

**《大数据平台及应用A》**

**平时作业**

**姓 名 钱晓楠**

**学 号 20221003675**

**班 号 1 9 5 2 2 1**

**专 业 数据科学与大数据技术**

**所 在 院 系 计算机学院**

**指 导 老 师 李 程 俊**

**日 期 2024年1月6日**

**目录**

[一、 第一次上机作业 2](#_Toc503)

[cd命令：切换目录 2](#_Toc15175)

[ls命令：查看文件与目录 2](#_Toc29537)

[mkdir命令：新建新目录 3](#_Toc8804)

[rmdir命令：删除空的目录 4](#_Toc17687)

[cp命令：复制文件或目录 5](#_Toc27835)

[mv命令：移动文件与目录，或更名 5](#_Toc10692)

[rm命令：移除文件或目录 6](#_Toc27968)

[cat命令：查看文件内容 6](#_Toc12213)

[tac命令：反向列示 7](#_Toc2953)

[more命令：一页一页翻动查看 9](#_Toc28835)

[head命令：取出前面几行 9](#_Toc28927)

[tail命令：取出后面几行 11](#_Toc21663)

[touch命令：修改文件时间或创建新文件 12](#_Toc24723)

[chown命令：修改件所有者权限 12](#_Toc10608)

[find命令：文件查找 13](#_Toc5407)

[tar命令：压缩命令 13](#_Toc6636)

[grep命令：查找字符串 13](#_Toc9188)

[HDFS相关命令 14](#_Toc11539)

[二、 第二次上机作业 15](#_Toc2997)

[2.1第一题 15](#_Toc29421)

[2.2 第二题 20](#_Toc7898)

[2.3 第三题 23](#_Toc30082)

[三、 第三次上机作业 25](#_Toc9393)

[3.1 第一题 25](#_Toc30305)

[3.2 第二题 29](#_Toc14730)

[3.3 第三题 32](#_Toc1426)

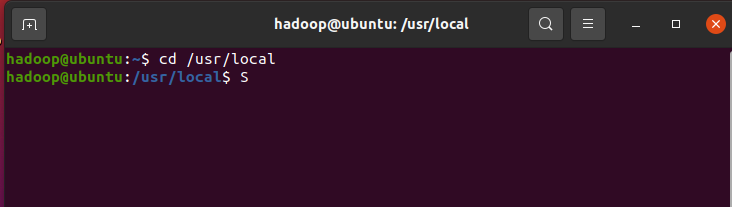
[四、 第四次上机作业 37](#_Toc26210)

# 第一次上机作业

## cd命令：切换目录

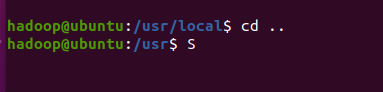
（1）切换到目录 /usr/local

cd /usr/local



（2）去到目前的上层目录

cd ..



