

**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年 10 月1日** |

# 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

④ Xsheel Xftp Xmanager vim

⑤ Vmware

## 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

以下为我在Ubuntu18上安装Hadoop的过程，安装Vmmare虚拟机和Linux系统的过程就省略了。

退出当前用户，使用root登陆系统

关闭防火墙



开机不启动防火墙



查看防火墙状态，显示为关闭状态即可



配置root用户远程连接

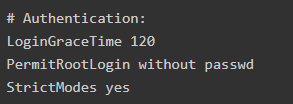
安装ssh-server



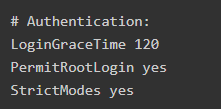
修改sshd\_config



找到

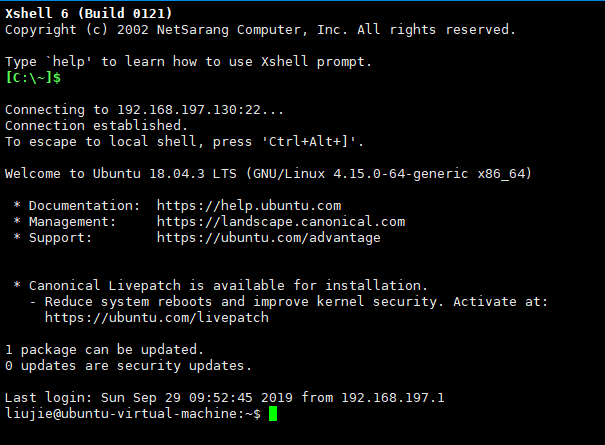


改成



然后重启ssh服务即可

重启后通过shell连接：

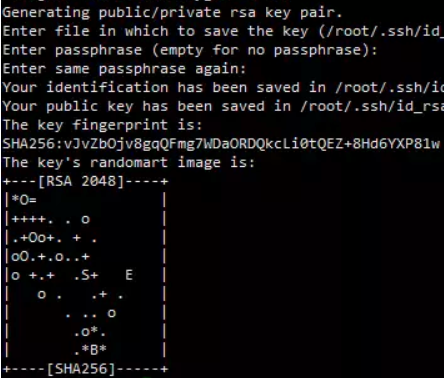


在进行了初次登陆后，会在当前家目录用户下有一个.ssh文件夹，进入该文件夹下：



使用rsa算法生成秘钥和公钥对：





运行后一路回车就可以了，其中第一个是要输入秘钥和公钥对的保存位置，默认是在：



然后把公钥加入到授权中：



安装JDK

解压jdk

创建一个java的文件夹，用户存放java文件



通过ftp工具把下载好的jdk安装包上传到java文件夹下。

解压：



解压成功后会在当前目录下看到：

jdk1.8.0\_181文件夹

然后删除安装包：

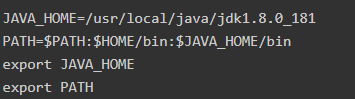


配置jdk

设置环境变量：



在末尾添加：

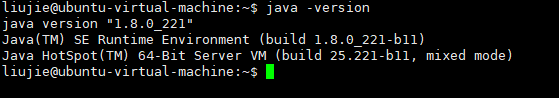


重新加载环境变量的配置文件：



检测：

Java -version



安装Hadoop

解压Hadoop

创建一个Hadoop的文件夹，用户存放在Hadoop文件



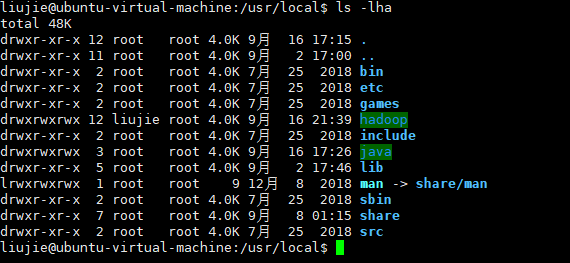
通过ftp工具把下载好的hadoop安装包上传到hadoop文件夹下。

解压：



解压完成

查看文件夹的所有者以及用户组，最重要的是拥有者是否是当前用户



如果拥有者不是当前用户，则修改为当前用户：

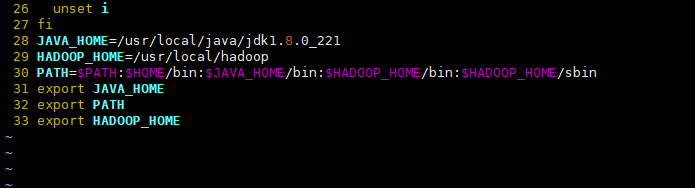
sudo chown -R liujie ./hadoop

配置Hadoop

设置环境变量：



在末尾添加：

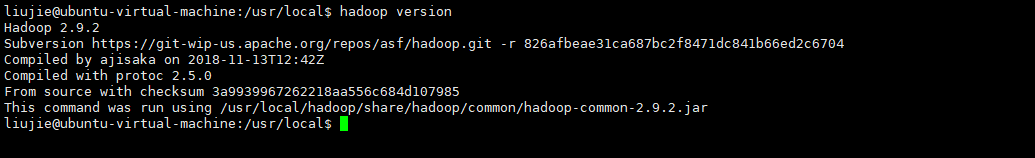


重新加载环境变量的配置文件：



检测：



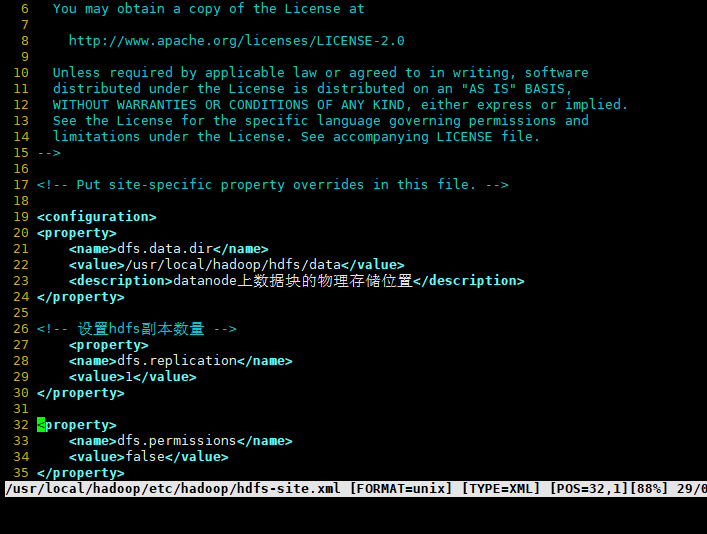


修改配置文件

在core-site.xml中添加：



在hdfs-site.xml中添加：



在Hadoop-env.sh中更改JAVA\_HOME

注释掉



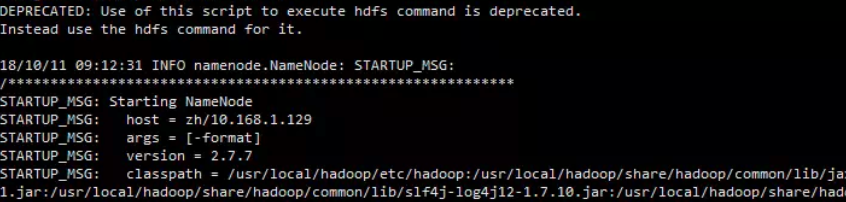
添加



测试启动

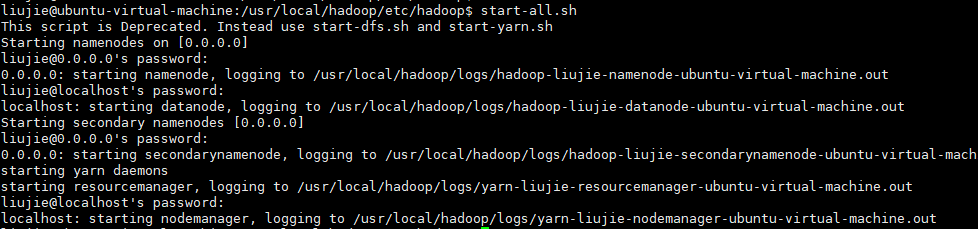
格式化namenode:





