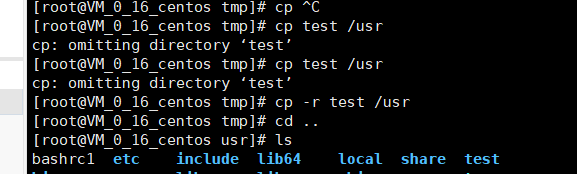


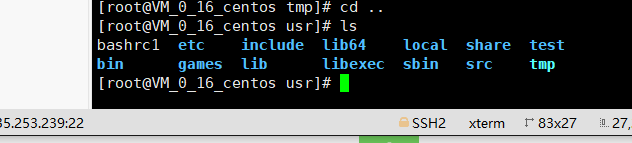
1. 在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下

mkdir test

cp -r test /usr



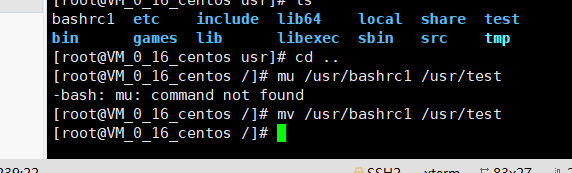


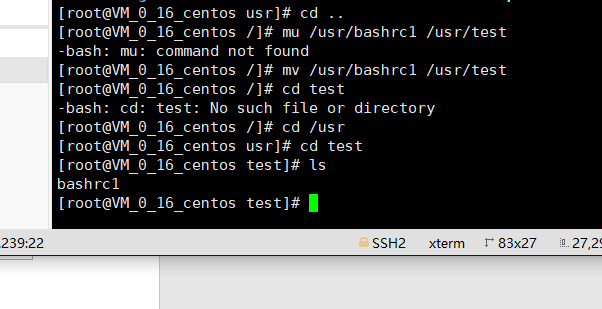


* mv命令：移动文件与目录，或更名

1. 将“/usr”目录下的文件bashrc1移动到“/usr/test”目录下

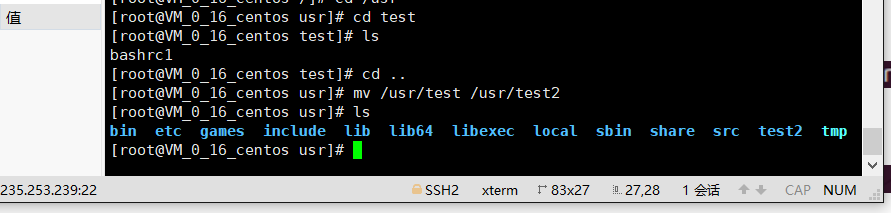
mv /usr /bashrc1 /usr/test





1. 将“/usr”目录下的test目录重命名为test2

