成都信息工程大学

大数据体系与技术

实验报告

实验题目： 熟悉常用的HDFS操作

系（院）： 软件工程学院

专业： 软件工程

班级： 203班

姓名： 胡昌豪

学号： 2020081114

2022年12月23日

一、实验目的

（1）理解HDFS在Hadoop体系结构中的角色；

（2）熟练使用HDFS操作常用的Shell命令；

（3）熟悉HDFS操作常用的Java API。

二、实验内容

（1）操作系统：Linux（建议Ubuntu16.04或Ubuntu18.04）；

（2）Hadoop版本：3.1.3；

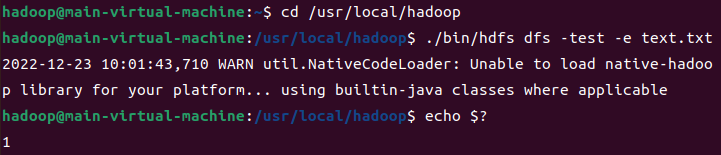
（3）JDK版本：1.8；

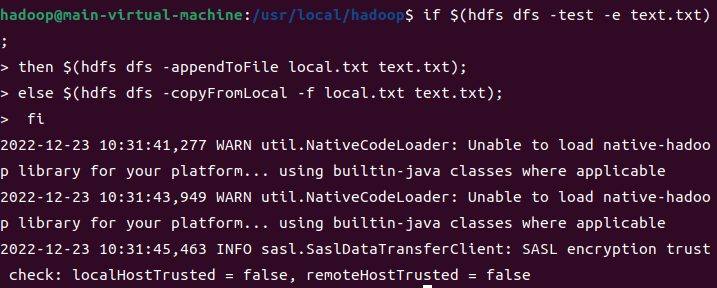
（4）Java IDE：Eclipse。

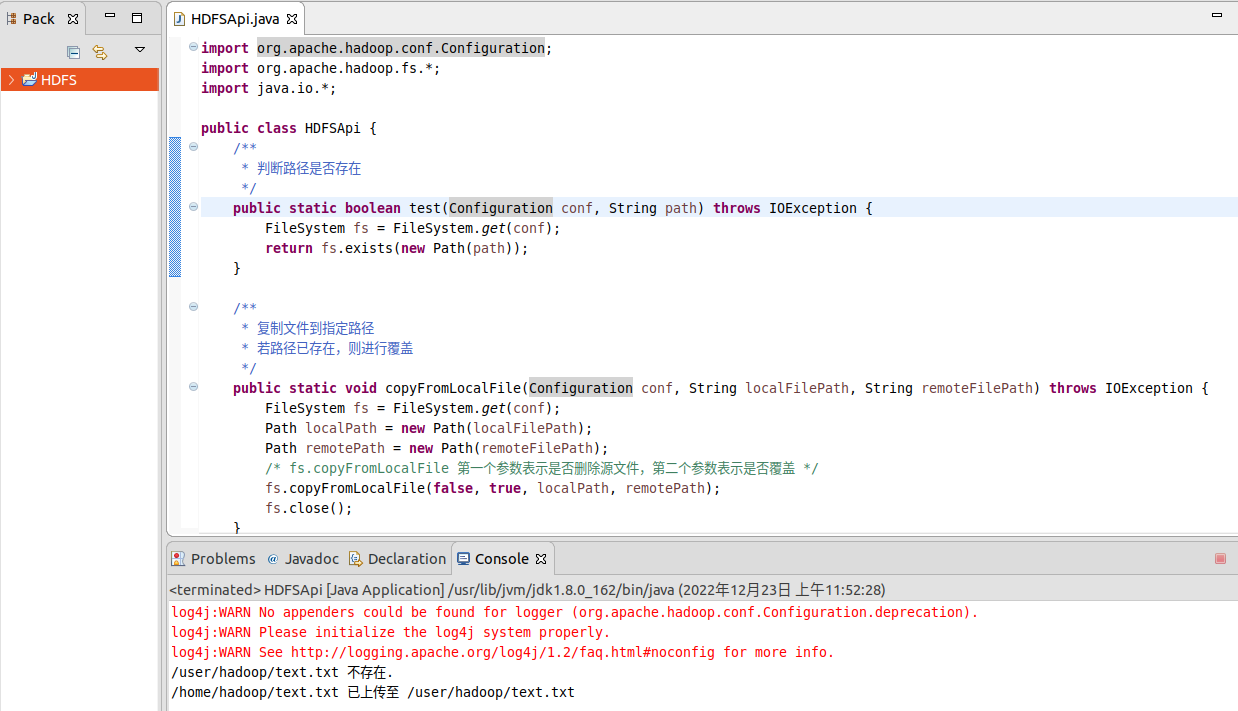
1. 实验步骤

（一）编程实现以下功能，并利用Hadoop提供的Shell命令完成相同任务：

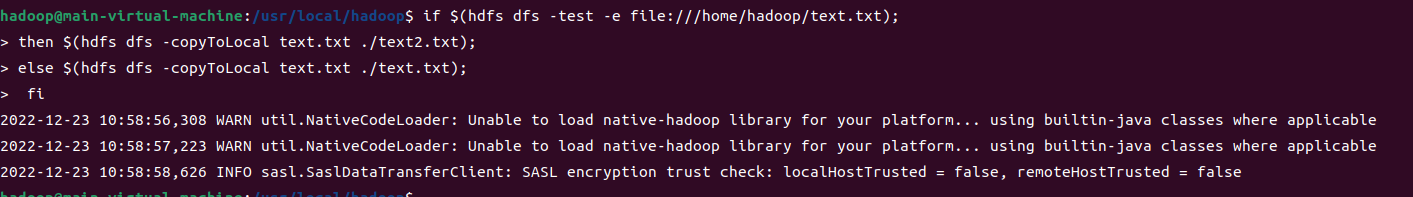
1. 向HDFS中上传任意文本文件，如果指定的文件在HDFS中已经存在，则由用户来指定是追加到原有文件末尾还是覆盖原有的文件；

