

**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** | **实验一** | | | **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 9 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 CentOS 7.6

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

2.1.1 安装ssh:

输入命令： yum install openssh-clients openssh-server

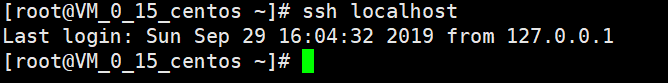
配置SSH免key登陆：

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

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

chmod 0600 ~/.ssh/authorized\_keys

安装完成输入ssh localhost测试，安装完成



2.1.2 配置JAVA环境

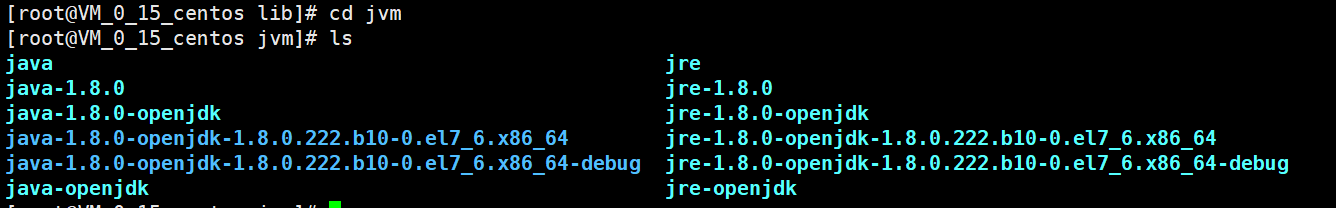
安装jdk，使用yum来安装1.8版本OpenJDK，输入：

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



配置java环境变量

找到自己的java版本，cd至/usr/lib/jvm，并ls



找到版本：java-1.8.0-openjdk-1.8.0.222.b10-0.el7\_6.x86\_64

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

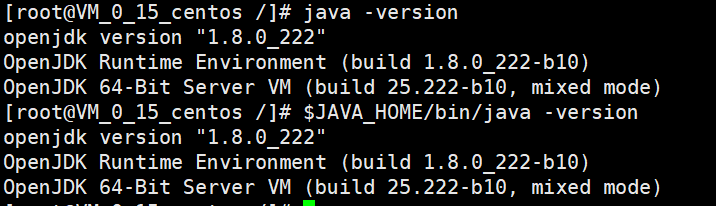
保存文件后执行下面命令使 JAVA\_HOME 环境变量生效:

source ~/.bashrc

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



2.1.3 安装hadoop

使用 wget 工具在线下载hadoop-3.2.0 版本：wget

<http://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/hadoop-3.2.0/hadoop-3.2.0.tar.gz>

这里使用清华的镜像，十分的快速

安装 Hadoop

将 Hadoop 安装到 /usr/local 目录下:

tar -zvxf hadoop-3.2.0.tar.gz -C /usr/local

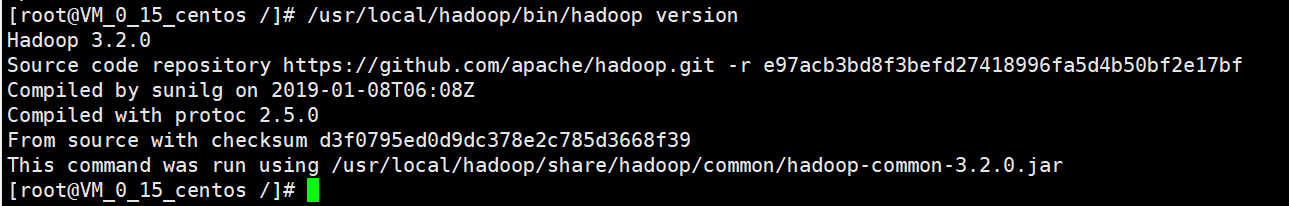
对安装的目录进行重命名，便于后续操作方便:

mv ./hadoop-3.2.0/ ./Hadoop

检查Hadoop是否已经正确安装:

输入/usr/local/hadoop/bin/hadoop version

显示版本信息则安装成功



2.1.4.hadoop伪分布式环境配置

设置 Hadoop 的环境变量

编辑 ~/.bashrc，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

并使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，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

