

**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 9 月** |

# 实验目的和要求

## 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。

## 实验软硬件环境

① 操作系统：Linux Ubuntu 16.04.6 LTS

② Java：jdk1.8.0\_221

③ Apache Hadoop：3.2.0

## 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。

# 实验记录

## 安装Hadoop

### 创建hadoop用户

①创建hadoop用户，并使用bash作为其shell

sudo useradd -m hadoop -s /bin/bash

②设置hadoop账户的密码

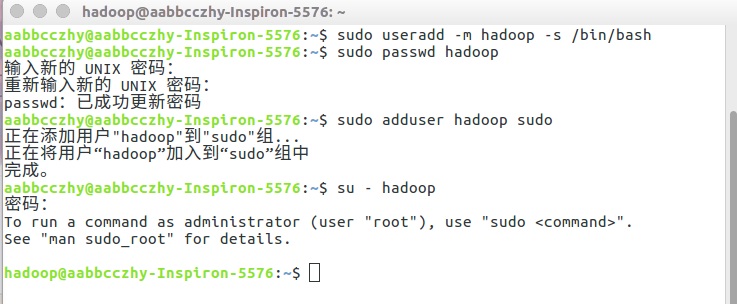
sudo passwd hadoop

③为hadoop用户添加管理员权限

sudo adduser hadoop sudo

④切换到hadoop用户

su - hadoop



如图，已经成功创建了hadoop用户并已成功切换到hadoop用户。

### Java的安装

①先从Oracle官网上下载jdk，我这边下载的是jdk1.8.0\_221.tar.gz；

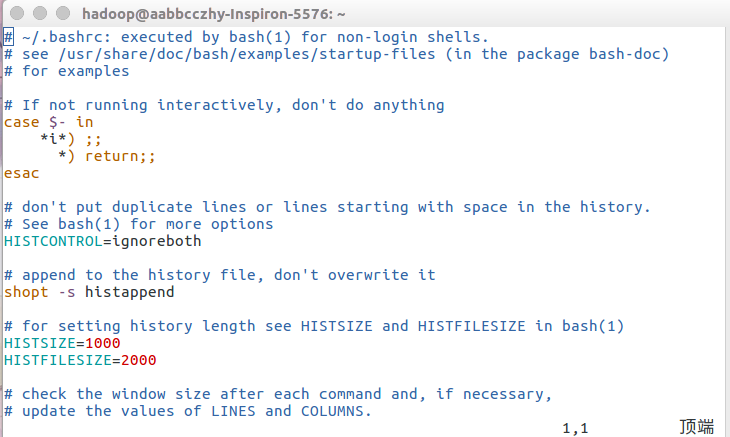
②将其解压到/usr/local目录下；

sudo tar -zxvf jdk1.8.0\_221.tar.gz -C /usr/local

③配置Java环境变量：

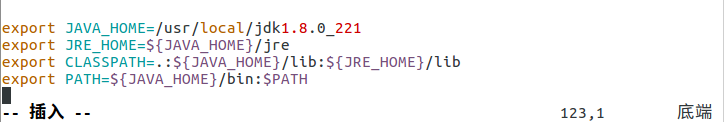
由于我使用的是bash（一般来说都是默认都是bash），bash 的配置文件在~/.bashrc，由于ubuntu自带编辑器是gedit，所以我们可以用gedit去打开这个配置文件。当然，你也可以使用vim，先使用sudo apt-get install vim安装。我这边已经装好了vim，所以还是用vim来修改配置文件。

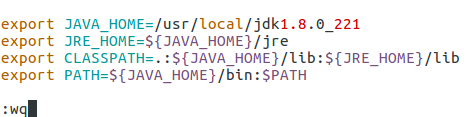
终端输入：sudo vim ~/.bashrc（这边需要注意的是，一定要带上sudo，虽然不带sudo也能进入到配置文件，但是那只是以只读的形式进入，会保存不了配置文件）



