

**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 cent os 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

**2.1.1安装ssh:**

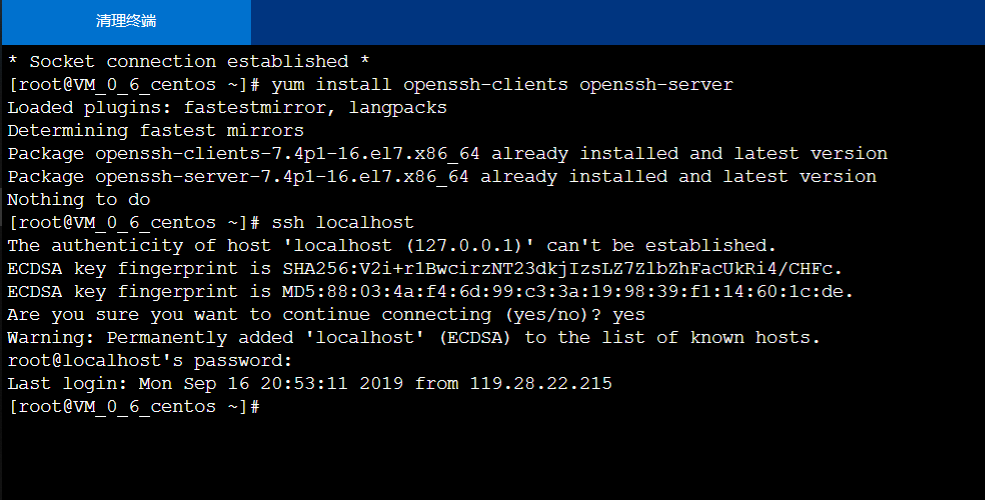
yum install openssh-clients openssh-server

安装完成后，使用以下命令测试：

ssh localhost

输入 root 账户的密码，如果可以正常登录，则说明SSH安装没有问题。

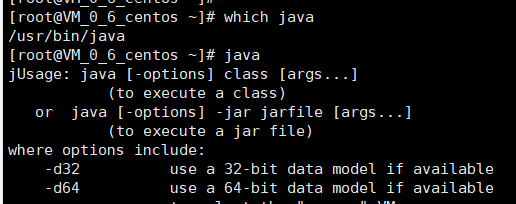


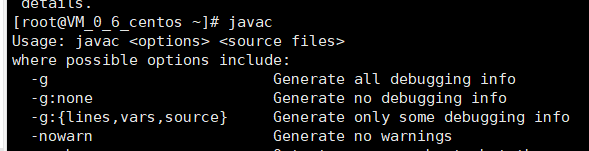


**2.1.2安装jdk**

使用 yum 来安装1.8版本 OpenJDK：

yum install java-1.8.0-openjdk java-1.8.0-openjdk-devel

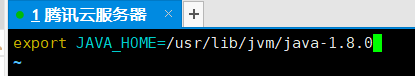




**2.1.3配置java环境变量**

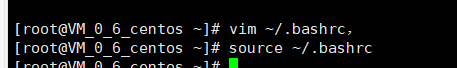
执行命令： vim ~/.bashrc，

在结尾追加：exportJAVA\_HOME=/usr/lib/jvm/java-1.8.0



保存文件后执行下面命令使 JAVA\_HOME 环境变量生效:

source ~/.bashrc

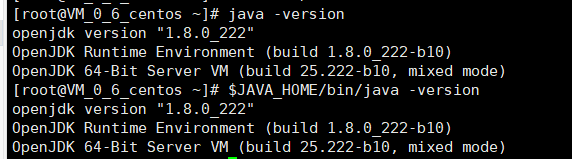


为了检测系统中java环境是否已经正确配置并生效，可以分别执行以下命令：

java -version

$JAVA\_HOME/bin/java -version

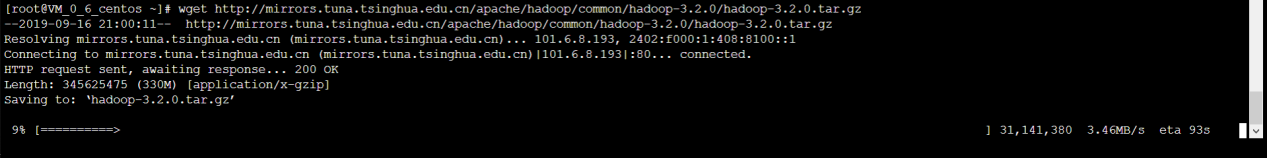
若两条命令输出的结果一致，且都为我们前面安装的 openjdk-1.8.0 的版本，则表明 JDK 环境已经正确安装并配置。



**2.1.4下载 Hadoop**

本教程使用 hadoop-3.2版本，使用 wget 工具在线下载

Wgethttp://mirrors.tuna.tsinghua.edu.cn/apache/hadoop/common/hadoop-3.2.0/hadoop-3.2.0.tar.gz



