

**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：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26** |

# 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

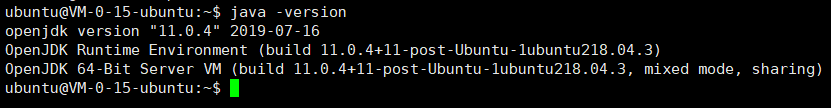
## 2.1安装Hadoop

### 2.1.1 服务器

在安装hadoop之前需要先有一个服务器或者一个linux系统的虚拟机。我用的是腾讯云的服务器。

### 2.1.2 必备软件

1、必须安装Java.

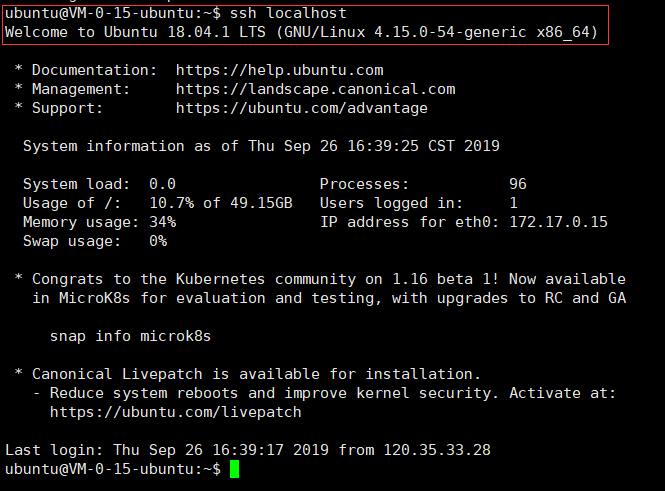


2、如果要使用可选的启动和停止脚本，则必须安装ssh且sshd必须正在运行以使用管理远程hadoop守护程序的hadoop脚本。另外，建议也安装pdsh以更好地进行ssh资源管理。

命令：

$ sudo apt-get install ssh

$ sudo apt-get install pdsh



### 2.1.3 准备启动hadoop集群

解压下载的hadoop发行版。在发行版中，编辑文件etc/Hadoop/Hadoop-env.sh定义一些参数。

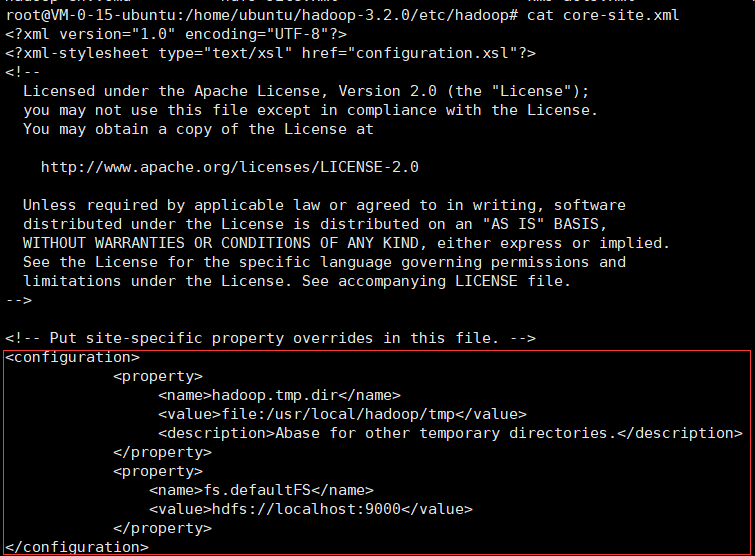
# set to the root of your Java installation