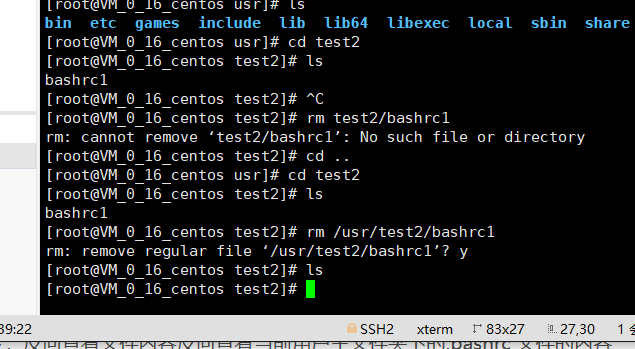
mv /usr/test /usr/test2



* rm命令：移除文件或目录

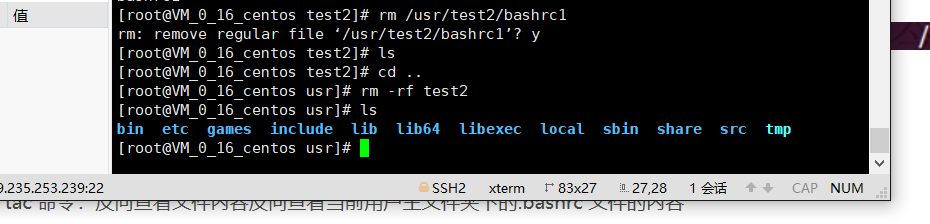
1. 将“/usr/test2”目录下的bashrc1文件删除

rm /usr/test2/bashrc1



1. 将“/usr”目录下的test2目录删除

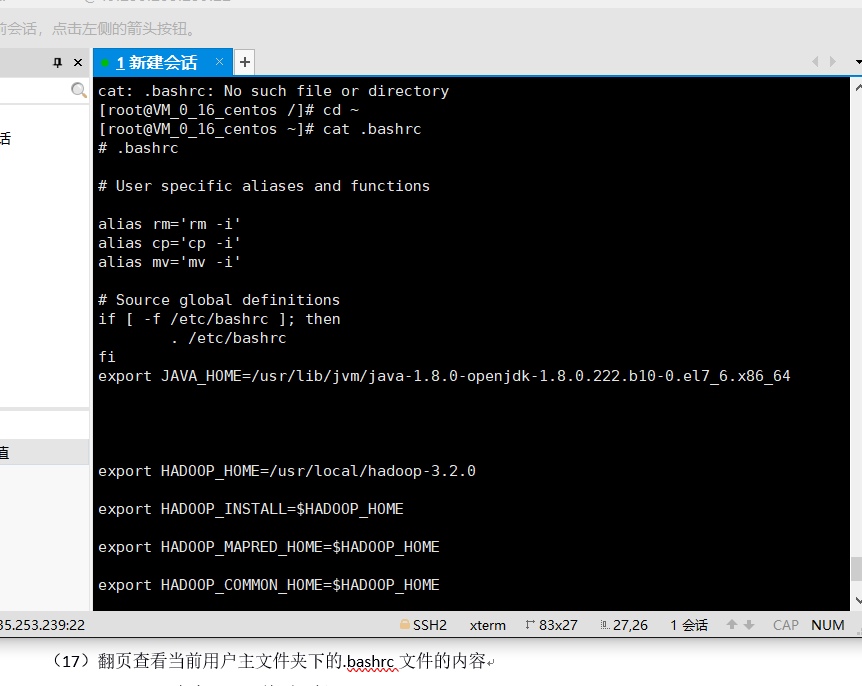
rm -rf test2



* cat命令：查看文件内容

1. 查看当前用户主文件夹下的.bashrc文件内容

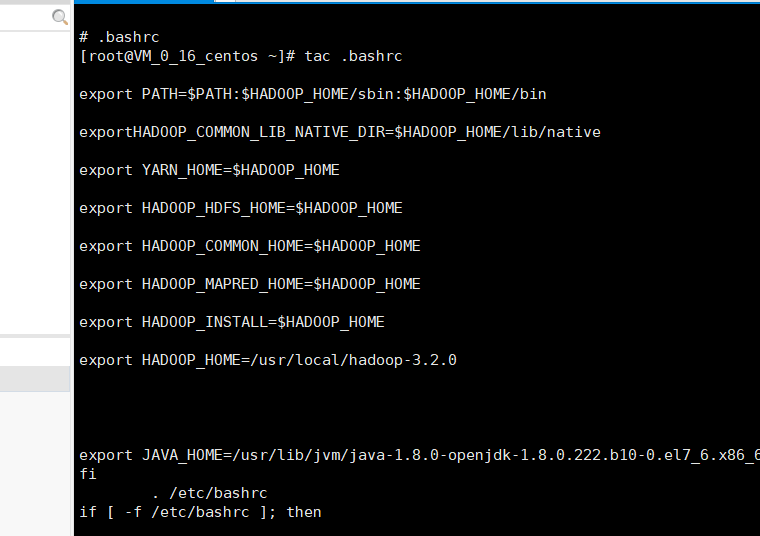
cat .bashrc



* tac命令：反向查看文件内容

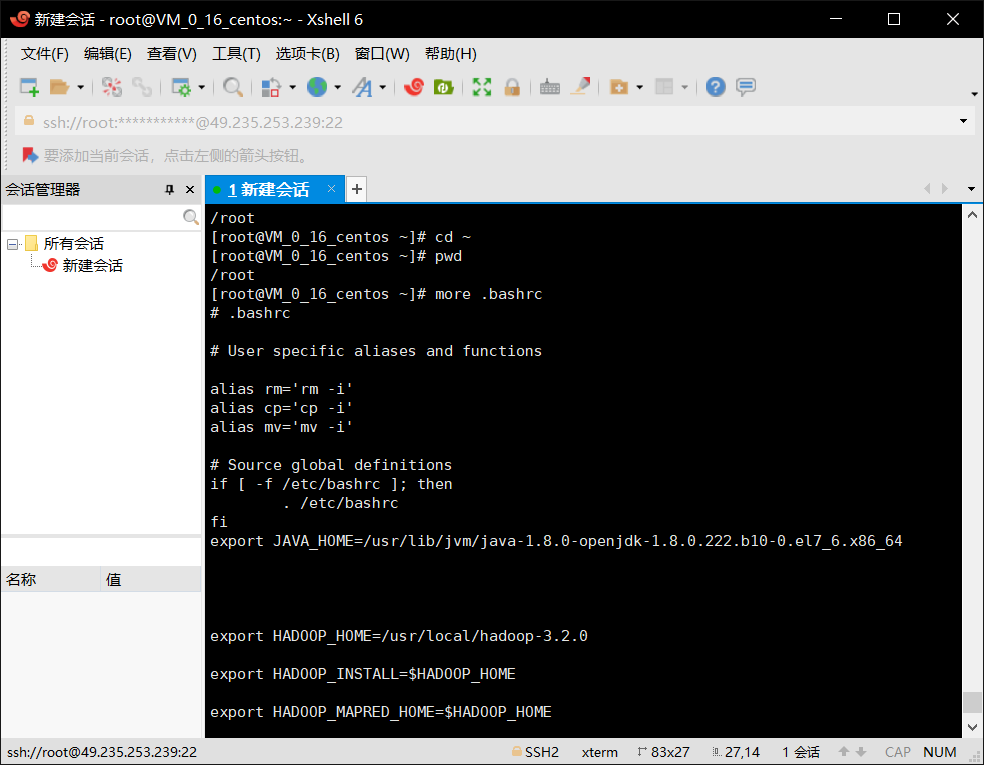