分别修改为

core-site.xml

<configuration>

<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>

</configuration>

hdfs-site.xml

<configuration>

<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>

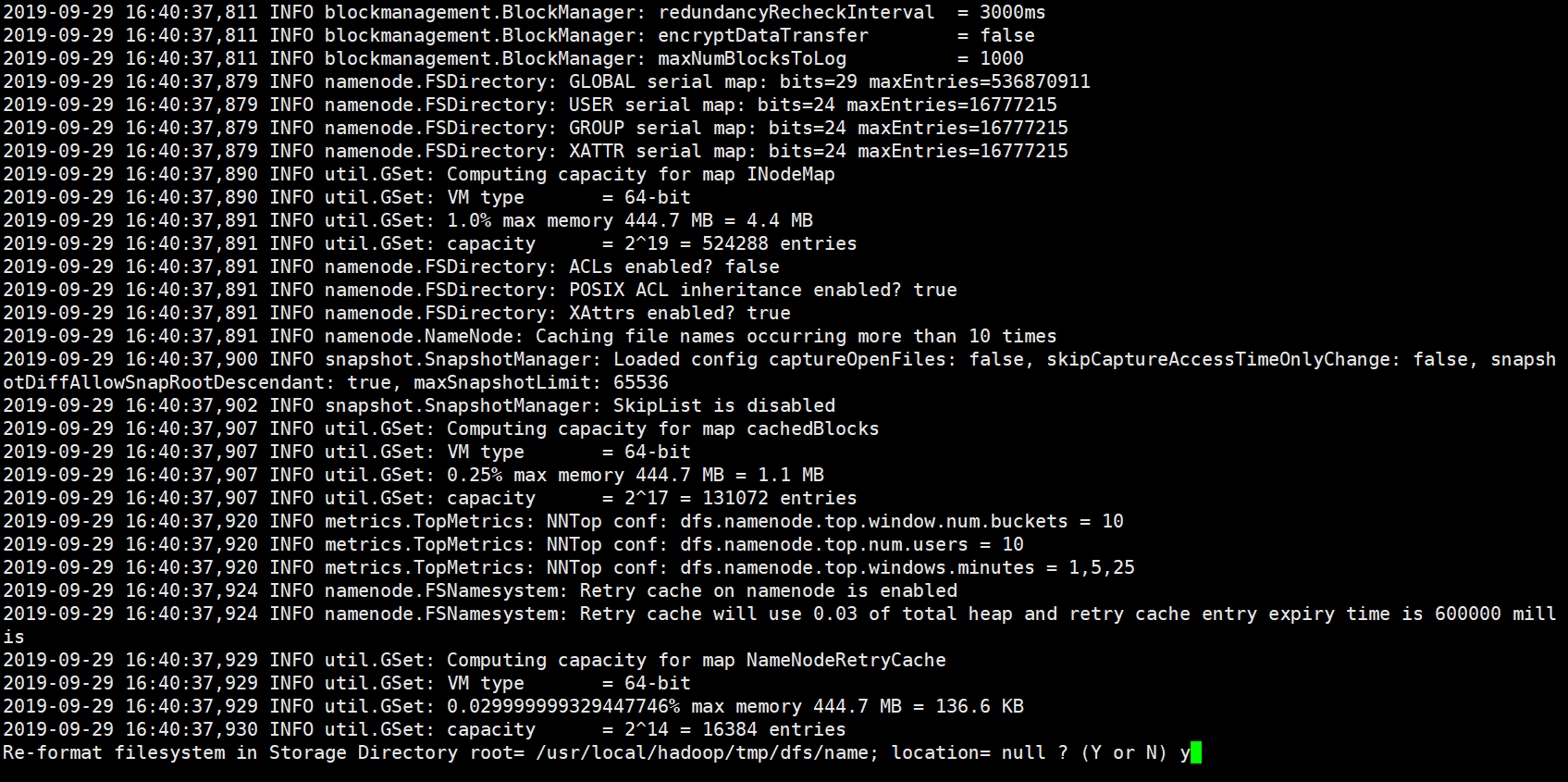
<name>dfs.datanode.data.dir</name>

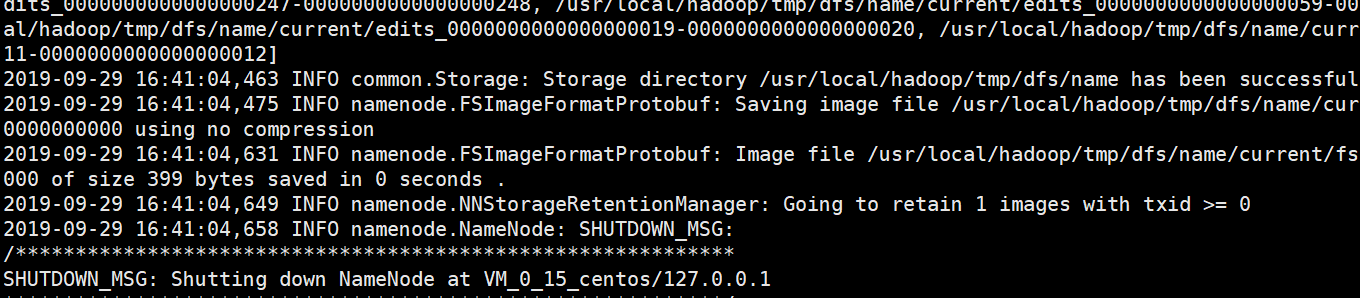
<value>file:/usr/local/hadoop/tmp/dfs/data</value>

</property>

</configuration>

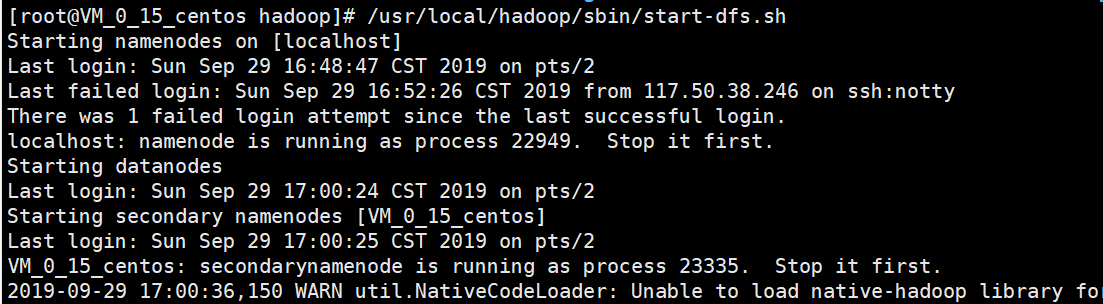
格式化NameNode:



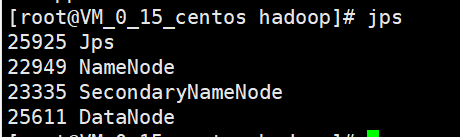


启动NameNode和DataNode进程：

/usr/local/hadoop/sbin/start-dfs.sh