export JAVA\_HOME=/usr/java/latest

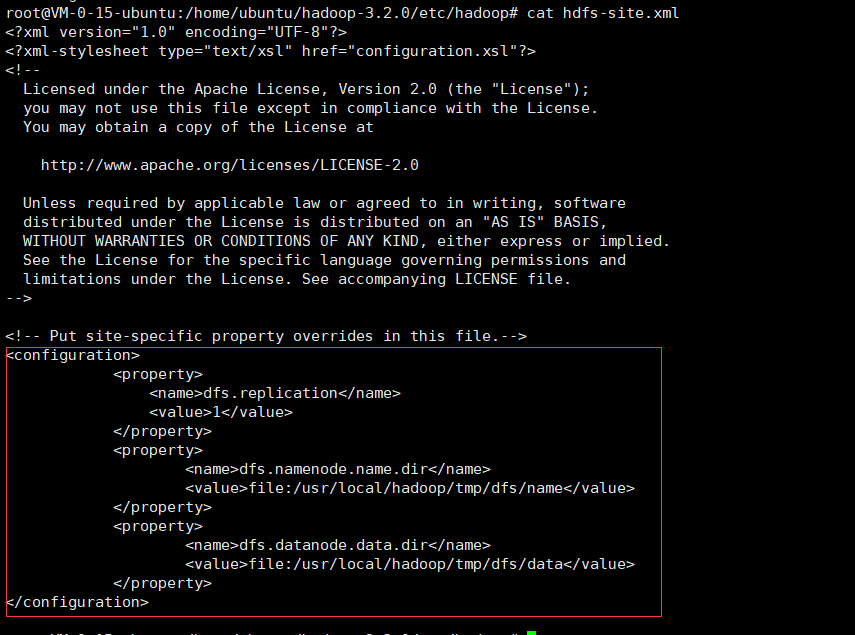
## 2.2 伪分布式安装

### 2.2.1 配置文件core-site.xml和hdfs-site.xml

core-site.xml文件



hdfs-site.xml



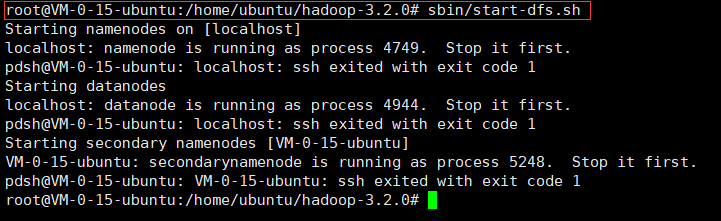
### 2.2.2 格式化文件系统

命令：$bin/hdfs namenode-format

如果前面的配置文件中出现了错误，将格式化失败，所以在配置文件输入代码时，要千万仔细才可以。

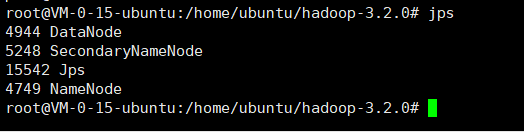
### 2.2.3 启动进程

命令：$./bin/start-all.sh



### 2.2.4 查看所有的Java进程

命令：jps

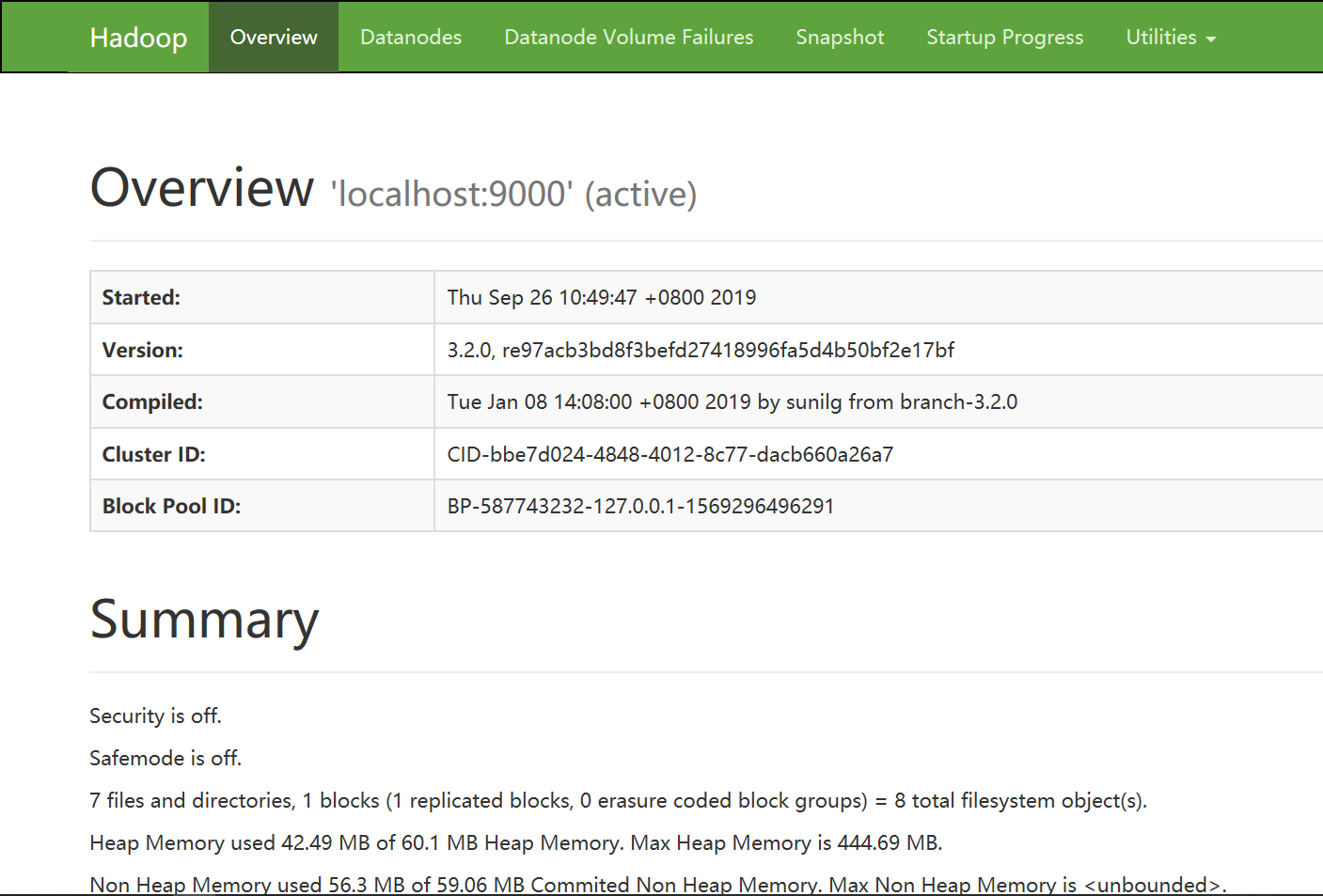


在进行jps检查时，如果前面的配置文件中没有配置好DataNode以及NameNode的话，在执行jps命令时，可能就只会出现Jps一条信息。

