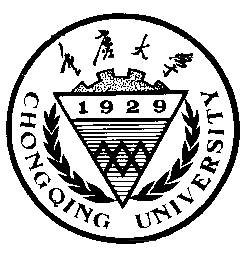
数据科学导论实验报告

实验一 python数据基础实践



学 生：严奕凡\_20220815

王国祥\_20221704

郑俊博\_20220652

张天议\_20220847

年 级：2022级

专 业：数据科学与大数据技术

重庆大学大数据与软件学院

2024年 5月 20日

## 实验目的：

为后续上机实验做准备，熟悉常用的Linux 操作和Hadoop 操作

## 实验要求

见实验内容

## 开发环境：

Windows下Vmware虚拟机搭载Linux Ubunto 64位

## 实验内容：

（一） 熟悉常用的**Linux** 操作

请按要求上机实践如下linux 基本命令。

cd 命令：切换目录

（1）切换到目录/usr/local

（2）去到目前的上层目录

（3）回到自己的主文件夹

ls 命令：查看文件与目录

（4）查看目录/usr 下所有的文件

mkdir 命令：新建新目录

（5）进入/tmp 目录，创建一个名为a 的目录,并查看有多少目录存在

（6）创建目录a1/a2/a3/a4

rmdir 命令：删除空的目录

（7）将上例创建的目录a（/tmp 下面）删除

（8）删除目录a1/a2/a3/a4，查看有多少目录存在

cp 命令：复制文件或目录

（9）将主文件夹下的.bashrc 复制到/usr 下，命名为bashrc1

（10）在/tmp 下新建目录test，再复制这个目录内容到/usr

mv 命令：移动文件与目录，或更名

（11）将上例文件bashrc1 移动到目录/usr/test

（12）将上例test 目录重命名为test2

rm 命令：移除文件或目录

（13）将上例复制的bashrc1 文件删除

（14）将上例的test2 目录删除

cat 命令：查看文件内容

（15）查看主文件夹下的.bashrc 文件内容

tac 命令：反向列示

（16）反向查看主文件夹下.bashrc 文件内容

more 命令：一页一页翻动查看

（17）翻页查看主文件夹下.bashrc 文件内容

head 命令：取出前面几行

（18）查看主文件夹下.bashrc 文件内容前20 行

（19）查看主文件夹下.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

（20）查看主文件夹下.bashrc 文件内容最后20 行

（21） 查看主文件夹下.bashrc 文件内容，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

（22）在/tmp 下创建一个空文件hello 并查看时间

（23）修改hello 文件，将日期调整为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/目录下新建文件夹test,然后在/目录下打包成test.tar.gz

（27）解压缩到/tmp 目录

grep 命令：查找字符串

（28）从～/.bashrc 文件中查找字符串'examples'