（3）回到自己的主文件夹

cd ~

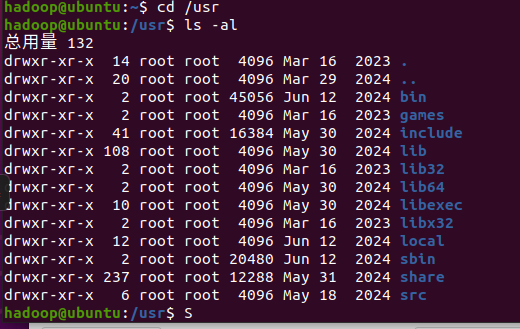


## ls命令：查看文件与目录

（4）查看目录/usr下所有的文件

cd /usr

ls -al



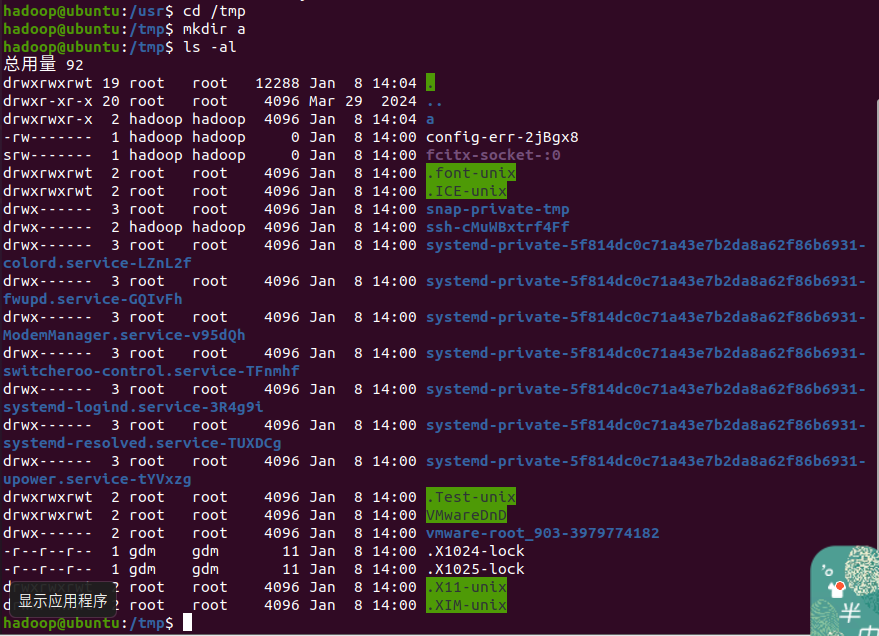
## mkdir命令：新建新目录

（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

cd /tmp

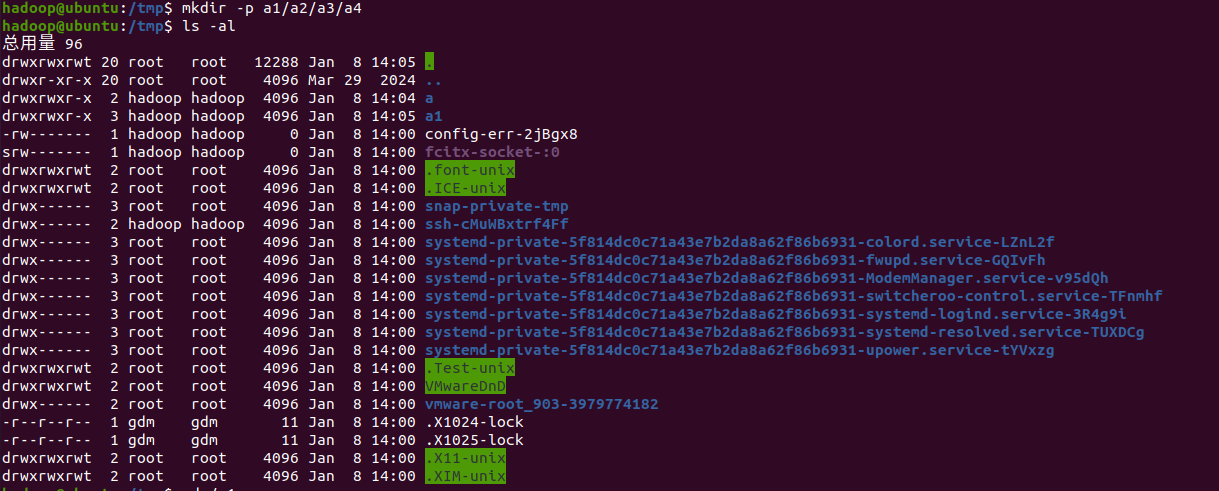
mkdir a

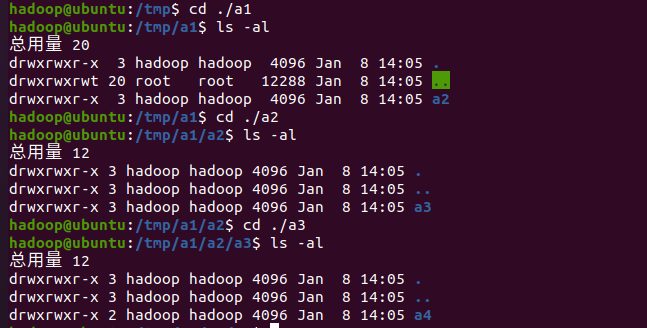
ls -al



（6）创建目录a1/a2/a3/a4

mkdir -p a1/a2/a3/a4

依次进入a1,a2,a3，可以看到文件夹均成功创建：



## rmdir命令：删除空的目录

（7）将上例创建的目录a（/tmp下面）删除

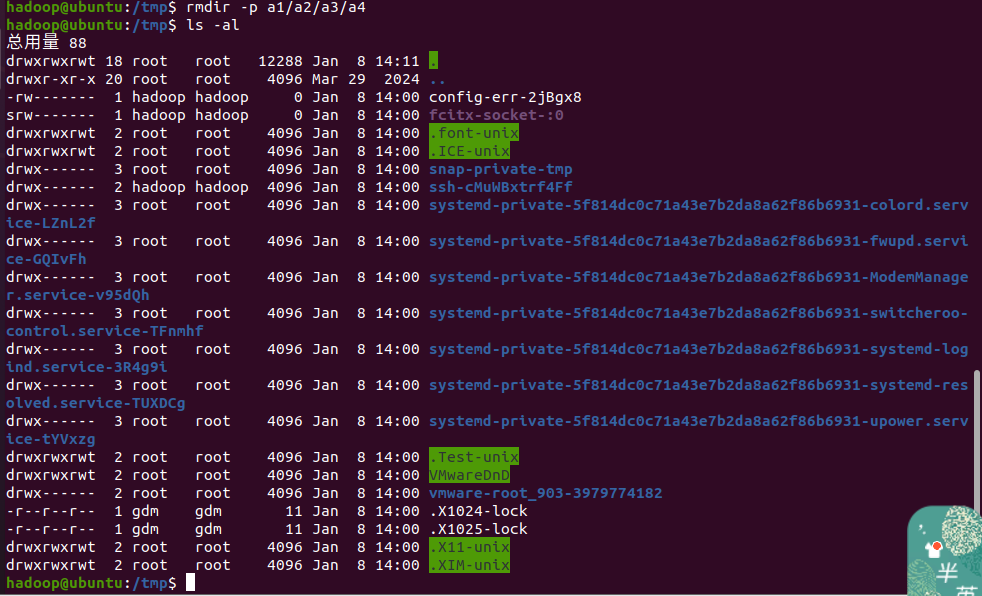
rmdir a



（8）删除目录a1/a2/a3/a4，查看有多少目录存在

rmdir -p a1/a2/a3/a4

ls -al



## 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

sudo cp ~/.bashrc /usr/bashrc1



（10）在/tmp下新建目录test，再复制这个目录内容到/usr

cd /tmp

mkdir test

sudo cp -r /tmp/test /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

sudo mv /usr/bashrc1 /usr/test



（12）将上例test目录重命名为test2