如图，我们就进入到了.bashrc这个文件里。

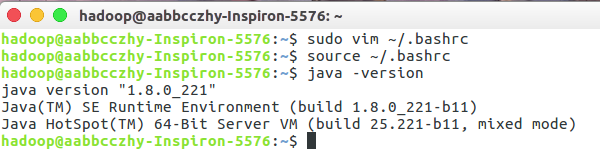
然后开始配置我们的Java环境变量：





如图，要进入编辑模式，按键盘上的i，然后配置环境变量如图，配置完成后按ESC退出编辑模式，然后输入:wq保存并退出。

终端输入source ~/.bashrc应用当前配置，最后输入java -version检查环境变量是否配置成功。



如图，JDK已经配置完成了。

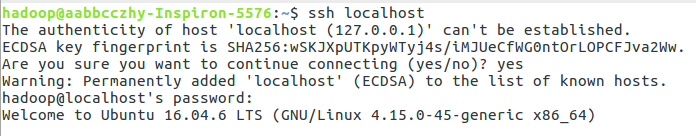
### SSH登录权限设置

①首先ubuntu只有ssh客户端，并没有服务端，需要先安装服务端。

sudo apt-get install ssh-server

②先进入ssh

ssh localhost



可以发现，需要输入密码才能进入，Hadoop的名称节点为了能启动集群中所有机器，需要配置好无密码登录。

③先退出ssh

exit

④切换到~/.ssh/目录下

cd ~/.ssh/

⑤生成SSH密钥

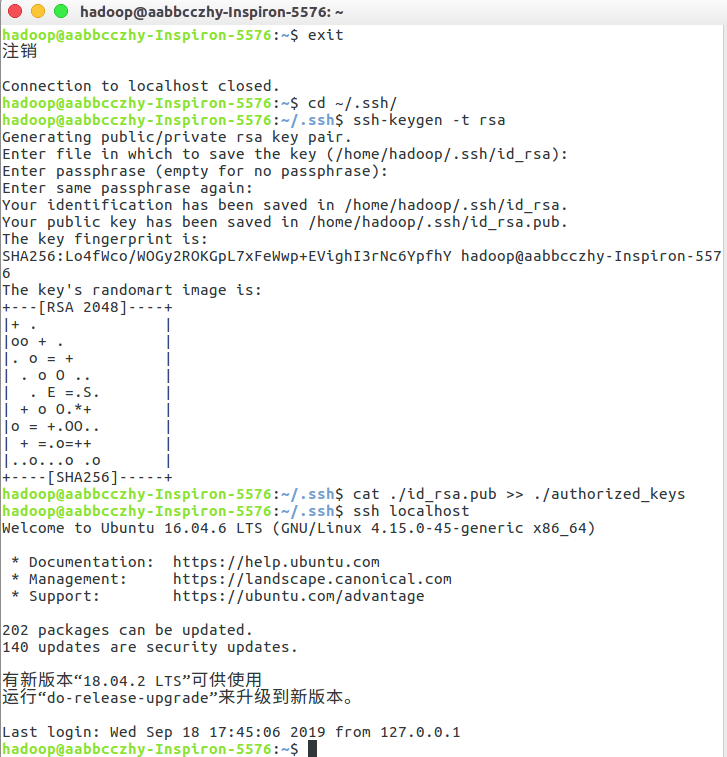
ssh-keygen -t rsa

三次回车后，成功生成密钥。

⑥加入授权

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

⑦这时候再运行ssh localhost就会发现不需要密码也能成功登录了。



### 安装Hadoop

经历了那么长时间，终于可以开始装hadoop了。首先先从官网上下载hadoop，我这边下载的是hadoop3.2.0.tar.gz

①首先解压到/usr/local/目录下并改名为hadoop

sudo tar -zxvf hadoop-3.2.0.tar.gz -C /usr/local/

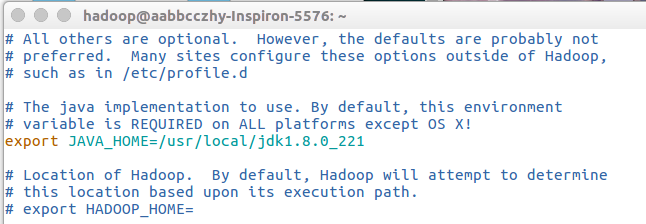
sudo mv hadoop-3.2.0 hadoop

②首先要更改hadoop-env.sh文件，加入JDK目录

老办法，使用vim

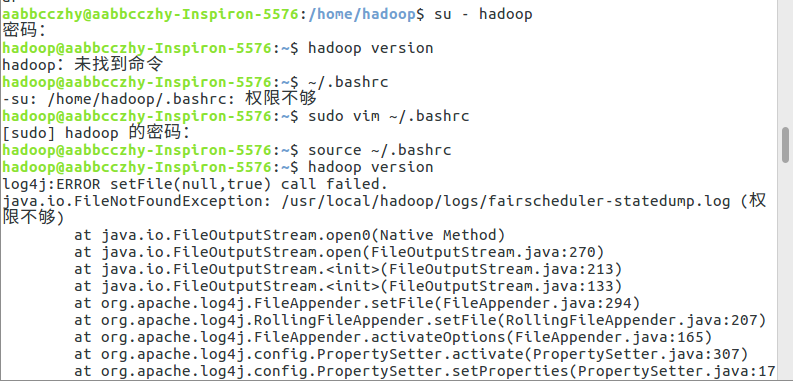
sudo vim /usr/local/hadoop/etc/hadoop/hadoop-env.sh