输入检查NameNode和DateNode是否正常启动

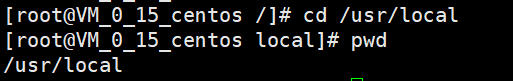


安装完成

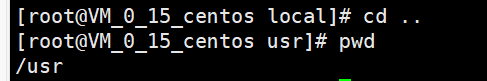
## 2.2 Linux操作

* cd命令：切换目录

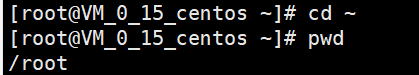
1. 切换到目录“/usr/local”



1. 切换到当前目录的上一级目录

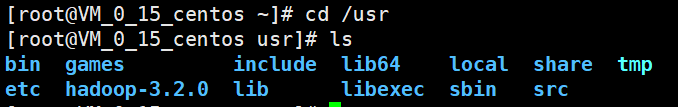
cd

1. 切换到当前登录Linux系统的用户的自己的主文件夹



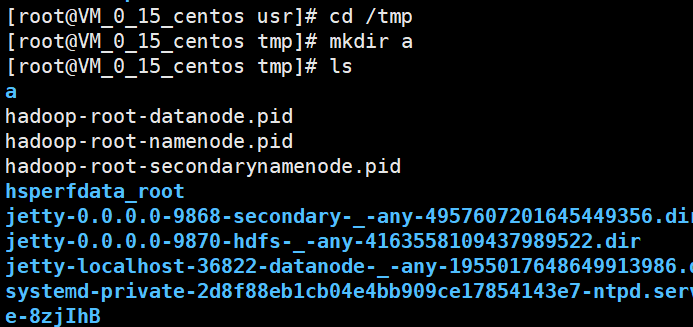
* ls命令：查看文件与目录

（4）查看目录“/usr”下的所有文件和目录

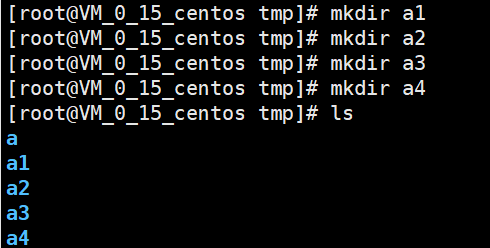


* mkdir命令：新建目录

（5）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



（6）进入“/tmp”目录，创建目录“a1/a2/a3/a4”