安装 Hadoop将 Hadoop 安装到 /usr/local 目录下:

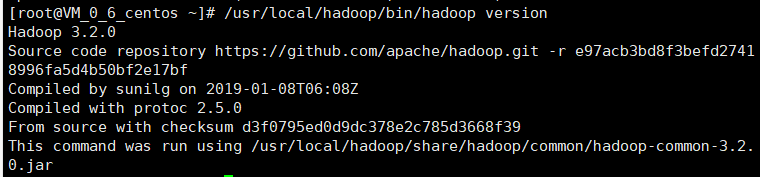
tar -zxf hadoop-3.2.0.tar.gz -C /usr/local

对安装的目录进行重命名，便于后续操作方便:

mv ./hadoop-3.2.0/ ./hadoop

检查Hadoop是否已经正确安装

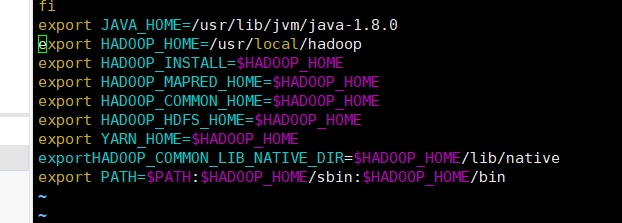
:/usr/local/hadoop/bin/hadoop version



**2.1.5hadoop伪分布式环境配置**

2.1.5.1设置 Hadoop 的环境变量

Vim ~/.bashrc，在结尾追加如下内容：



使Hadoop环境变量配置生效:

source ~/.bashrc

修改 Hadoop 的配置文件Hadoop的配置文件位于安装目录的 /etc/hadoop目录下，在本教程中即位于 /usr/local/hadoop/etc/hadoop 目录下，需要修改的配置文件为如下两个:

/usr/local/hadoop/etc/hadoop/core-site.xml

/usr/local/hadoop/etc/hadoop/hdfs-site.xml

编辑core-site.xml，修改<configuration></configuration>节点的内容为如下所示：<configuration>

<!--配置存放Datenode数据-->

<property>

<name>hadoop.tmp.dir</name>

<value>file:/usr/local/hadoop/tmp</value>

<description>location to store temporary files </description></property>

<!-- 配置主机NameNode的主机地址-->

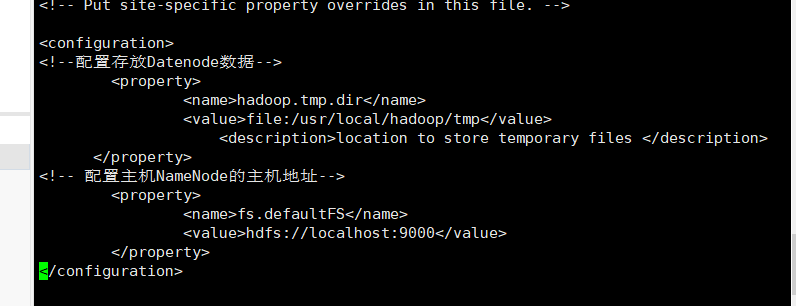
<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>



同理，编辑 hdfs-site.xml，修改<configuration></configuration>节点的内容为如下所示：

<configuration>

<!--配置HDFS的副本，冗余度-->

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/usr/local/hadoop/tmp/dfs/name</value>