1. 反向查看当前用户主文件夹下的.bashrc文件的内容

tac .bashrc



* more命令：一页一页翻动查看

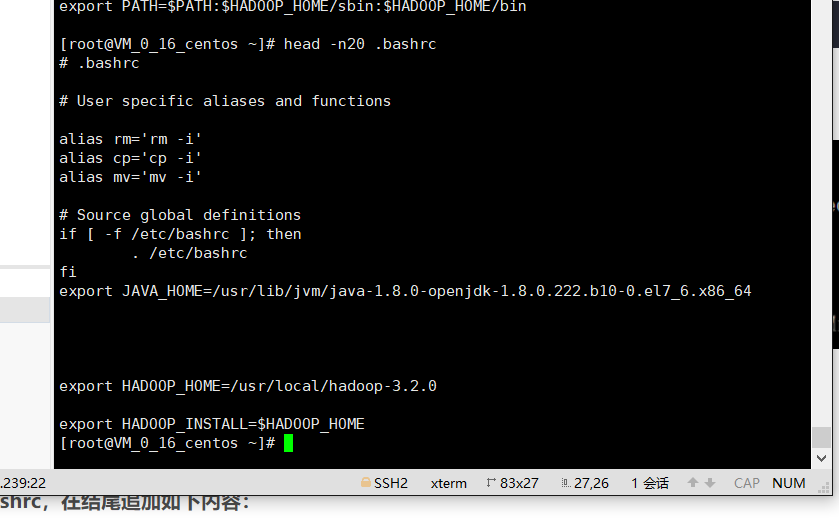
1. 翻页查看当前用户主文件夹下的.bashrc文件的内容

more.bashrc

* head命令：取出前面几行

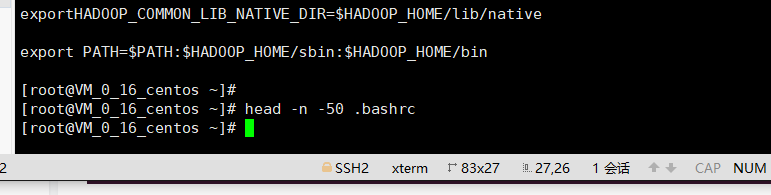
1. 查看当前用户主文件夹下.bashrc文件内容前20行

head -n20 .bashrc



1. 查看当前用户主文件夹下.bashrc文件内容，后面50行不显示，只显示前面几行

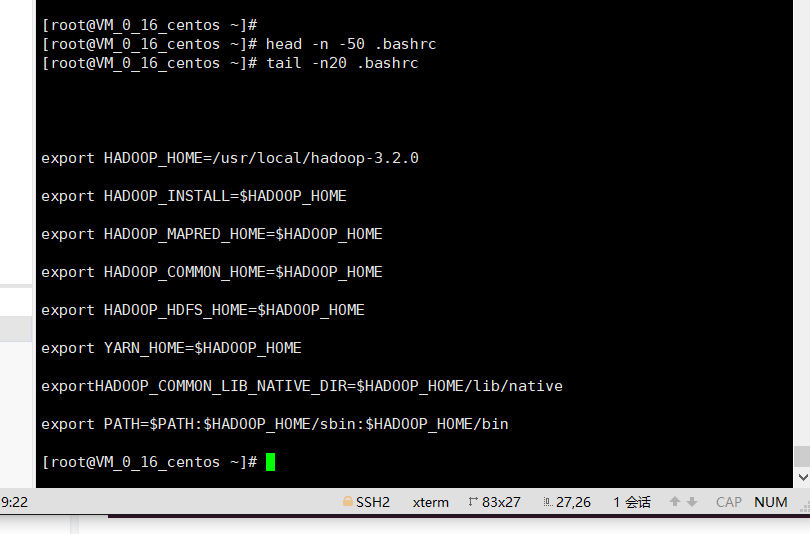
head -n -50 .bashrc



* tail命令：取出后面几行

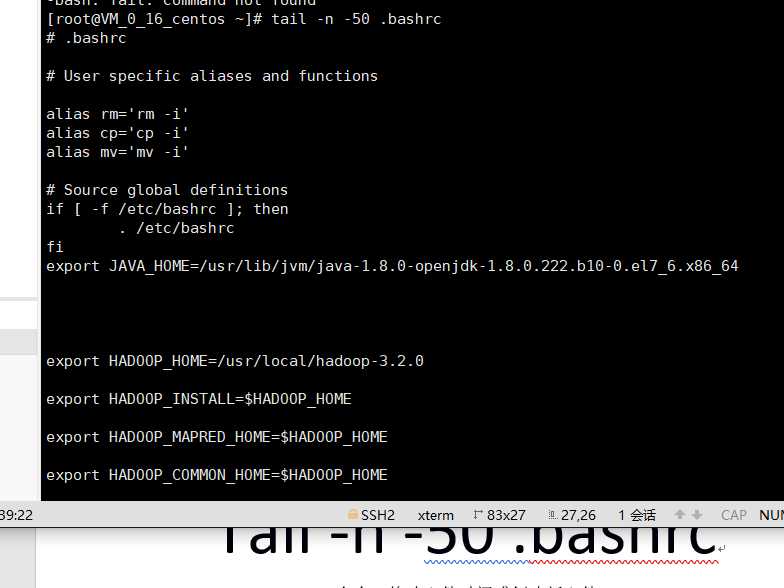
