

**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** | **熟悉常用的Linux操作和Hadoop操作** | | | **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** | **2019.10.07** |

# 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

### 2.1.1软件安装

（1）ssh和pdsh安装

sudo apt-get install ssh

sudo apt-get install pdsh

（2）java安装

sudo apt-get install openjdk-8-jdk

打开 ~/.bashrc

在最底部添加如下内容：

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export JRE\_HOME=${JAVA\_HOME}/jre

export CLASSPATH=.:${JAVA\_HOME}/lib:${JRE\_HOME}/lib

export PATH=${JAVA\_HOME}/bin:$PATH

### 2.1.2下载hadoop

Swget https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.0/hadoop-3.2.0.tar.gz

### 2.1.3 hadoop修改配置文件

（1）编辑etc/hadoop/hadoop-env.sh

添加export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

### 2.1.4 伪分布式环境配置

etc/hadoop/core-site.xml添加：

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

etc/hadoop/hdfs-site.xml添加:

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

### 2.1.5 设置无密码ssh

（1）检验是否需要密码：

ssh localhost

（2）修改为不需要密码

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

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

chmod 0600 ~/.ssh/authorized\_keys

### 2.1.6 hadoop初始化

（1）格式化FS

bin/hdfs namenode -format

### 2.1.7 运行hadoop

（1）开启名称节点和数据节点

sbin/start-dfs.sh

（2）访问http://localhost:9870

## 2.2 Linux操作

### 2.2.1 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



1. 切换到当前登录Linux系统的用户的自己的主文件夹



结果符合预期。

### 2.2.2 ls命令：查看文件与目录

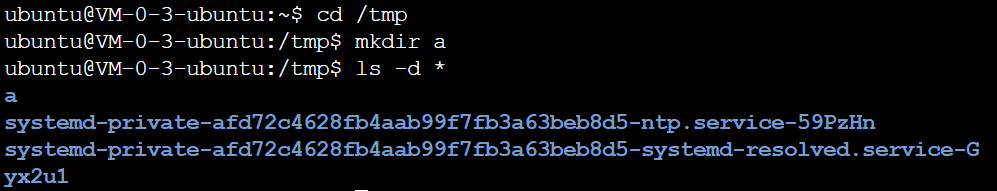
1. 查看目录“/usr”下的所有文件和目录



结果符合预期。

### 2.2.3 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



结果符合预期。

1. 进入“/tmp”目录，创建目录“a1/a2/a3/a4”



结果符合预期。

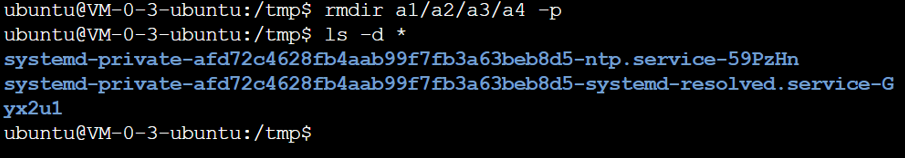
### 2.2.4 rmdir命令：删除空的目录

1. 将上面创建的目录a（在“/tmp”目录下面）删除



结果符合预期。

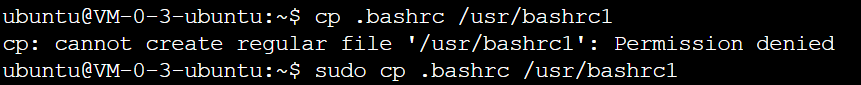
（8）删除上面创建的目录“a1/a2/a3/a4” （在“/tmp”目录下面），然后查看“/tmp”目录下面存在哪些目录



