

**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** |

# 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。

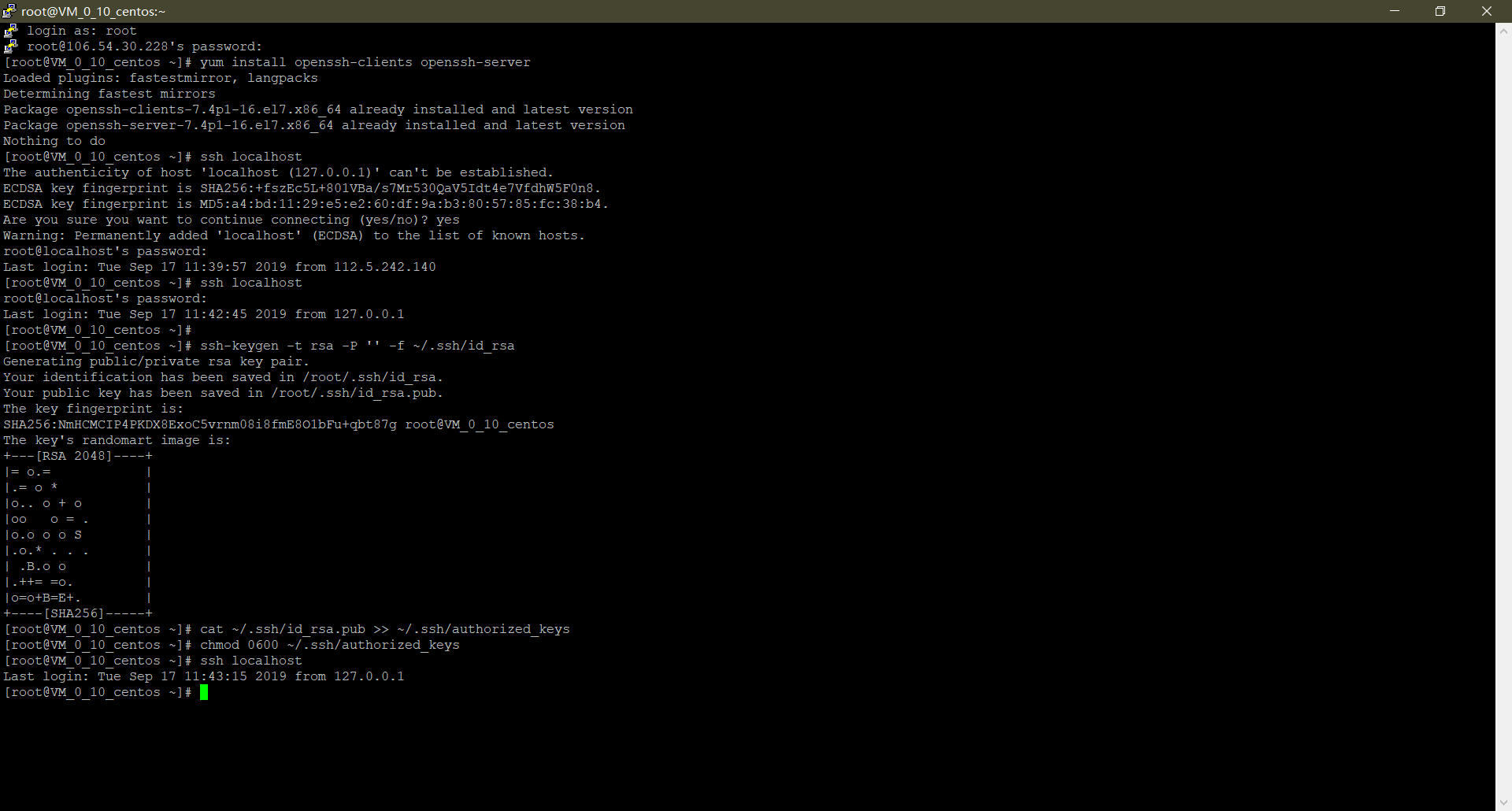
② 熟悉常用的Linux操作，实现并记录30项操作的过程及结果。

③ 熟悉常用的Hadoop操作，实现并记录4项操作的过程及结果。

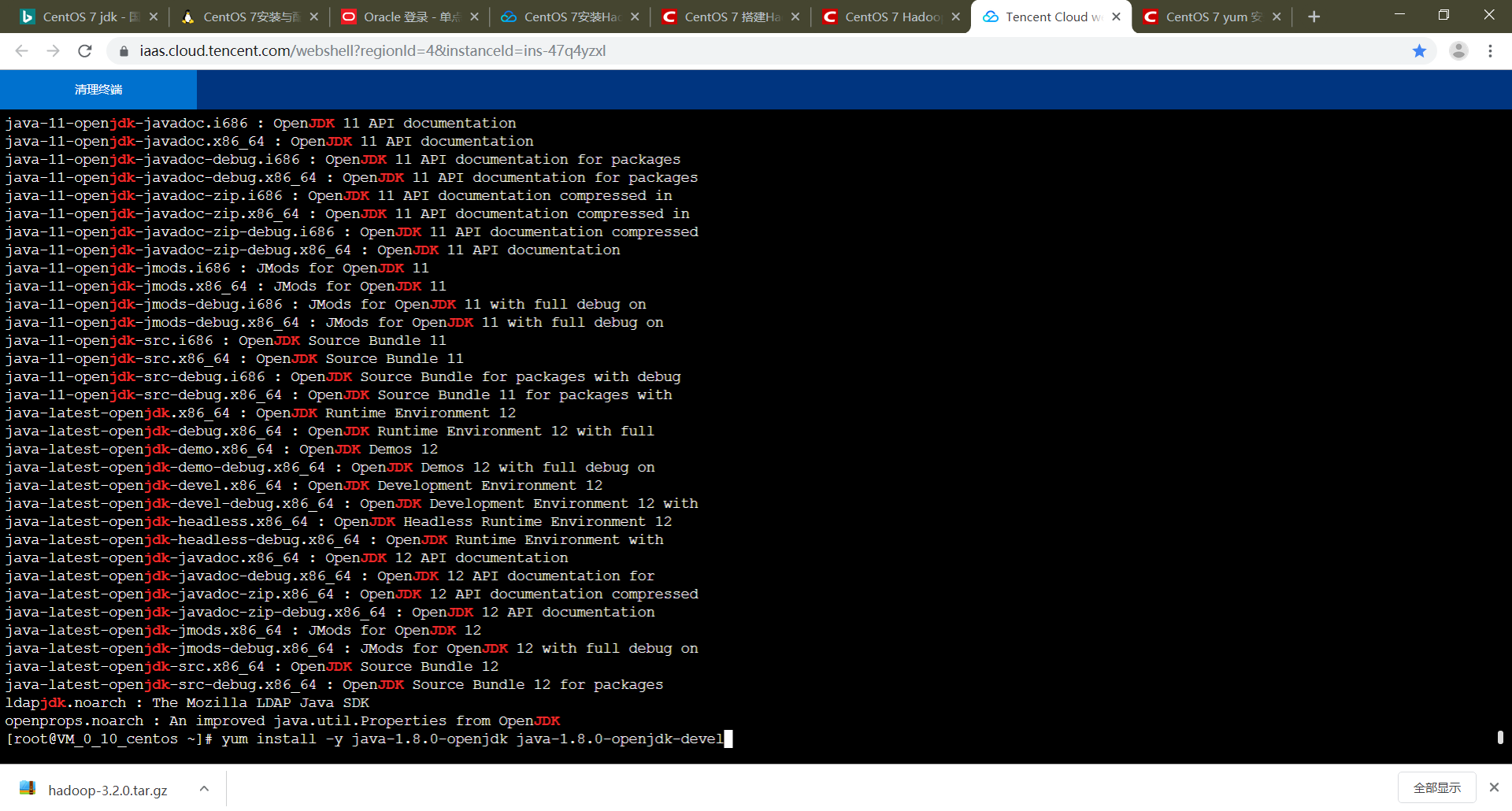
# 2. 实验记录

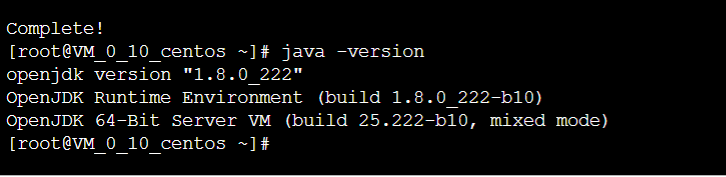