## 2.3 激动人心的时刻

在浏览器中的网址栏输入自己的主机公网:9870

如：106.54.76.26:9870/

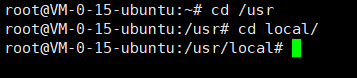


出现以上页面，就代表hadoop安装成功！！！喜大普奔！

## 2.2 Linux操作

### 2.2.1 cd命令

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



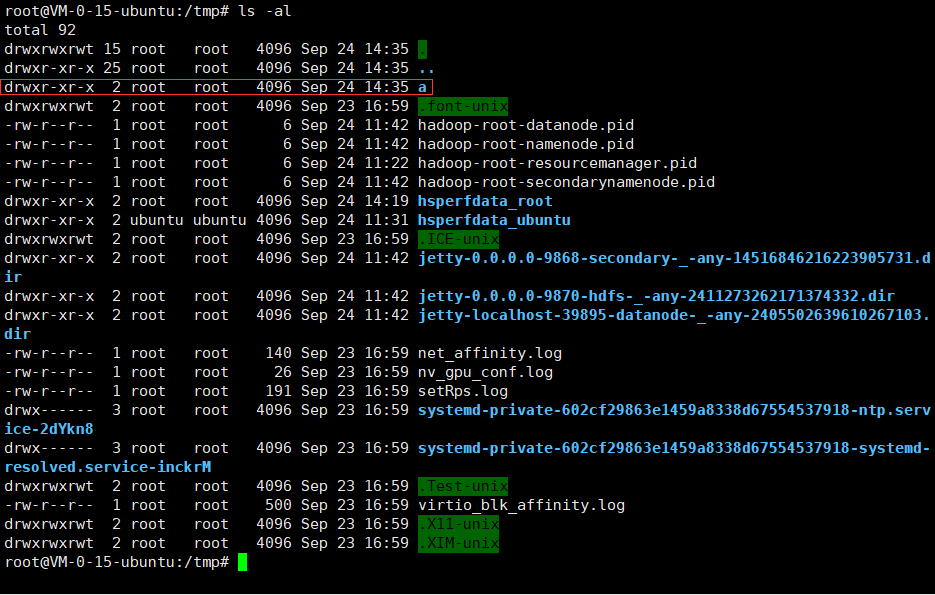
### 2.2.2 ls命令

1. 查看目录“/usr”下的所有文件和目录

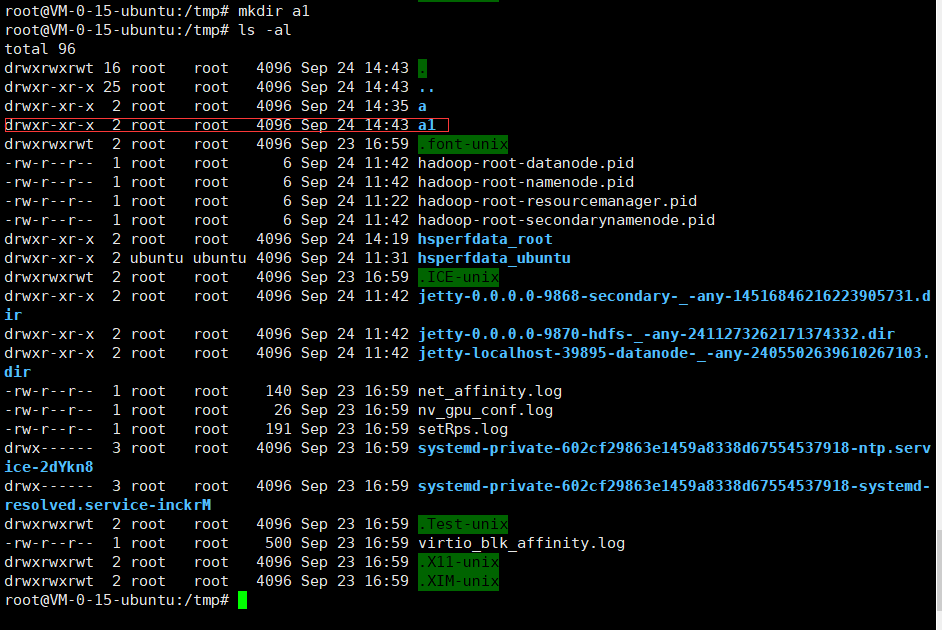


### 2.2.3 mkdir命令

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

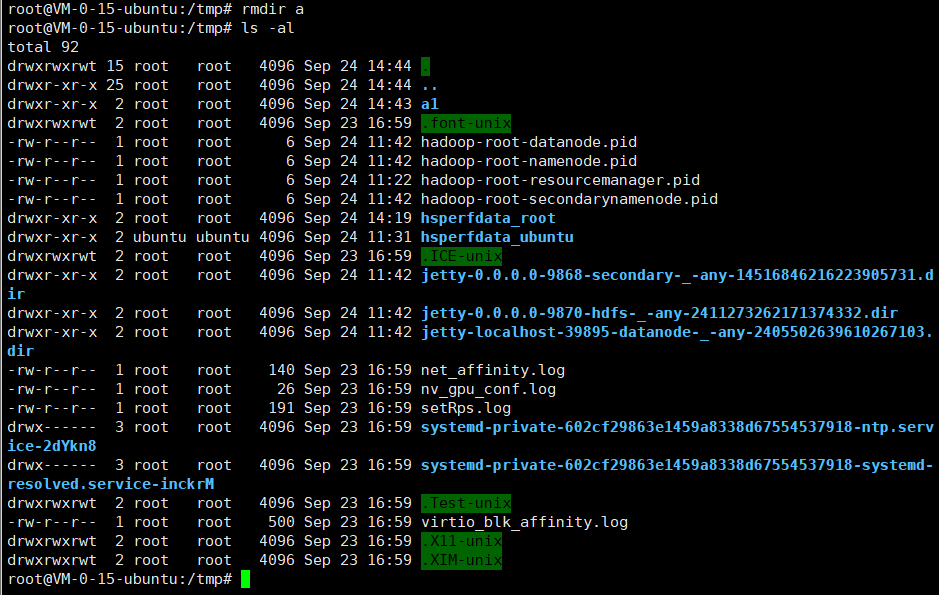


1. 进入“/tmp”目录，创建目录“a1/a2/a3/a4”