加入JDK目录的配置，保存退出



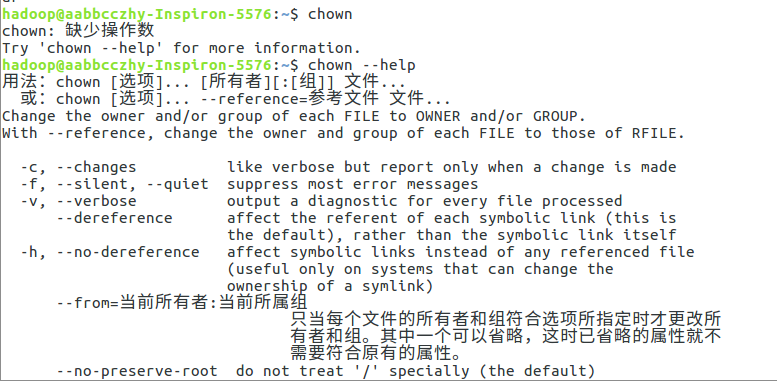
③试着查看版本信息

hadoop version



这边报了个错误，抛出了一个FileNotFoundException的异常，后面写着权限不够，说明我们没有将权限交给hadoop这个用户

我只记得授权是chown，但不知道咋用，先help一下，然后成功完成将hadoop文件夹及其子文件夹授权给hadoop用户。

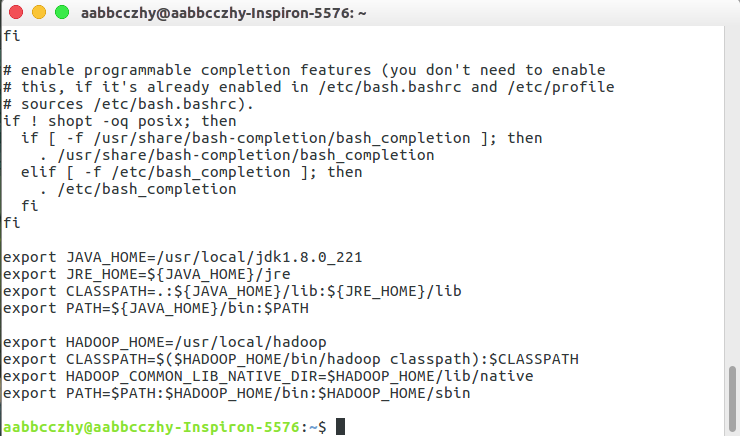




如图所示，hadoop version显示出正确的版本信息。

这里我直接hadoop version是因为我配置了hadoop的环境变量，如果没有配置环境变量，需要执行/usr/local/hadoop/bin/hadoop version 才可以的。

还是在~/.bashrc修改，环境变量如下：



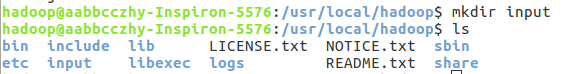
需要注意的是，不用用户的.bashrc文件是不一样的，像我一开始就是把环境变量配置在我自己的aabbcczhy用户下，当我切换到hadoop用户的时候，环境变量就没有效果了，这点需要特别注意！

### 测试Hadoop是否安装成功

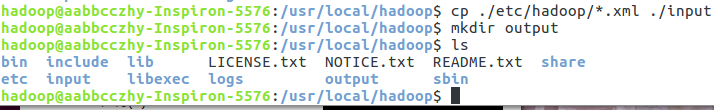
Hadoop文档提供了例子可以用来测试，下面运行WordCount的例子来检测Hadoop是否安装成功。

①首先，在hadoop目录下新建文件夹input

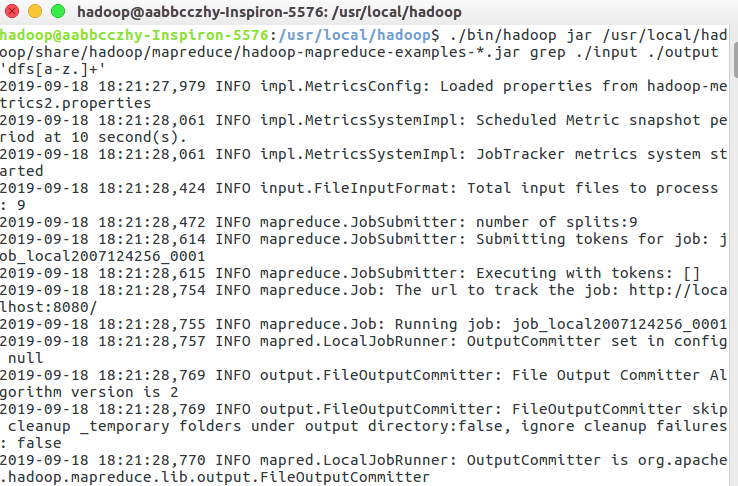
mkdir input



②将etc/hadoop文件夹下的配置文件拷贝进input文件夹



③执行字符串的查找，并输出到output



④输入output文件夹里的内容，得到结果 1 dfsadmin，该结果符合正则表达式‘dfs[a-z.]+’