结果符合预期。

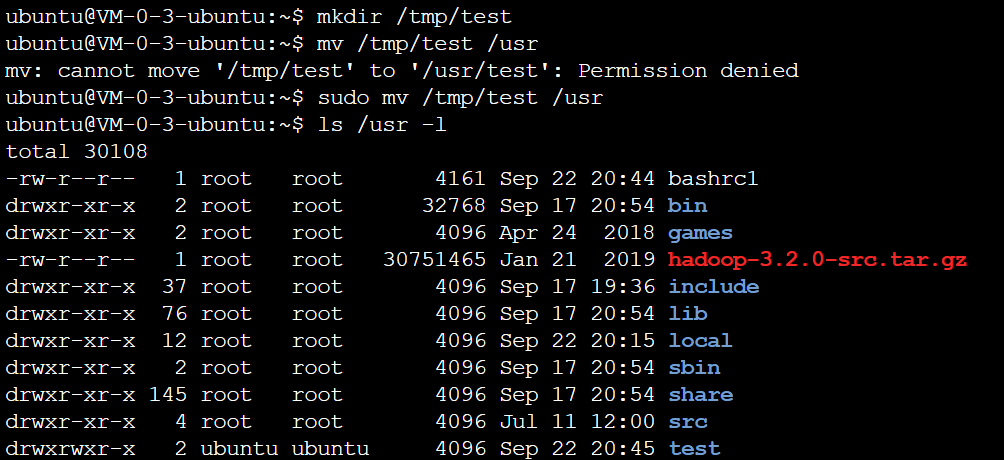
### 2.2.5 cp命令：复制文件或目录

（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



结果符合预期。

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



结果符合预期。

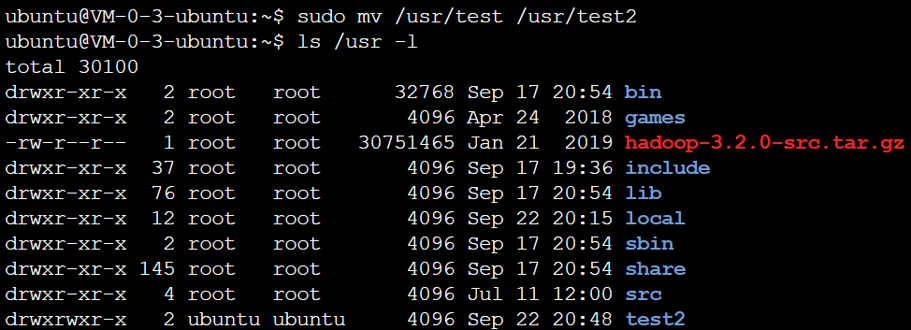