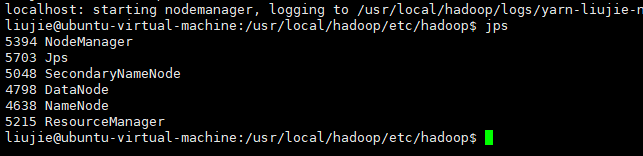
启动hdfs



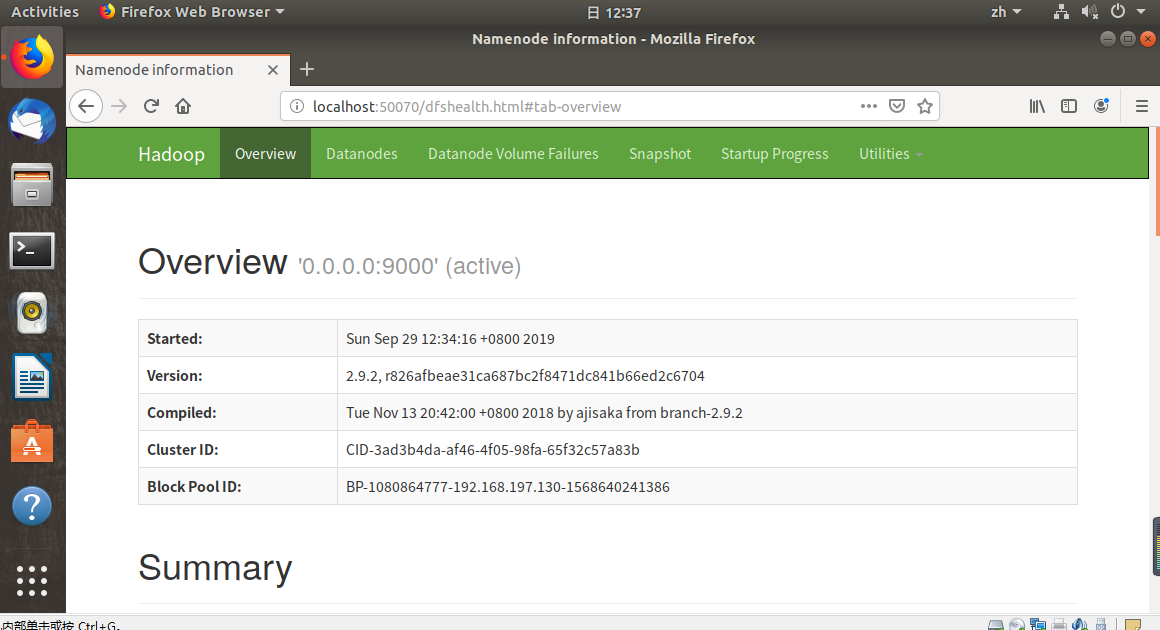


查看相应的进程：

Jps



访问测试：



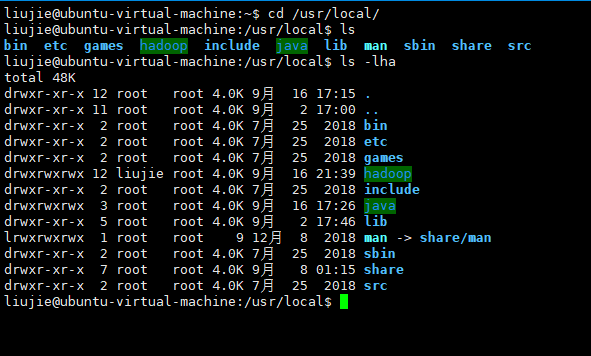
## 2.2 熟悉常用Linux操作

### 2.2.1 cd命令

① 切换到目录“/usr/local”

输入命令“cd /usr/local/”

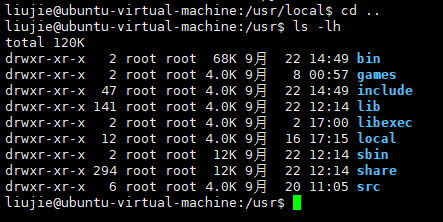
命令运行结果如下图所示，目录切换到“/usr/local”，结果符合预期。



② 切换到当前目录的上一级目录

输入命令 cd ..