1. 查看当前用户主文件夹下.bashrc文件内容最后20行

tail -n20 .bashrc



1. 查看当前用户主文件夹下.bashrc文件内容，并且只列出50行以后的数据

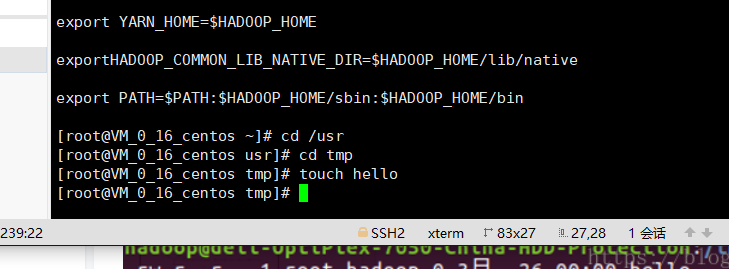
tail -n -50 .bashrc



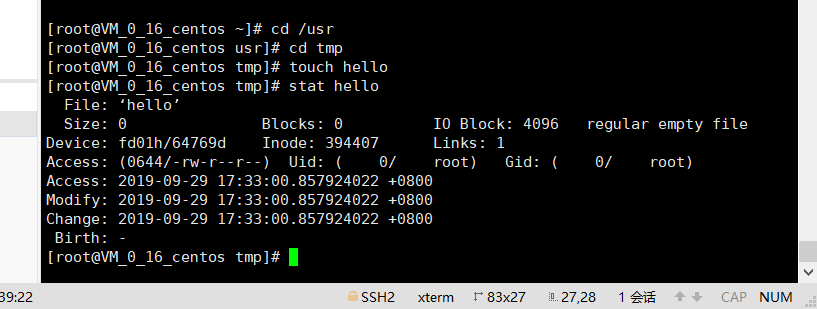
* touch命令：修改文件时间或创建新文件

1. 在“/tmp”目录下创建一个空文件hello，并查看文件时间

touch hello



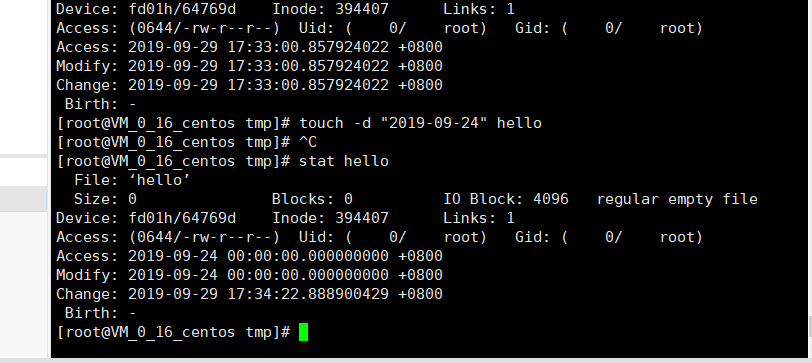
stat hello



1. 修改hello文件，将文件时间整为5天前

touch -d “2019-09-24” hello

stat hello

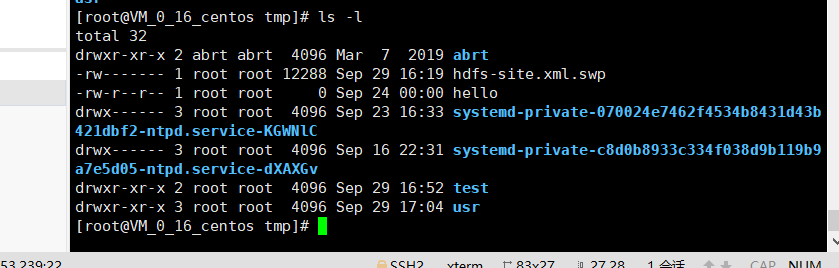