### 2.2.6 mv命令：移动文件与目录，或更名

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



结果符合预期。

（12）将“/usr”目录下的test目录重命名为test2



结果符合预期。

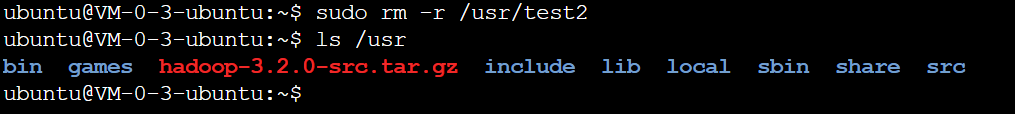
### 2.2.7 rm命令：移除文件或目录

（13）将“/usr/test2”目录下的bashrc1文件删除



结果符合预期。

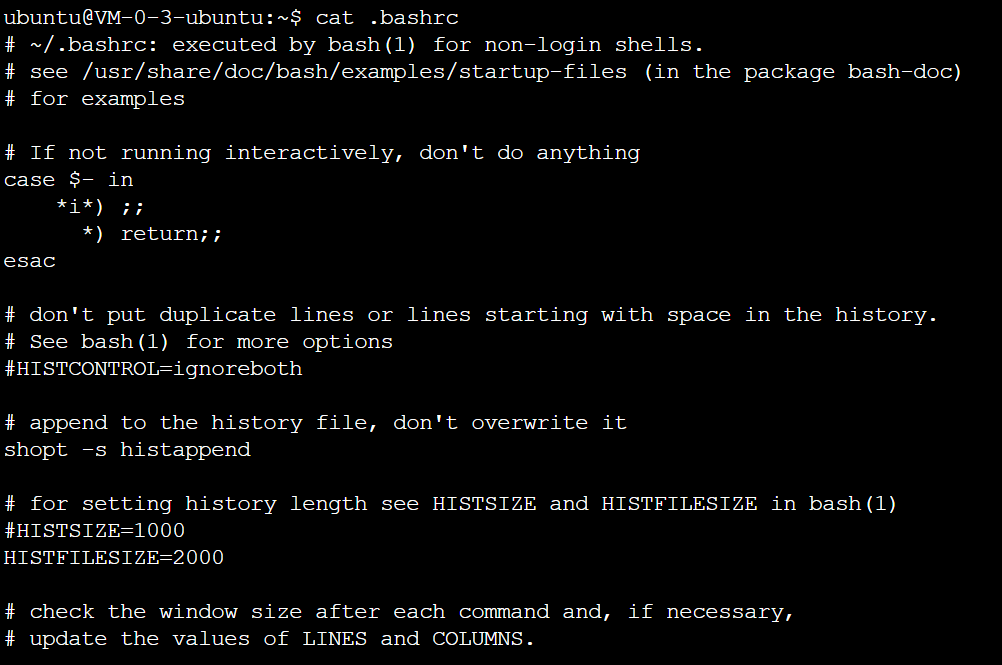