（29）配置Java 环境变量，在～/.bashrc 中设置

（30）查看JAVA\_HOME 变量的值

（二）熟悉常用的**Hadoop** 操作

（31）启动hadoop，在HDFS 中创建用户目录（现在已经在hadoop 目录/usr/local/hadoop）

（32）接着在此用户目录下创建text 文件夹，并查看文件列表

（33）将～/.bashrc 文件上传到HDFS 的test 文件夹，并查看test

（34）将HDFS 文件夹test 拷到本机

## 实验结果分析(重点)

（一） 熟悉常用的**Linux** 操作

请按要求上机实践如下linux 基本命令。

cd 命令：切换目录

1. 切换到目录/usr/local



1. 去到目前的上层目录



1. 回到自己的主文件夹



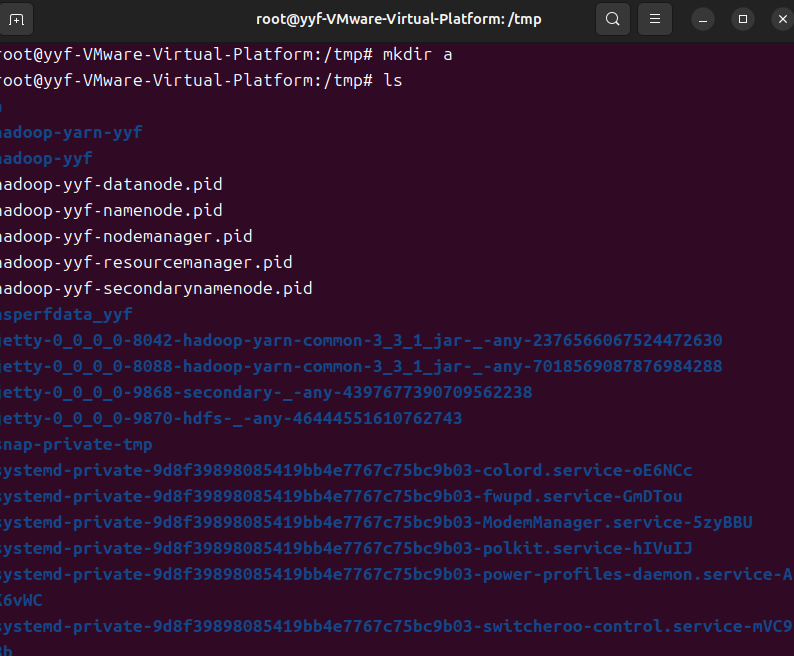
ls 命令：查看文件与目录

1. 查看目录/usr 下所有的文件

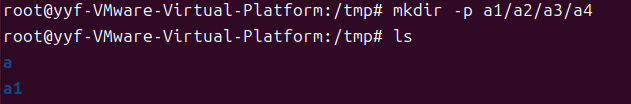


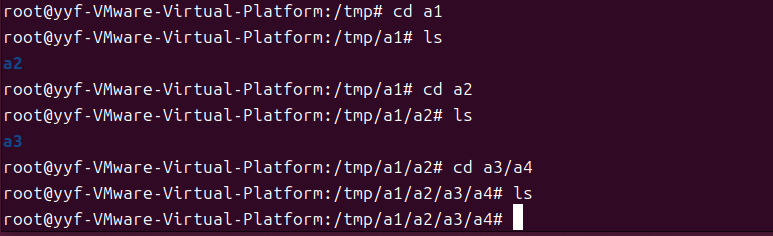
mkdir 命令：新建新目录

1. 进入/tmp 目录，创建一个名为a 的目录,并查看有多少目录存在



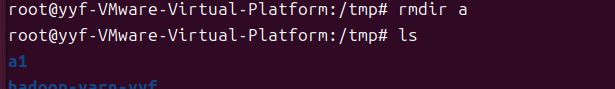