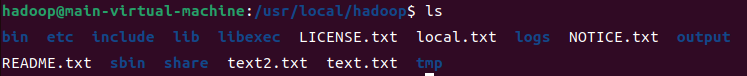


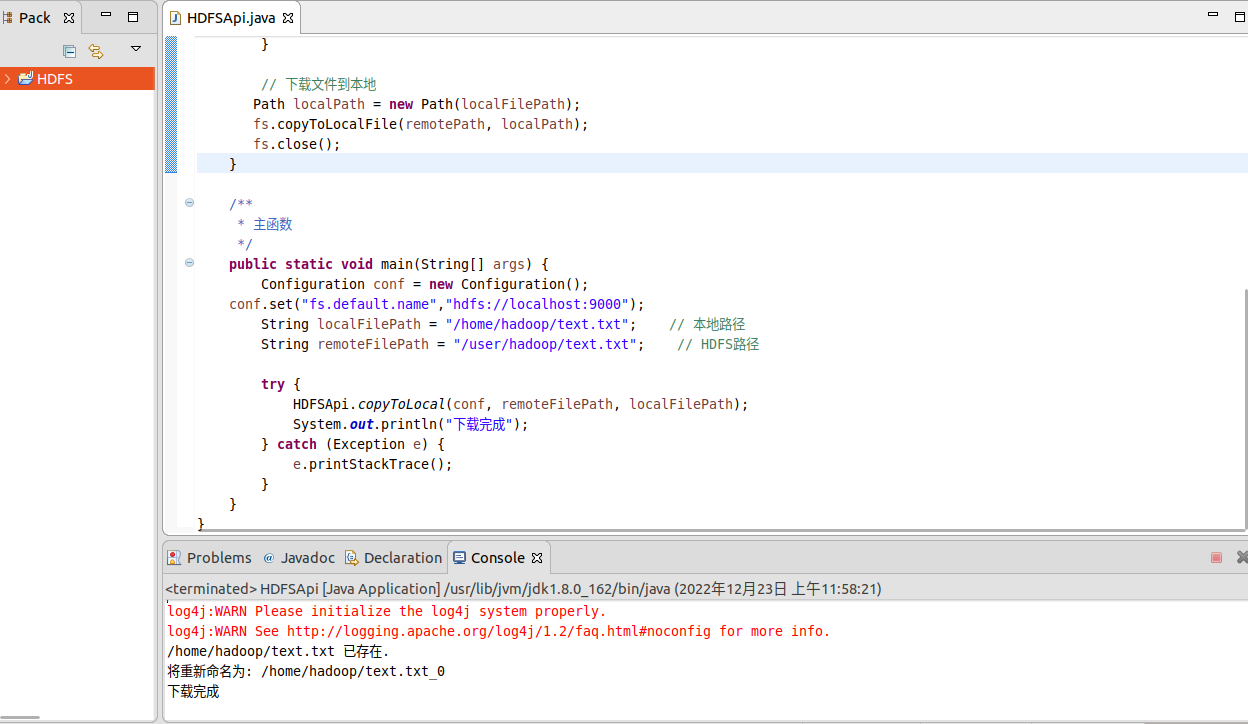




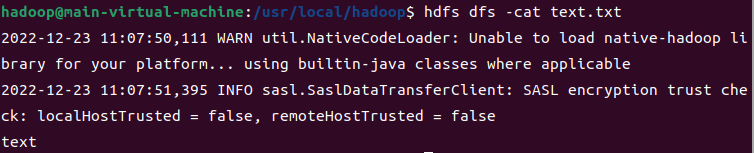