## 2.1 安装Hadoop

Step1：安装openSSH server和openSSH client实现免密码登录

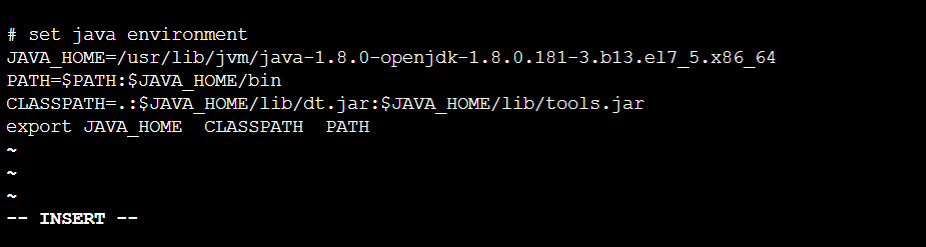


Step2：安装jdk



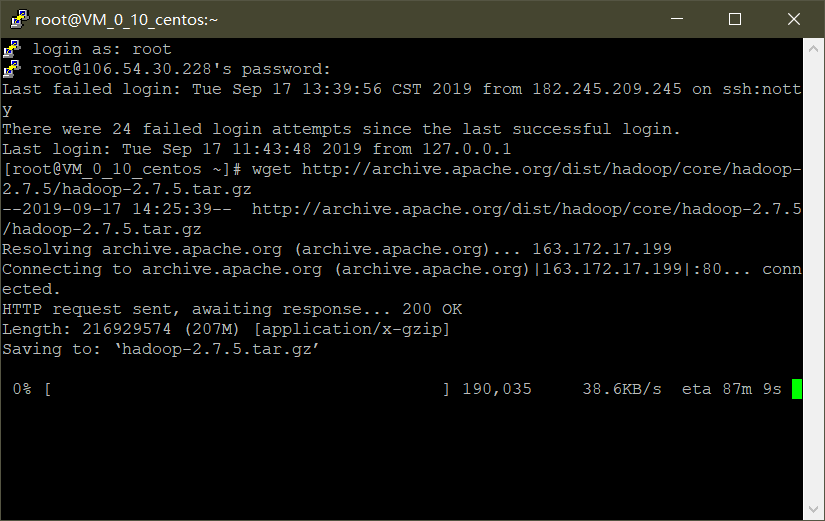


Step3：配置JAVA环境



Step4：hadoop安装

下载.tar.gz文件

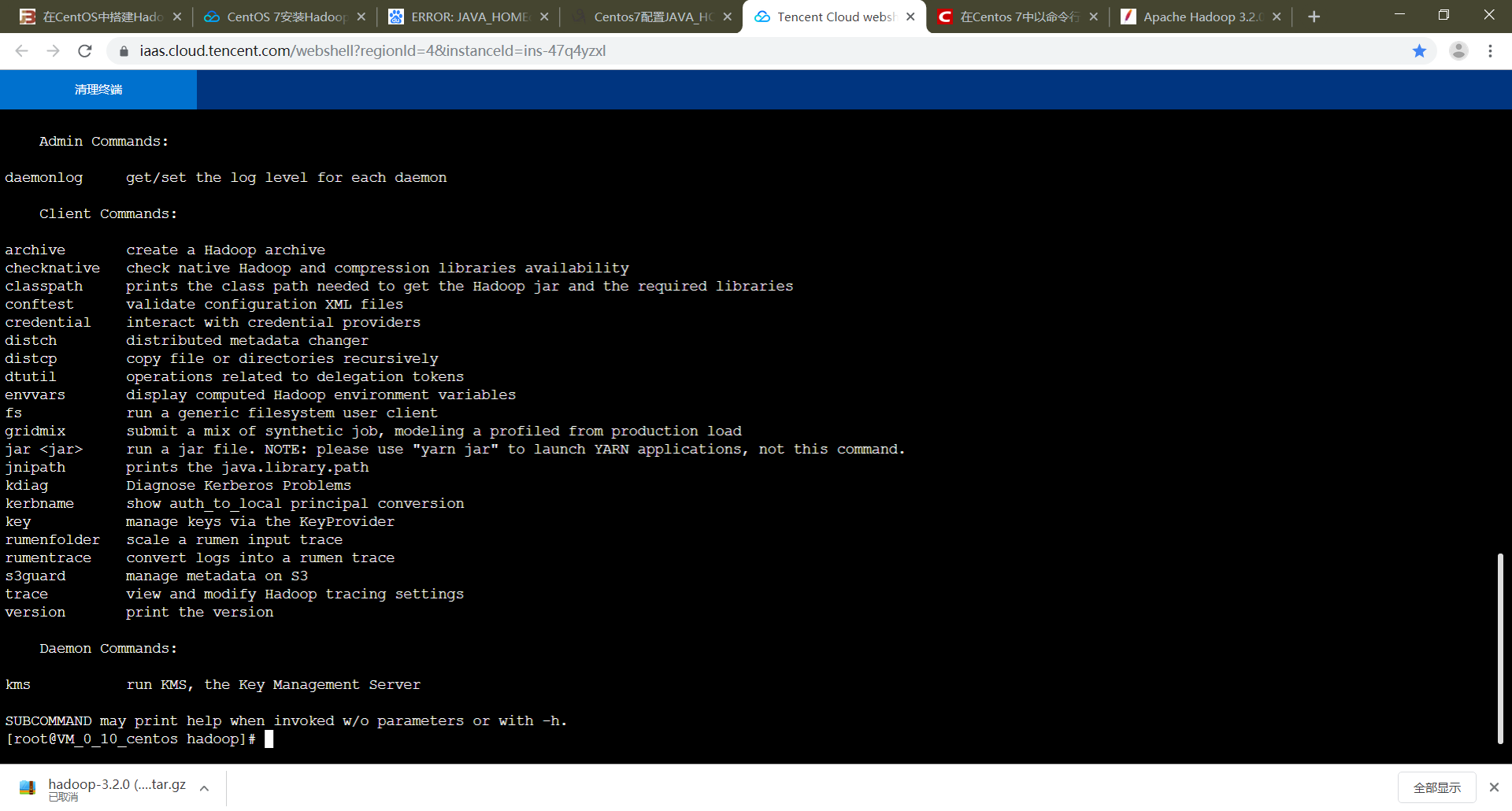


**将 Hadoop 安装到 /usr/local 目录下:**

tar -zxf hadoop-3.2.0.tar.gz -C /usr/**local**

**对安装的目录进行重命名:**

mv ./hadoop-3.2.0/ ./hadoop（忘记截图了）