经过上述测试，说明hadoop已经安装成功。

### Hadoop伪分布式安装

①修改配置文件core-site.xml

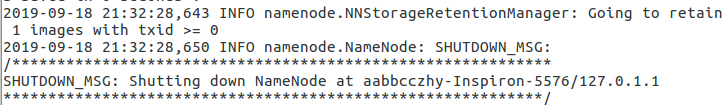


②修改配置文件hdfs-site.xml

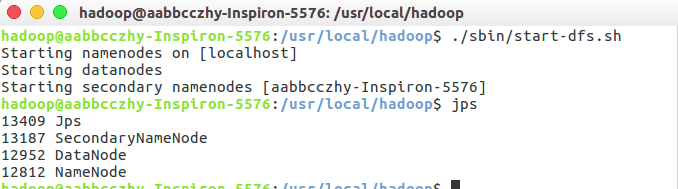


③初始化文件系统

./bin/hadoop namenode -format



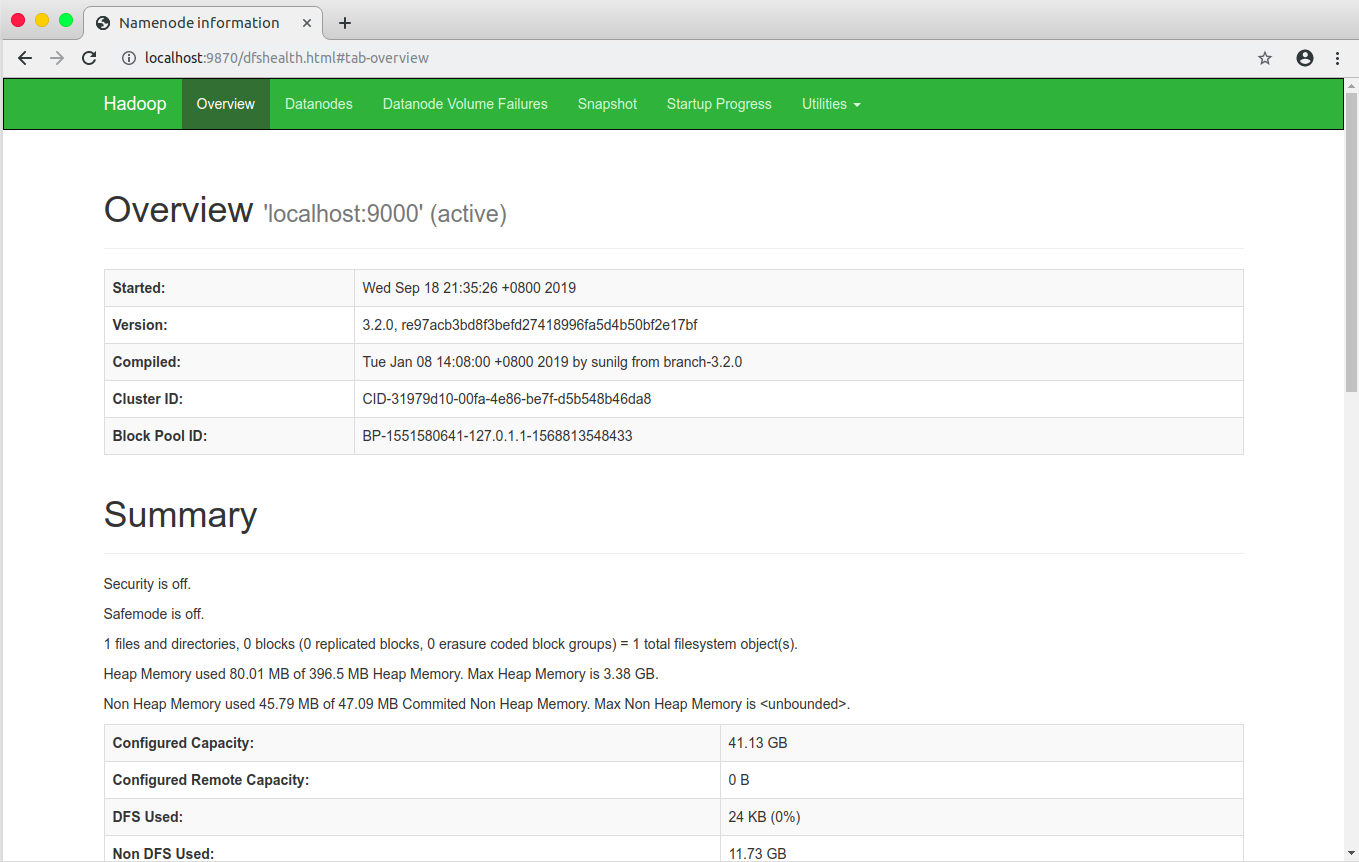
④启动集群，并查看所有Java进程



⑤访问web页面，可以查看Hadoop信息

打开浏览器，输入http://localhost:9870

（需要注意的是，在hadoop3.x版本，端口号已经变为9870，不再是2.x版本的50070。这个问题困扰了我很久，一直以为我没安装成功）



至此，Hadoop伪分布式安装完成。

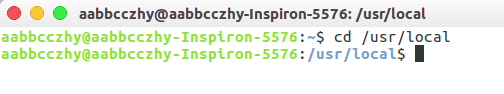
## Linux操作

### cd命令

① 切换到目录“/usr/local”

输入命令“cd /usr/local/”

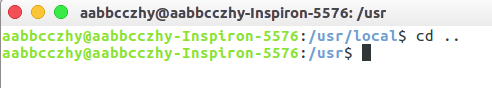
命令运行结果如下图所示，目录切换到“/usr/local”，结果符合预期。



②切换到当前目录的上一级目录

输入命令“cd ..”