sudo mv /usr/test /usr/test2



## rm命令：移除文件或目录

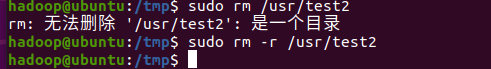
（13）将上例复制的bashrc1文件删除

sudo rm /usr/test2/bashrc1



（14）将上例的test2目录删除

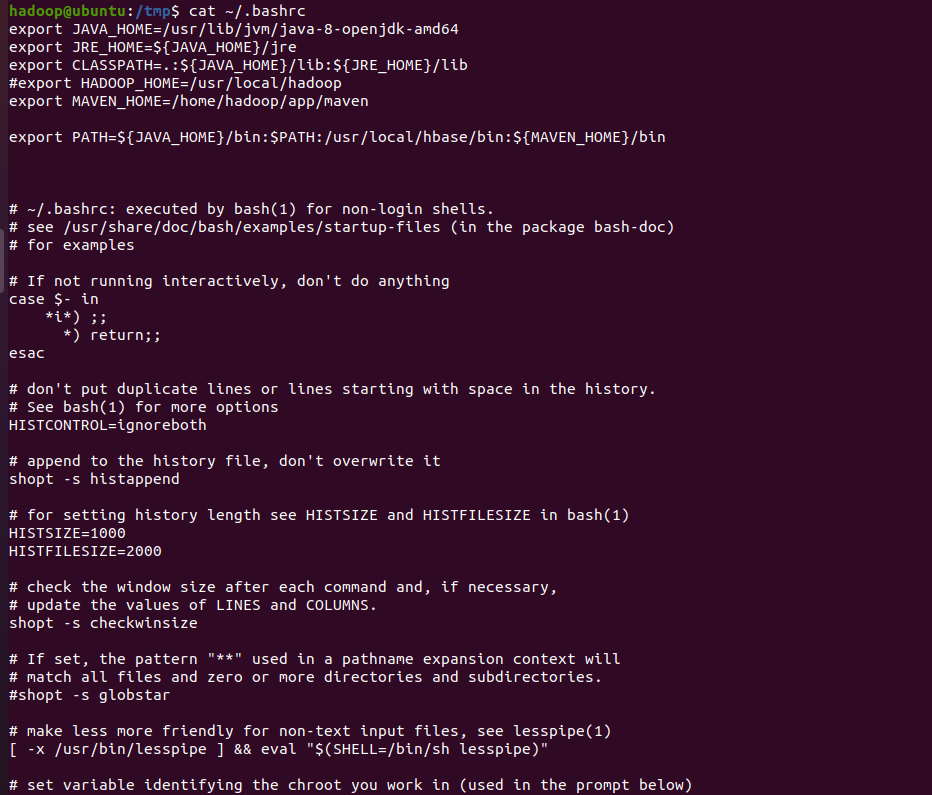
sudo rm -r /usr/test2

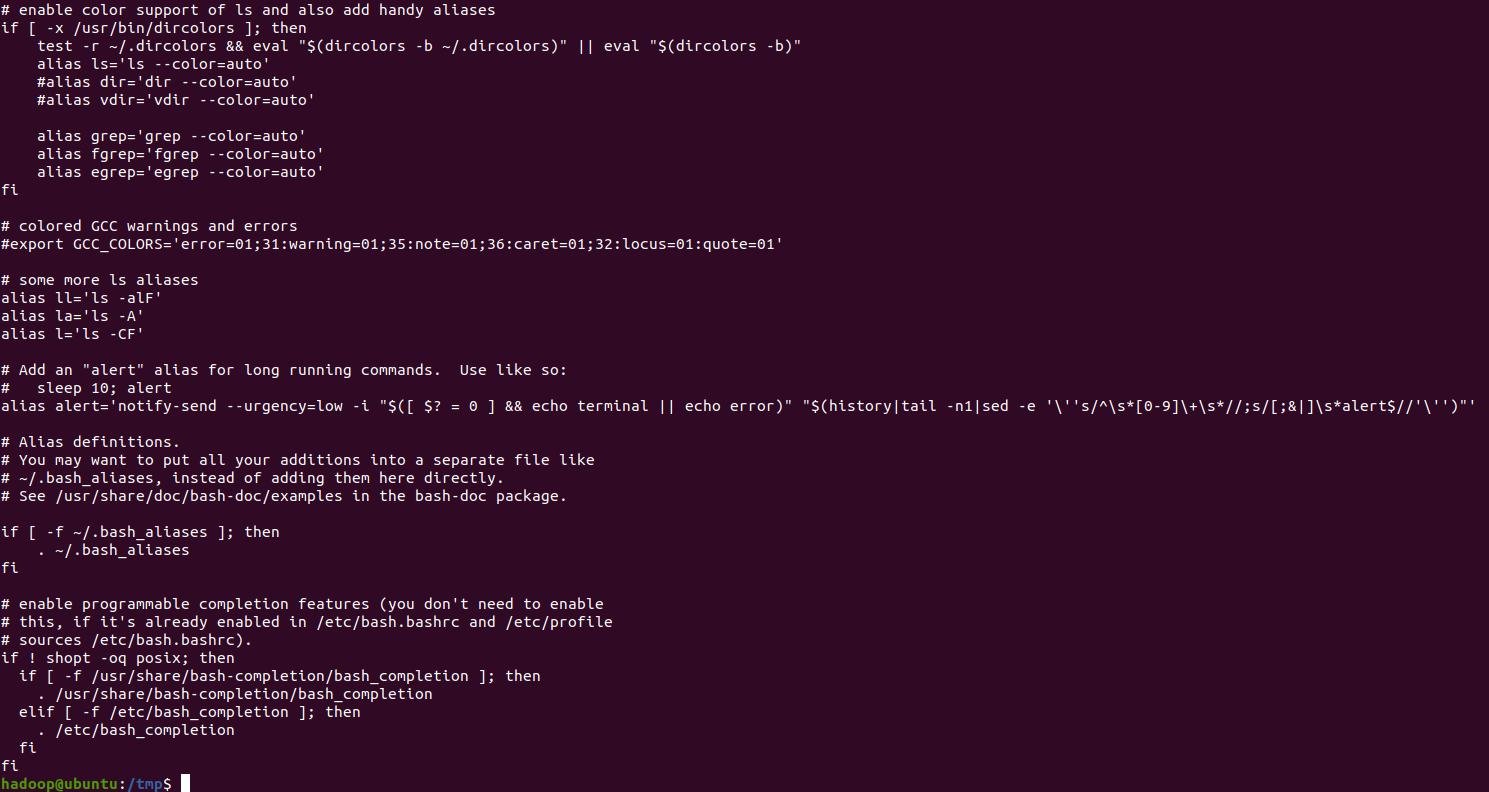


## cat命令：查看文件内容

（15）查看主文件夹下的.bashrc文件内容

cat ~/.bashrc

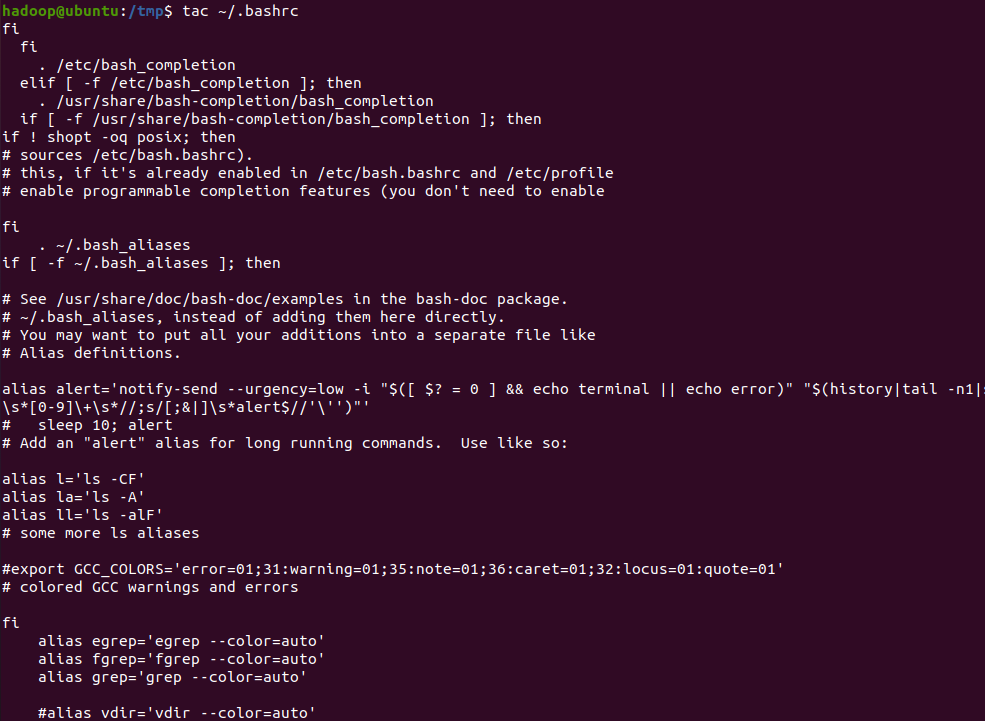


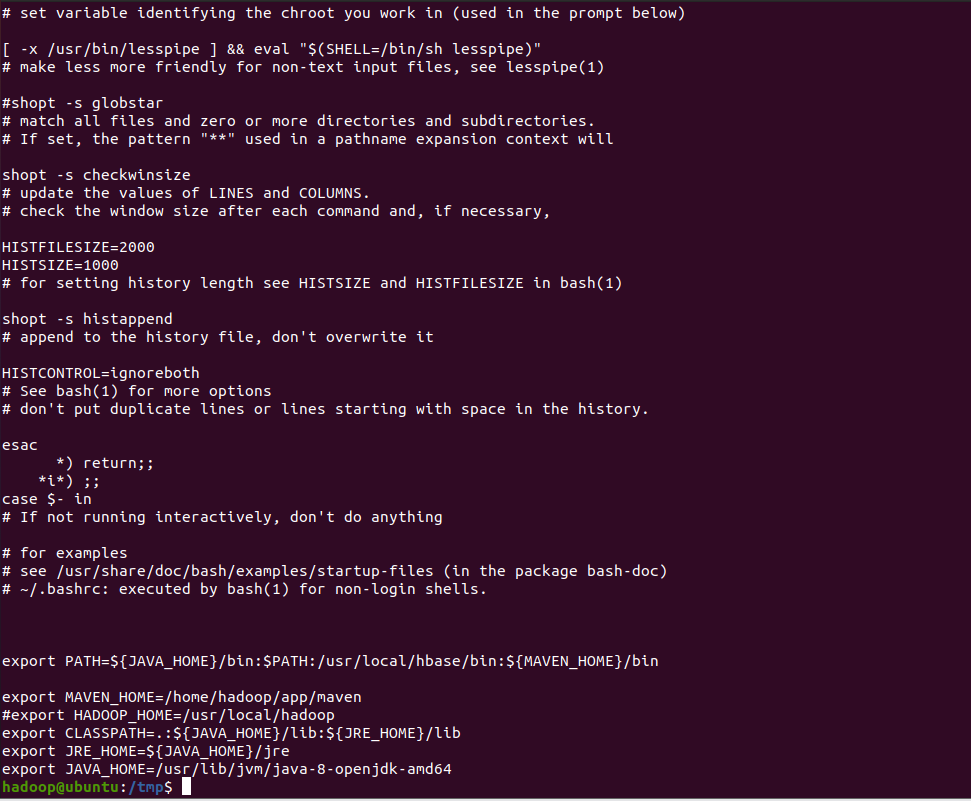


## tac命令：反向列示

（16）反向查看主文件夹下.bashrc文件内容

tac ~/.bashrc

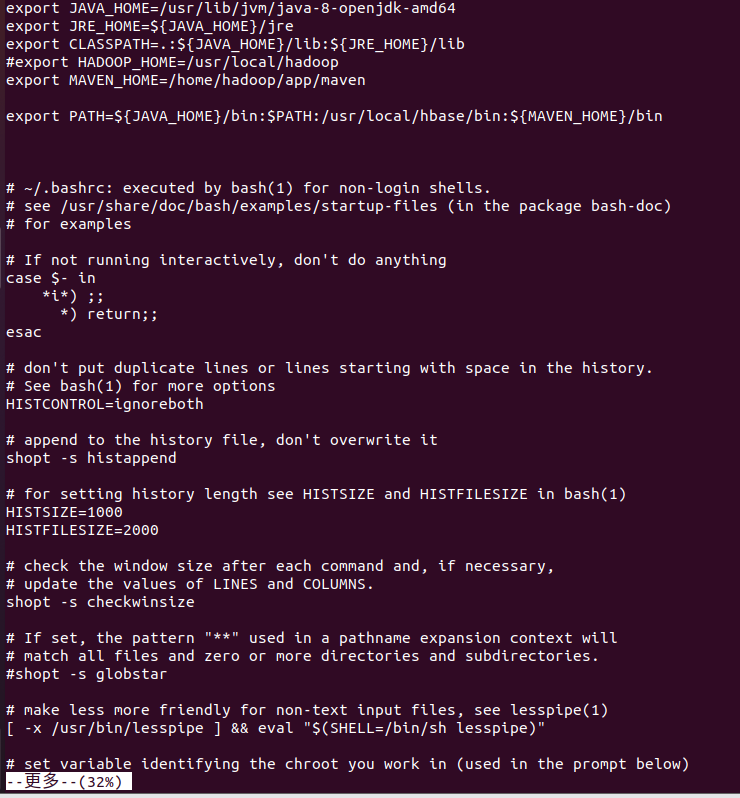




## more命令：一页一页翻动查看

（17）翻页查看主文件夹下.bashrc文件内容

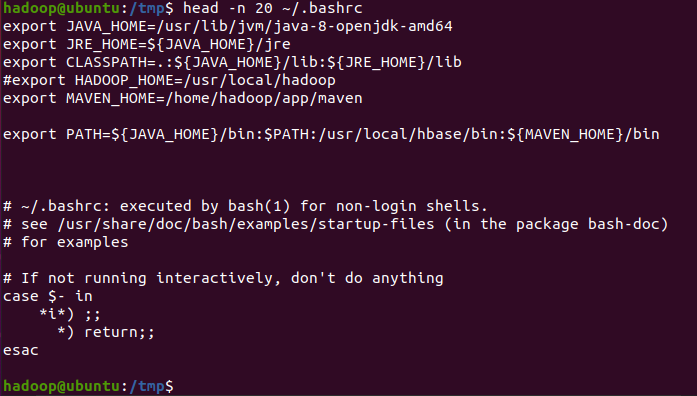
more ~/.bashrc



## head命令：取出前面几行

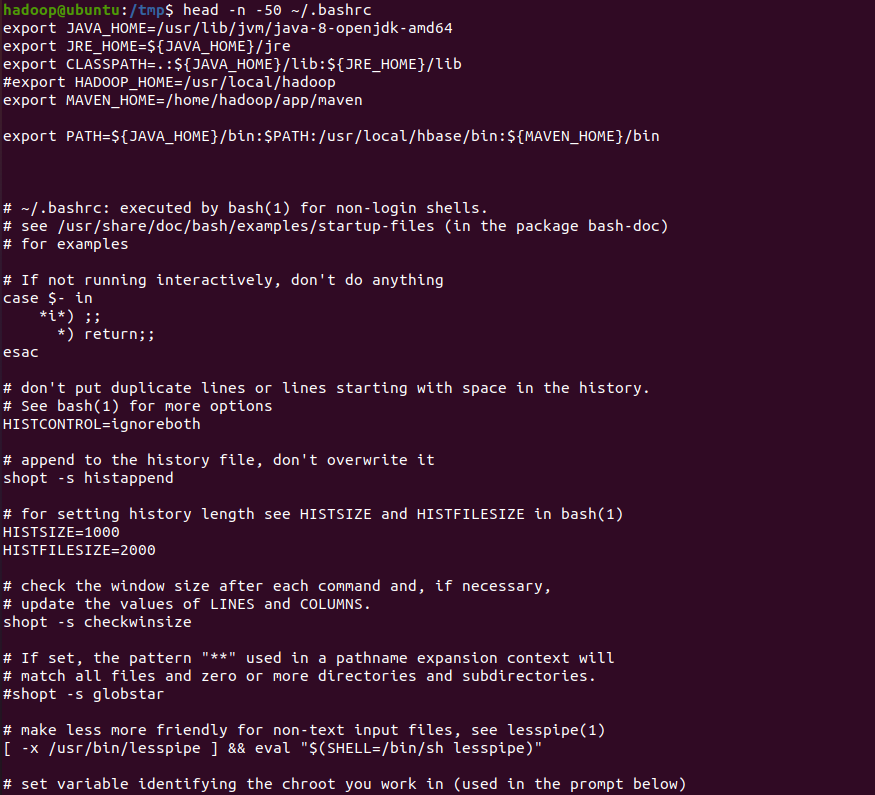
（18）查看主文件夹下.bashrc文件内容前20行

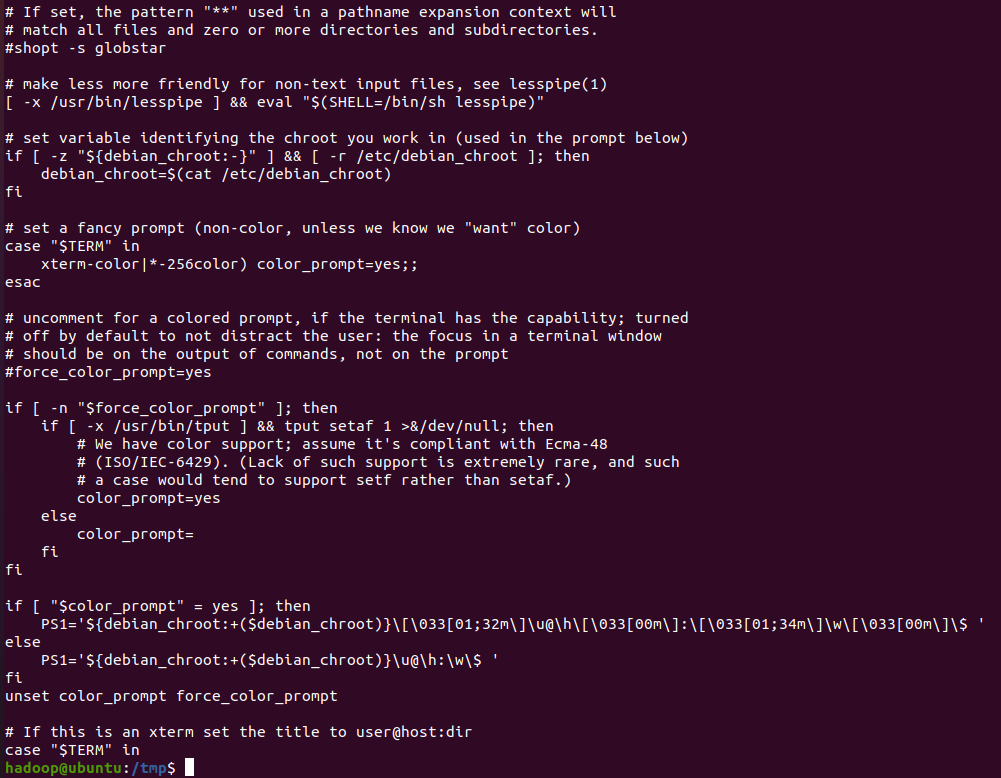
head -n 20 ~/.bashrc



（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

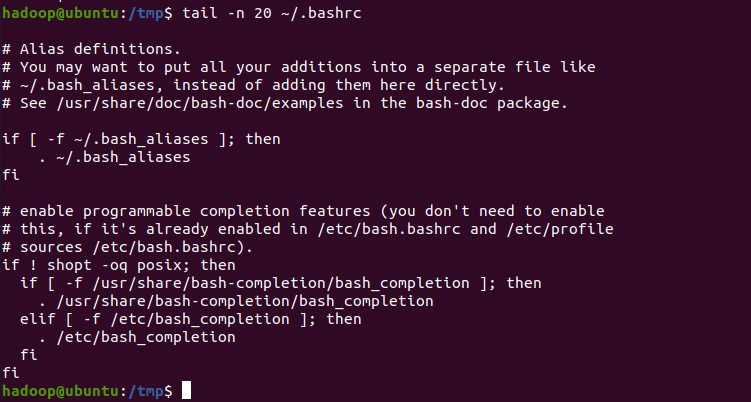