1. 创建目录a1/a2/a3/a4



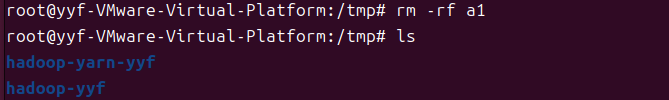


rmdir 命令：删除空的目录

1. 将上例创建的目录a（/tmp 下面）删除



1. 删除目录a1/a2/a3/a4，查看有多少目录存在

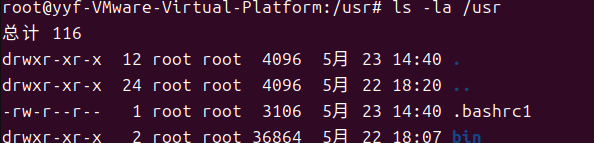


-r递归地删除 -f强制删除

cp 命令：复制文件或目录

1. 将主文件夹下的.bashrc 复制到/usr 下，命名为bashrc1





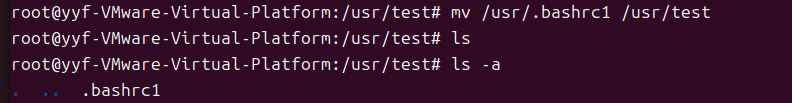
1. 在/tmp 下新建目录test，再复制这个目录内容到/usr





mv 命令：移动文件与目录，或更名

1. 将上例文件bashrc1 移动到目录/usr/test



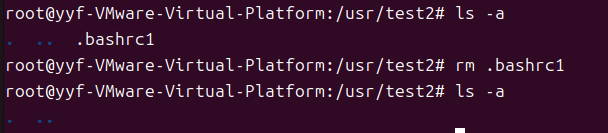
（12）将上例test 目录重命名为test2



这里有点意思：即使我在test目录将test目录重命名为test2 我依旧显示在test 但实际上这个目录已经被重命名了

rm 命令：移除文件或目录

（13）将上例复制的bashrc1 文件删除

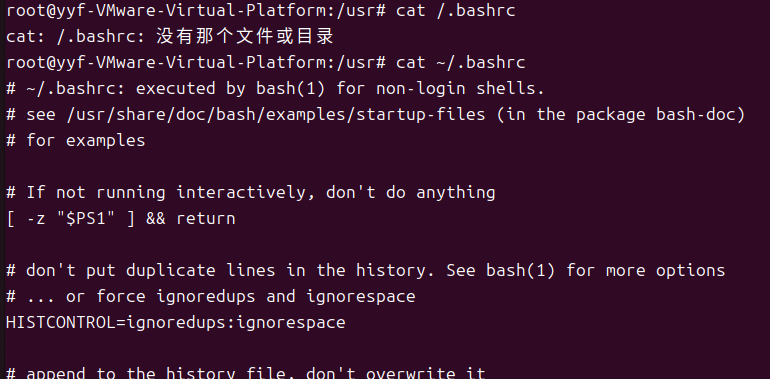