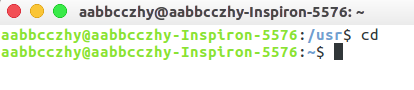
命令运行结果如下图所示，目录切换到“/usr”，结果符合预期。



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”

命令运行结果如下图所示，目录切换到主文件夹“~”，结果符合预期。

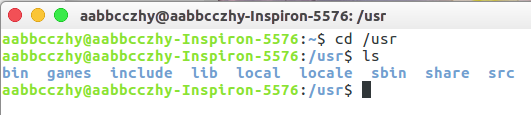


### ls命令：查看文件与目录

①查看目录“/usr”下的所有文件和目录

先进入“/usr”目录，输入命令“cd /usr”

然后输入“ls”，运行结果如下图，结果符合预期。



### mkdir命令：新建目录

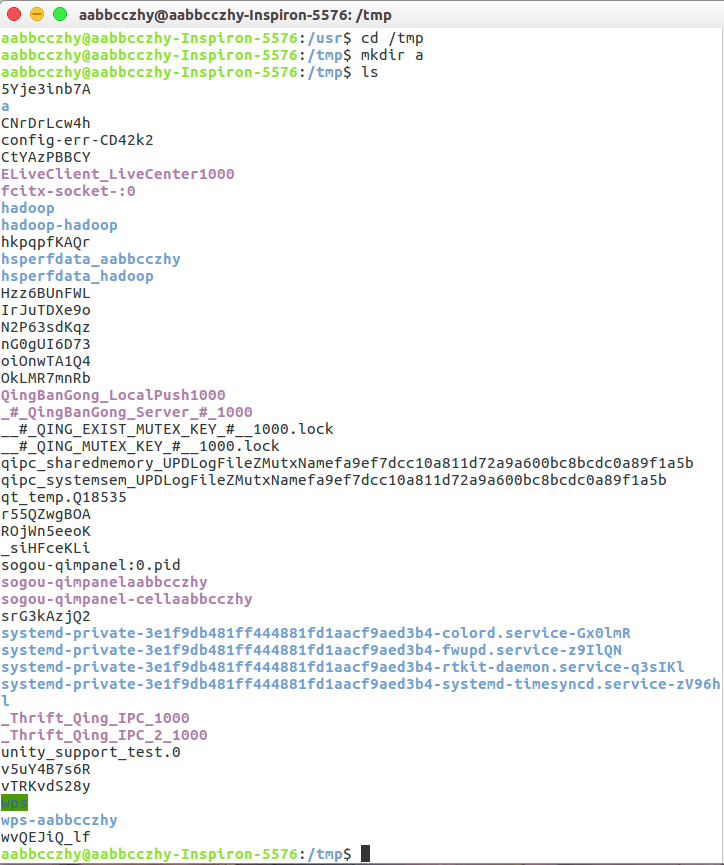
①进入“/tmp”目录，创建一个名为“a”的目录，并查看“/tmp”目录下已经存在哪些目录

输入命令“cd /tmp”，切换到“/tmp”目录

输入命令“mkdir a”，创建一个名为“a”的目录

输入命令“ls”,查看“/tmp”目录下已经存在的目录

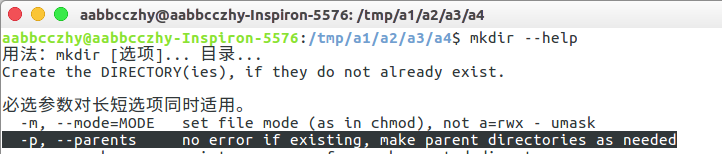
运行结果如下图，实验结果符合预期。



②进入“/tmp”目录，创建目录“a1/a2/a3/a4”

输入命令“mkdir -p a1/a2/a3/a4”，创建4级文件夹，其中-p参数表示从父目录开始，如果存在则继续，不存在则创建文件夹。

输入命令“cd a1/a2/a3/a4”，可以进入到a4文件夹，实验结果符合预期。





### 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“cd /tmp”，切换到“/tmp”目录

输入命令“rmdir a”，删除目录a

执行“ls”，发现a目录已经被删除，实验结果符合预期。

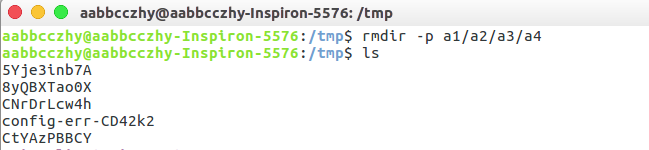


②删除上面创建的目录“a1/a2/a3/a4” （在“/tmp”目录下面），然后查看

“/tmp”目录下面存在哪些目录

输入命令“rmdir -p a1/a2/a3/a4”，删除创建的目录“a1/a2/a3/a4”

输入命令“ls”，a1目录已经不存在，实验结果符合预期。

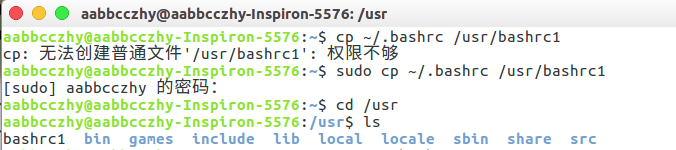


### cp命令：复制文件或目录

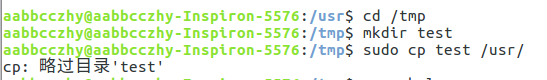
①将当前用户的主文件夹下的文件.bashrc复制到目录“/usr”下，并重命名为bashrc1

cp ~/.bashrc /usr/bashrc1

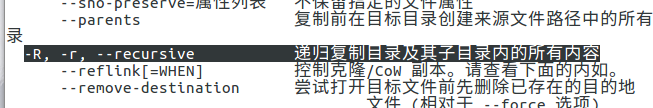
(可能会出现权限不足的问题，在命令前加上sudo即可，下同)