## tail命令：取出后面几行

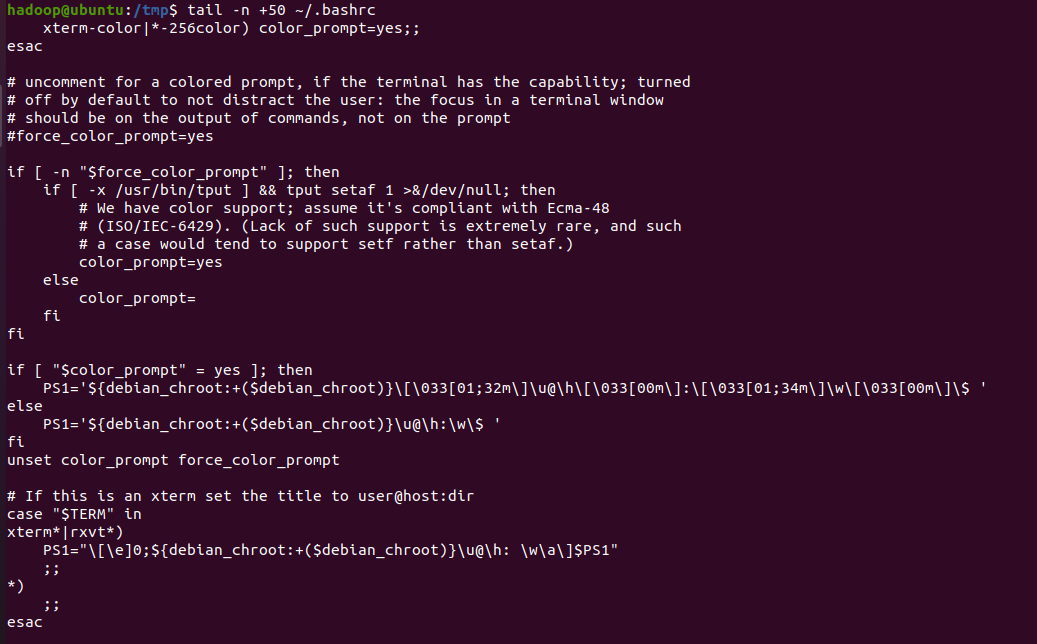
（20）查看主文件夹下.bashrc文件内容最后20行

tail -n 20 ~/.bashrc



（21） 查看主文件夹下.bashrc文件内容，只列出50行以后的数据

tail -n +50 ~/.bashrc



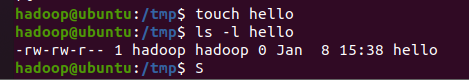
## touch命令：修改文件时间或创建新文件

（22）在/tmp下创建一个空文件hello并查看时间

cd /tmp

touch hello

ls -l hello



（23）修改hello文件，将日期调整为5天前

touch -d “5 days ago” hello

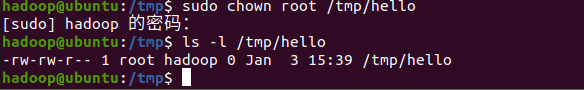


## chown命令：修改件所有者权限

（24）将hello文件所有者改为root帐号，并查看属性

sudo chown root /tmp/hello

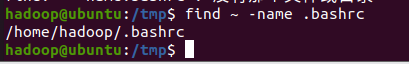
ls -l /tmp/hello



## find命令：文件查找

（25）找出主文件夹下文件名为.bashrc的文件

find ~ -name .bashrc



## 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

sudo mkdir /test

sudo tar -zcv -f /test.tar.gz test



（27）解压缩到/tmp目录

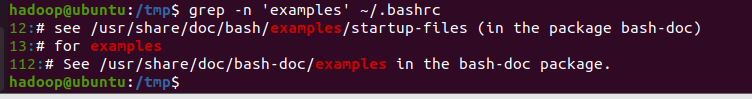
sudo tar -zxv -f /test.tar.gz -C /tmp



## grep命令：查找字符串

（28）从～/.bashrc文件中查找字符串'examples'