</property>

<!--配置HDFS的权限-->

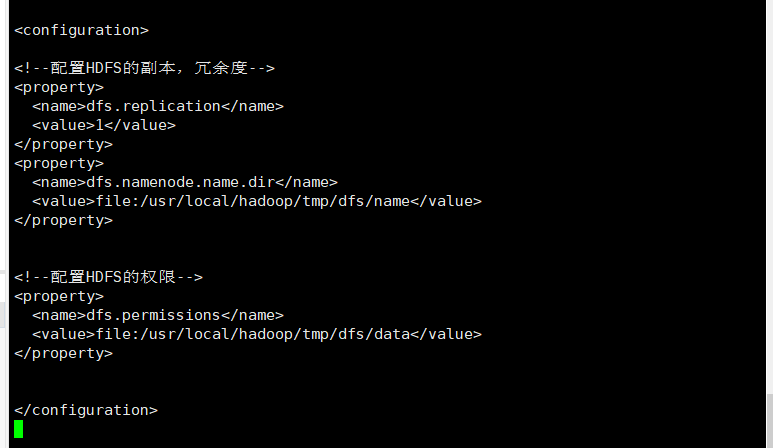
<property>

<name>dfs.permissions</name>

<value>file:/usr/local/hadoop/tmp/dfs/data</value>

</property>

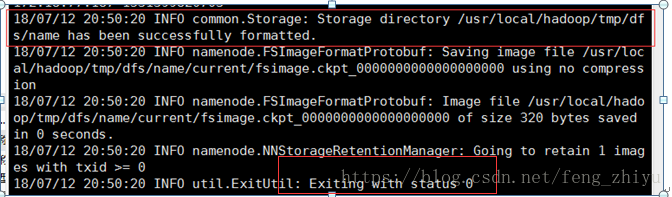
</configuration>



2.1.5.2格式化

NameNode:/usr/local/hadoop/bin/hdfs namenode -format

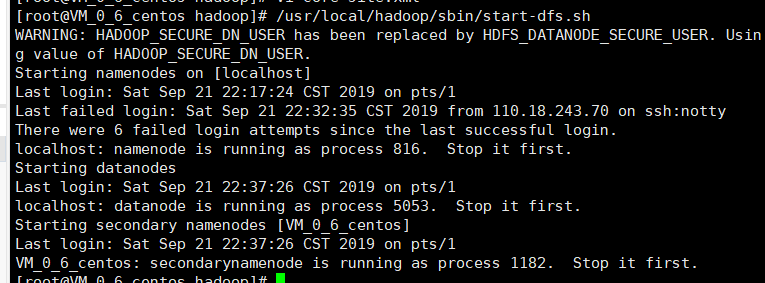
输出以下信息表示格式化成功：



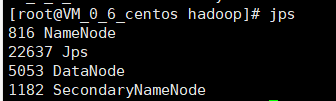
（因为前面没有截图，所以这张图是百度的）

2.1.5.3启动NameNode和DataNode进程：

执行过程中会提示输入用户密码，输入 root 用户密码即可。另外，启动时ssh会显示警告提示是否继续连接，输入 yes 即可检查NameNode和DateNode是否正常启动。



如果NameNode和DataNode已经正常启动，会显示NameNode、DataNode和SecondaryNameNode的进程信息:



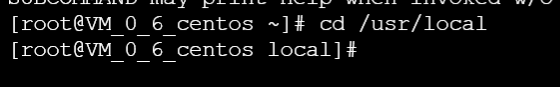
## 2.2 Linux操作

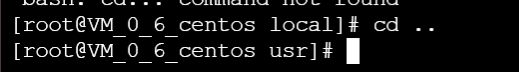
**2.2.1 cd命令：切换目录**

（1）切换到目录“/usr/local”

（2）切换到当前目录的上一级目录

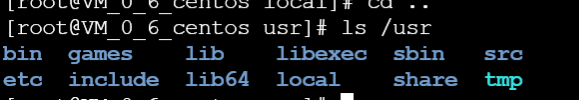
（3）切换到当前登录Linux系统的用户的自己的主文件夹





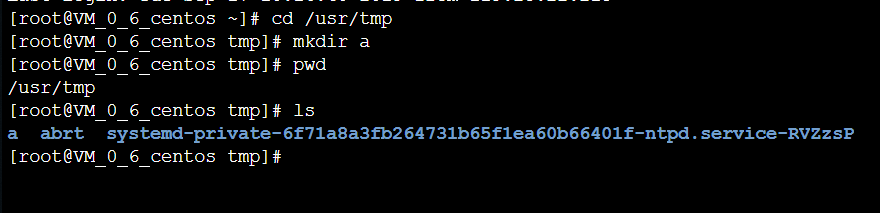
**2.2.2 ls命令：查看文件与目录**

（4）查看目录“/usr”下的所有文件和目录

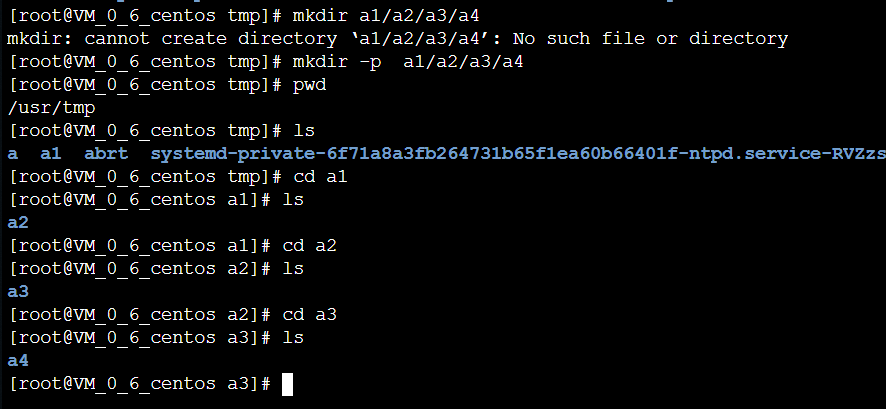


**2.2.3 mkdir命令：新建目录**

（5）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



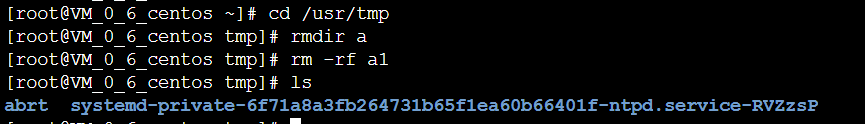
（6）进入“/tmp”目录，创建目录“a1/a2/a3/a4”