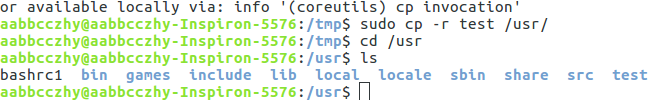


②在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下



看似一切顺利，突然就略过目录了，看了一下帮助解决了问题

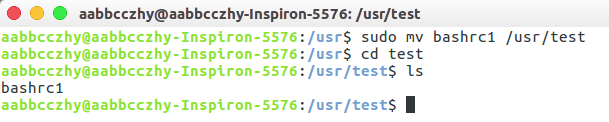




### mv命令：移动文件与目录，或更名

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

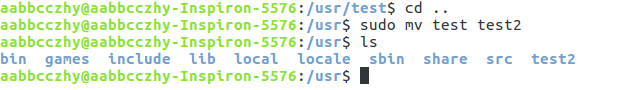
sudo mv bashrc1 /usr/test



②将“/usr”目录下的test目录重命名为test2

cd ..

sudo mv test test2



### rm命令：移除文件或目录

①将“/usr/test2”目录下的bashrc1文件删除

cd test2

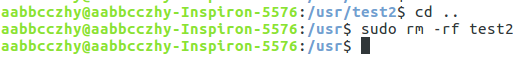
sudo rm bashrc1



②将“/usr”目录下的test2目录删除

cd ..