* chown命令：修改文件所有者权限

1. 将hello文件所有者改为root帐号，并查看属性

chown root hello

ls -l

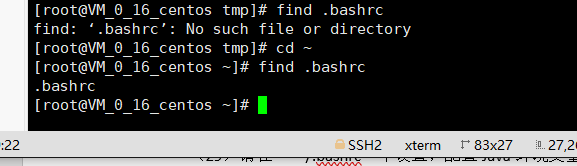


* find命令：文件查找

1. 找出主文件夹下文件名为.bashrc的文件

cd ~

find.bashrc

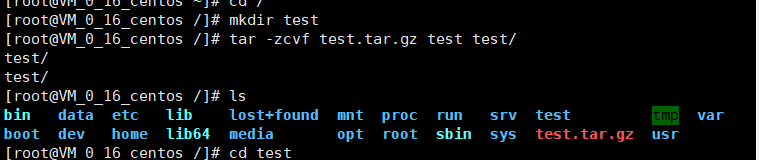


* tar命令：压缩命令

1. 在根目录“/”下新建文件夹test，然后在根目录“/”下打包成test.tar.gz

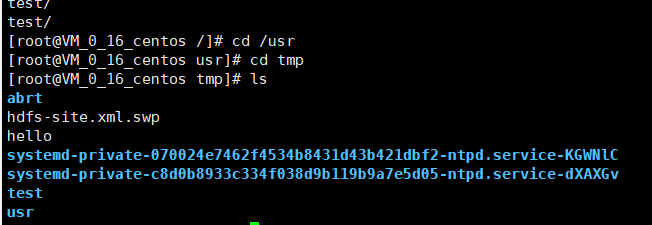
mkdir test

tar -zcvf test.gar.gz test test/



1. 把上面的test.tar.gz压缩包，解压缩到“/tmp”