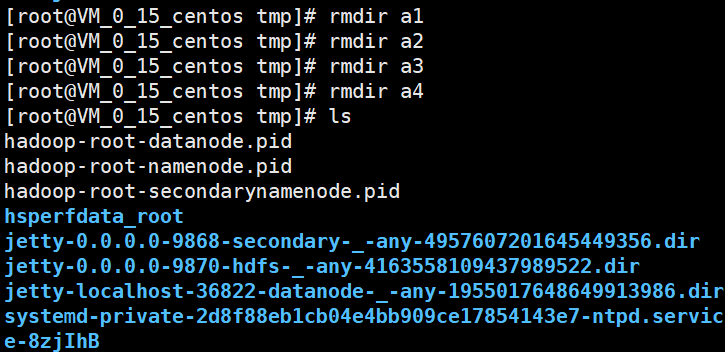


* rmdir命令：删除空的目录

（7）将上面创建的目录a（在“/tmp”目录下面）删除



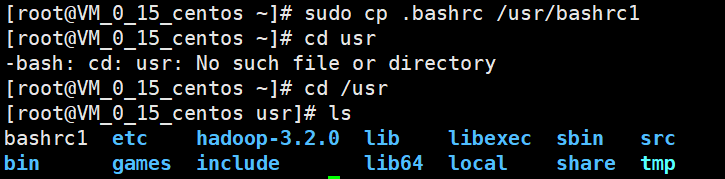
（8）删除上面创建的目录“a1/a2/a3/a4” （在“/tmp”目录下面），然后查看“/tmp”目录下面存在哪些目录



* cp命令：复制文件或目录

（9）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

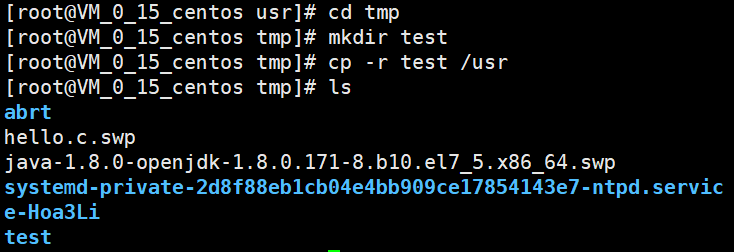
sudo cp .bashrc /usr/bashrc1

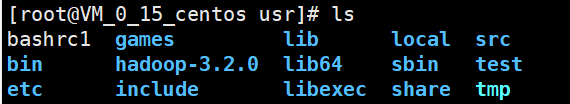


（10）在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下

mkdir test

cp -r test /usr

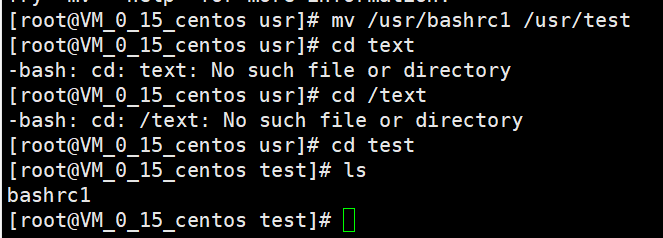




* mv命令：移动文件与目录，或更名

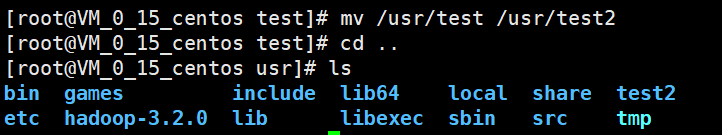
（11）将“/usr”目录下的文件bashrc1移动到“/usr/test”目录下