（14）将上例的test2 目录删除



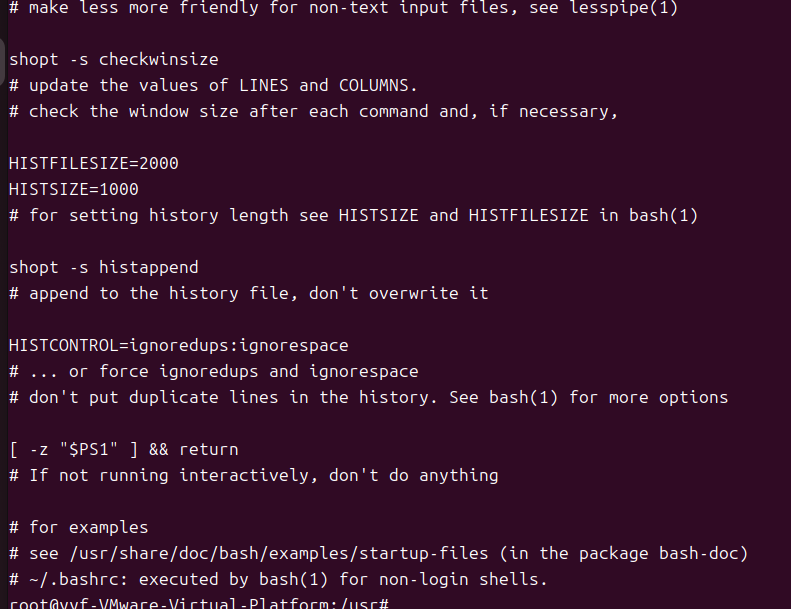
cat 命令：查看文件内容

（15）查看主文件夹下的.bashrc 文件内容



tac 命令：反向列示

（16）反向查看主文件夹下.bashrc 文件内容

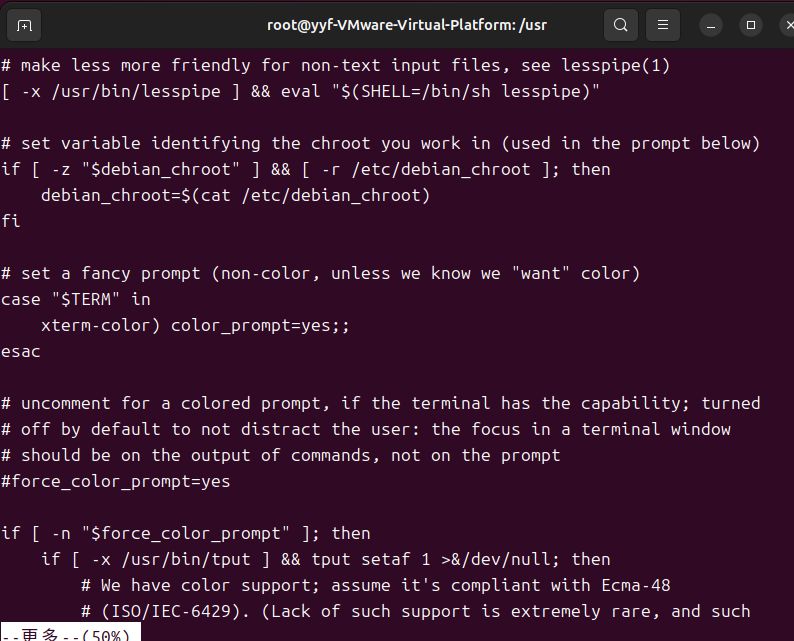


more 命令：一页一页翻动查看

（17）翻页查看主文件夹下.bashrc 文件内容

空格是向下 b是向上 q是退出

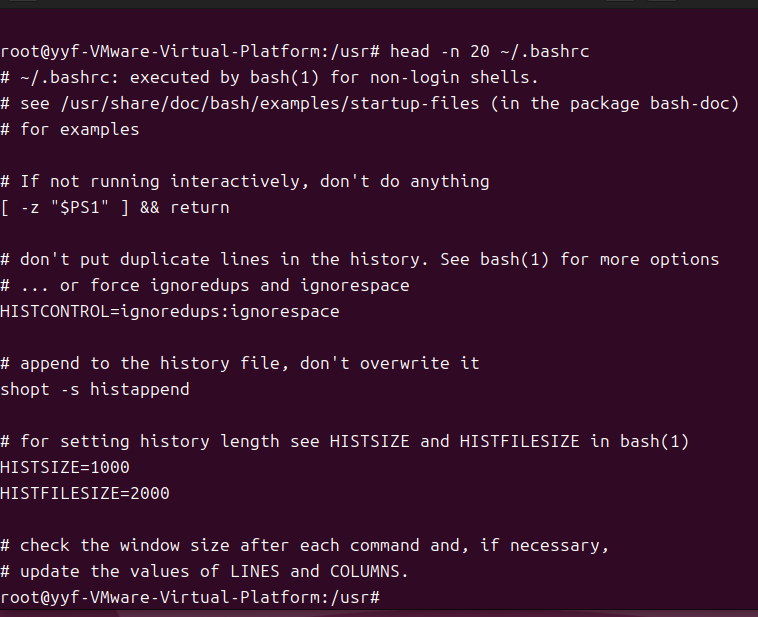
more ~/.bashrc



head 命令：取出前面几行

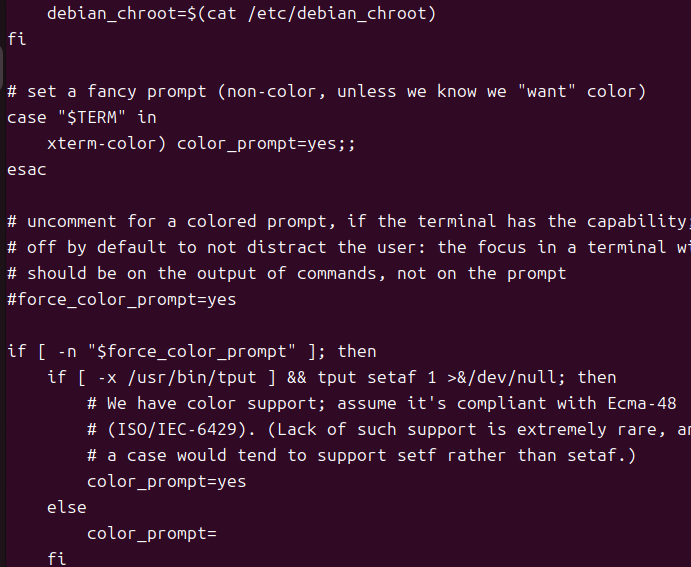
（18）查看主文件夹下.bashrc 文件内容前20 行

head -n 20 ~/.bashrc



（19）查看主文件夹下.bashrc 文件内容，后面50 行不显示，只显示前面几行

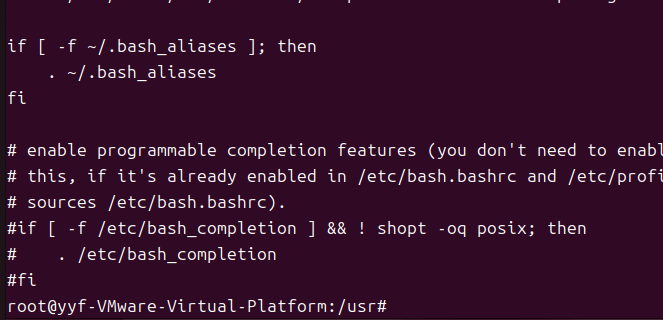
head -n -50 ~/.bashrc



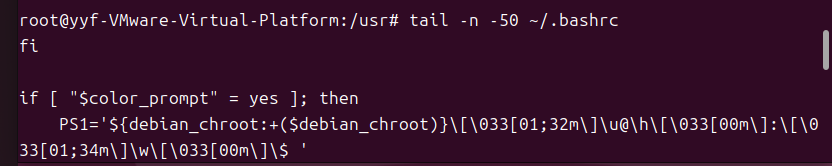
tail 命令：取出后面几行