安装完成后 cd /usr/local/hadoop/bin/hadoop version检验

Step5：配置hadoop环境

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

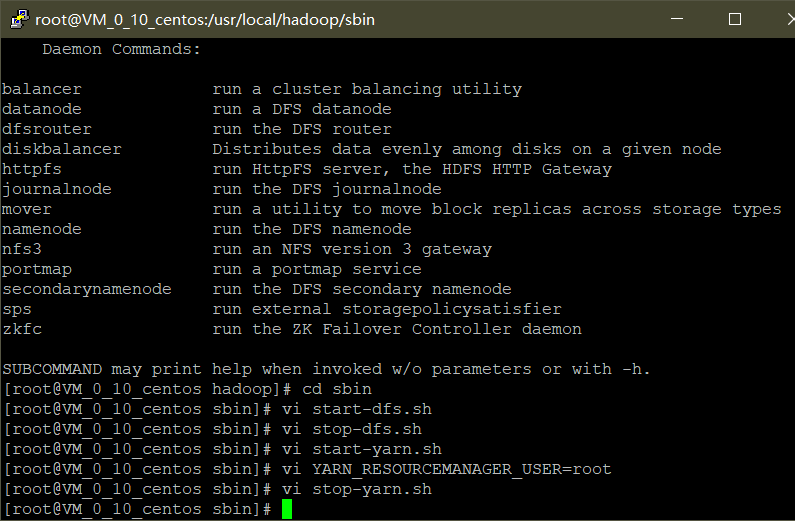
Export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

export PATH=$PATH:$HADOOP\_HOME/sbin:$HADOOP\_HOME/bin

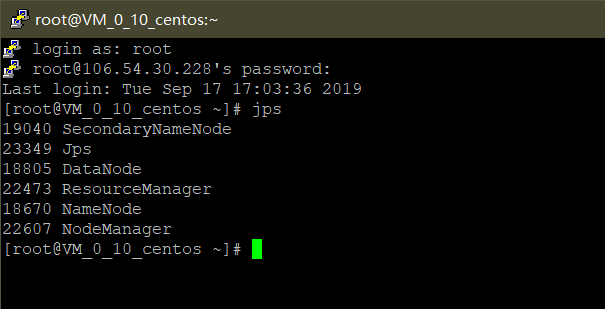
 使Hadoop环境变量配置生效:

source ~/.bashrc

Step6：配置distributed环境

按照hadoop.apache.org的文档，修改以上的四个.sh文件

输入jps，有四个参数正常启动（后面变成了六个，我好像打开了什么东西）



依次：

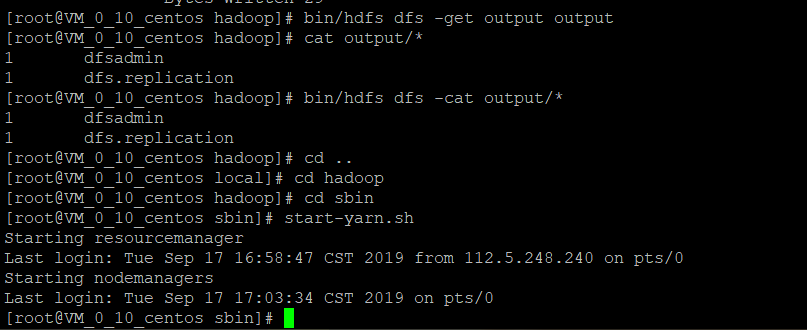
bin/hdfs namenode -format

sbin/start-dfs.sh

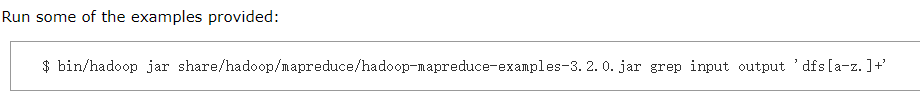
bin/hdfs dfs -mkdir /user/root 如果这里不是当前用户会显示没有这个file

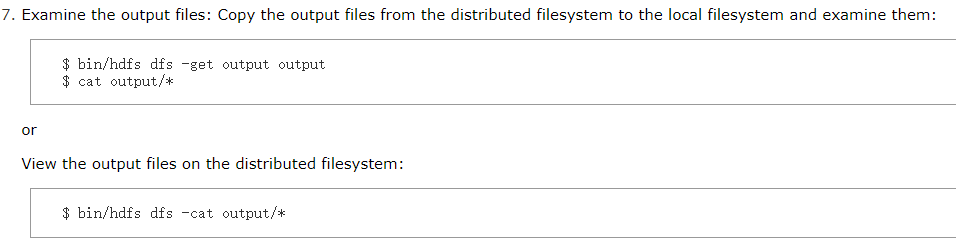
bin/hdfs dfs -mkdir input

$ bin/hdfs dfs -put etc/hadoop/\*.xml input



Step7：检验





以上是apache提供的检验方法，第一个是例子。

关闭：



到此配置结束

## 2.2 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。





## 2.3 Hadoop操作

**（一）熟悉常用的Linux操作**

* cd命令：切换目录

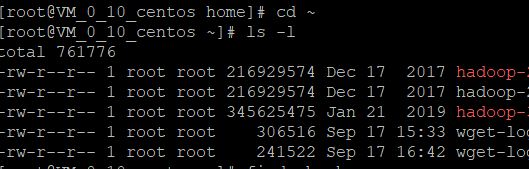
1. 切换到目录“/usr/local”



1. 切换到当前目录的上一级目录

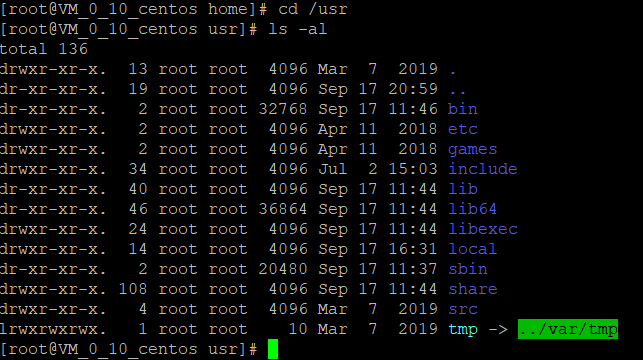


1. 切换到当前登录Linux系统的用户的自己的主文件夹



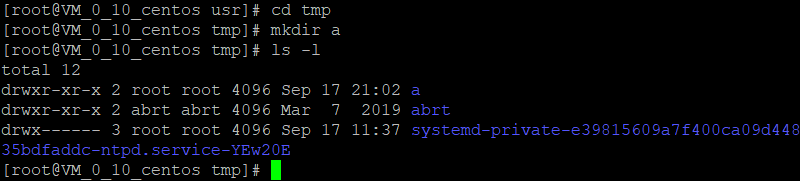
* ls命令：查看文件与目录

1. 查看目录“/usr”下的所有文件和目录

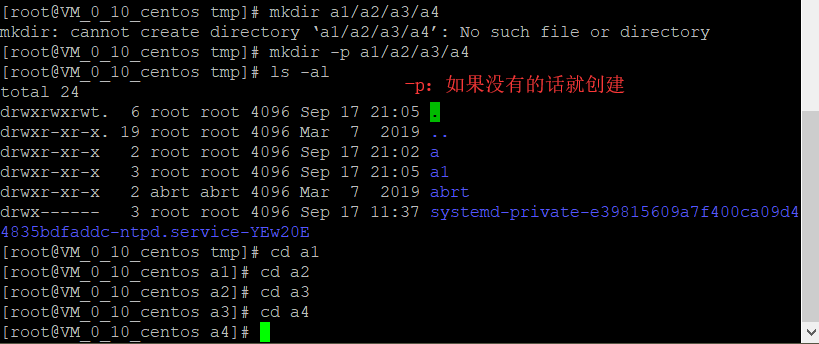


* mkdir命令：新建目录

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



1. 进入“/tmp”目录，创建目录“a1/a2/a3/a4”