mv /usr /bashrc1 /usr/test



（12）将“/usr”目录下的test目录重命名为test2

mv /usr /test/usr/test2

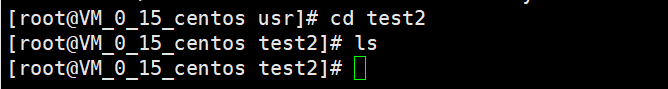


* rm命令：移除文件或目录

（13）将“/usr/test2”目录下的bashrc1文件删除

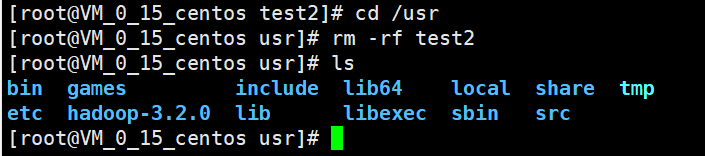
rm /usr/test2/bashrc1



cd

（14）将“/usr”目录下的test2目录删除

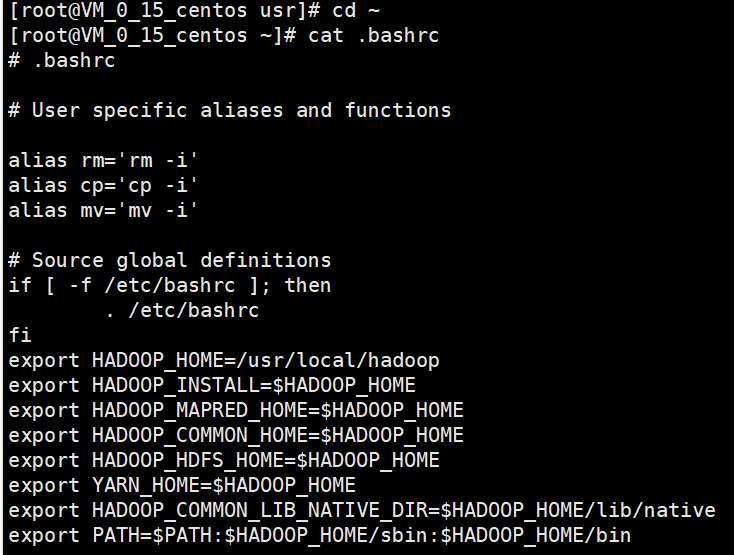
rm -rf test2



* cat命令：查看文件内容

（15）查看当前用户主文件夹下的.bashrc文件内容

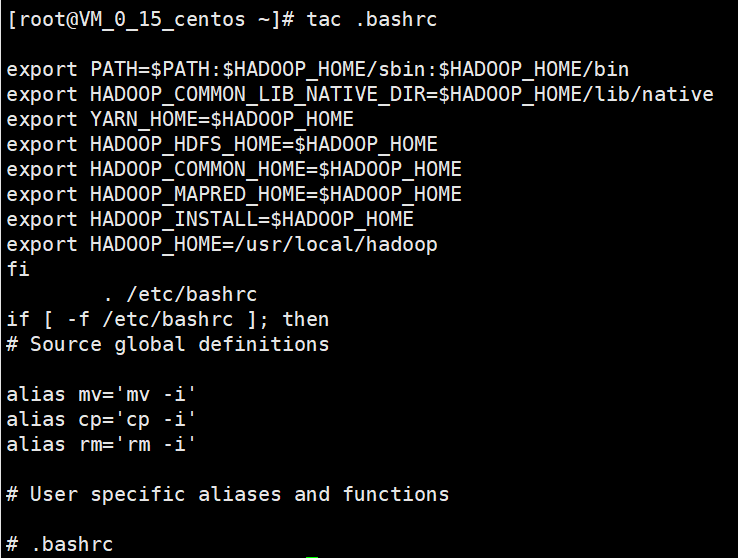
cat .bashrc



* tac命令：反向查看文件内容

（16）反向查看当前用户主文件夹下的.bashrc文件的内容

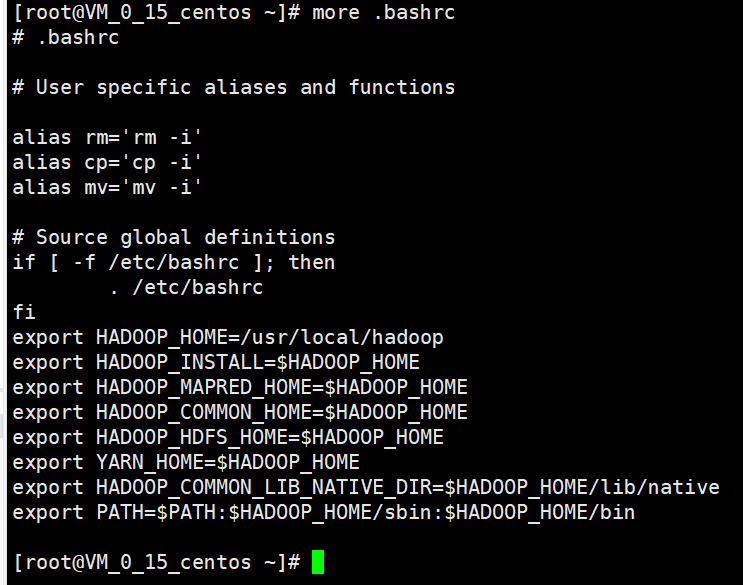
tac .bashrc



* more命令：一页一页翻动查看

（17）翻页查看当前用户主文件夹下的.bashrc文件的内容

more .bashrc

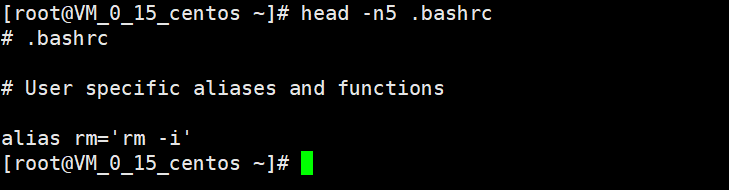