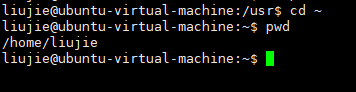
命令运行结果如下图所示，目录切换到“/usr”，结果符合预期。



③ 切换到当前登录Linux系统的用户的自己的主文件夹

输入命令：cd ~

命令运行结果运行如下图所示，目录切换到“/home/liujie”,结果符合预期

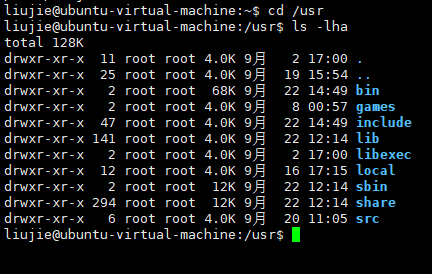


### 2.2.2 ls命令：查看文件与目录

① 查看目录“/usr”下的所有文件和目录

输入命令：cd /usr ls -lha

命令运行结果如下图所示，结果符合预期



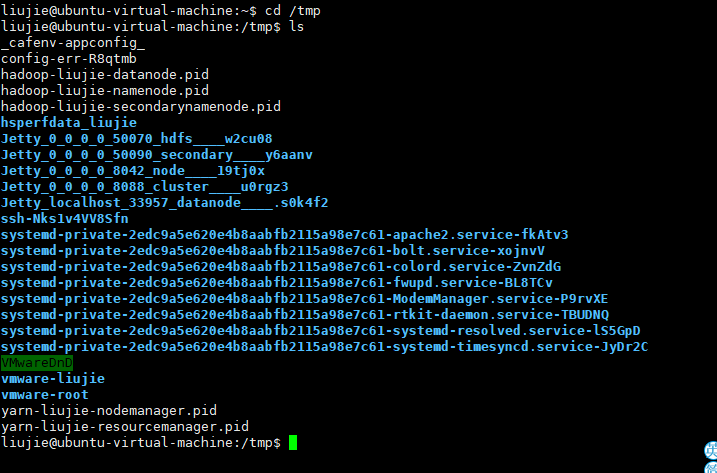
### 2.2.3 mkdir命令：新建目录

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

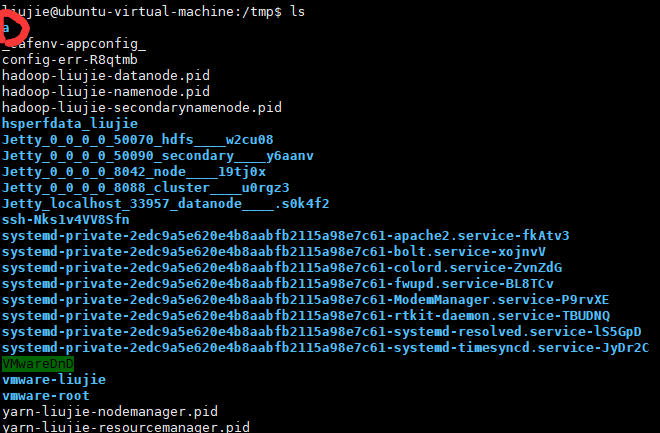
输入命令：cd /tmp mkdir a ls -d # -d只显示目录

命令结果如下图所示，结果符合预期

没有创建a目录之前的截图：



创建a目录之后的截图：

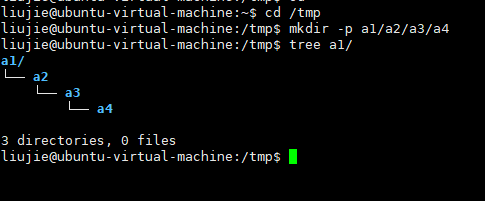


② 进入“/tmp”目录，创建目录“a1/a2/a3/a4”

输入命令：cd /tmp mkdir -p a1/a2/a3/a4 # -p是递归创建目录

最后用tree命令查看了a1目录结构

运行命令结果如下所示：



### 2.2.4 rmdir命令: 删除空的目录

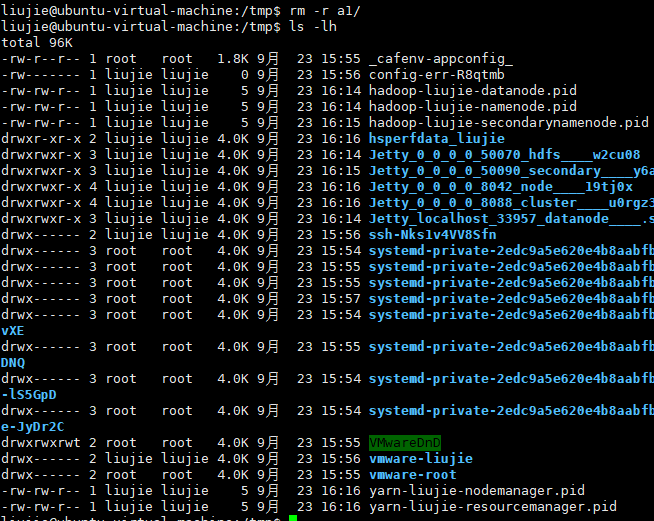
① 将上面创建的目录a（在“/tmp”目录下面）删除

输入命令：rm -r a

② 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

输入命令 rm -r a1/ #递归删除 cd /tmp ls -d

运行结果如下：



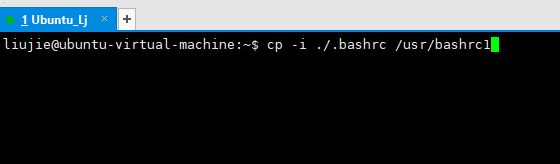
### 2.2.5 cp命令：复制文件或目录

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

输入命令：cd ~ #先切换到主目录

cp -i ./.bashrc /usr/bashrc1 # -i 覆盖文件前会提示

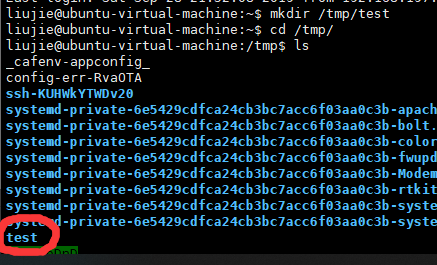
运行结果如下：



② 在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下

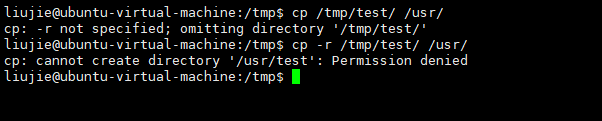
输入命令：mkdir /tmp/test cp -r /tmp/test /usr

以下为在/tmp目录下新建的test目录截图:



以下为将test复制到 /usr目录下，-r选项为复制目录

同时完成操作还需要root权限，否则会有如下结果:



### 2.2.6 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

输入命令：mv /usr/bashrc1 /usr/test

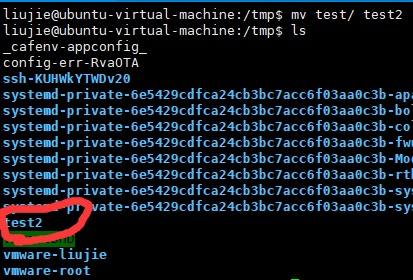
完成以上命令也需要管理员权限

② 将“/usr”目录下的test目录重命名为test2

mv /usr/test /usr/test2

由于/usr目录总需要root用户，我索性在/tmp目录下执行以上操作，道理相同

截图如下,将test重命名为test2,:



### 2.2.7 rm命令：移除文件或目录

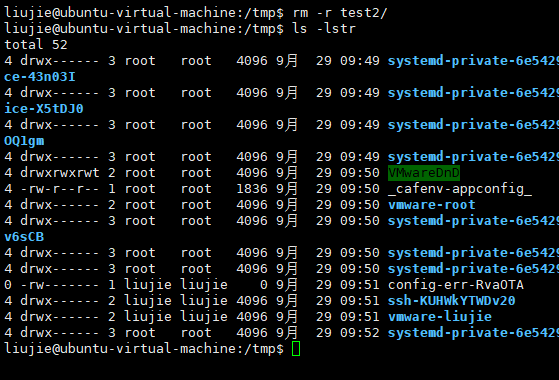
① 将“/usr/test2”目录下的bashrc1文件删除

输入命令：rm -r /usr/test2/bashrc1

② 将“/usr”目录下的test2目录删除

输入命令：rm -r /usr/test2

删除后该目录下就没有test2了，截图如下:



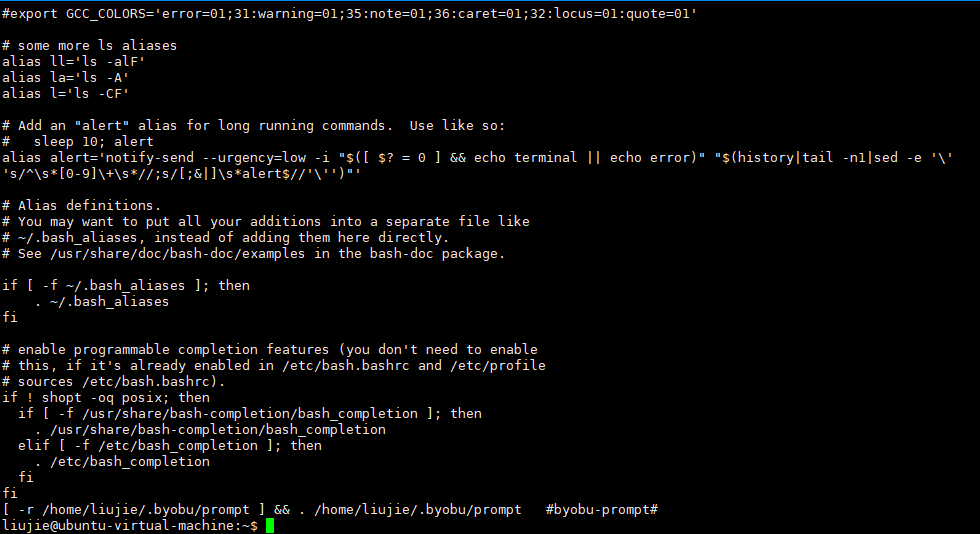
### 2.2.8 cat命令：查看文件内容

① 查看当前用户主文件夹下的.bashrc文件内容

输入命令：首先切换到用户目录 cd /home/liujie

然后执行：cat ./.bashrc

截图如下，以下为文件内容:

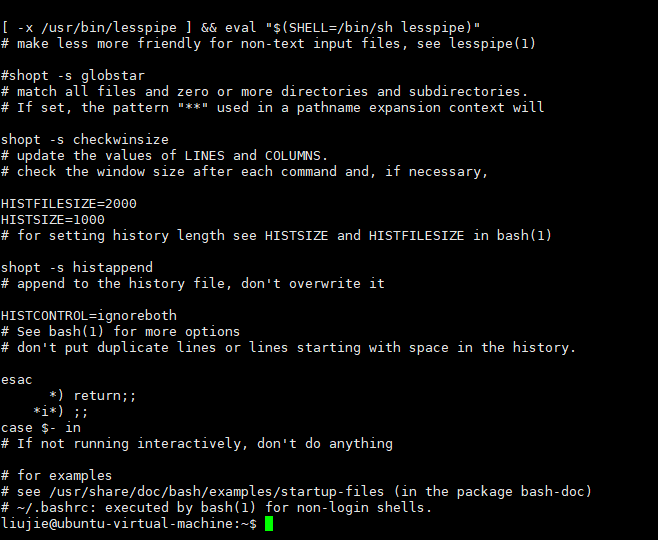


### 2.2.9 tac命令：反向查看文件内容

① 反向查看当前用户主文件夹下的.bashrc文件的内容

输入命令：tac /home/liujie/.bashrc

显示结果如下，与上面的cat命令明显不同，顺序相反

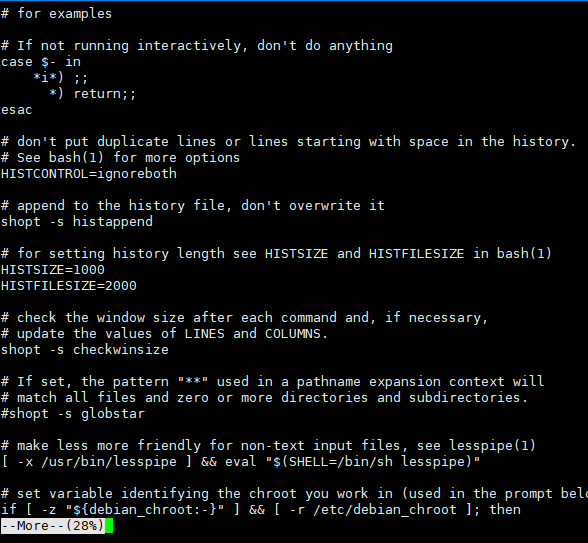


### 2.2.10 more命令：一页一页翻动查看

① 翻页查看当前用户主文件夹下的.bashrc文件的内容

输入命名：more /home/liujie/.bashrc

运行结果如下:

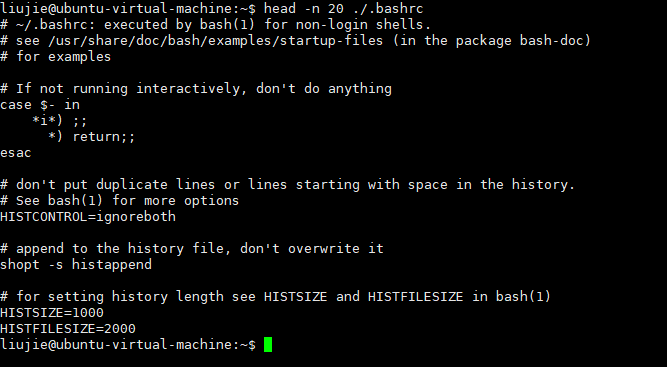


### 2.2.11 head命令：取出前面几行

① 查看当前用户主文件夹下.bashrc文件内容前20行

输入命令：head -n 20 /home/liujie/.bashrc

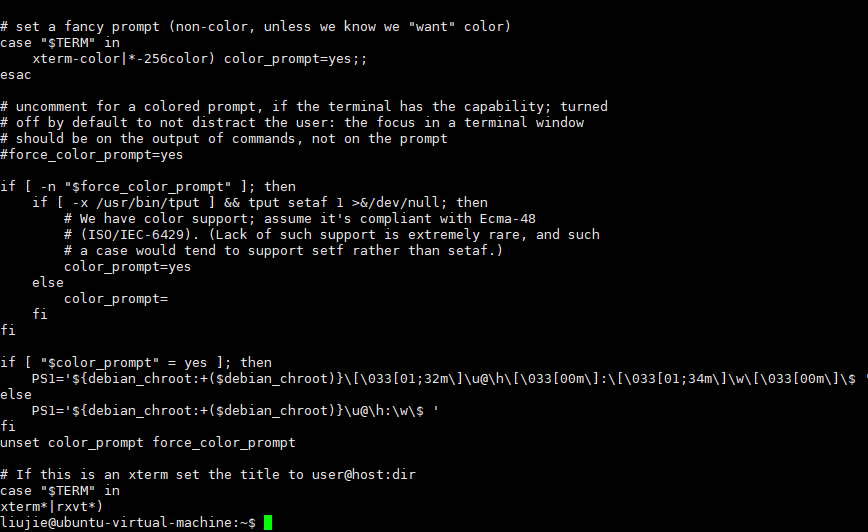
运行结果如下所示：



② 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

输入命令：head -n -50 /home/liujie/.bashrc

截图如下所示，符合预期结果，最后50行没有显示：

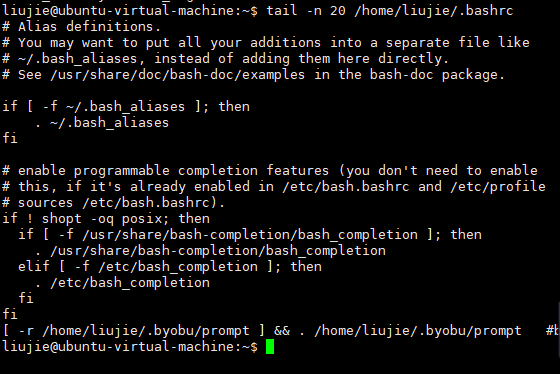


### 2.2.12 tail命令：取出后面几行

① 查看当前用户主文件夹下.bashrc文件内容最后20行

输入命令：tail -n 20 /home/liujie/.bashrc

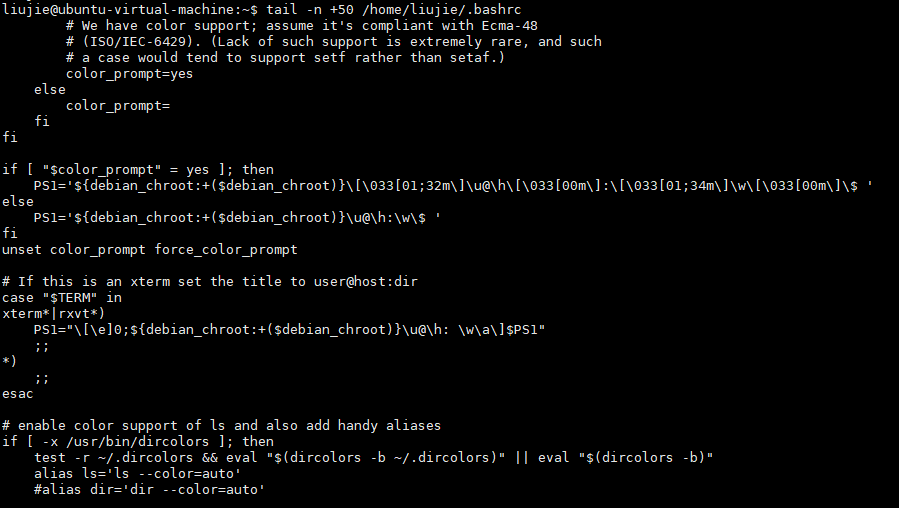
运行结果如下图所示，符合预期：



② 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

输入命令：tail -n +50 /home/liujie/.bashrc

截图如下，由于内容过多，只截取了开头部分：

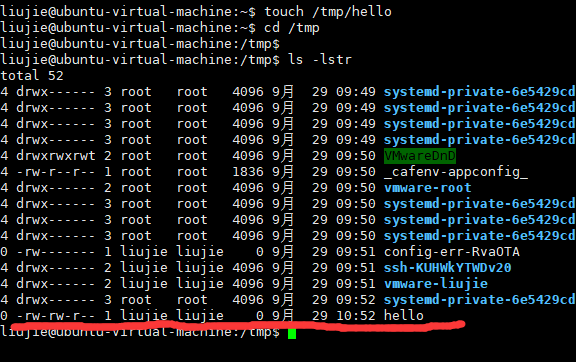


### 2.2.13 touch命令：修改文件时间或创建新文件

① 在“/tmp”目录下创建一个空文件hello，并查看文件时间

输入命令：touch /tmp/hello ls -lstr