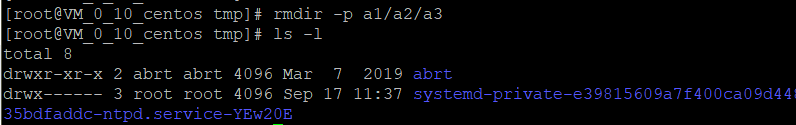


* rmdir命令：删除空的目录

1. 将上面创建的目录a（在“/tmp”目录下面）删除

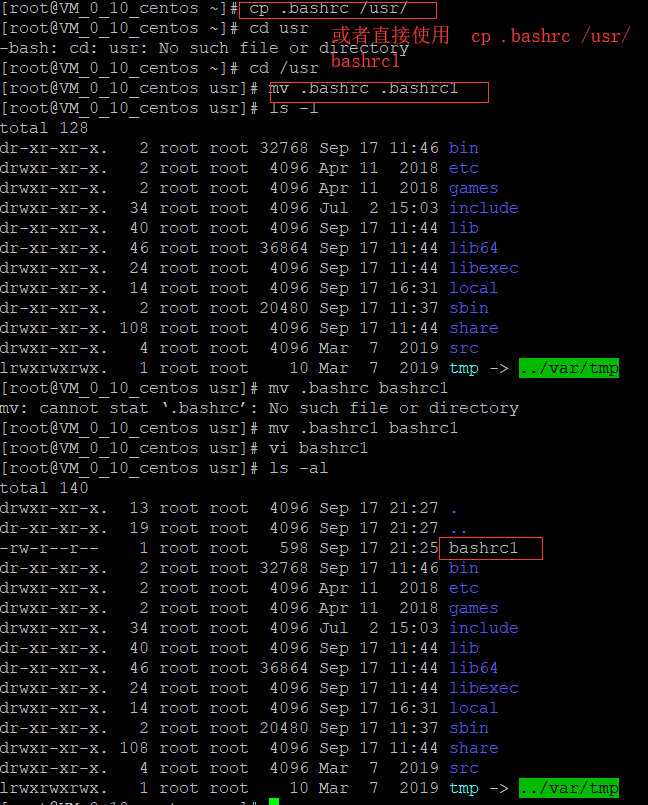


1. 删除上面创建的目录“a1/a2/a3/a4” （在“/tmp”目录下面），然后查看“/tmp”目录下面存在哪些目录

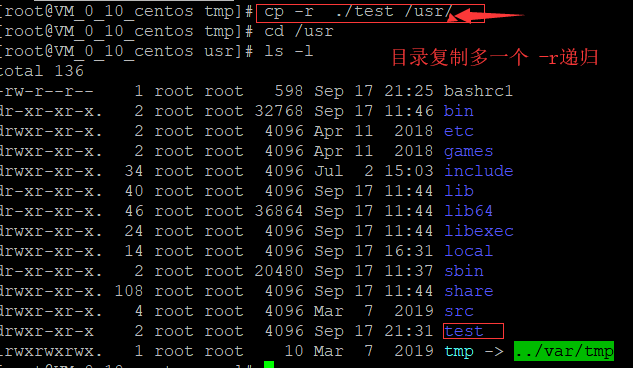


* cp命令：复制文件或目录

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

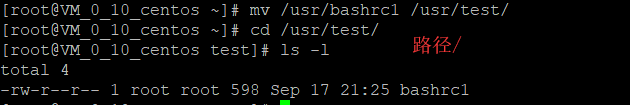


1. 在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下

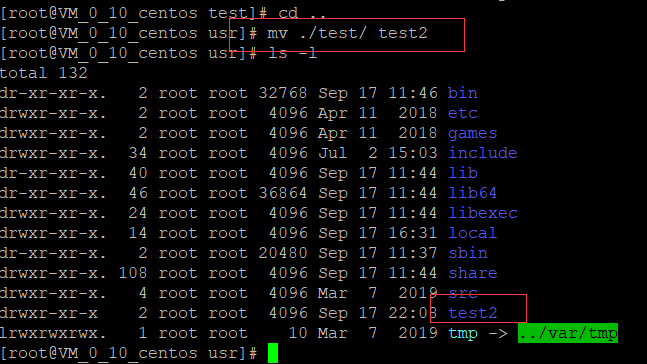


* mv命令：移动文件与目录，或更名

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

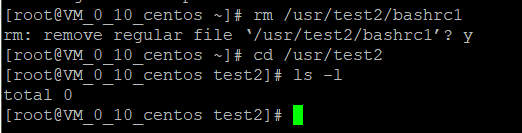


1. 将“/usr”目录下的test目录重命名为test2

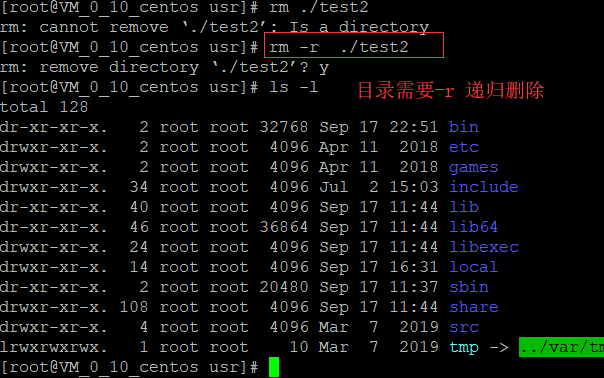


* rm命令：移除文件或目录

1. 将“/usr/test2”目录下的bashrc1文件删除