tar -zxvf test.tar.gz -C /usr/tmp

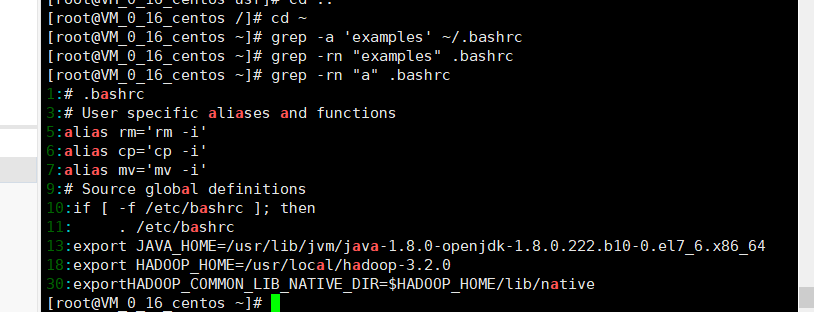


* grep命令：查找字符串

1. 从“～/.bashrc”文件中查找字符串'examples'

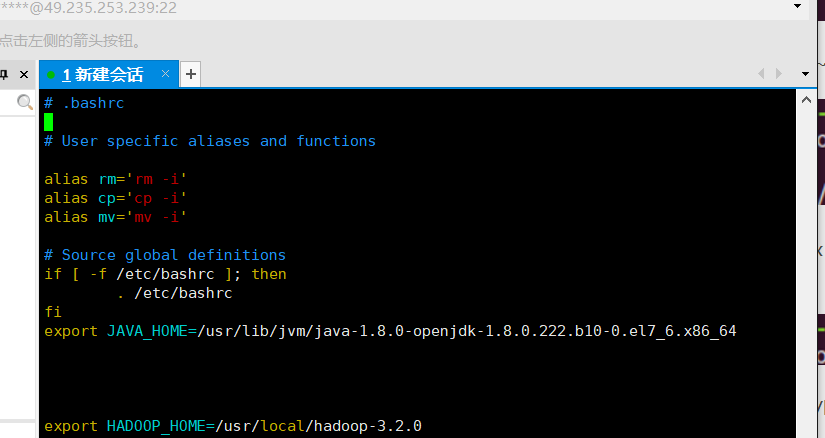
文件中无examples所以换了个a查找

grep -rn "a" .bashrc

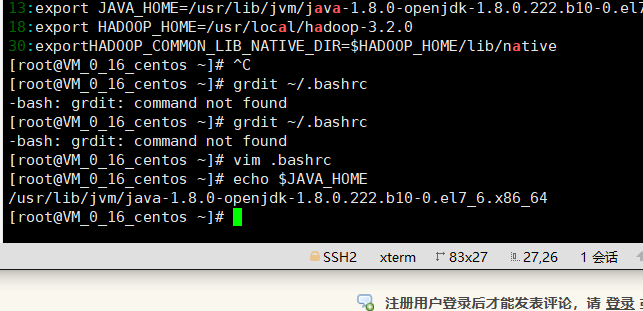


* 配置环境变量

（29）请在“～/.bashrc”中设置，配置Java环境变量



（30）查看JAVA\_HOME变量的值





## 2.3Hadoop操作

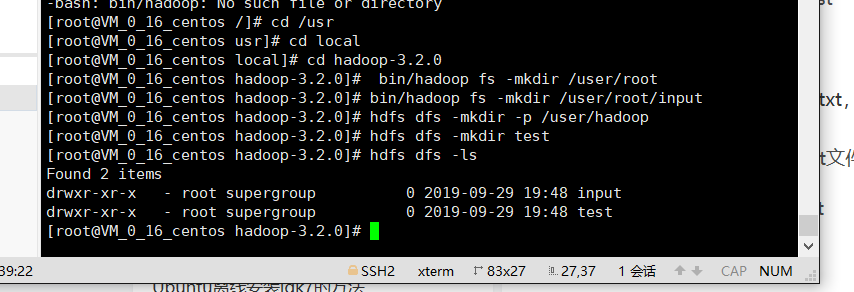
（31）使用hadoop用户登录Linux系统，启动Hadoop（Hadoop的安装目录为“/usr/local/hadoop”），为hadoop用户在HDFS中创建用户目录“/user/hadoop”

hdfsdfs -mkdir -p /user/hadoop

（32）接着在HDFS的目录“/user/hadoop”下，创建test文件夹，并查看文件列表

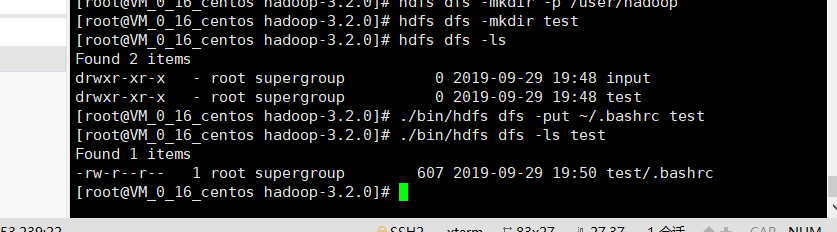
hdfs dfs -mkdir test

hdfs dfs -ls .