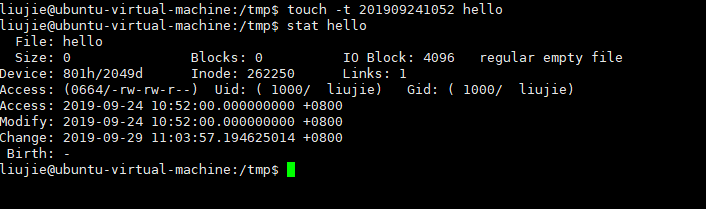
截图如下所示，时间显示为9月29 10:52



② 修改hello文件，将文件时间整为5天前

输入命令：touch -t 201909241052 hello

结果如下，时间变为了5天前

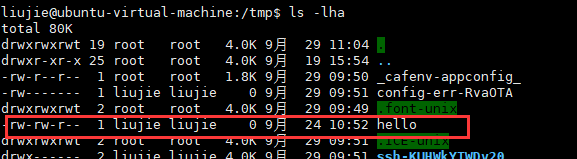


### 2.2.14 chown命令：修改文件所有者权限

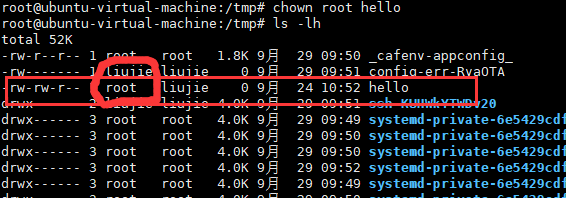
① 将hello文件所有者改为root帐号，并查看属性

输入命令：chown root hello

Hello之前权限及属性截图：



修改后的截图，此操作需要在root权限下执行：

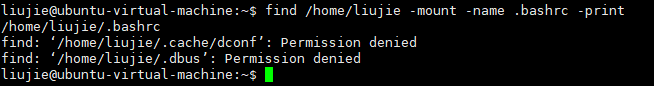


### 2.2.15 find命令：文件查找

① 找出主文件夹下文件名为.bashrc的文件

输入命令：find /home/liujie -mount -name .bashrc -print

运行结果如下，结果符合预期：



### 2.2.16 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

输入命令：mkdir /test tar -zcvf test.tar.gz /test

② 把上面的test.tar.gz压缩包，解压缩到“/tmp”目录

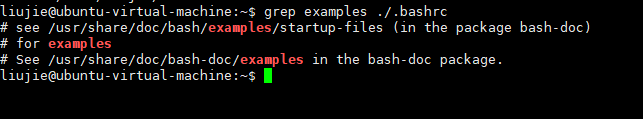
输入命令：tar -zxvf test.tar.gz -C /tmp

### 2.2.17 grep命令：查找字符串

① 从“～/.bashrc”文件中查找字符串'examples'