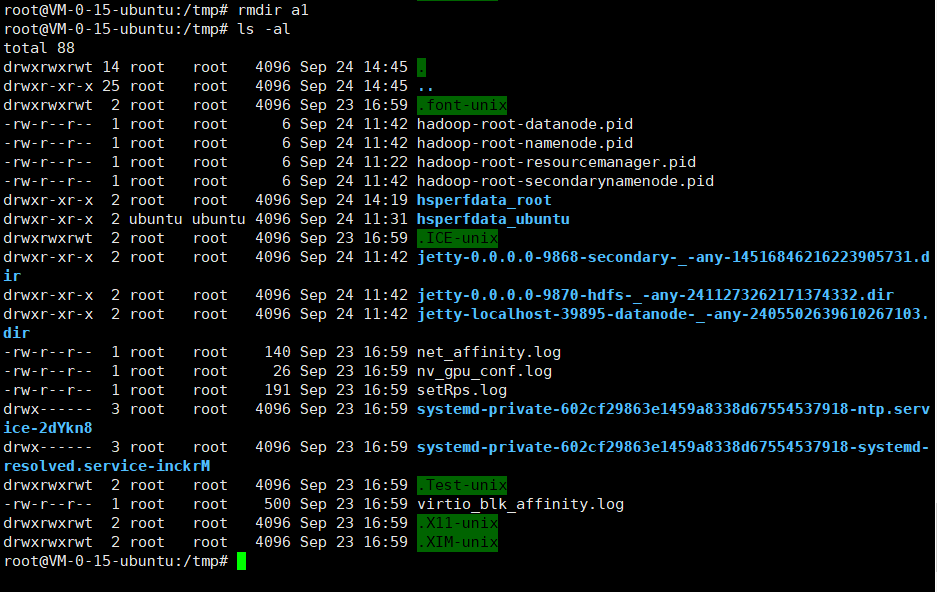


### 2.2.4 rmdir命令

1. 将上面创建的目录a（在“/tmp”目录下面）删除



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



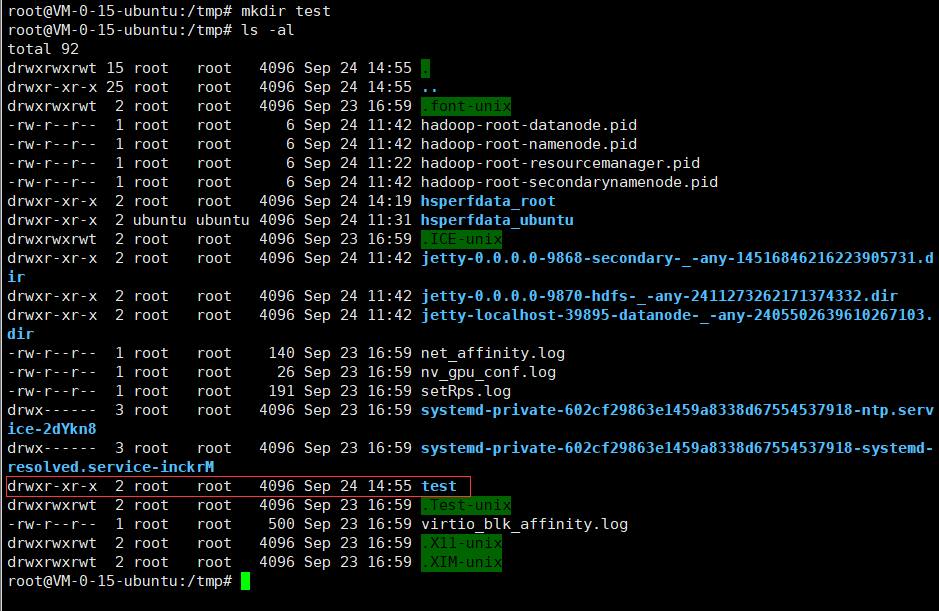
### 2.2.5 cp命令

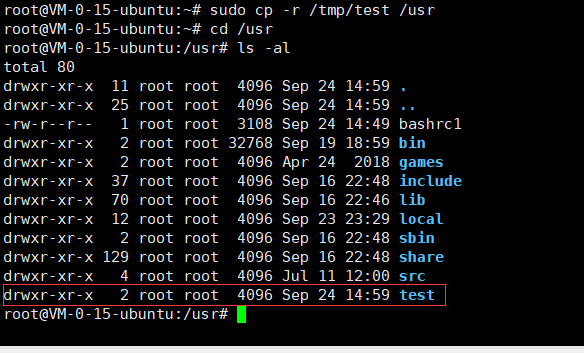
1. 将当前用户的主文件夹下的文件.bashrc复制到目录“/usr”下，并重命名为bashrc1





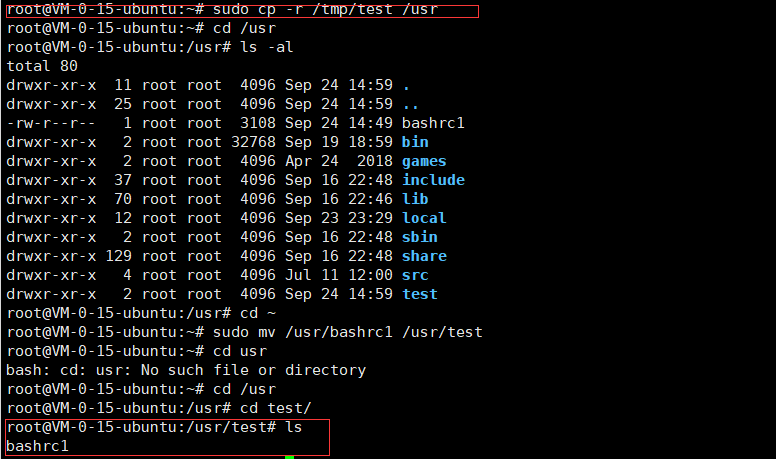
1. 在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下