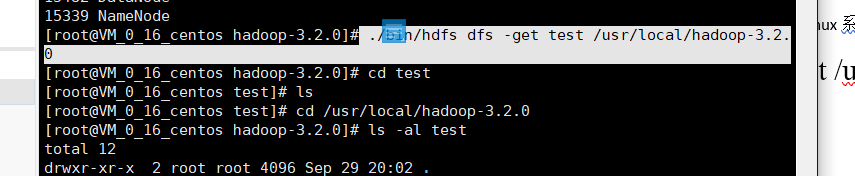
（33）将Linux系统本地的“～/.bashrc”文件上传到HDFS的test文件夹中，并查看test

./bin/hdfsdfs -put ~/.bashrc test

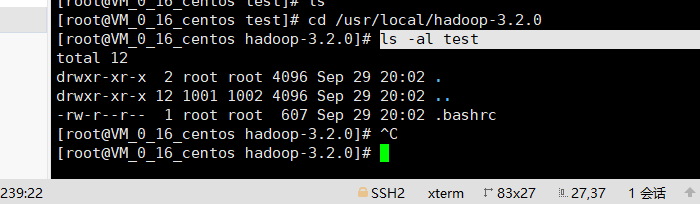


（34）将HDFS文件夹test复制到Linux系统本地文件系统的“/usr/local/hadoop”目录下

./bin/hdfsdfs -get test /usr/local/hadoop-3.2.0



ls -al test



# 3. 实验总结

/\*实验总结是对整个实验过程的回顾和讨论，如果实验过程当中出现问题，实验总结是最恰当的位置用于记录问题的由来和改正过程。（以下内容仅供参考）\*

安装hadoop云电脑重装了3次 虚拟机重装了3次

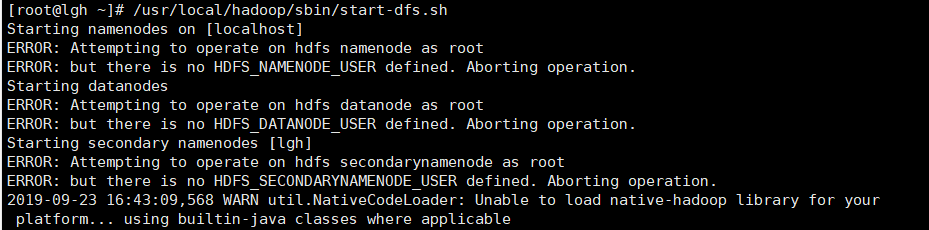
我太难了

**修改 Hadoop 的配置文件**

配置了好几遍才配好，老是有字母写错

**启动NameNode和DataNode进程：/usr/local/hadoop/sbin/start-dfs.sh**

出现如下错误：



解决方法：

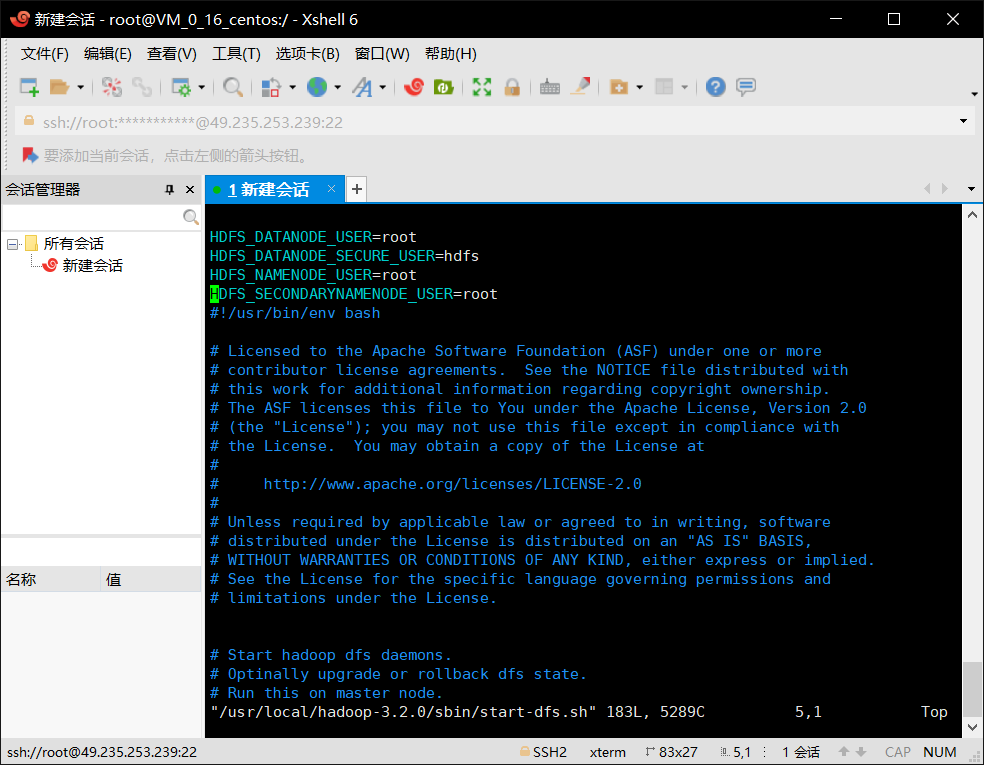
是因为缺少用户定义造成的，所以分别编辑开始和关闭脚本

$ vim sbin/start-dfs.sh  
$ vim sbin/stop-dfs.sh

Sbin是hadoop文件下的一个文件夹、

在顶部空白处添加内容：  