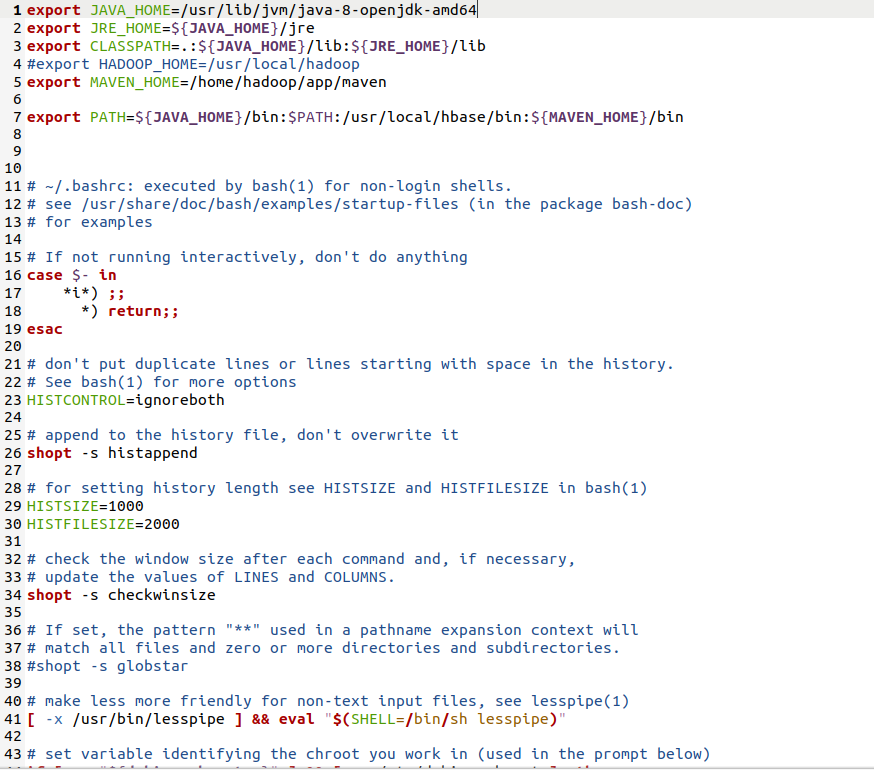
grep -n 'examples' ~/.bashrc



（29）配置Java环境变量，在～/.bashrc中设置

gedit ~/.bashrc

将语句 export JAVA\_HOME=JDK安装路径 加入文件第一行并保存



source ~/.bashrc



（30）查看JAVA\_HOME变量的值

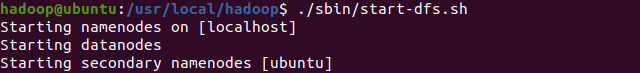
echo $JAVA\_HOME



## HDFS相关命令

1. 启动hadoop

./sbin/start-dfs.sh



1. 在HDFS中创建用户目录（现在已经在hadoop目录/usr/local/hadoop）

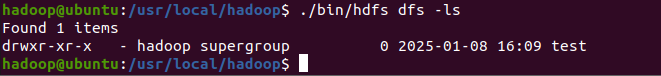
./bin/hdfs dfs -mkdir -p /user/hadoop



（32）接着在此用户目录下创建text文件夹，并查看文件列表

./bin/hdfs dfs -mkdir test

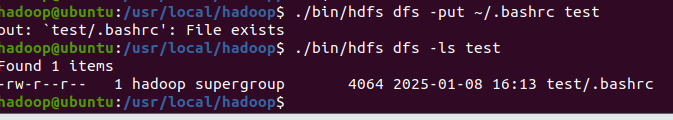
./bin/hdfs dfs -ls .



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

./bin/hdfs dfs -put ~/.bashrc test

./bin/hdfs dfs -ls test



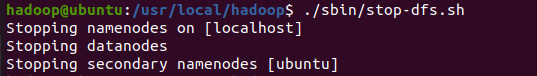
（34）将HDFS文件夹test拷到本机

./bin/hdfs dfs -get test ./test



查阅资料发现，2025-01-08 16:28:17,048 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false属于正常现象，无视即可。

最后关闭hdfs：



# 第二次上机作业

## 2.1第一题

1.编程实现以下指定功能，并利用Hadoop提供的Shell命令完成相同任务：

1. 向HDFS中上传任意文本文件，如果指定的文件在HDFS中已经存在，由用户指定是追加到原有文件末尾还是覆盖原有的文件；

|  |
| --- |
| Shell命令:  检查文件是否存在: hdfs dfs -test -e text.txt(执行完这一句不会输出结果，需要继续输入命令  " echo $?")  屏幕截图 2025-01-09 074139  可以看到文件不存在，故采用覆盖命令  追加命令: hdfs dfs -appendToFile local.txt text.txt  覆盖命令1: hdfs dfs -copyFromLocal -f local.txt text.txt  覆盖命令2: hdfs dfs -cp -f file:///home/hadoop/local.txt text.txt  创建local.txt并传入hdfs：    查阅资料发现，2025-01-08 16:28:17,048 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false属于正常现象，无视即可。  也可以使用如下命令实现：  （如下代码可视为一行代码，在终端中输入第一行代码后，直到输入 fi 才会真正执行）：  if $(hdfs dfs -test -e text.txt);  then $(hdfs dfs -appendToFile local.txt text.txt);  else $(hdfs dfs -copyFromLocal -f local.txt text.txt);  fi |

1. 从HDFS中下载指定文件，如果本地文件与要下载的文件名称相同，则自动对下载的文件重命名；

|  |
| --- |
| Shell命令：  if $(hdfs dfs -test -e file:///home/hadoop/text.txt);  then $(hdfs dfs -copyToLocal text.txt ./text2.txt);  else $(hdfs dfs -copyToLocal text.txt ./text.txt);  fi |

1. 将HDFS中指定文件的内容输出到终端中；

|  |
| --- |
| Shell命令：  hdfs dfs -cat text.txt |

1. 显示HDFS中指定的文件的读写权限、大小、创建时间、路径等信息；

|  |
| --- |
| Shell命令：  hdfs dfs -ls -h text.txt |

1. 给定HDFS中某一个目录，输出该目录下的所有文件的读写权限、大小、创建时间、路径等信息，如果该文件是目录，则递归输出该目录下所有文件相关信息；

|  |
| --- |
| Shell命令：  hdfs dfs -ls -R -h /user/hadoop |

1. 提供一个HDFS内的文件的路径，对该文件进行创建和删除操作。如果文件所在目录不存在，则自动创建目录；

|  |
| --- |
| Shell命令：  if $(hdfs dfs -test -d dir1/dir2);  then $(hdfs dfs -touchz dir1/dir2/filename);  else $(hdfs dfs -mkdir -p dir1/dir2 && hdfs dfs -touchz dir1/dir2/filename);  fi    删除文件：hdfs dfs -rm dir1/dir2/filename |

1. 提供一个HDFS的目录的路径，对该目录进行创建和删除操作。创建目录时，如果目录文件所在目录不存在则自动创建相应目录；删除目录时，由用户指定当该目录不为空时是否还删除该目录；

|  |
| --- |
| Shell命令：  创建目录：hdfs dfs -mkdir -p dir1/dir2  删除目录（如果目录非空则会提示not empty，不执行删除）：hdfs dfs -rmdir dir1/dir2  强制删除目录：hdfs dfs -rm -R dir1/dir2    由于文件为空，所以执行hdfs dfs -rmdir dir1/dir2时目录已经被删除    当在目录中创建文件后，执行hdfs dfs -rmdir dir1/dir2时目录不会被删除    执行强制删除命令 |

1. 向HDFS中指定的文件追加内容，由用户指定内容追加到原有文件的开头或结尾；

|  |
| --- |
| Shell命令：  追加到文件末尾：hdfs dfs -appendToFile local.txt text.txt    可以看到文件被成功追加到末尾  追加到文件开头：  （由于没有直接的命令可以操作，方法之一是先移动到本地进行操作，再进行上传覆盖）：  hdfs dfs -get text.txt  cat text.txt >> local.txt  hdfs dfs -copyFromLocal -f local.txt text.txt    由于本地已经存在text.txt，所以将hdfs上的text.txt的传到本地重命名为text1.txt。同时local.txt的文件内容已经追加上了“1212”，所以可以看到，文件的内容成功被追加到文件开头。 |