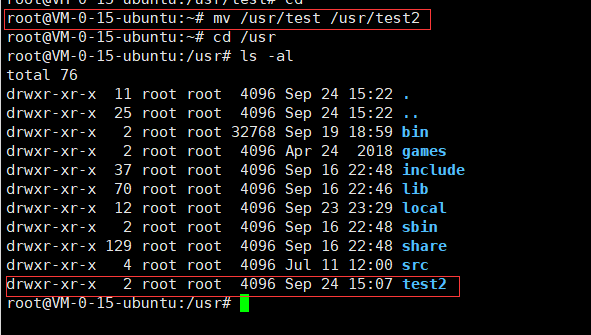


### 2.2.6 mv命令

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

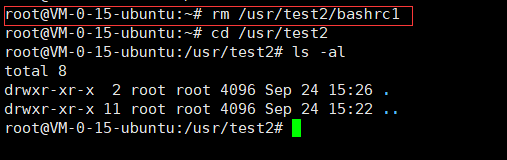


1. 将“/usr”目录下的test目录重命名为test2

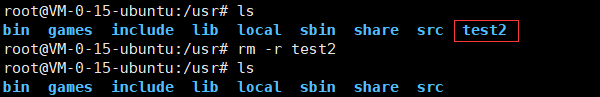


### 2.2.7 rm命令

1. 将“/usr/test2”目录下的bashrc1文件删除

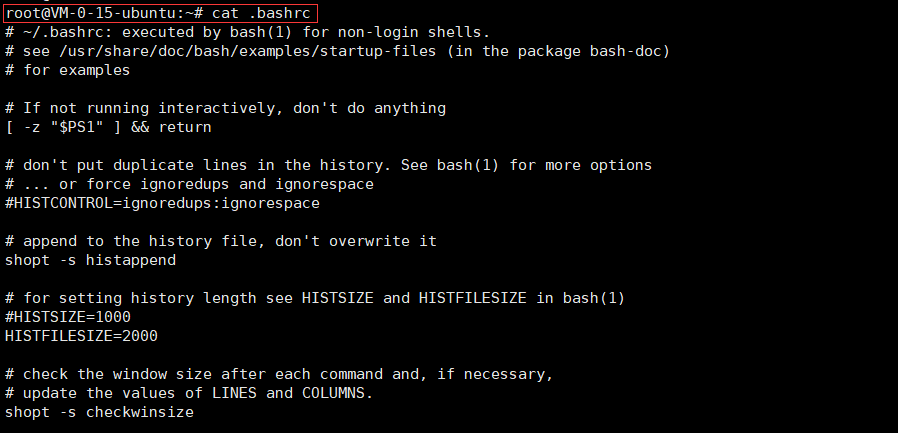


1. 将“/usr”目录下的test2目录删除



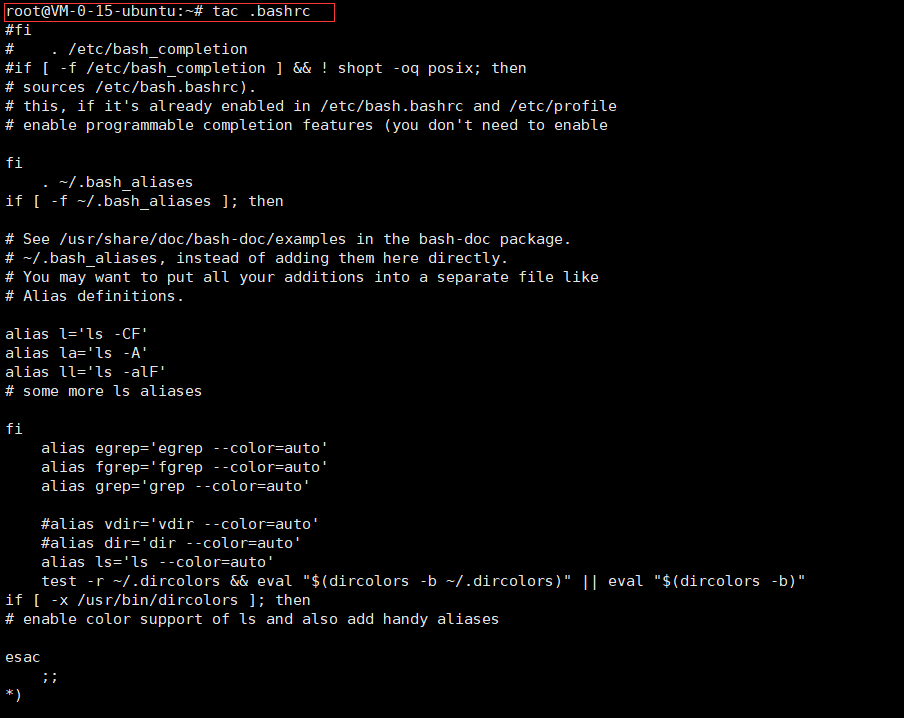
### 2.2.8 cat命令

1. 查看当前用户主文件夹下的.bashrc文件内容



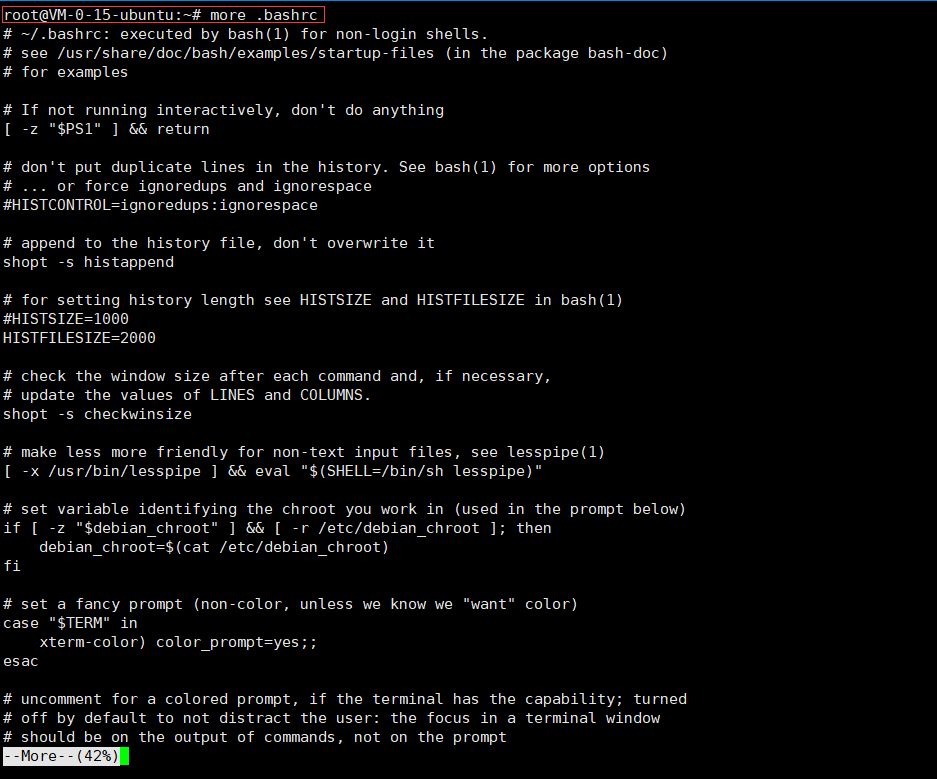