**2.2.4 rmdir命令：删除空的目录**

（7）将上面创建的目录a（在“/tmp”目录下面）删除

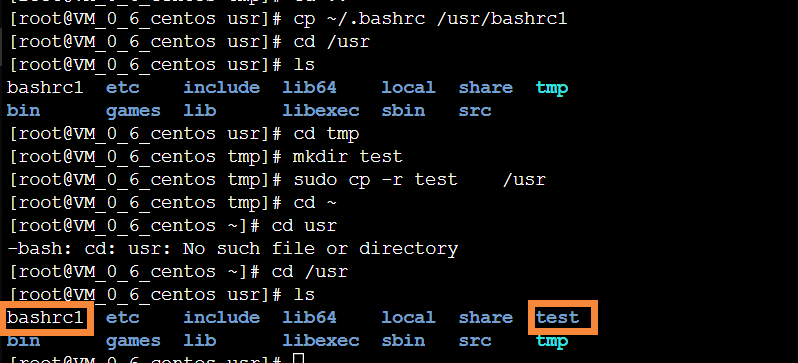
（8）删除上面创建的目录“a1/a2/a3/a4” （在“/tmp”目录下面），然后查看“/tmp”目录下面存在哪些目录



**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

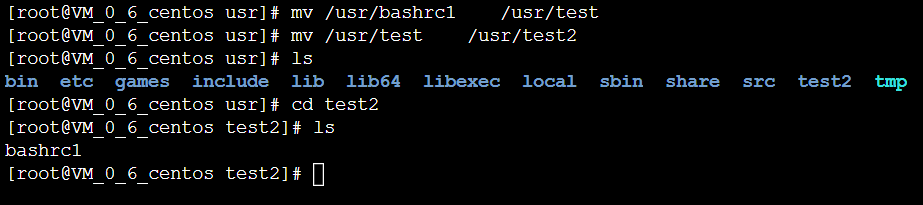
（10）在目录“/tmp”下新建目录test，再把这个目录复制到“/usr”目录下