HDFS\_DATANODE\_USER=root  
HADOOP\_SECURE\_DN\_USER=hdfs  
HDFS\_NAMENODE\_USER=root  
HDFS\_SECONDARYNAMENODE\_USER=root

vim /usr/local/Hadoop-3.2.0/sbin/start-dfs.sh



vim /usr/local/hadoop-3.2.0/sbin/stop-dfs.sh

**还有一次修改 Hadoop 的配置文件**

**忘了保存，我人都傻掉了。又不知道是哪里错了**

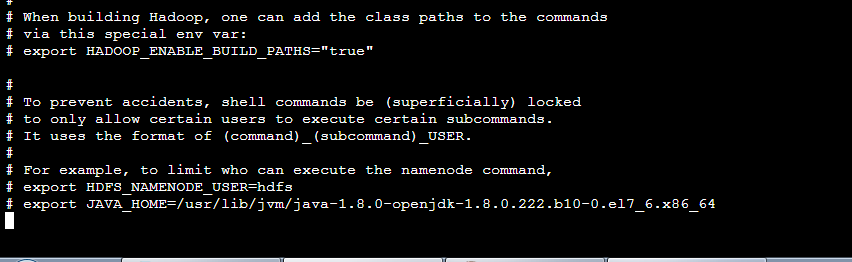
**最后由我们宿舍的毛少帮我检查出来**

jps只有一行

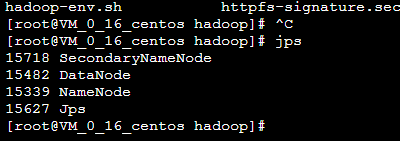


## 后来检查是java配置出现了问题

最后手动配置hadoop-env.sh



然后再重新启动



HDFS的目录“/user/hadoop”下，创建test文件夹

hdfsdfs -mkdir test

hdfsdfs -ls .

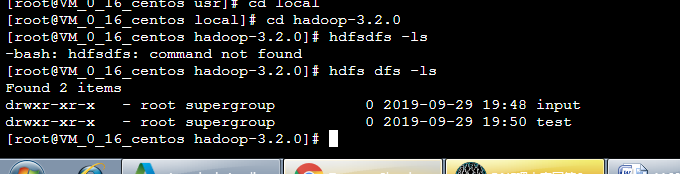
出现错误

Input path does not exist: hdfs://localhost:9000/user/root/input

错误原因： 文件存储系统中没有这个文件

解决方法：创建文件夹 bin/hadoop fs -mkdir /user/root

bin/hadoop fs -mkdir /user/root/input



就可以了