1. 删除HDFS中指定的文件；

|  |
| --- |
| Shell命令：  hdfs dfs -rm text.txt |

1. 删除HDFS中指定的目录，由用户指定目录中如果存在文件时是否删除目录；

|  |
| --- |
| Shell命令：  删除目录（如果目录非空则会提示not empty，不执行删除）：hdfs dfs -rmdir dir1/dir2  强制删除目录：hdfs dfs -rm -R dir1/dir2    由于文件为空，所以执行hdfs dfs -rmdir dir1/dir2时目录已经被删除    当在目录中创建文件后，执行hdfs dfs -rmdir dir1/dir2时目录不会被删除    执行强制删除命令 |

1. 在HDFS中，将文件从源路径移动到目的路径。

|  |
| --- |
| Shell命令：  hdfs dfs -mv text.txt text2.txt |

## 2.2 第二题

2.编程实现一个类“MyFSDataInputStream”，该类继承“org.apache.hadoop.fs.FSDataInputStream”，要求如下：实现按行读取HDFS中指定文件的方法“readLine()”，如果读到文件末尾，则返回空，否则返回文件一行的文本。

|  |
| --- |
| Maven的配置如下：    输出结果： |

## 2.3 第三题

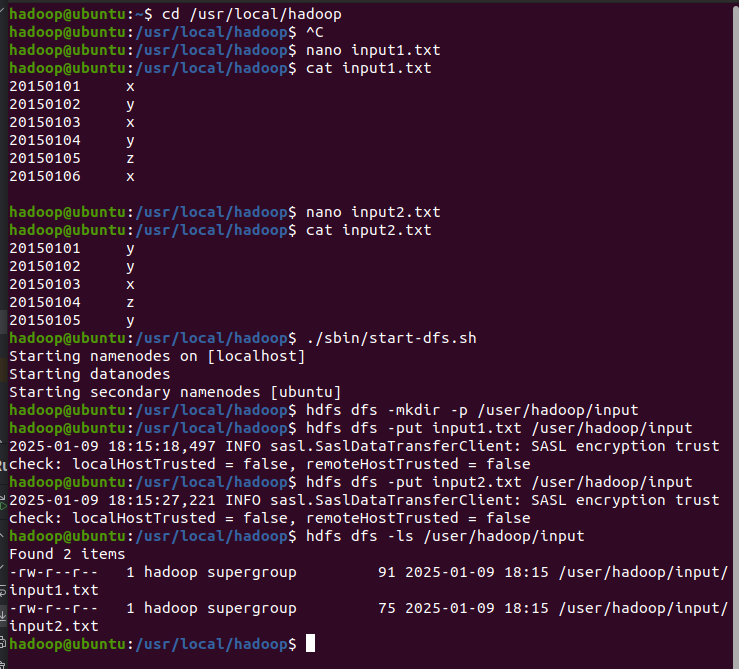
3.查看Java帮助手册或其它资料，用“java.net.URL”和“org.apache.hadoop.fs.FsURLStreamHandlerFactory”编程完成输出HDFS中指定文件的文本到终端中。

|  |
| --- |
| 输出内容： |

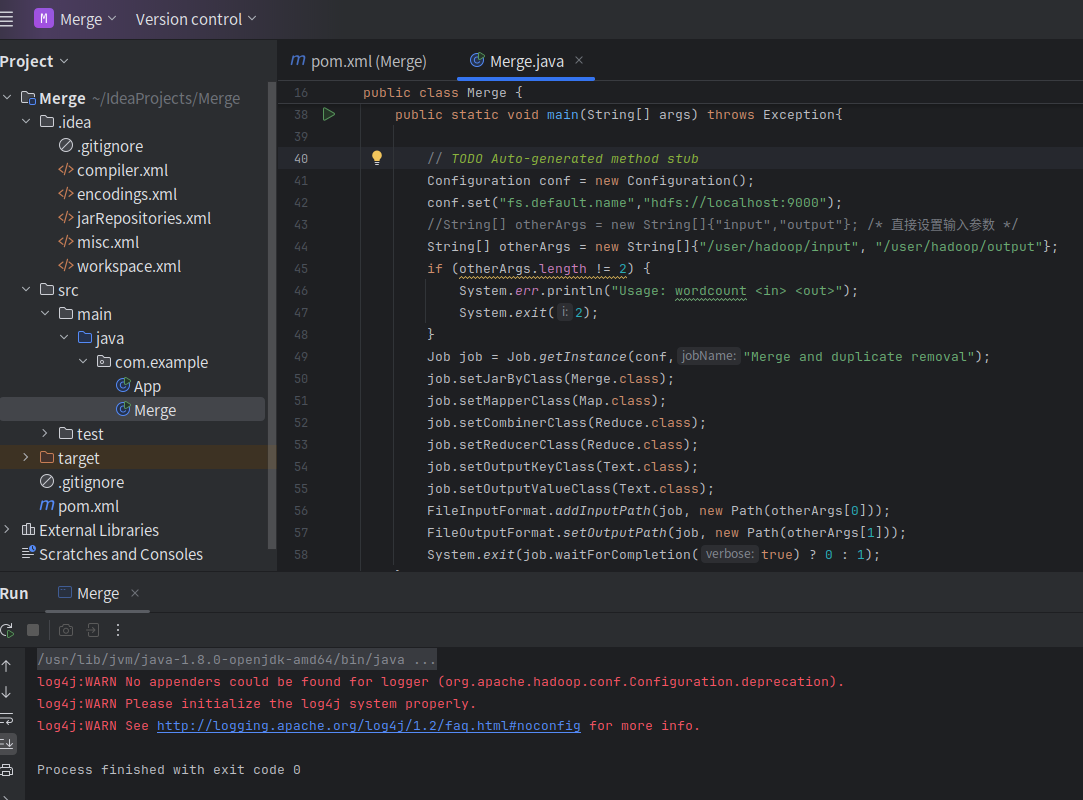
# 第三次上机作业

## 3.1 第一题

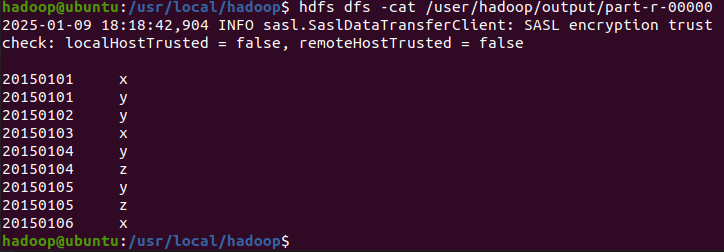
开启hdfs并创建input1.txt与input2.txt，同时将其上传至hdfs中。