**2.2.6 mv命令：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

（12）将“/usr”目录下的test目录重命名为test2

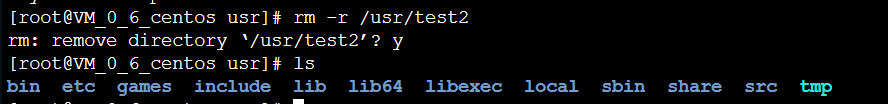


**2.2.7 rm命令：移除文件或目录**

（13）将“/usr/test2”目录下的bashrc1文件删除

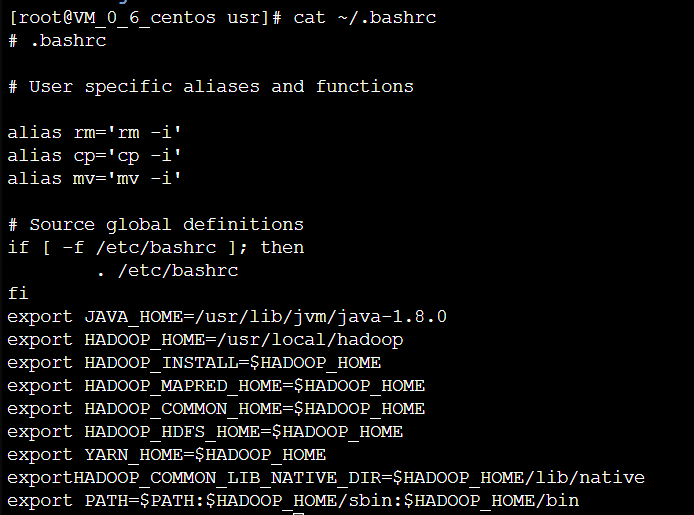
（14）将“/usr”目录下的test2目录删除





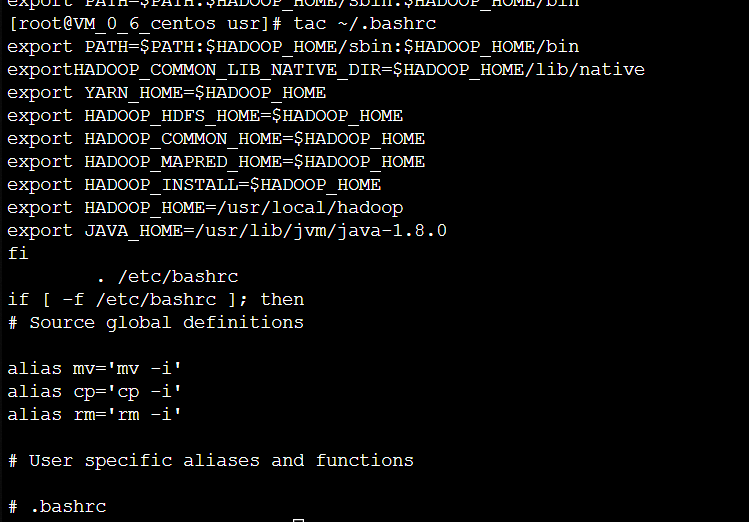
**2.2.8 cat命令：查看文件内容**

（15）查看当前用户主文件夹下的.bashrc文件内容



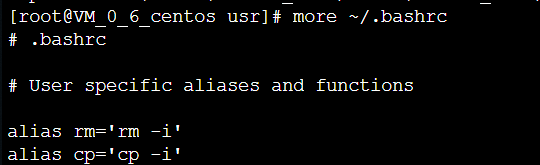
**2.2.9 tac命令：反向查看文件内容**

（16）反向查看当前用户主文件夹下的.bashrc文件的内容



**2.2.10 more命令：一页一页翻动查看**

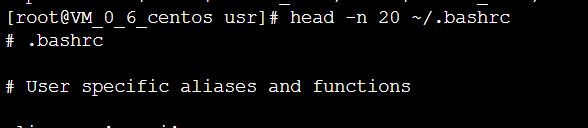
（17）翻页查看当前用户主文件夹下的.bashrc文件的内容



**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

（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

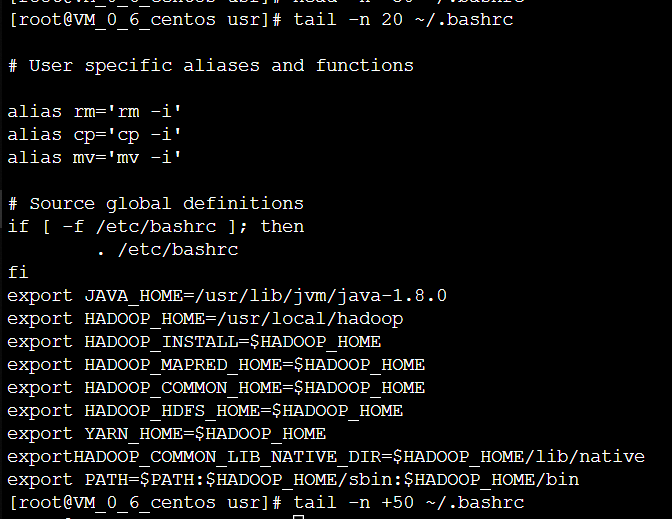




**2.2.12 tail命令：取出后面几行**

（20）查看当前用户主文件夹下.bashrc文件内容最后20行

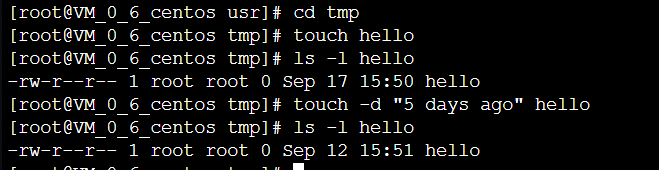
（21）查看当前用户主文件夹下.bashrc文件内容，并且只列出50行以后的数据



**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

（23）修改hello文件，将文件时间整为5天前



**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





(我一直都是root用户。。。。)

**2.2.15 find命令：文件查找**

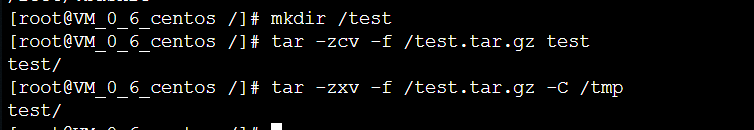
（25）找出主文件夹下文件名为.bashrc的文件



**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

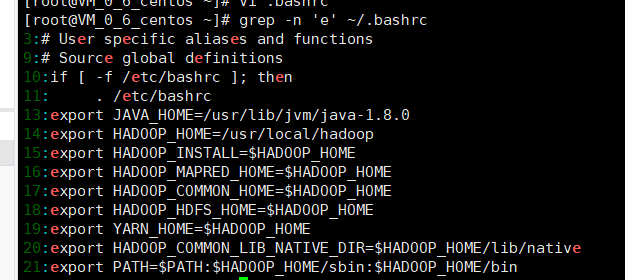
（27）把上面的test.tar.gz压缩包，解压缩到“/tmp”目录



**2.2.17 grep命令：查找字符串**

（28）从“～/.bashrc”文件中查找字符串'examples'