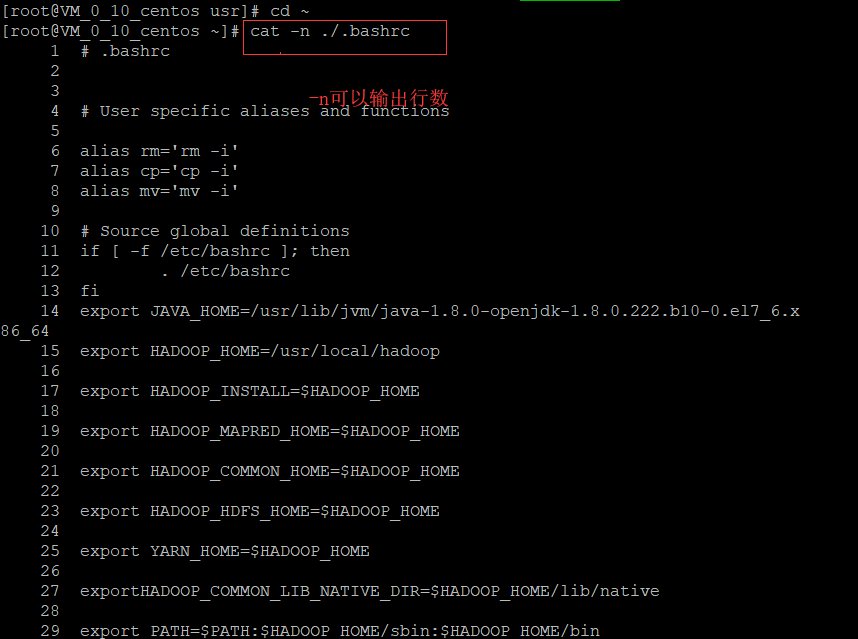


1. 将“/usr”目录下的test2目录删除



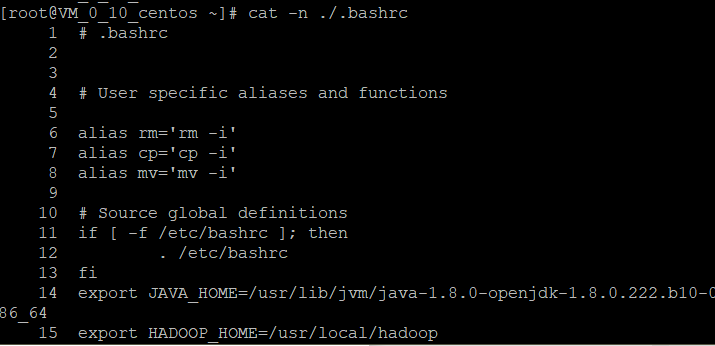
* cat命令：查看文件内容

1. 查看当前用户主文件夹下的.bashrc文件内容



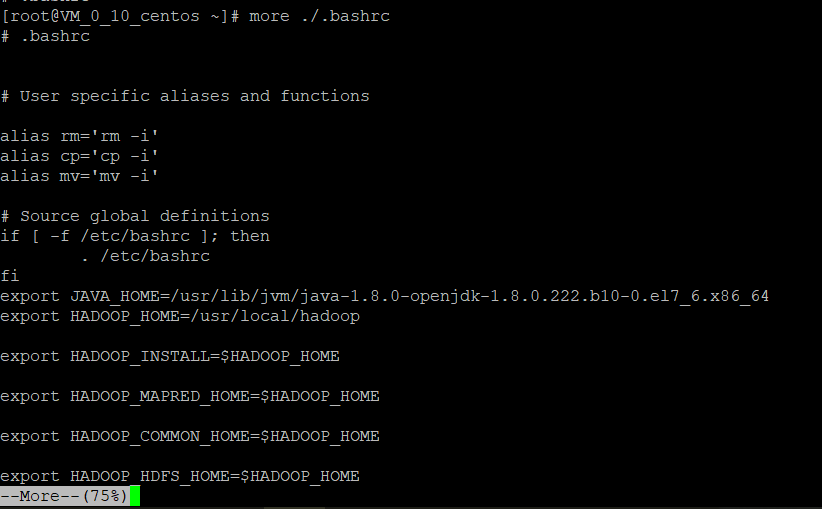
* tac命令：反向查看文件内容

1. 反向查看当前用户主文件夹下的.bashrc文件的内容



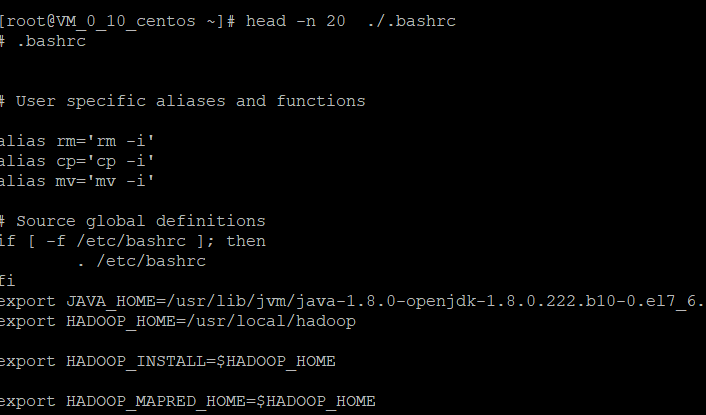
* more命令：一页一页翻动查看

1. 翻页查看当前用户主文件夹下的.bashrc文件的内容



* head命令：取出前面几行

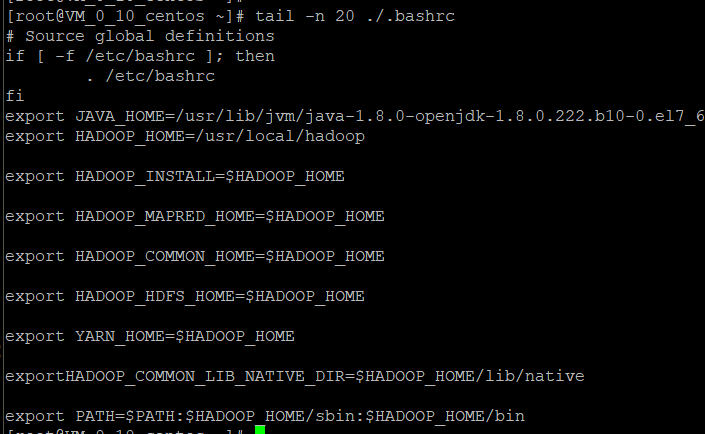
1. 查看当前用户主文件夹下.bashrc文件内容前20行

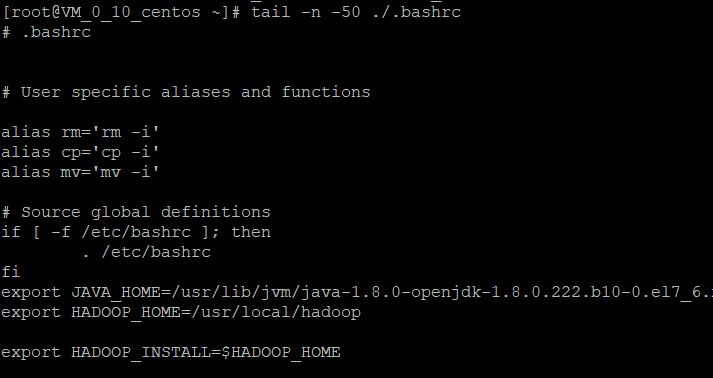


（19）查看当前用户主文件夹下.bashrc文件内容，后面50行不显示，只显示前面几行

* tail命令：取出后面几行

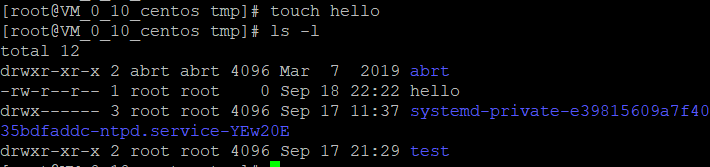
1. 查看当前用户主文件夹下.bashrc文件内容最后20行



（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

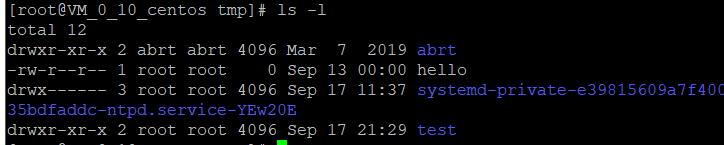