（14）将“/usr”目录下的test2目录删除



结果符合预期。

### 2.2.8 cat命令：查看文件内容

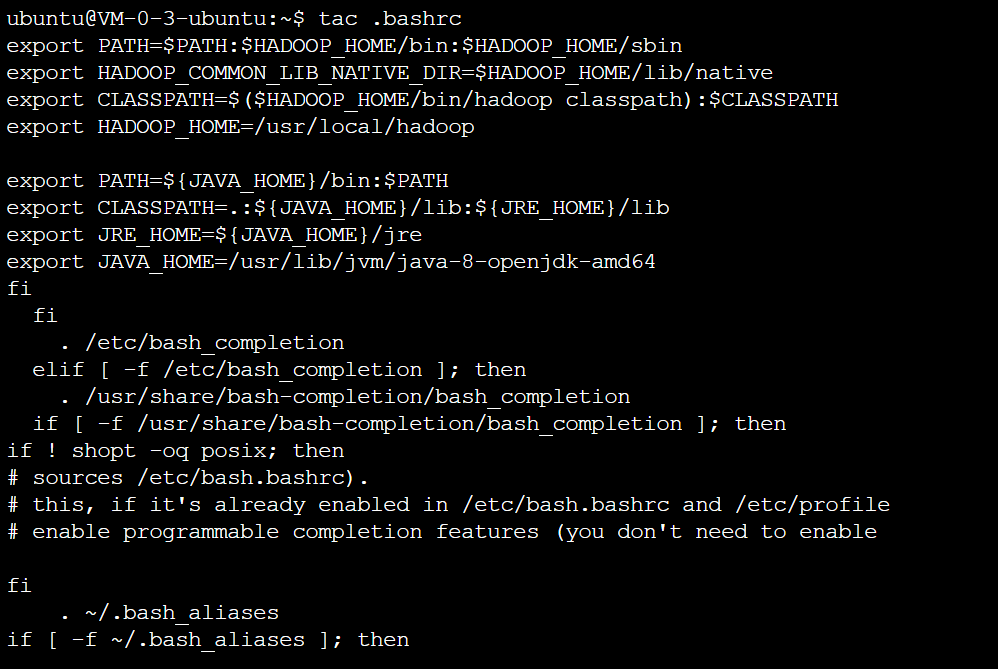
（15）查看当前用户主文件夹下的.bashrc文件内容



结果符合预期。

### 2.2.9 tac命令：反向查看文件内容

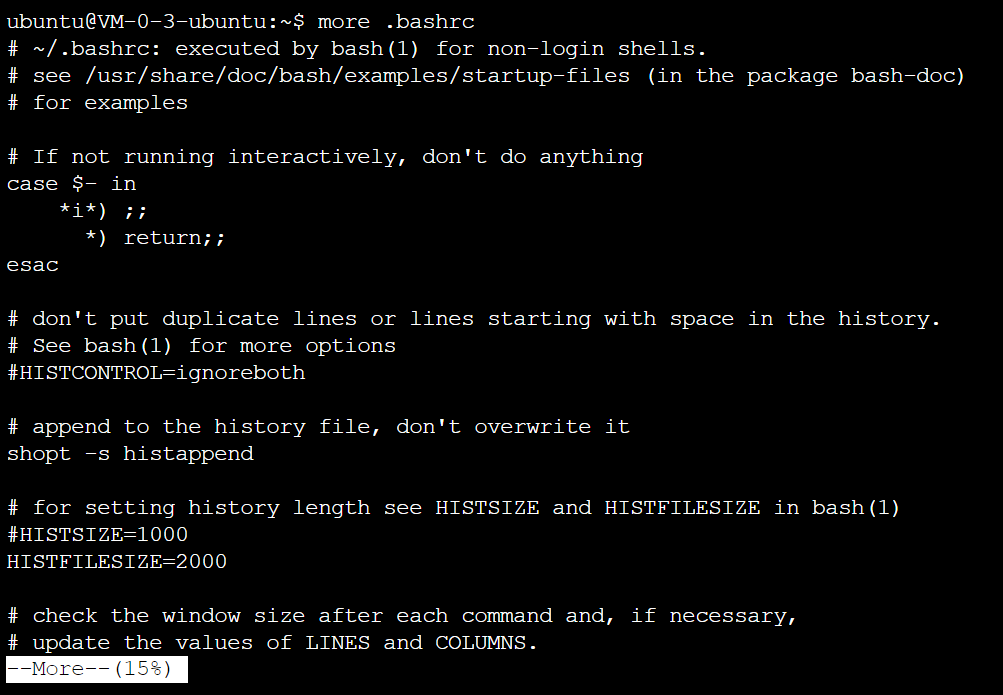
（16）反向查看当前用户主文件夹下的.bashrc文件的内容



结果符合预期。

### 2.2.10 more命令：一页一页翻动查看

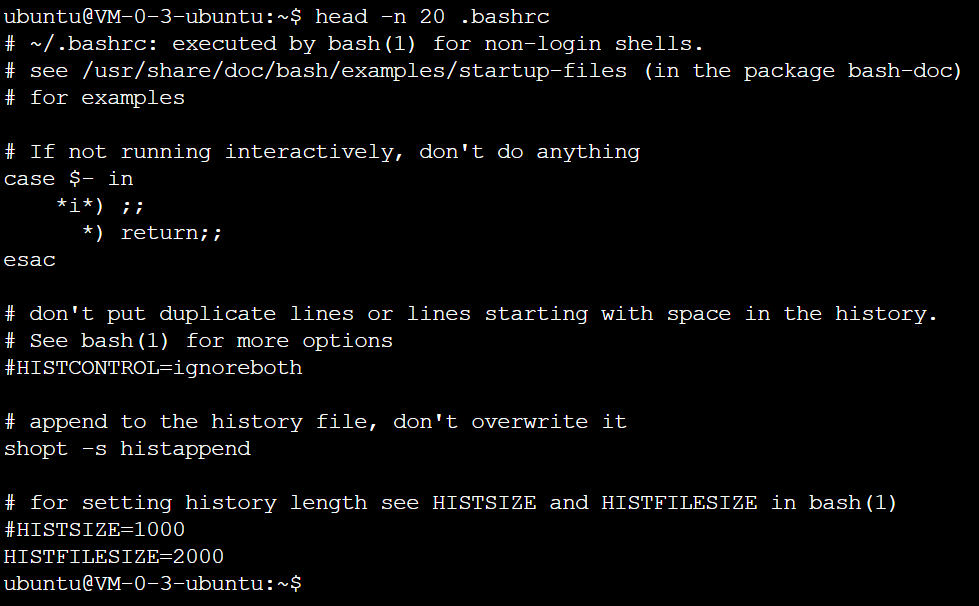