* head命令：取出前面几行

（18）查看当前用户主文件夹下.bashrc文件内容前20行

head -n20 .bashrc

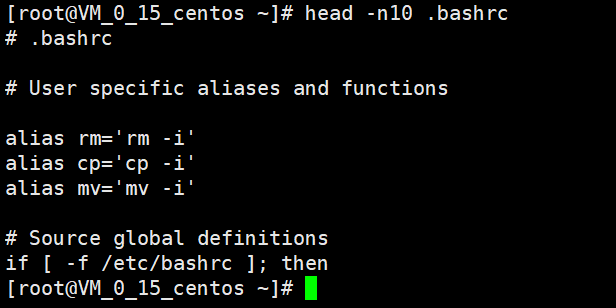
(改为5行) head -n5.bashrc



（19）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 -n50 .bashrc

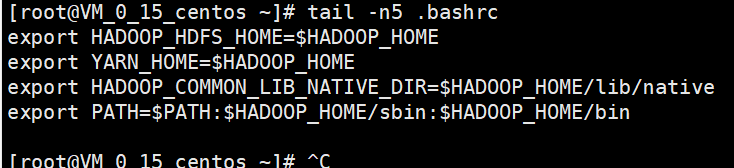
（改为10行）head -n10.bashrc



* tail命令：取出后面几行

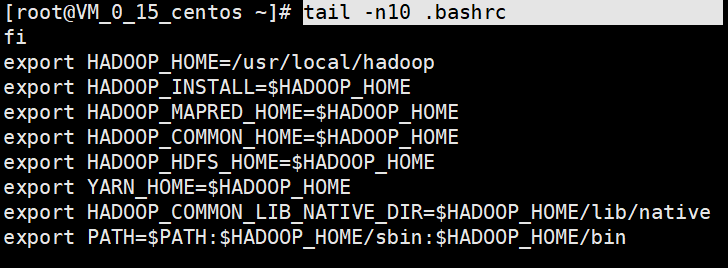
（20）查看当前用户主文件夹下.bashrc文件内容最后20行

(改为5行) tail -n5 .bashrc



（21）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

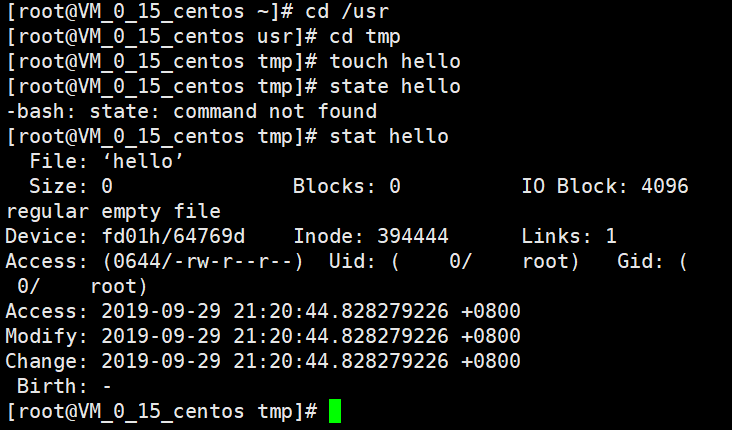
（改为10行）tail -n10 .bashrc



* touch命令：修改文件时间或创建新文件

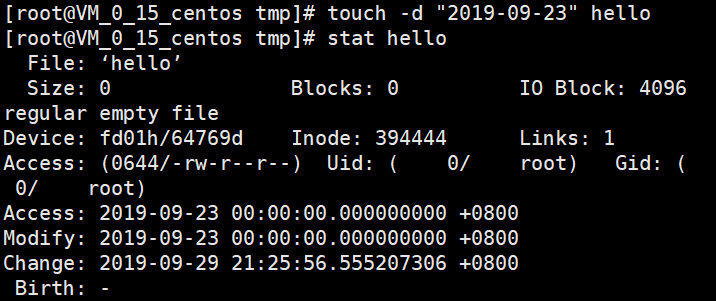
（22）在“/tmp”目录下创建一个空文件hello，并查看文件时间

touch hello并且stat hello