1. 修改hello文件，将文件时间整为5天前



* chown命令：修改文件所有者权限

1. 将hello文件所有者改为root帐号，并查看属性



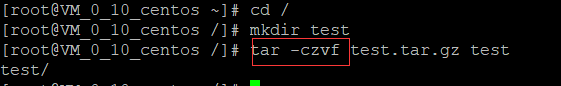
* find命令：文件查找

1. 找出主文件夹下文件名为.bashrc的文件

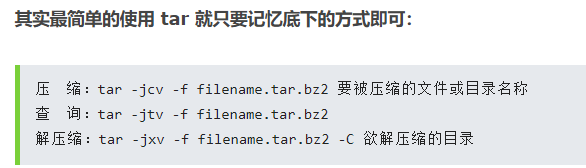


* tar命令：压缩命令

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

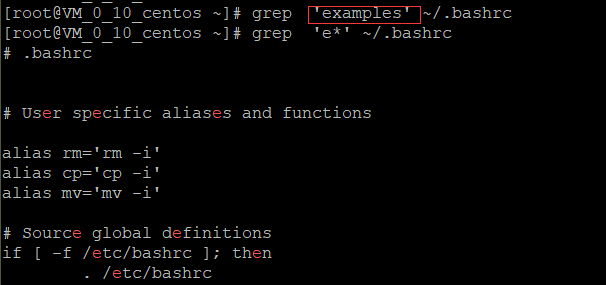


1. 把上面的test.tar.gz压缩包，解压缩到“/tmp”目录



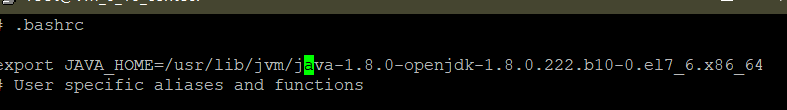
* grep命令：查找字符串

1. 从“～/.bashrc”文件中查找字符串'examples'



* 配置环境变量

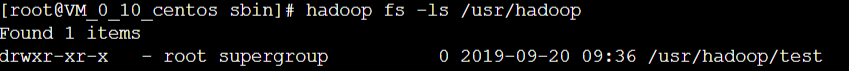
1. 请在“～/.bashrc”中设置，配置Java环境变量

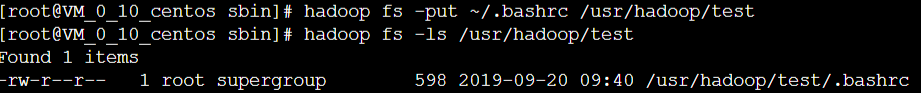


1. 查看JAVA\_HOME变量的值

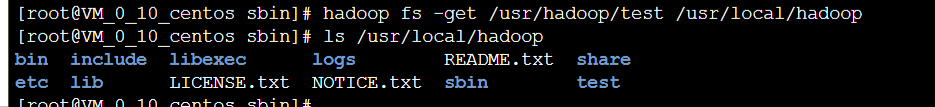
**（二）熟悉常用的Hadoop操作**

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”