sudo rm -rf test2

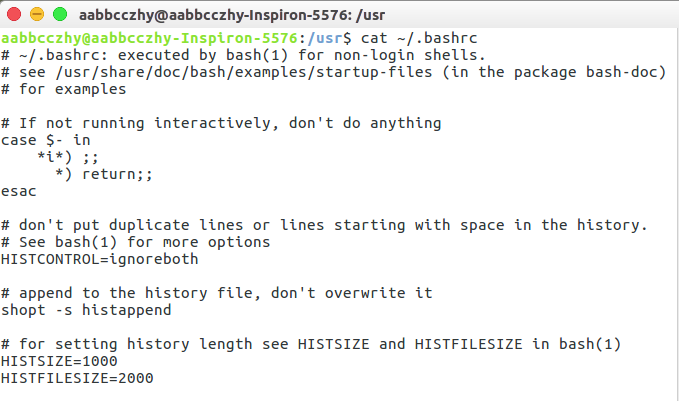


如果test2是空目录，用rmdir效果一样，如果不是空目录，则要用rm -rf删。

### 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

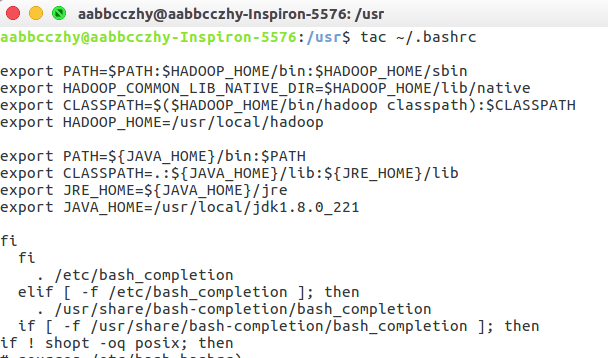


（.bashrc的内容有点多，没有都截图下来）

### 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

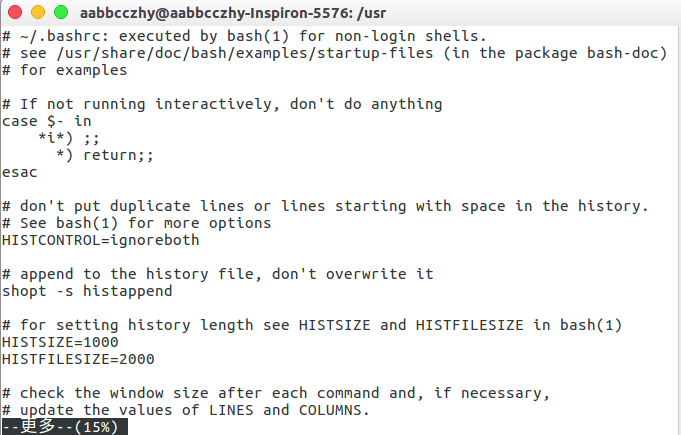


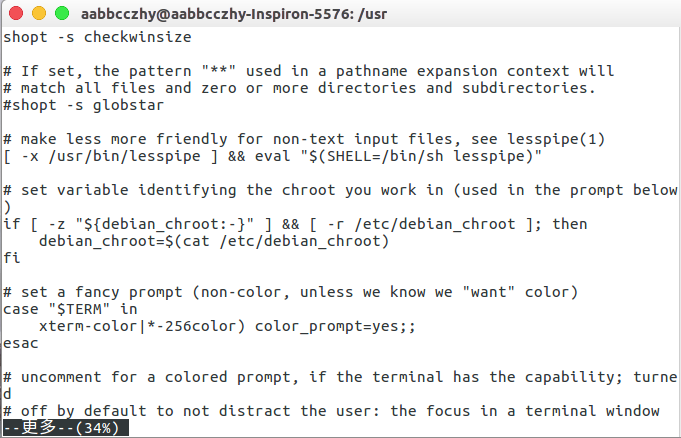
反向查看，可以很明显看到我配好的Java和Hadoop的环境变量。

### 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



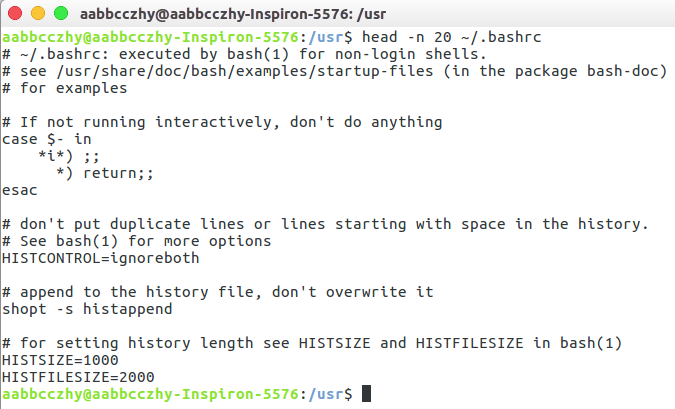


可以按空格键进入下一页查看更多内容。

### head命令：取出前面几行

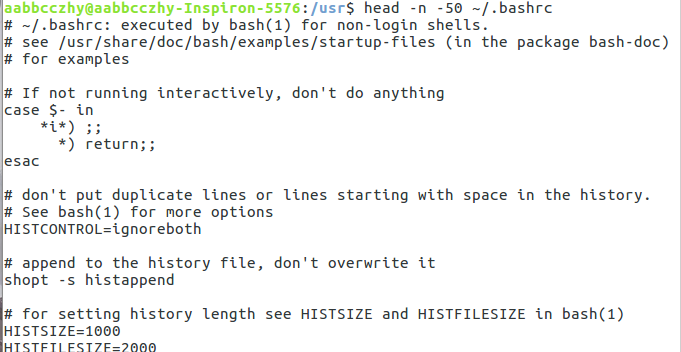
①查看当前用户主文件夹下.bashrc文件内容前20行

head -n 20 ~/.bashrc



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

head -n -50 ~/.bashrc

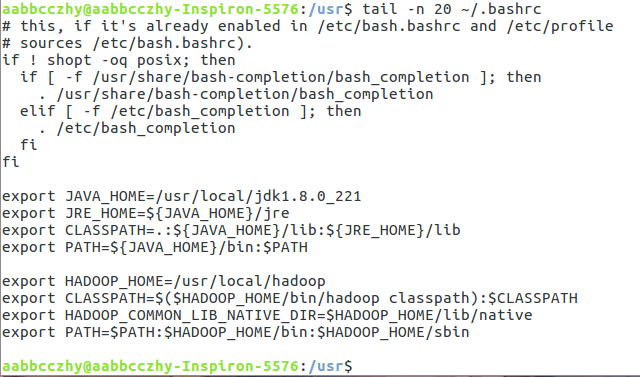


（只截取一部分，经验证，确实为跳过后50行输出）

### tail命令：取出后面几行

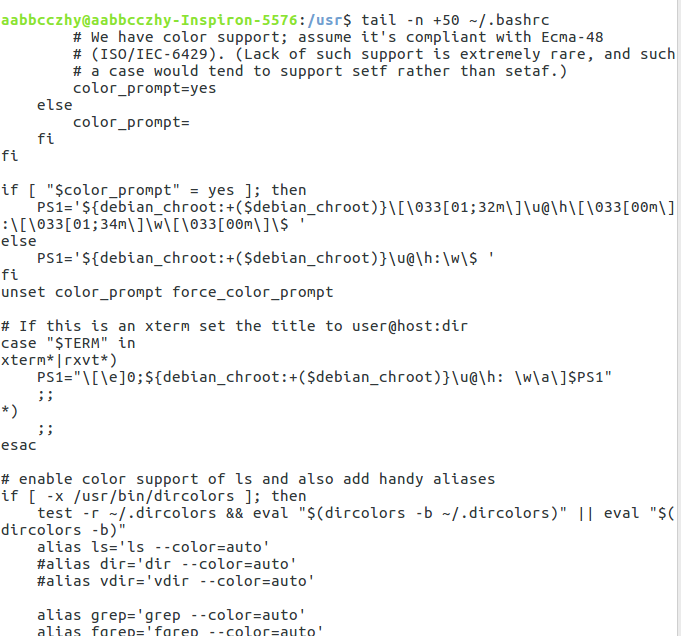
①查看当前用户主文件夹下.bashrc文件内容最后20行

tail -n 20 ~/.bashrc



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

tail -n +50 ~/.bashrc



（只截取一部分，经验证，确实是从第50行开始输出）

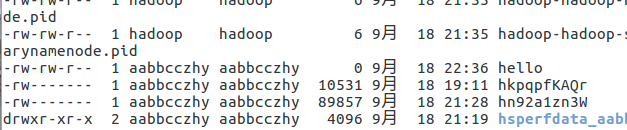
### touch命令：修改文件时间或创建新文件

①在“/tmp”目录下创建一个空文件hello，并查看文件时间

touch hello

ll



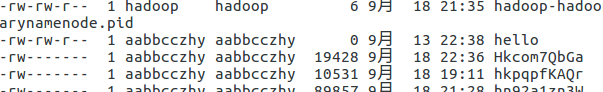


②修改hello文件，将文件时间整为5天前

touch -d “5 days ago” hello

ll





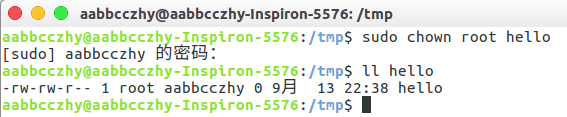
hello文件时间变为5天前，符合实验预期。

### 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

sudo chown root hello

ll hello



hello文件的所有者变为root，符合实验预期。

### find命令：文件查找

①找出主文件夹下文件名为.bashrc的文件

sudo find ~ -name .bashrc