（20）查看主文件夹下.bashrc 文件内容最后20 行



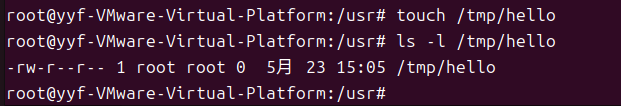


（21） 查看主文件夹下.bashrc 文件内容，只列出50 行以后的数据



touch 命令：修改文件时间或创建新文件

（22）在/tmp 下创建一个空文件hello 并查看时间



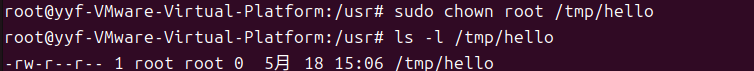
（23）修改hello 文件，将日期调整为5 天前





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



因为我一开始用了sudo -I 所以所有的都是root权限

所以看不出来变化 事实上已经是root了

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/目录下新建文件夹test,然后在/目录下打包成test.tar.gz

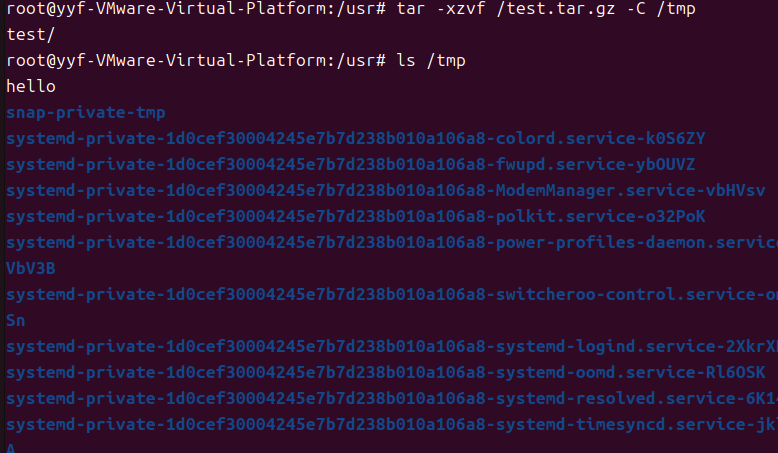


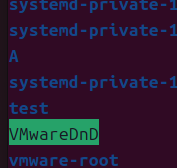
另外：

如果我想把 /usr/test 用tar打包到 /home 命名为a

sudo tar -czvf /home/a.tar.gz -C /usr/test .