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

（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

/\*实验总结是对整个实验过程的回顾和讨论，如果实验过程当中出现问题，实验总结是最恰当的位置用于记录问题的由来和改正过程。（以下内容仅供参考）\*/

## 事实证明，找到不好的教程会走很多的弯路。

遇到了几个比较主要的问题：

1. JAVA\_HOME的路径混乱，第一次比较傻，不熟悉vi指令，不小心多配了一次；后面又陆续出现了两三次，本以为可以逢凶化吉，没想到在最后一步namenode的时候一怒之下重装系统了。
2. 不知道怎么回事，SSH验证不通过，甚至不能再腾讯云上直接操作，后面查资料，换了一种登录方式才上去。

一开始用的是腾讯云网页上登录，后面改用了putty登录，其实都一样，就是截图会比较混乱，请多多多多多多多多包涵。

1. 我吐了，我重装了两次系统，ssh和javahome轮流出错

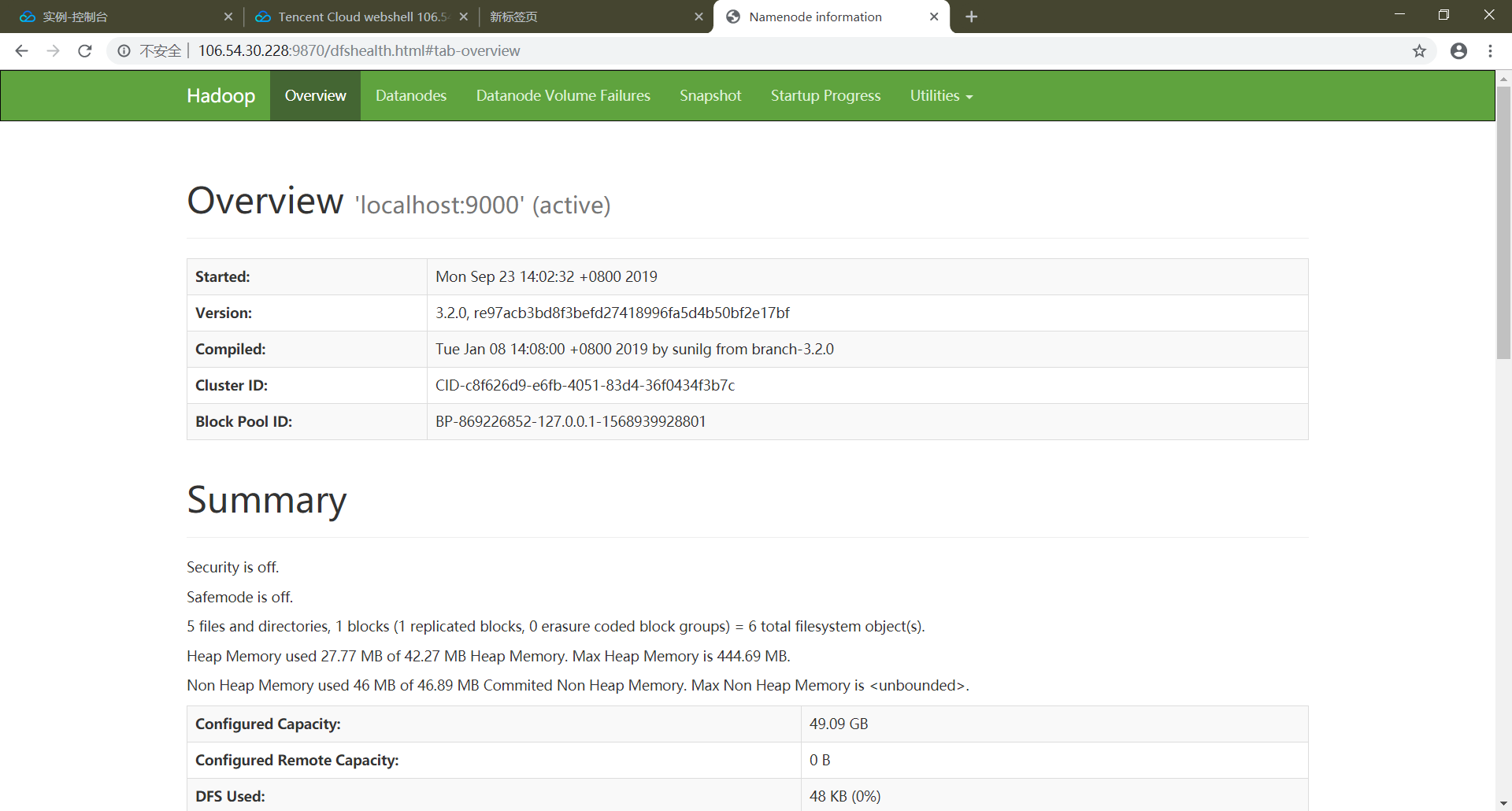
然后有一个问题，如果我切换普通用户的话就 没有办法开启hdfs，ssh持续显示permission denied，我怀疑是之前的start-dfs.sh和stop-dfs.sh文件里面的namenode只写了root，但是改了之后也不行，网上的资料都是只改成root，所以也可能是我想错了，我把错误放到网上去搜，好像没人有这个错的，我服了

后来我又想了一下，是不是hdfs的权限（hadoop文件夹）的权限没开，只给了root用户，

然后我把普通用户加到root的group里面。目前还是不行，打算看下下节课实验做什么，慢慢想。

然后问了老师，说用root就行，

验证



舒服了