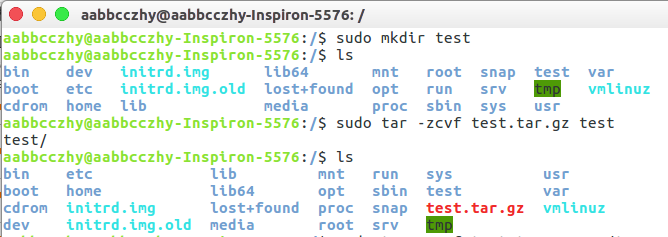
输出.bashrc的文件路径，符合实验预期。

### 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

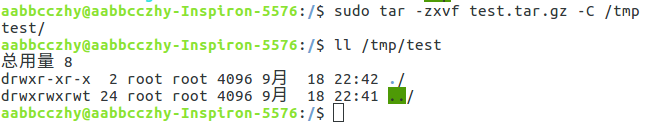
sudo mkdir test

sudo tar -zcvf test.tar.gz test



②把上面的test.tar.gz压缩包，解压缩到“/tmp”目录

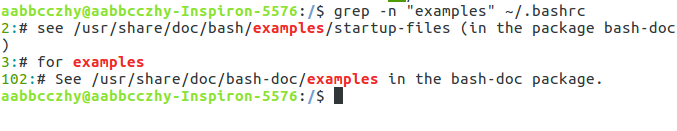
sudo tar -zxvf test.tar.gz -C /tmp



### grep命令：查找字符串

①从“～/.bashrc”文件中查找字符串'examples'

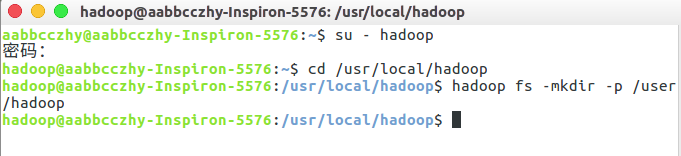
grep -n “examples” ~/.bashrc



（环境变量配置在Hadoop安装的时候已经详细说明，此处不再重复）

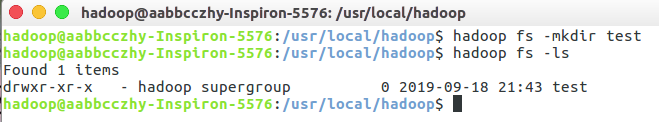
## Hadoop操作

### 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”



由于我这边已经配置好了环境变量，所以直接可以使用hadoop fs ...，如果没有配环境变量则需要在/usr/local/hadoop目录下使用命令“./bin/hadoop fs ...”

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



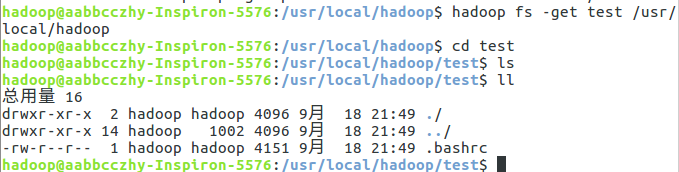
创建文件夹就比较简单了，需要注意的就是如果你没有执行第一步，也就是创建用户目录/user/hadoop，那么你直接运行“hadoop fs -ls”会提示你没有那个文件或目录。

通过这一步也可以看出，hadoop的HDFS目录将/user/hadoop作为其主目录，因此，你执行“hadoop fs -mkdir test”的时候，test文件夹是被创建在/user/hadoop下的。

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



### 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



这边是将我刚刚上传的.bashrc文件复制到本地/usr/local/hadoop文件夹，为了验证成功与否，直接去这个目录下查看，果然成功复制到了该目录。（.bashrc是隐藏文件，如果看不到的话可以用ll命令或者ls -a,如果你是在图形界面看的话，可以用Ctrl + H显示隐藏文件）

# 实验总结

本次实验遇到的问题主要有：

①权限问题。由于权限没有分配给hadoop用户，以至于在运行hadoop version的时候程序抛出了FileNotFound Exception的异常，这个问题在网上找了很久的资料才得以解决。

②低级错误。在进行实验的第三题的时候，我忘记自己没有进行Hadoop的伪分布式安装，使得我一度怀疑我装的Hadoop是不是有问题。

总而言之，本次实验收获不少，也懂了很多Linux与Windows之间不同的地方，希望在以后的学习中，能对Linux 的使用更为熟悉。