### 2.2.9 tac命令

1. 反向查看当前用户主文件夹下的.bashrc文件的内容



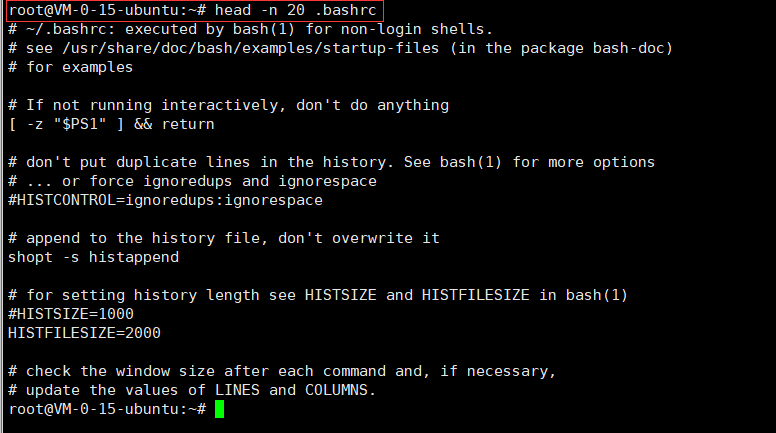
### 2.2.10 more命令

1. 翻页查看当前用户主文件夹下的.bashrc文件的内容

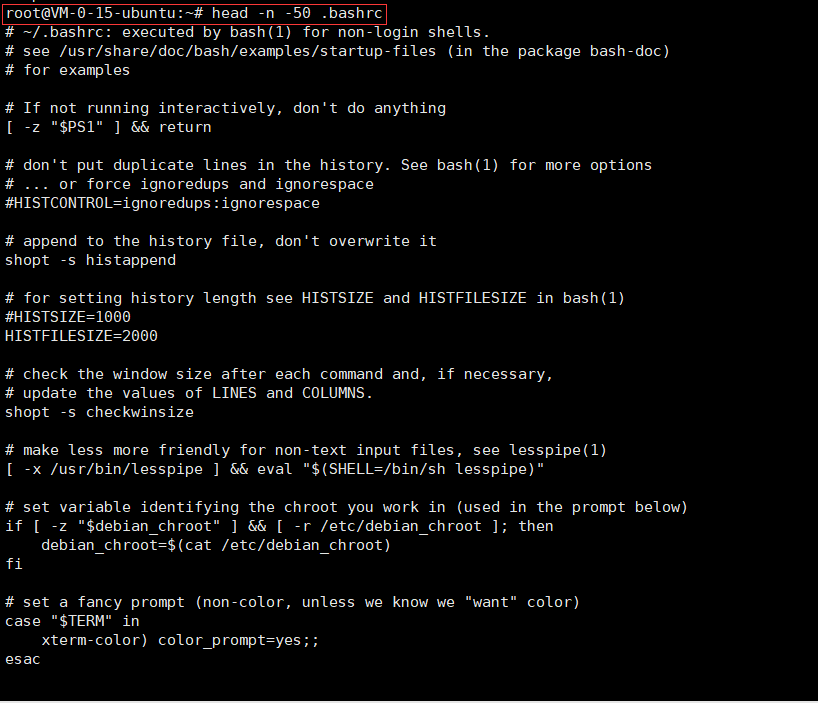


### 2.2.11 head命令：取出前几行

（18）查看当前用户主文件夹下.bashrc文件内容前20行

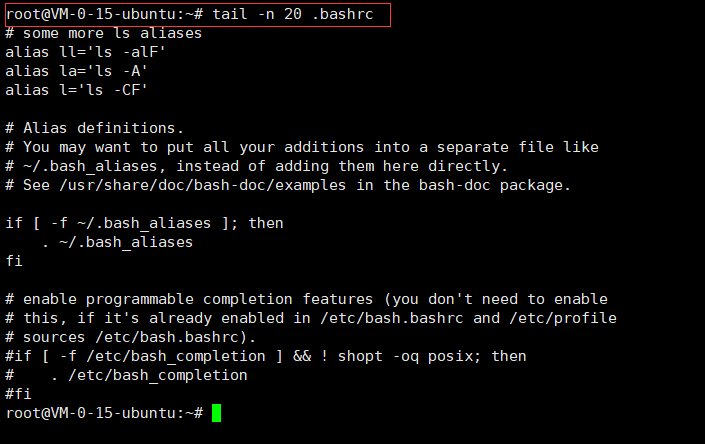


（19）查看当前用户主文件夹下.bashrc文件内容，后面50行不显示，只显示前面几行

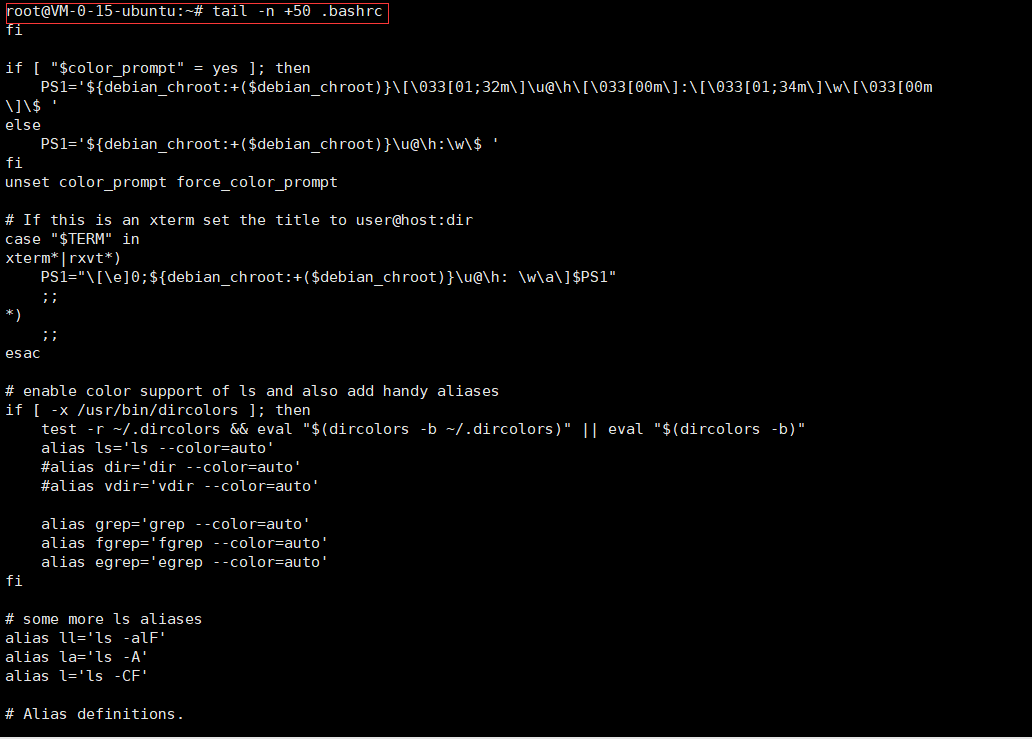


### tail命令：取出后面几行

1. 查看当前用户主文件夹下.bashrc文件内容最后20行



（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