（17）翻页查看当前用户主文件夹下的.bashrc文件的内容



结果符合预期。

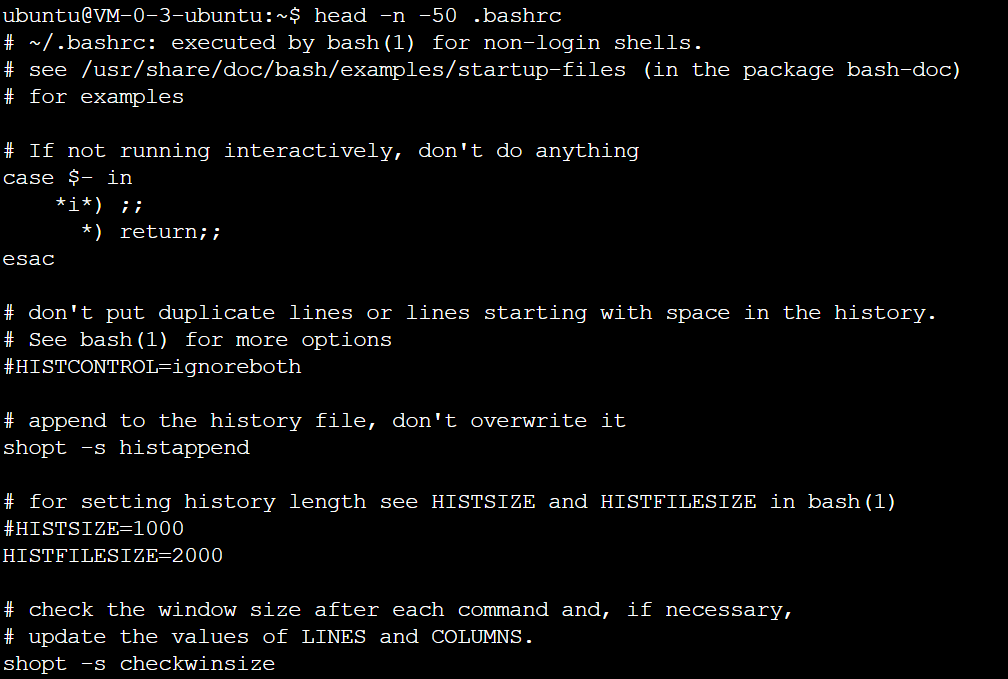
### 2.2.11 head命令：取出前面几行

（18）查看当前用户主文件夹下.bashrc文件内容前20行



结果符合预期。

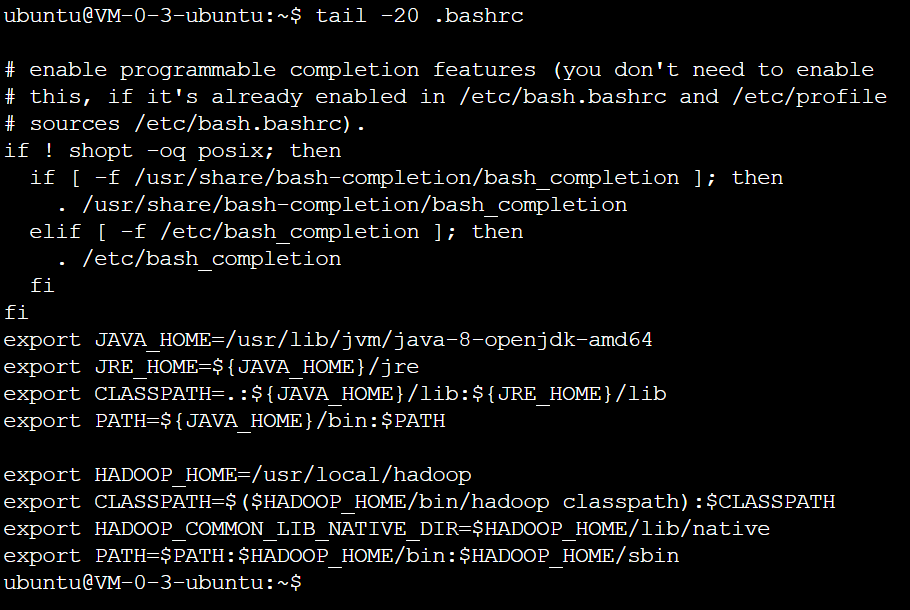
（19）查看当前用户主文件夹下.bashrc文件内容，后面50行不显示，只显示前面几行