**grep -n 'examples' ~/.bashrc【但是我的.bashrc文件里好像没有example这个单词所以我搜索了 e】**



**2.2.18 配置环境变量**

（29）请在“～/.bashrc”中设置，配置Java环境变量

vim ~/.bashrc



（30）查看JAVA\_HOME变量的值



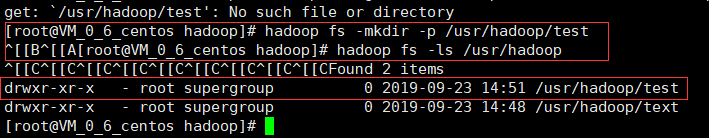


## 2.3 Hadoop操作

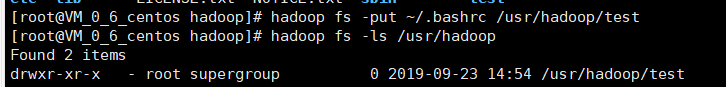
（31）使用hadoop用户登录Linux系统，启动Hadoop（Hadoop的安装目录为“/usr/local/hadoop”），为hadoop用户在HDFS中创建用户目录“/user/hadoop”



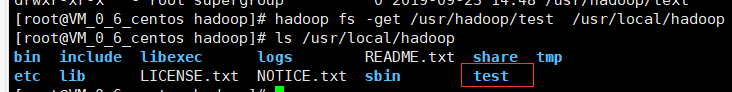
（32）接着在HDFS的目录“/user/hadoop”下，创建test文件夹，并查看文件列表



（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

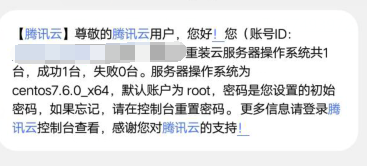
/\*实验总结是对整个实验过程的回顾和讨论，如果实验过程当中出现问题，实验总结是最恰当的位置用于记录问题的由来和改正过程。（以下内容仅供参考）\*/

上网搜索是个好东西，大家都要爱它。

首先，听到老师说下午的实验很简单，就只要点下载点下一步就可以了，就满心期待。结果，听怎么做就听了一节课，然后我也不知道应该怎么做，我觉得我和老师心目中的简单可能不是一个概念！！

然后在快下课的时候，老师说：还没安装好的同学要加快了！！我就在想难道全世界就我没装好吗？(很明显不是)但是我还是决定先去食堂抢一个位置再说，吃饱最重要！！！

然后吃完饭回宿舍就开始装了，同时很明显我一直看五六个教程，给我的配置环境写了几十条，甚至烦躁的敲了一大堆乱七八糟的东西进去，导致我登入还没做任何操作的时候就已经显示error了，这时候我在想的是，说不定我应该去买一个Ubuntu，说不定和老师一样的系统我就可以和老师一样了，但是我舍友及时敲醒了我。她:万一买了也不行岂不是更难过？

于是我开始了另一条路，名叫 重装，在重装之前我也百度了【云服务器可以重装吗？重装要注意什么东西吗？】等问题，但是好像没有什么回答，本着大不了重新买一个的心态，我点了 重装 。

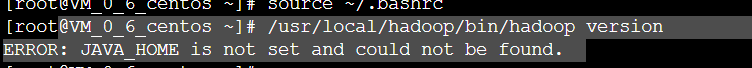
结果就是 云服务器的重装真简单，不用自己下载系统，不要用u盘，只要重置一下密码就可以了。

重装好了之后，我就从头开始，首先装一杯开水，然后调整心情开始，一切都很顺利，和教程的截图几乎一模一样。我在想！重装真好，我爱重装。Hadoop真好装

然后，现实就给我重重的一击，明明和教程一模一样，但是教程告诉我

检查Hadoop是否已经正确安装输入:/usr/local/hadoop/bin/hadoop version

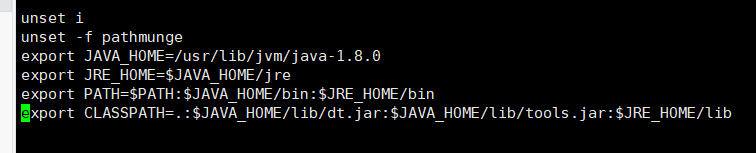
应该就会出现版本信息，但是我出现的是:



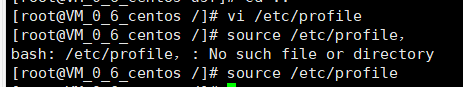
用我最新墙头的口癖那就是： 出大问题！！很没有排FACE！！

我本来准备去睡觉的，但是回过头我的舍友都在努力，我觉得我不能背叛她们，于是我就打开了x度，，原来不止我一个人出现过这样的问题，哈哈哈哈哈哈哈哈哈哈哈哈 教程：[https://blog.csdn.net/mrdonghe/article/details/92371790?tdsourcetag=s\_pcqq\_aiomsg]