修改代码中的文件路径，运行代码：

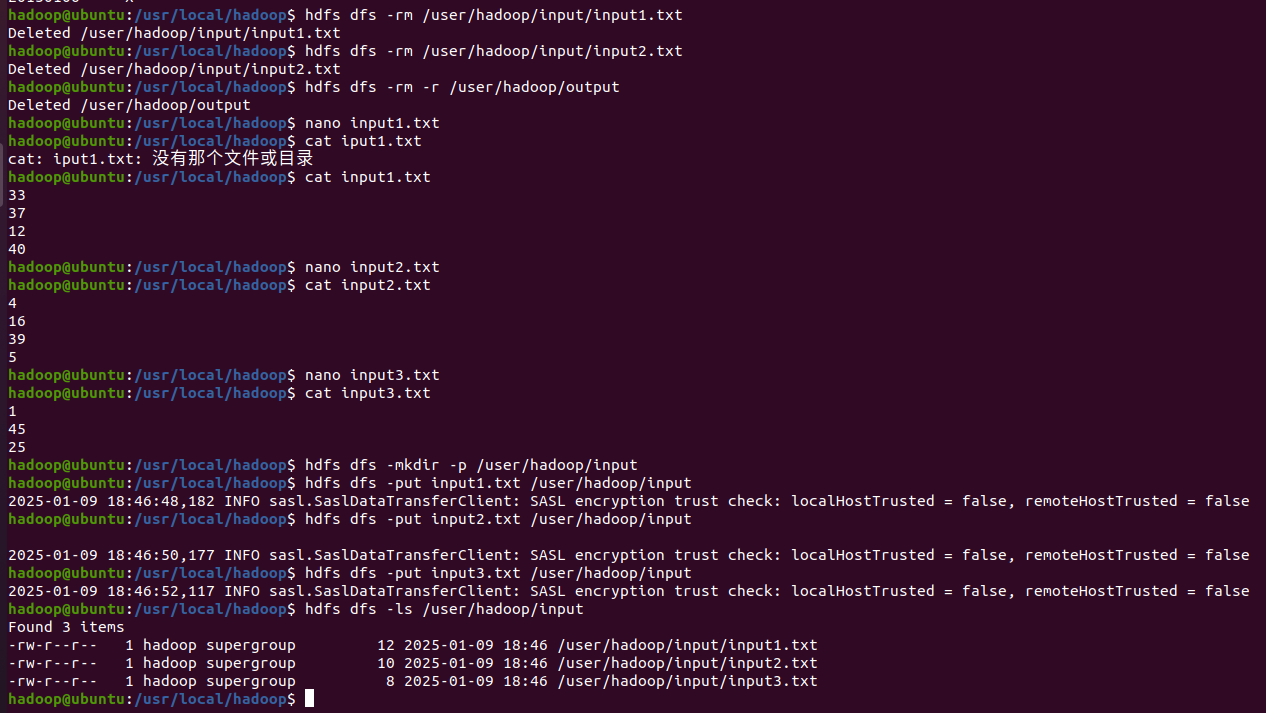


查看输出结果：

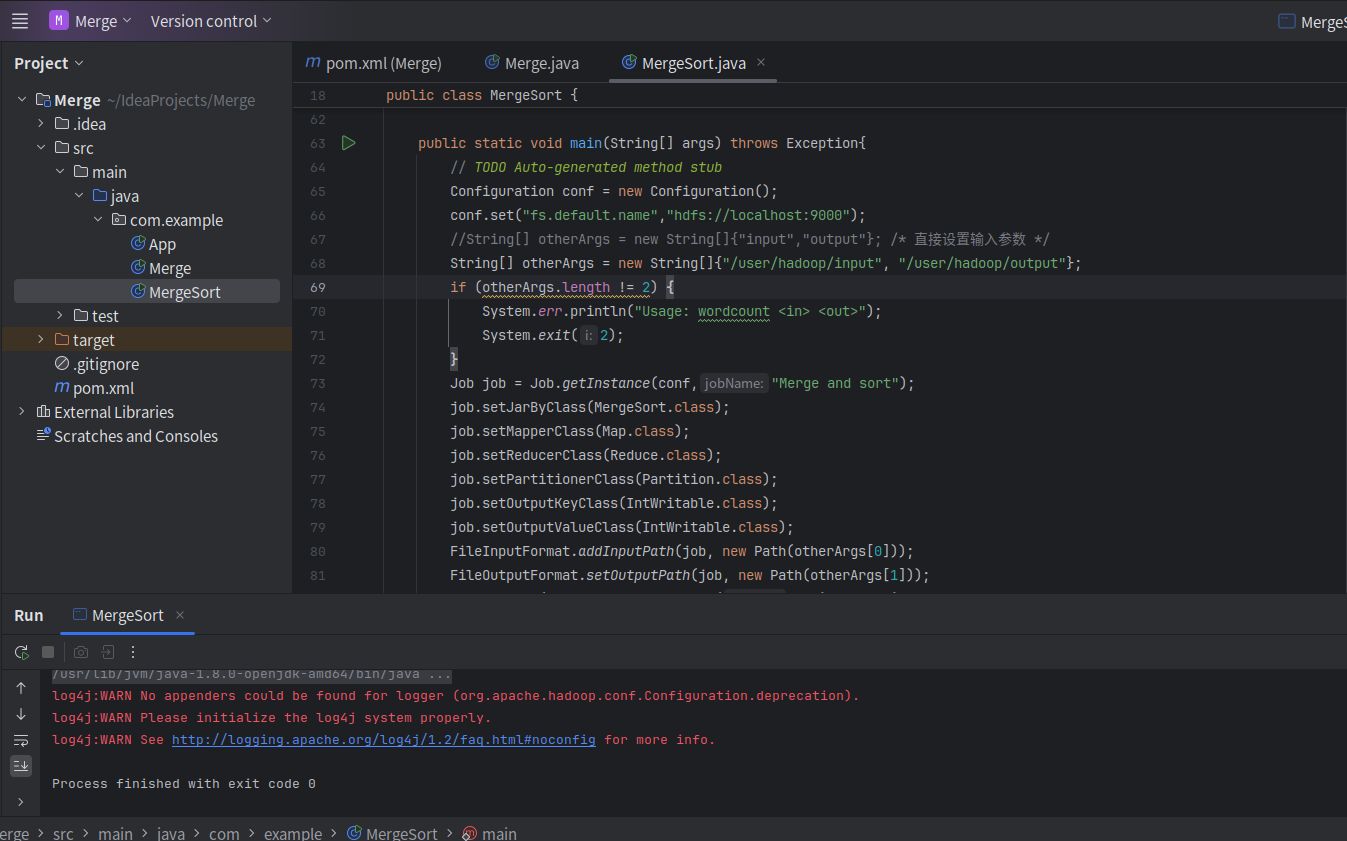


## 3.2 第二题

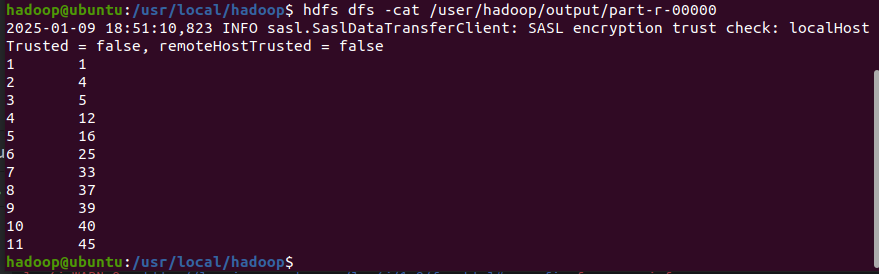
1. 首先准备数据，删除原来旧的文件夹、文件，上传input1.txt、input2.txt、input3.txt：