结果符合预期。

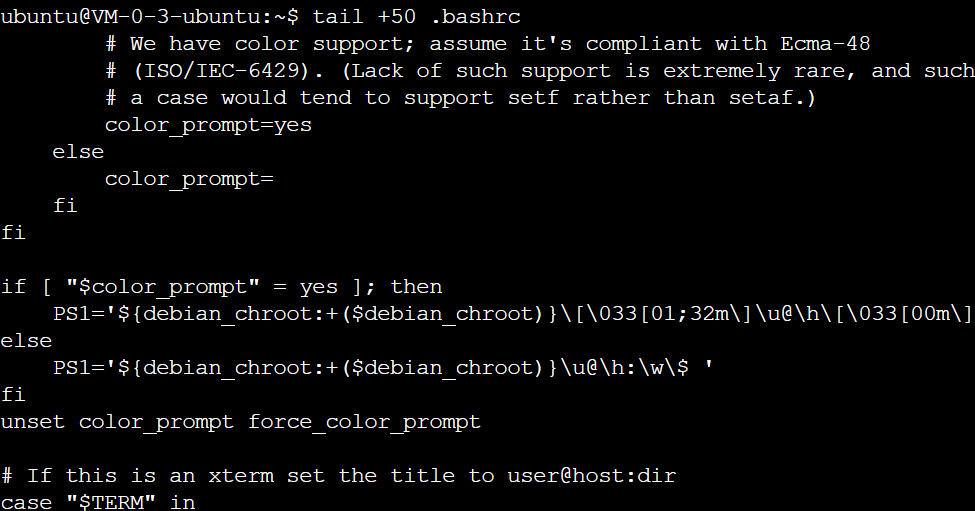
### 2.2.12 tail命令：取出后面几行

（20）查看当前用户主文件夹下.bashrc文件内容最后20行



结果符合预期。

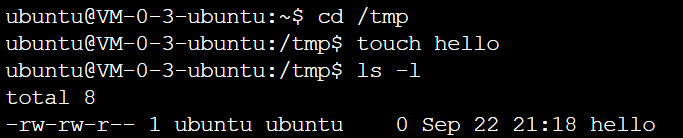
（21）查看当前用户主文件夹下.bashrc文件内容，并且只列出50行以后的数据