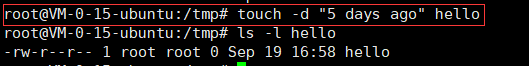
### touch命令：修改文件时间和创建新的文件

（22）在“/tmp”目录下创建一个空文件hello，并查看文件时间



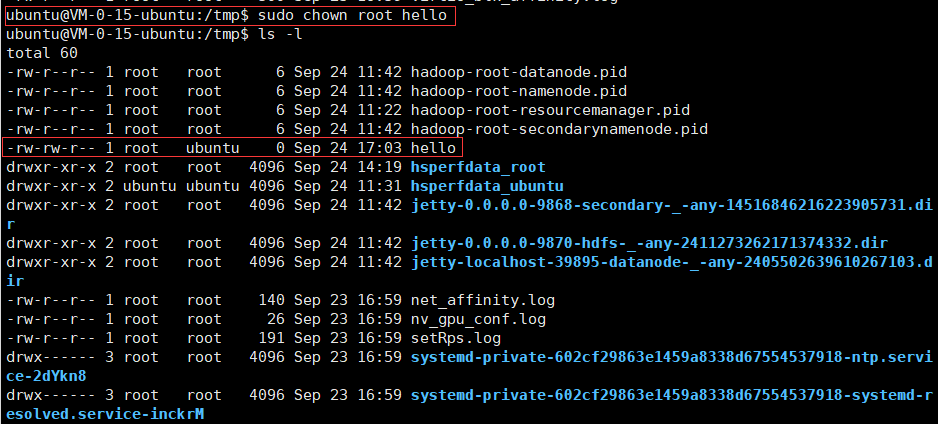


（23）修改hello文件，将文件时间整为5天前



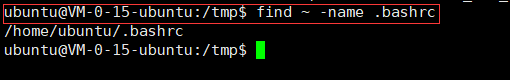
### chown命令：修改文件所有者权限

（24）将hello文件所有者改为root帐号，并查看属性



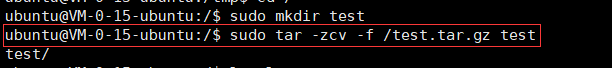
### find命令：文件查找

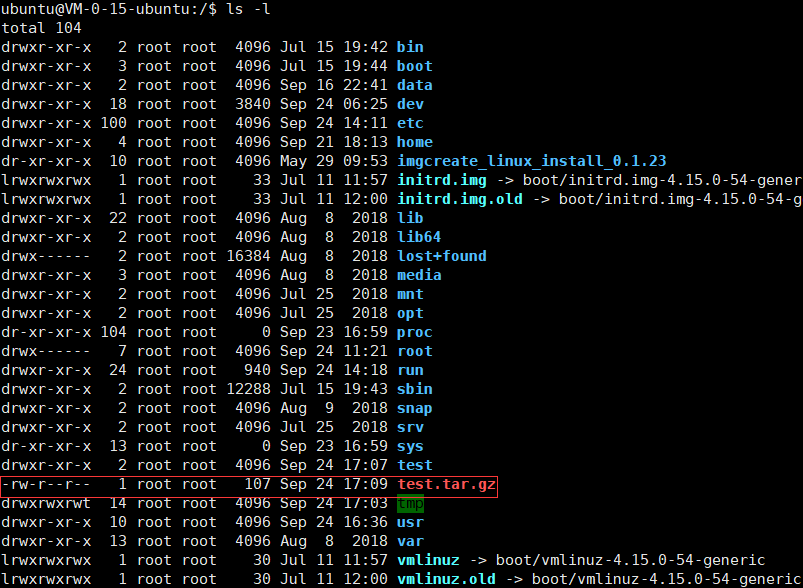
（25）找出主文件夹下文件名为.bashrc的文件



### tar命令：压缩命令

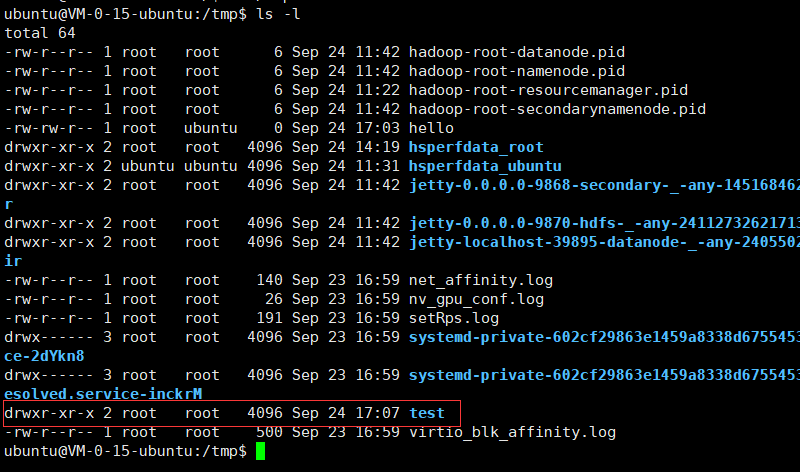
（26）在根目录“/”下新建文件夹test，然后在根目录“/”下打包成test.tar.gz