（23）修改hello文件，将文件时间整为5天前

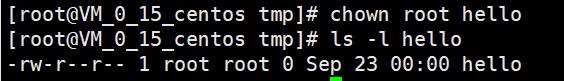
touch -d “2019-09-23” hello



* chown命令：修改文件所有者权限

（24）将hello文件所有者改为root帐号，并查看属性

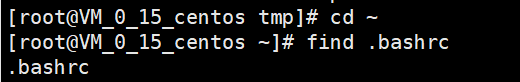
chown root hello并 ls -l hello



* find命令：文件查找

（25）找出主文件夹下文件名为.bashrc的文件

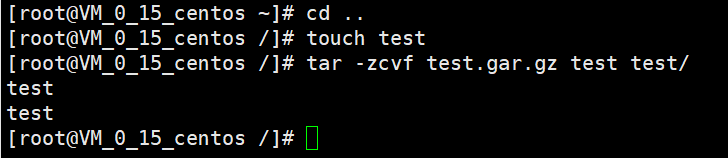
find .bashrc

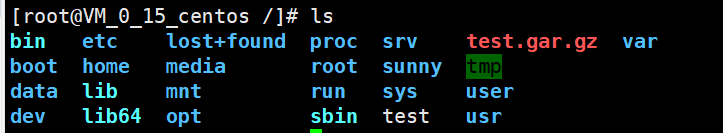


* tar命令：压缩命令

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

tar -zcvf test.gar.gz test test/





（27）把上面的test.tar.gz压缩包，解压缩到“/tmp”目录

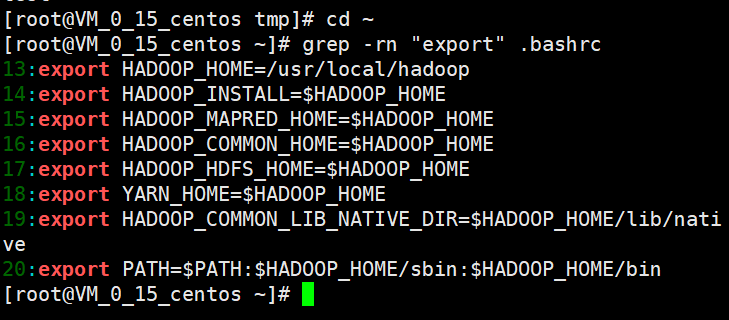
tar -zxvf test.gar.gz -C /usr/tmp



* grep命令：查找字符串

（28）从“～/.bashrc”文件中查找字符串'export'