结果符合预期。

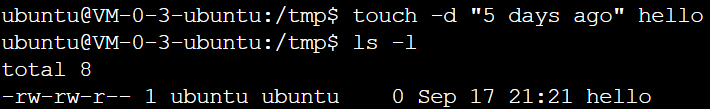
### 2.2.13 touch命令：修改文件时间或创建新文件

（22）在“/tmp”目录下创建一个空文件hello，并查看文件时间



结果符合预期。

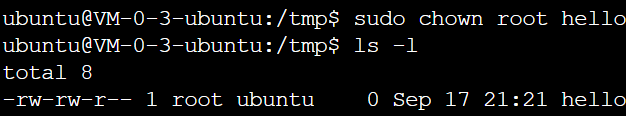
（23）修改hello文件，将文件时间整为5天前



结果符合预期。

### 2.2.14 chown命令：修改文件所有者权限

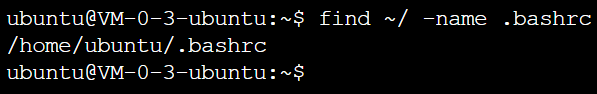
（24）将hello文件所有者改为root帐号，并查看属性



结果符合预期。

### 2.2.15 find命令：文件查找

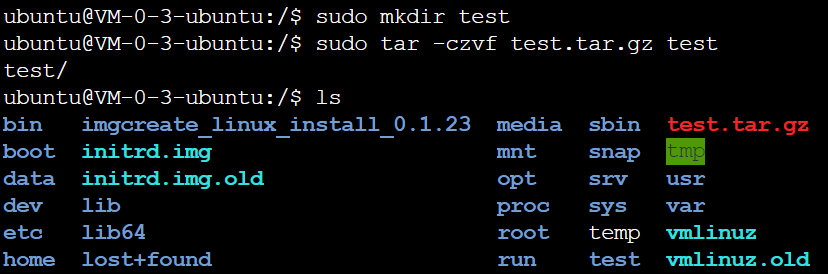
（25）找出主文件夹下文件名为.bashrc的文件



结果符合预期。

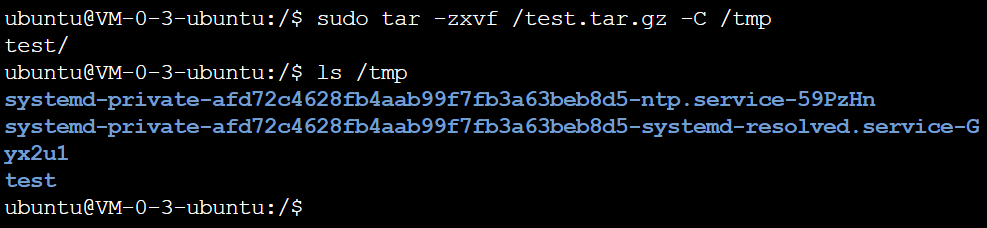
### 2.2.16 tar命令：压缩命令

（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



结果符合预期。

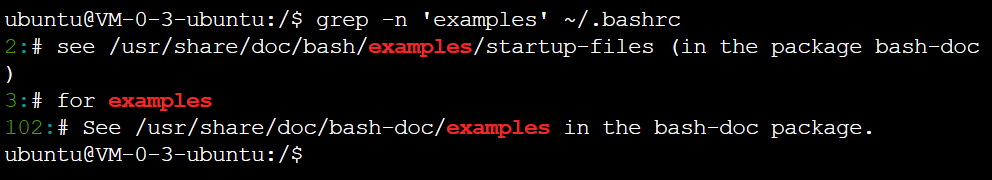
（27）把上面的test.tar.gz压缩包，解压缩到“/tmp”目录



结果符合预期。

### 2.2.17 grep命令：查找字符串

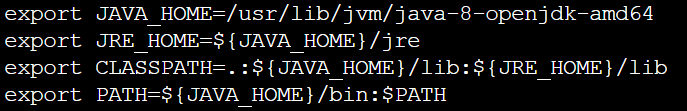
（28）从“～/.bashrc”文件中查找字符串'examples'