输入命令：grep examples ~/.bashrc

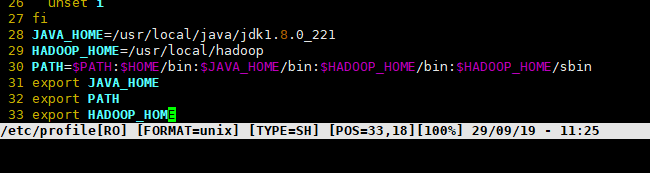
运行结果如下，结果符合预期：



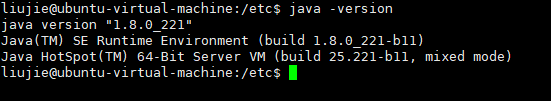
### 2.2.18 配置环境变量

① 请在“～/.bashrc”中设置，配置Java环境变量

截图如下所示：



Java版本为：



② 查看JAVA\_HOME变量的值

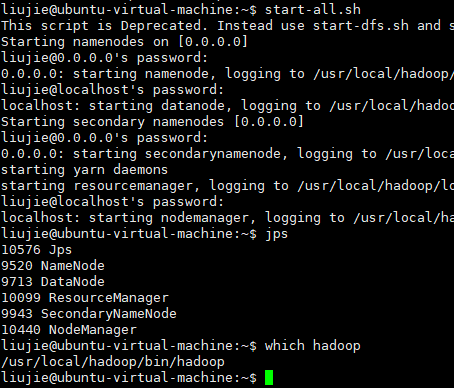
变量值如下,我将java存放到了usr/local中，在/etc/profile中配置的，不过道理都相同。



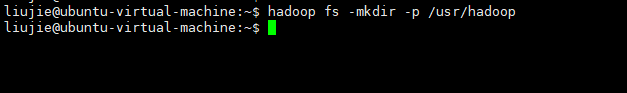
## 2.3 熟悉常用的Hadoop操作

### 2.3.1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”

**输入start-all.sh启动Hadoop**



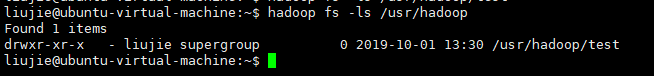
hadoop fs -mkdir -p /usr/hadoop



### 2.3.2接着在HDFS的目录“/user/hadoop”下，创建test文件夹，并查看文件列表

hadoop fs -mkdir -p /usr/hadoop/test

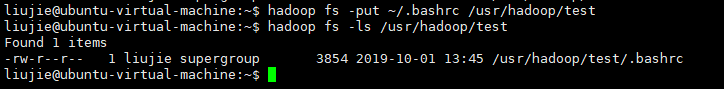
hadoop fs -ls /usr/hadoop



### 2.3.3将Linux系统本地的“～/.bashrc”文件上传到HDFS的test文件夹中，并查看test

hadoop fs -put ~/.bashrc /usr/hadoop/test

hadoop fs -ls /usr/hadoop/test



### 2.3.4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

hadoop fs -get /usr/hadoop/test /usr/local/hadoop/

ls /usr/local/hadoop/



# 3. 实验总结

总体来说还是比较顺利的，Ubuntu和虚拟机以前都安装过了，主要安装了Hadoop，按照网上的教程进行了Hadoop伪分布式的安装。在安装的过程中应该注意你安装时所使用的的用户，有些文件需要授予权限，不然后面运行Hadoop时会出现很多意想不到的问题，还有在安装的过程中应该认真仔细，跟着步骤一步一步来，要注意细节，如果漏掉一些步骤的话可能就不会安装成功了。常用的Linux操作命令多实践一下，动手敲一敲就可以掌握了。想吐槽一下在虚拟机上运行Linux没有老师说的那么不堪(哈哈，没有冒犯之意)，下载慢的问题改一下下载源就和本机下载速度差不多，而且也不需要在虚拟上下载东西，用xsheel，ssh到Ubuntu就可以了，在Windows端控制就好了，下载文件的话，在Windows下载好，用xftp传输过去就好了，相关应用程序界面等问题xmanager也可以解决。

自己在Vultr租服务器，速度方面确实很快，但是bash输入命令反应太慢了，所以还是感觉虚拟机用起来比较好。不过有个好处是可以fq，还是可以学到很多东西的。总之这次实验学到了很多，也锻炼了自己解决问题的能力，受益匪浅。