感谢这些无私奉献的好人，它告诉我可能是协议不够多，然后就 vi /etc/profile



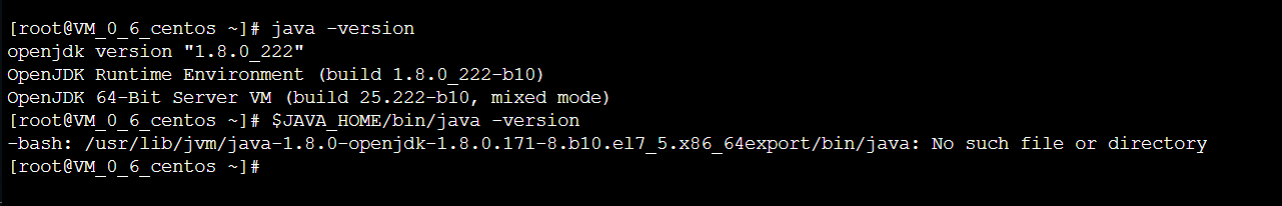
改完之后，我自信满满的重新Hadoop version了一次，但是结果还是一样的，结果我又哭了，反反复复五六次。



然后之后，我就可以了。嗯

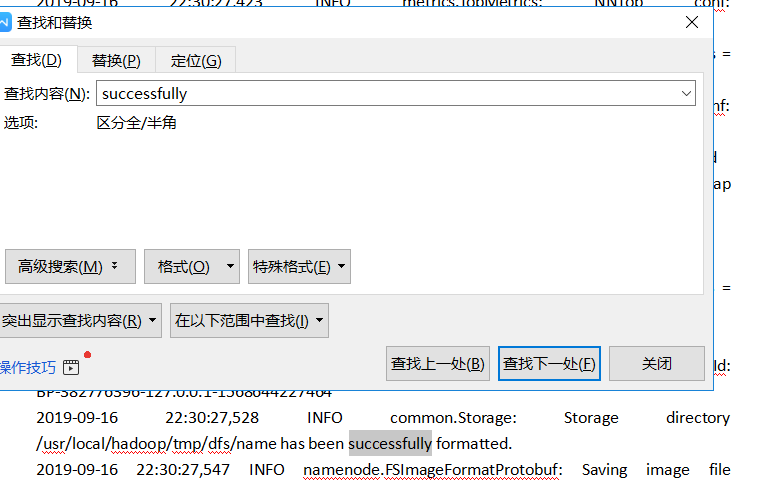


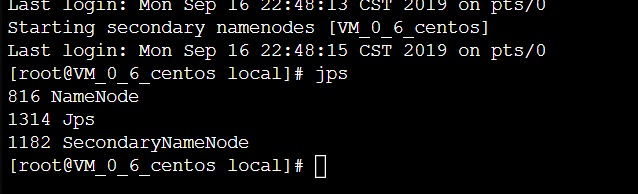
但是其实这个时候又出现了一个新问题



就是我的 java -version $JAVA\_HOME/bin/java -version 两个显示的不一样了，其实又感觉到被背叛了！！！我发现是因为，我忘记了要协议生效！！

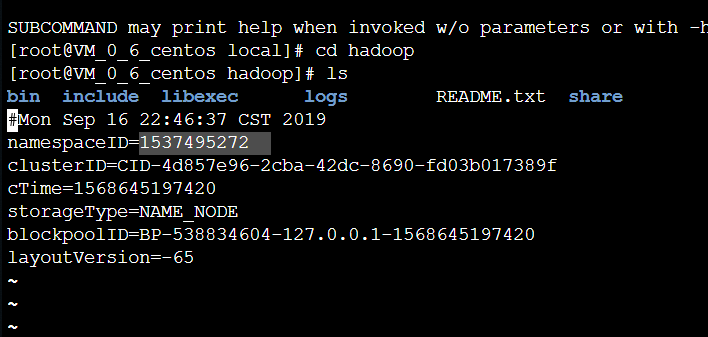
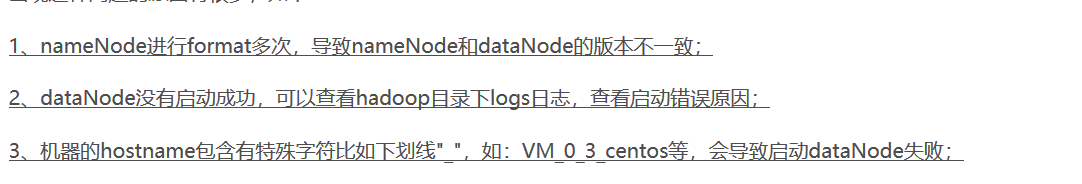
然后按照教程我做到了格式化的那一步，但是我格式化后出来的东西真的是太多了，但是我灵机一动，突然想起来有一个叫做word的办公软件，有一个查找功能（赞）



但是很明显我的结果和网上的截图不一样，不然我的实验报告到此就结束了，嘻嘻嘻嘻嘻。于是我就重新初始化了一次，结果这次还缺少一个结点 。说真的心态有点崩了，还不如去睡觉。