结果符合预期。

### 2.2.18 配置环境变量

（29）请在“～/.bashrc”中设置，配置Java环境变量



结果符合预期。

（30）查看JAVA\_HOME变量的值

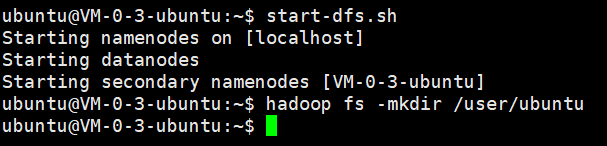


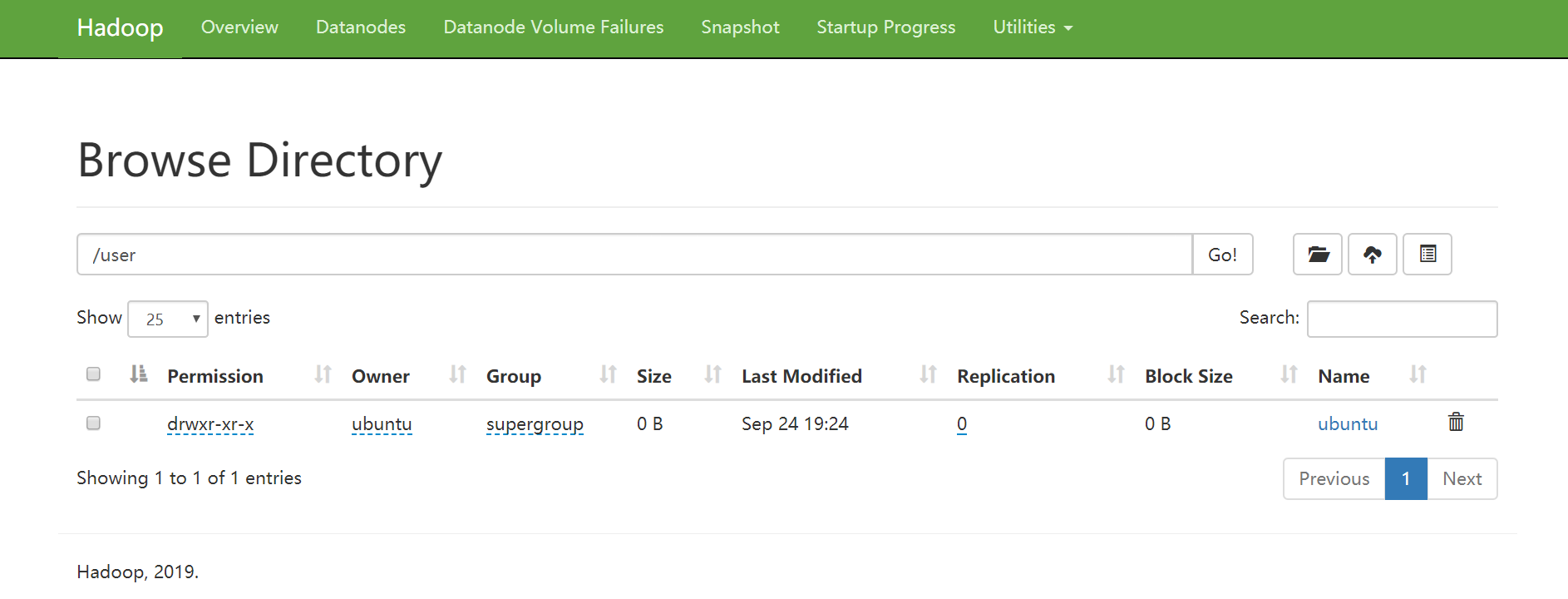
结果符合预期。



## 2.3 Hadoop操作

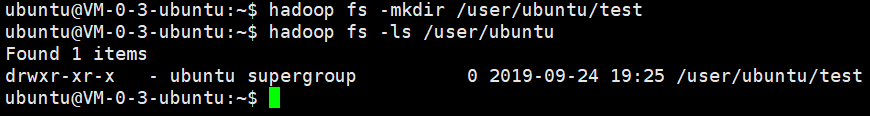
（31）使用hadoop用户登录Linux系统，启动Hadoop（Hadoop的安装目录为“/usr/local/hadoop”），为hadoop用户在HDFS中创建用户目录“/user/hadoop”





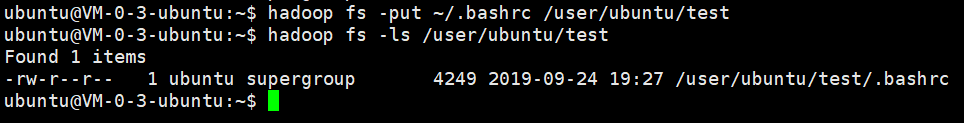
结果符合预期。

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



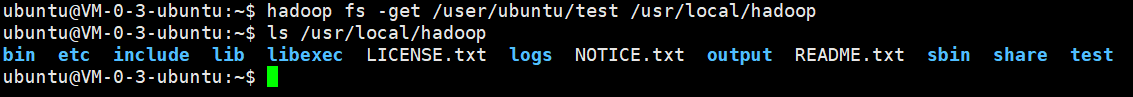
结果符合预期。

（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



结果符合预期。

（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



结果符合预期。

# 3. 实验总结

## 3.1 Hadoop安装

关于hadoop的安装，选择的是通过xshell访问云服务器，进而在云服务器上安装hadoop。整体安装过程是参考官网的教程，比较顺利。在安装hadoop前，我是没有接触过Linux的，在hadoop的安装过程中，Linux的常用操作也熟悉了起来。其中安装java的时候，我是用apt-get install openjdk-8-jdk 来进行安装，安装完也不知道环境路径配好了没有，也不清楚安装在哪了，就捣腾了挺久，后来用还是用linux的查找命令where找到了java的安装路径，重新配置了系统环境路径。

1. 问题：启动hdfs时报错pdsh@VM-0-3-ubuntu: localhost: rcmd: socket: Permission denied

原因：没有设置pdsh的rcmd

解决：设置pdsh的rcmd，执行命令echo "ssh" > /etc/pdsh/rcmd\_default

或创建/etc/pdsh/rcmd\_default，里面填入ssh

成功解决。

（2）问题：启动hdfs时警告WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...

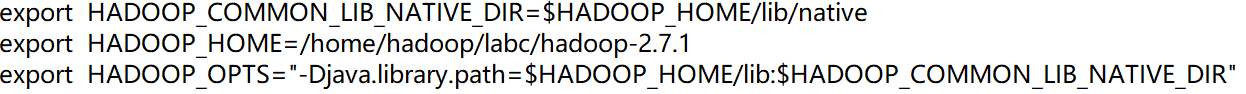
原因：1、可能hadoop版本64位于操作系统版本是32位

2、hadoop依赖库不正常

3、配置环境问题

解决：排除了1、2的可能，我修改了配置文件

在/etc/profile中和在hadoop-env.sh文件末尾，添加下面配置：



成功解决。

## 3.2 Linux 操作

如果题目是hadoop安装过程中用过的命令，盲敲命令都没问题。如果不是，百度一下，查好后敲几次，截图交实验报告时，顺便看了几次也记得差不多了。

## 3.3 Hadoop 操作

在上传文件的时候出问题了（报错：  
org.apache.hadoop.ipc.RemoteException(java.io.IOException)），原因是运行命令hadoop namenode -format格式化多次导致造成一个叫spaceID的不一致了，网上的解决方案是停止集群后，删除core-site.xml中配置的hadoop.tmp.dir对应文件夹。然而，我在core-site.xml中设置的是默认路径（详情请见实验记录2.1.4），我找不着。灵机一动，我就改配置文件的路径到一个指定路径上，再次格式化一次，没问题了。