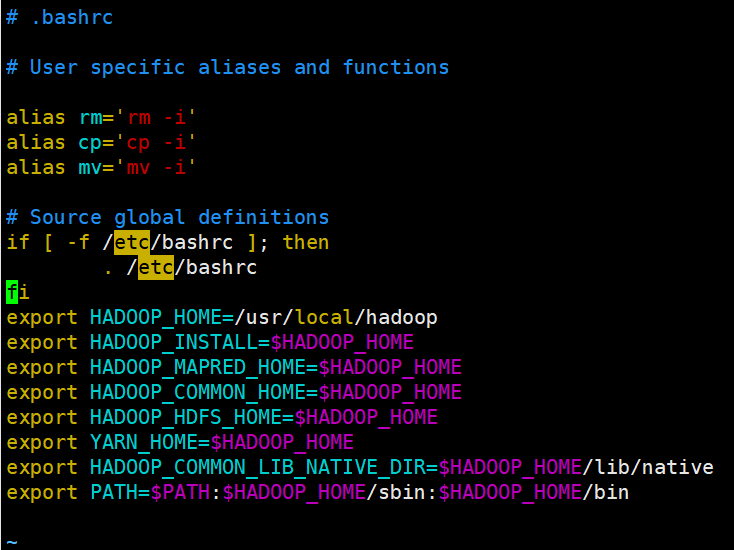
grep -rn "export" .bashrc



* 配置环境变量

（29）请在“～/.bashrc”中设置，配置Java环境变量

vim ~/bashrc



（30）查看JAVA\_HOME变量的值

echo $JAVA\_HOME



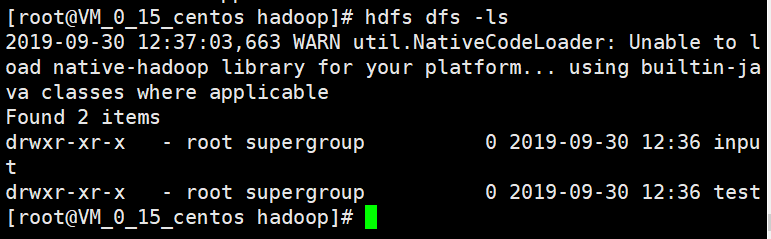


## 2.3 Hadoop操作

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

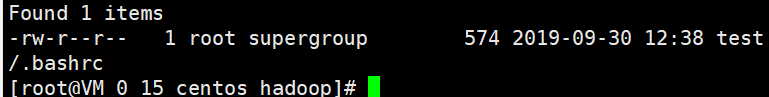
hdfs dfs -mkdir test

hdfs dfs -ls .



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

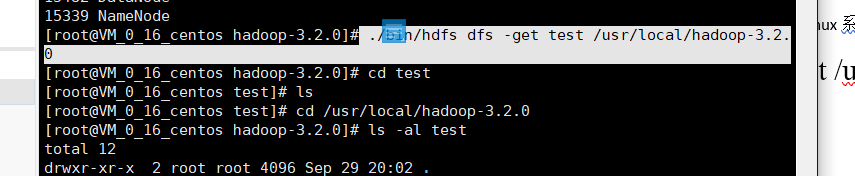
./bin/hdfs dfs -put ~/.bashrc test



（33）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/hdfs dfs -put ~/.bashrc test

./bin/hdfs dfs -ls test



ls -al test

# 3. 实验总结

一开始启动hadoop的时候，出现WARNING: HADOOP\_SECURE\_DN\_USER has been replaced by HDFS\_DATANODE\_SECURE\_USER.

解决方案，在$ vim sbin/start-dfs.sh $ vim sbin/stop-dfs.sh 将

HDFS\_DATANODE\_USER=root

HADOOP\_SECURE\_DN\_USER=hdfs

HDFS\_NAMENODE\_USER=root

HDFS\_SECONDARYNAMENODE\_USER=root

改为

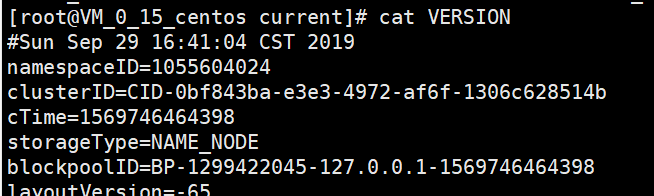
HDFS\_DATANODE\_USER=root

HDFS\_DATANODE\_SECURE\_USER=hdfs

HDFS\_NAMENODE\_USER=root

HDFS\_SECONDARYNAMENODE\_USER=root

后来写实验报告的时候，没注意再次格式化了一次格式化NameNode，导致DataNode和它的clusterID不同，jps显示不出来，后找到/usr/local/hadoop/tmp/dfs/name/current里获取clusterID



再把/usr/local/hadoop/tmp/dfs/data/current目录下的clusterID修改一致才成功

## 3.1 加油问题