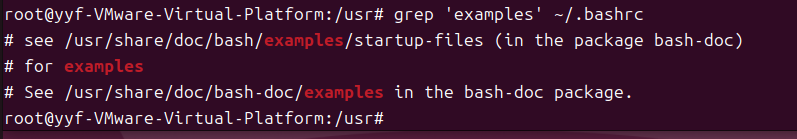
（27）解压缩到/tmp 目录





grep 命令：查找字符串

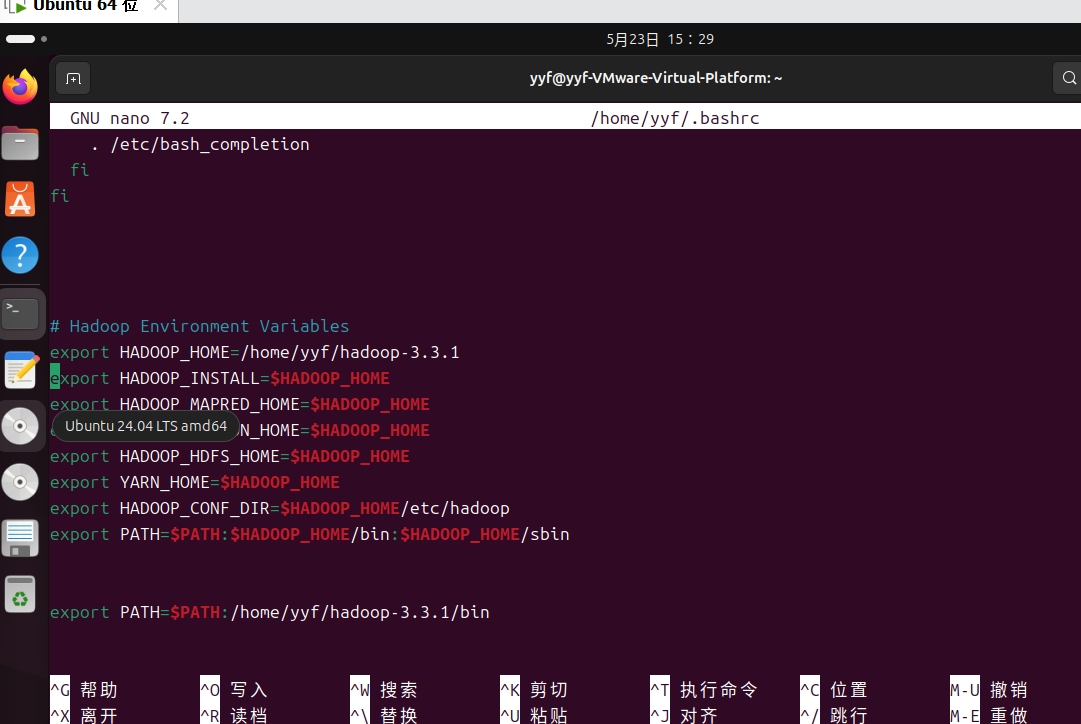
（28）从～/.bashrc 文件中查找字符串'examples'



（29）配置Java 环境变量，在～/.bashrc 中设置

同下

（30）查看JAVA\_HOME 变量的值



注：当我使用sudo -i进入root权限再查看~/.bashrc时 不会有上述信息 因为这些信息是我之前在非root 情况下编辑的 。

通常情况下，root用户的.bashrc文件是空的，或者只包含了一些最基本的设置。这是因为root用户的交互式Shell会话通常被认为是一个高级操作，因此不会加载用户的个性化设置，以避免潜在的安全风险。

要注意的是，root用户的.bashrc文件通常不包含像Hadoop和Java这样的应用程序的环境变量设置，因为这些设置通常是与普通用户相关的，而不是root用户。