1. 编写代码并指定路径



1. 查看运行结果如下：

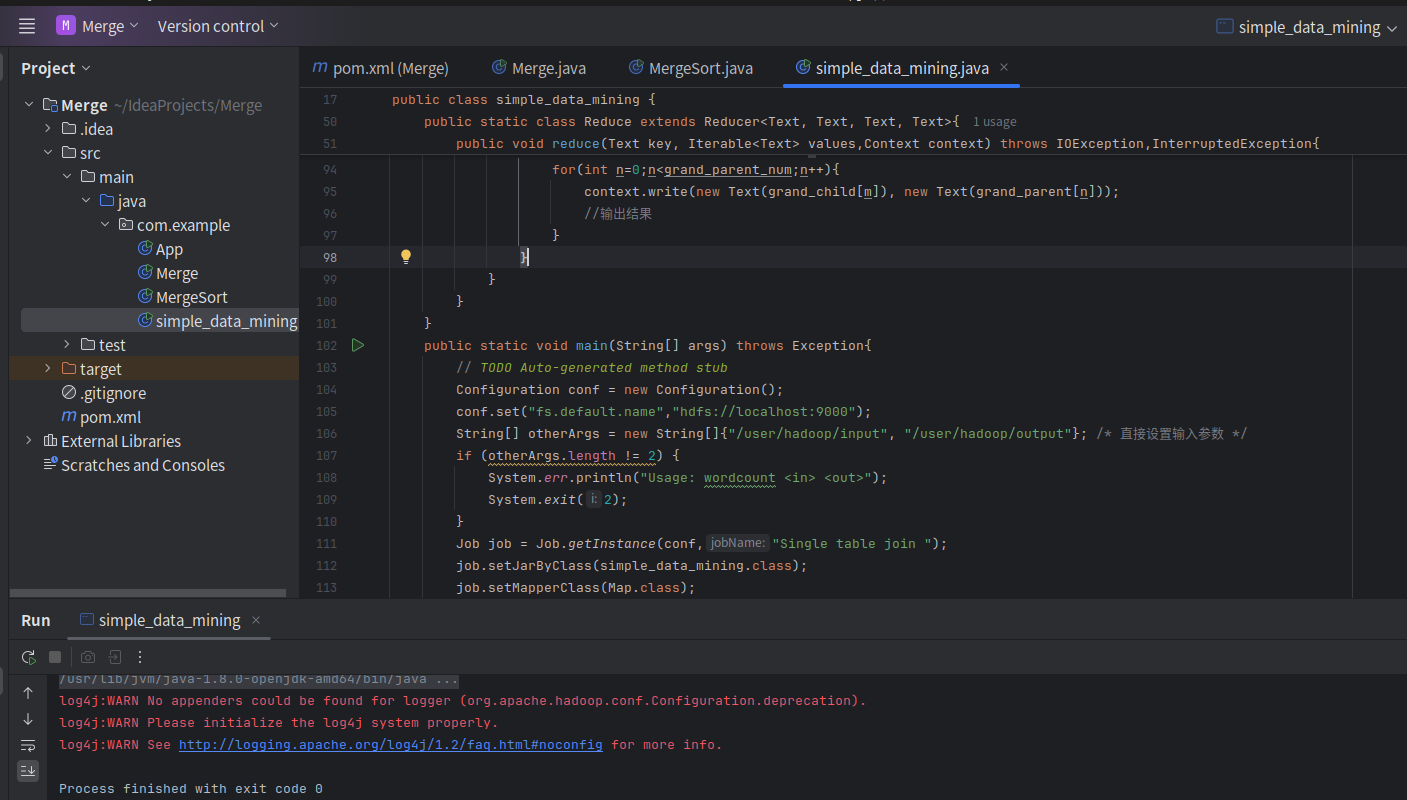


## 3.3 第三题

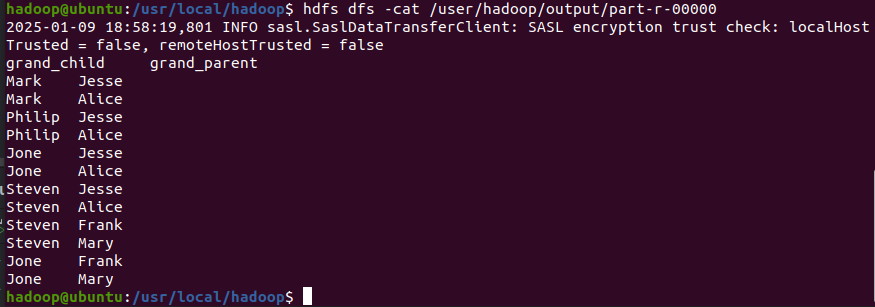
1. 首先将数据上传到hdfs中



1. 编写代码并指定文件路径，代码成功运行：



1. 查看输出结果：



# 第四次上机作业

主题层和裁剪层多边形如下，其中黑色的为主题层，橙色的为裁剪层，用不同颜色表示以示区别。

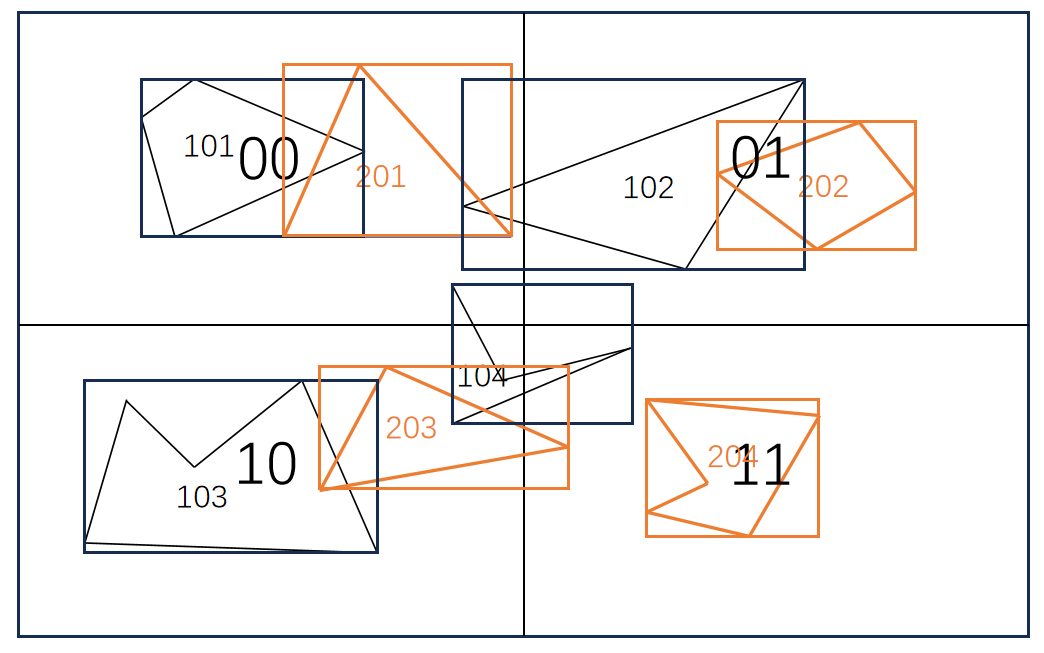


图 1 主题层和裁剪层多边形

多边形划分到网格以及多边形叠加操作的示意图如下：

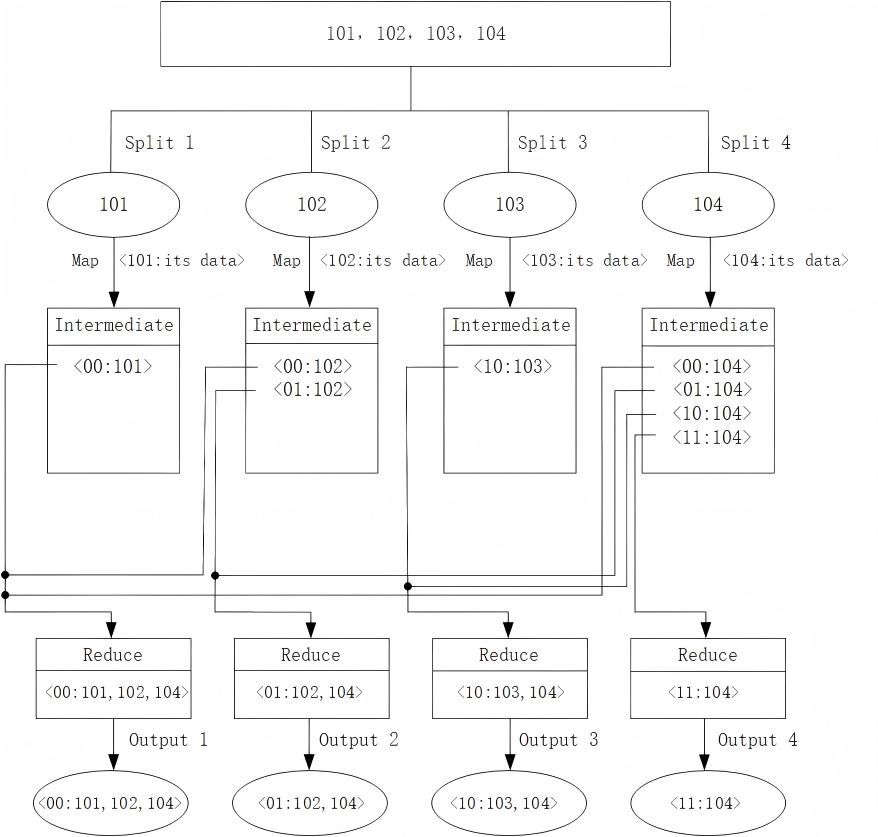


图 2 多边形划分到网格示意图

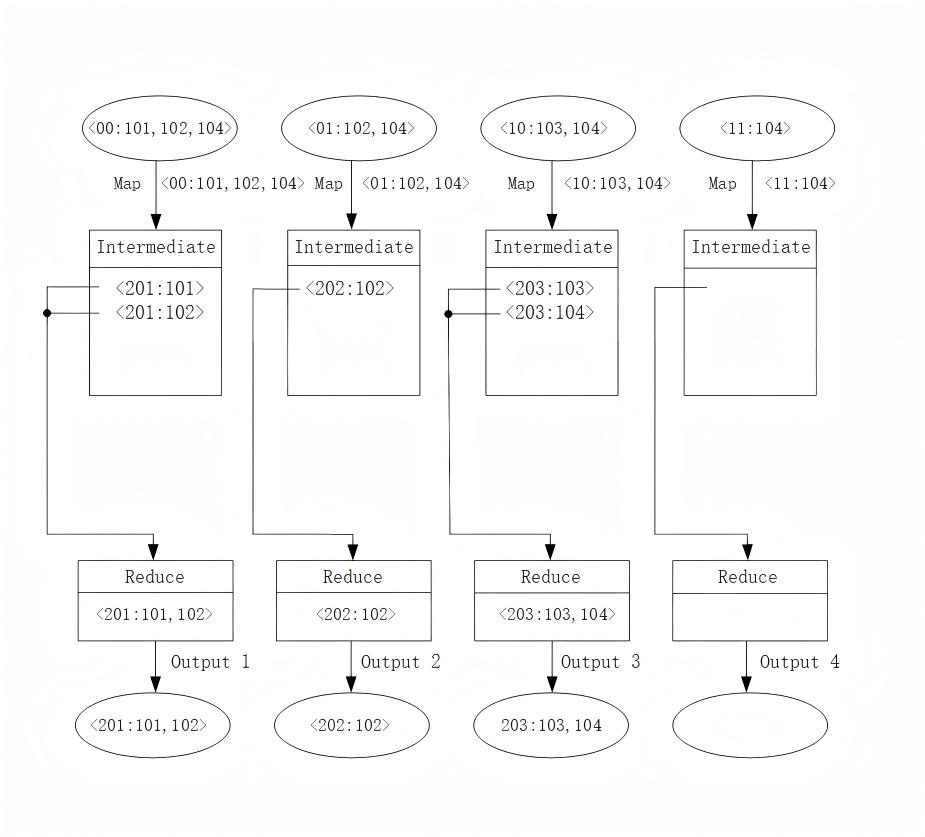


图 3 多边形叠加操作示意图