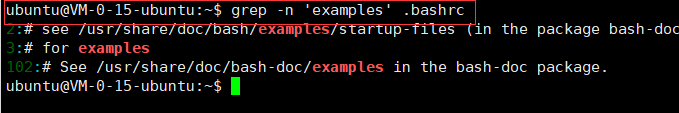
（27）把上面的test.tar.gz压缩包，解压缩到“/tmp”目录





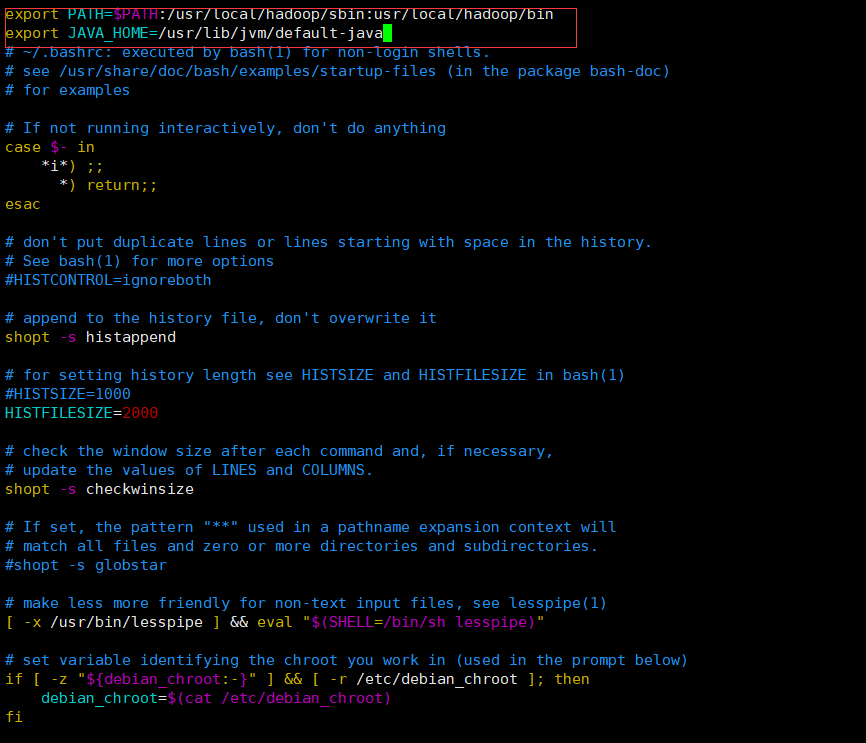
### grep命令：查找字符串命令

（28）从“～/.bashrc”文件中查找字符串'examples'



* 配置环境变量

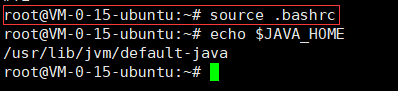
（29）请在“～/.bashrc”中设置，配置Java环境变量



（30）查看JAVA\_HOME变量的值

在输入命令echo $JAVA\_HOME之前，需要先输入命令source .bashrc，运行.bashrc文件，这样在输入命令时echo $JAVA\_HOME时才能能得到JAVA\_HOME的路径，否则就会无法输出任何内容。





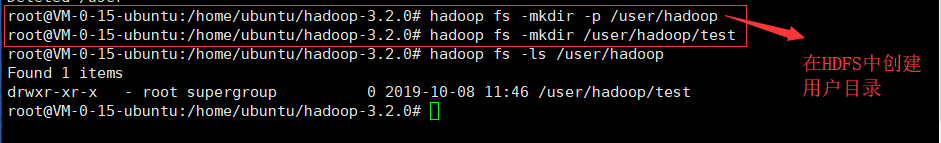
## Hadoop操作

**一、熟悉常用的Hadoop操作**

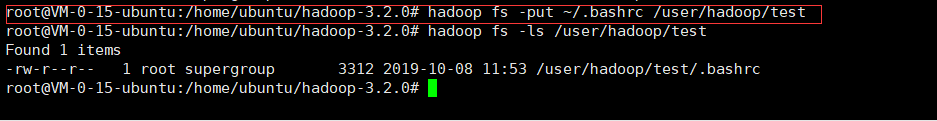
（31）使用hadoop用户登录Linux系统，启动Hadoop（Hadoop的安装目录为“/usr/local/hadoop”），为hadoop用户在HDFS中创建用户目录“/user/hadoop”

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

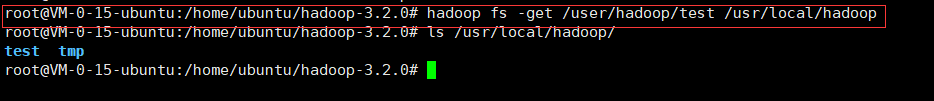
在/user/hadoop下创建test文件夹时的前提是/user/hadoop是存在的，所以要先在HDFS中创建用户目录:/user/hadoop，用的命令是：hadoop dfs -mkdir -p /user/hadoop，刚开始时我就没有创建这个用户目录，导致接下来的两个小时里都在苦恼为什么我就是无法把文件写进去，创建了用户目录之后一切都顺利了。



（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



（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



# 3. 实验总结

做这份实验之前一直在装hadoop，从第一节实验课开始就在弄hadoop，刚开始的时候是准备就在虚拟机上进行hadoop的操作，但是由于老师说在虚拟机上会有很多麻烦而且很慢，就和班上许多同学一样，买了一个服务器，在服务器的控制台上进行Linux以及hadoop的相关操作。从开始接触hadoop到安装好整整花了两个星期的时间，在这个过程中遇到了许多问题，十分的稀奇古怪，可能和hadoop的缘分不是很深，以至于好像每一步都是坑，在好不容易以为就要见到曙光的时候，止步于初始化配置文件，让自己不得不怀疑是不是自己在前面的安装过程中是不是有错误，最后在林老师的帮忙下顺利的安装好了hadoop，喜大普奔。终于高高兴兴的开始做起了实验，实验前部分的Linux命令操作没有很大的问题，但是在进行hadoop命令操作时，又出现了问题，第一个问题就是在输入hadoop fs -mkdir命令时，会提示找不到hadoop这个命令，我查了百度，后面发现是因为没有配置HADOOP\_HOME。按照百度上的步骤找到了profile文件，添加了hadoop的环境变量，然后保存文件，退出后，一定要先执行命令source profile让刚刚修改过的profile文件运行起来，这是再输入hadoop fs -mkdir命令时就不会再提示找不到hadoop命令了。接下来的第二个问题，这个问题花了我两个多小时，缘由是因为没有认真看书和实验报告上的题目，在（32）中，要在/user/hadoop文件中创建一个test文件，因为跳过了（31）题，所以我并没有想到要先用hadoop fs -mkdir -p /user/hadoop命令创建/user/hadoop，直接输入了命令hadoop fs -mkdir /user/hadoop/test，然后就一直提示没有这个路径，这就是不认真看书看题的后果啊，这个问题解决之后后面就一路通畅，顺利的做完了实验报告。

最后，正确的hadoop命令是①hadoop fs … ②hdfs dfs …