（二）熟悉常用的**Hadoop** 操作

启动 star-all.sh  
停止 stop-all.sh

（注意要在非root内执行，root不包含环境变量文件）   
直接启动和停止all可能有一些问题，但暂时不会遇到

（31）启动hadoop，在HDFS 中创建用户目录（现在已经在hadoop 目录/usr/local/hadoop）

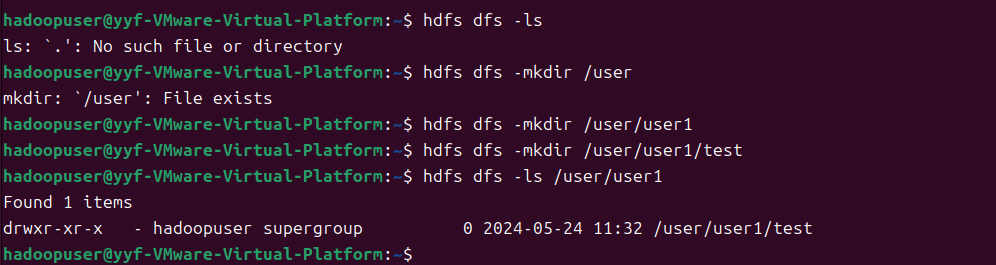
注 在hadoop里面创建目录必须得有父级目录 他不会自动创建的

hdfs dfs -mkdir /user

hdfs dfs -mkdir /user/user1

（32）接着在此用户目录下创建text 文件夹，并查看文件列表

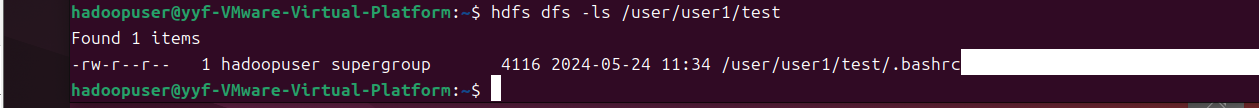
hdfs dfs -mkdir /user/user1/test



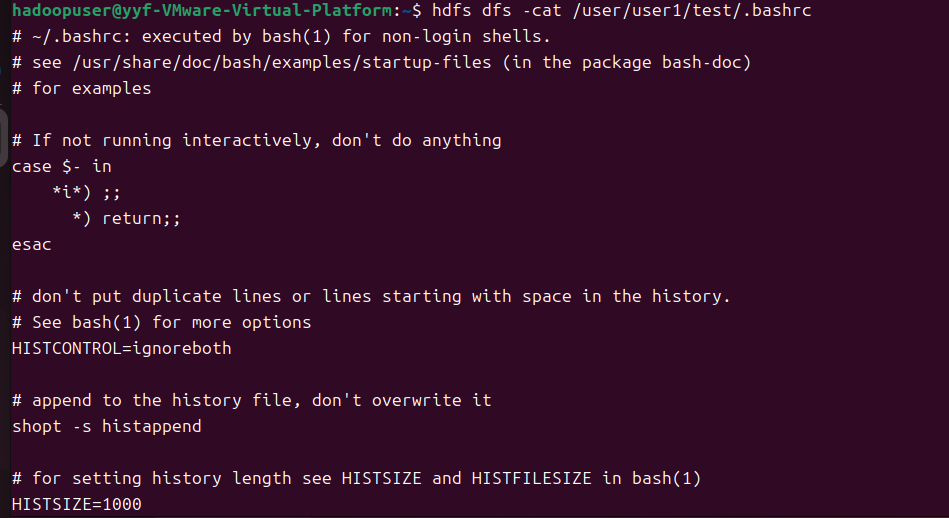
（33）将～/.bashrc 文件上传到HDFS 的test 文件夹，并查看test

hdfs dfs -put ~/.bashrc /user/user1/test



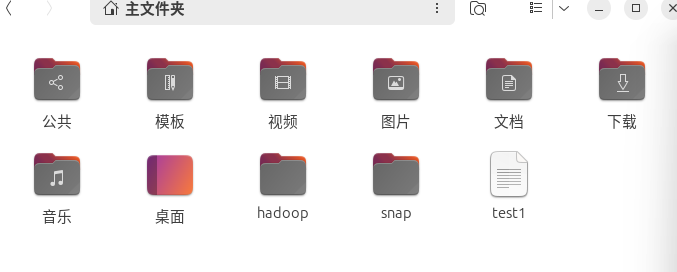


hdfs dfs -cat /path 可以查看文件的



（34）将HDFS 文件夹test 拷到本机

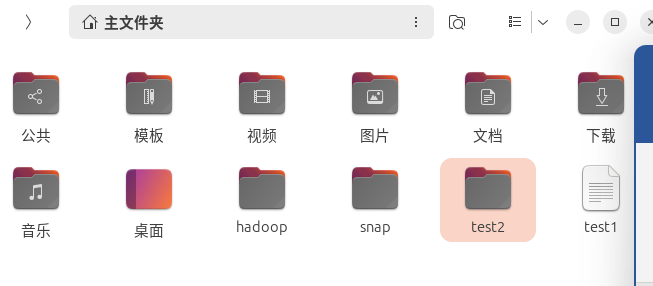




上面是拷贝文件

下面是拷贝文件夹



****

****

## 问题和解决

Linux Ubunto 64位

1、需要用到vmware tool ，以便于在虚拟机和主机之间交流   
  
2、常用的快捷方式   
control+shift+c复制

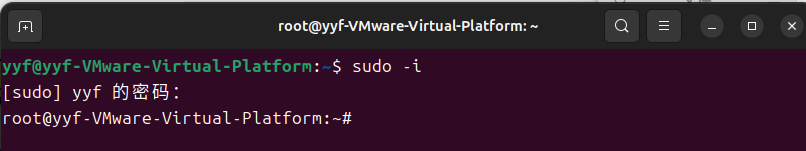
Control+shift+v粘贴

注：在有vmware tool之后 复制粘贴可以参考windows 但是在终端中有所不同

Control+H查看隐藏文件

Control+alter+t = 新建一个终端

Sudo -I 并输入密码 使得当前终端获得管理员权限

如图

3、一些问题

Windows主机通过vm-tool像ubunto虚拟机传文件

新建并切换一个hadoop专用用户

ssh配置

JAVA\_HOME配置

Hadoop PATH配置

Hadoop初始化

Hadoop操作

我的代码仓库地址：[mozhongzhou/CQU\_Python\_experiment (github.com)](https://github.com/mozhongzhou/CQU_Python_experiment)