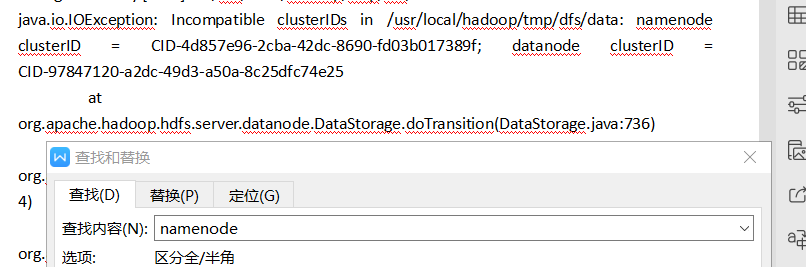
但是这时候用我旧墙头的话就是：如果你练了10遍，你准备放弃了，你就再跳一次，你看看这一次会不会比第一次好，如果好的话，那就说明你的练习有用，那就可以继续再练10遍（ps：大概意思是这个，具体怎么描述我忘了，真的是一个很称职的IDOL！！！）

于是乎受到偶像的激励，我就又去百度了一下，果真还是有很多和我一样的笨蛋呢



namespaceID=1537495272

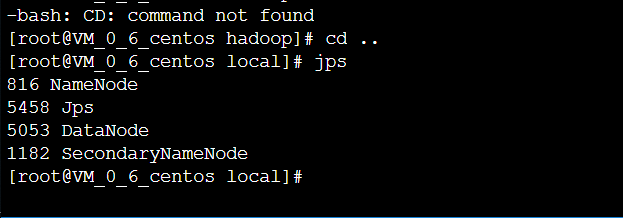
（查找功能真的非常好用！！！！！）



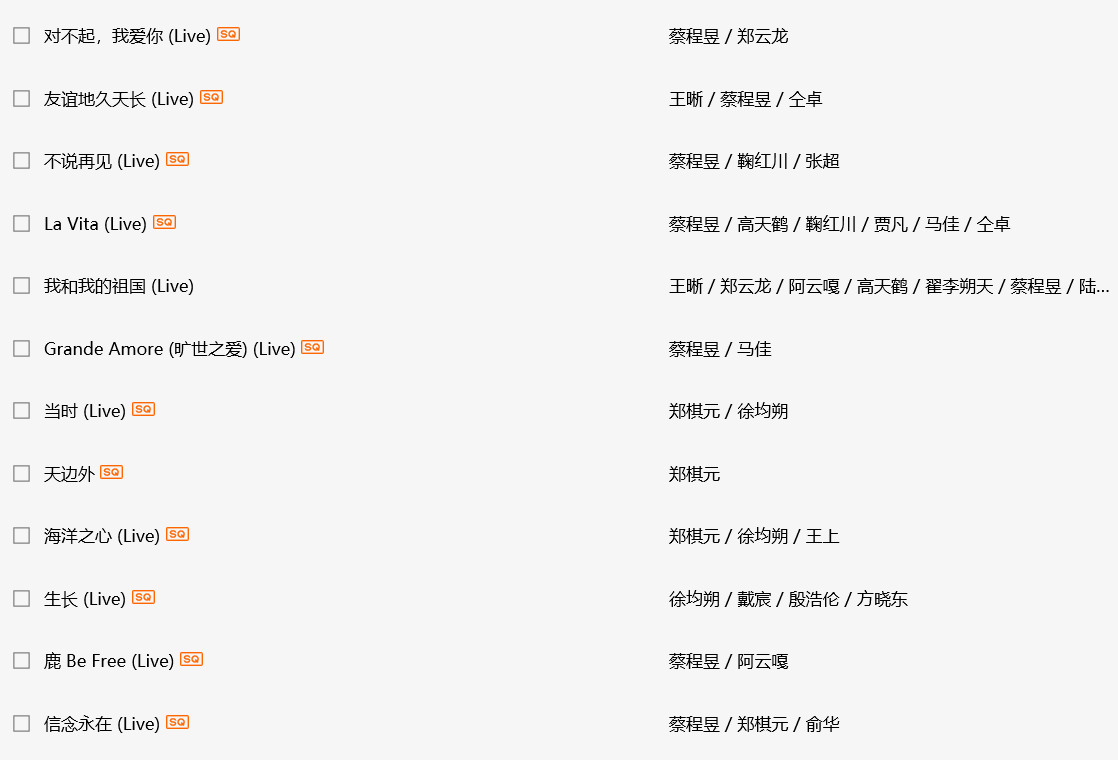
CID-4d857e96-2cba-42dc-8690-fd03b017389f

Incompatible clusterIDs in /usr/local/hadoop/tmp/dfs/data: namenode clusterID = CID-4d857e96-2cba-42dc-8690-fd03b017389f;

datanode clusterID = CID-97847120-a2dc-49d3-a50a-8c25dfc74e25



分享一下这次实验的歌单



分享一下百分之一的搜索记录