1. 从HDFS中下载指定文件，如果本地文件与要下载的文件名称相同，则自动对下载的文件重命名；

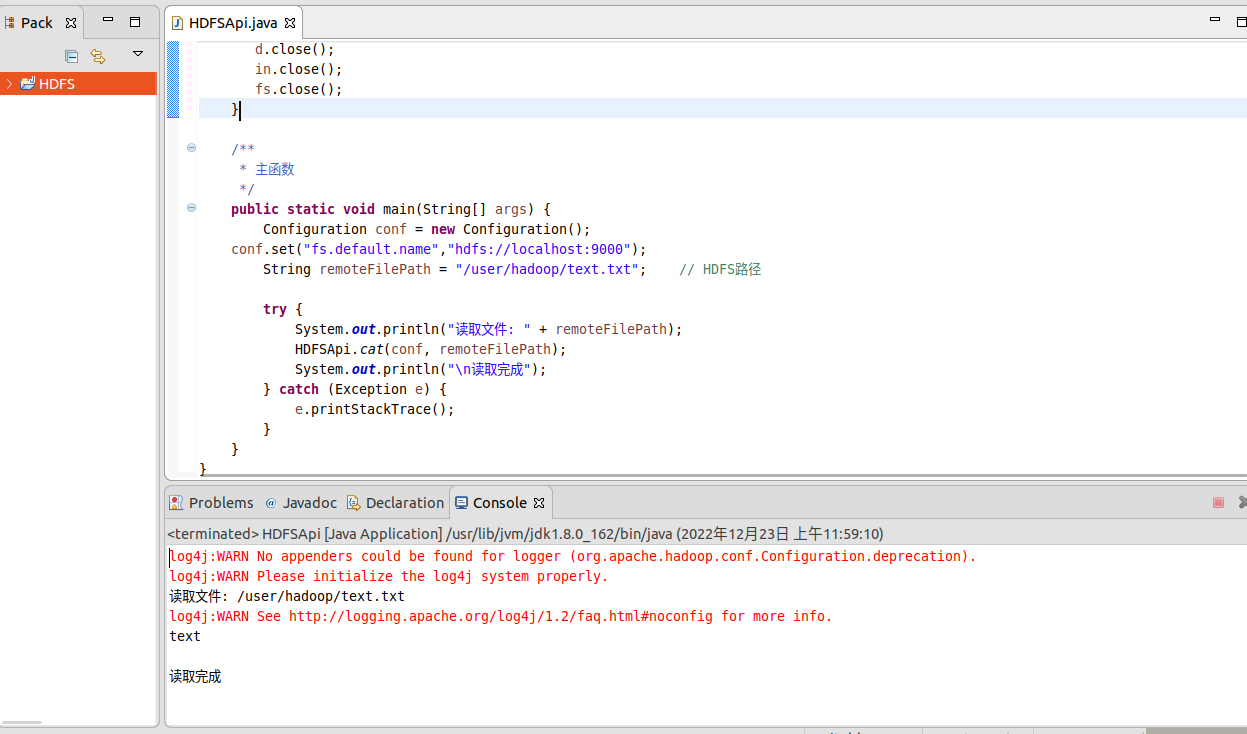




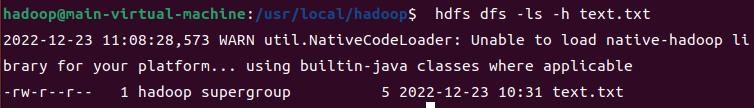


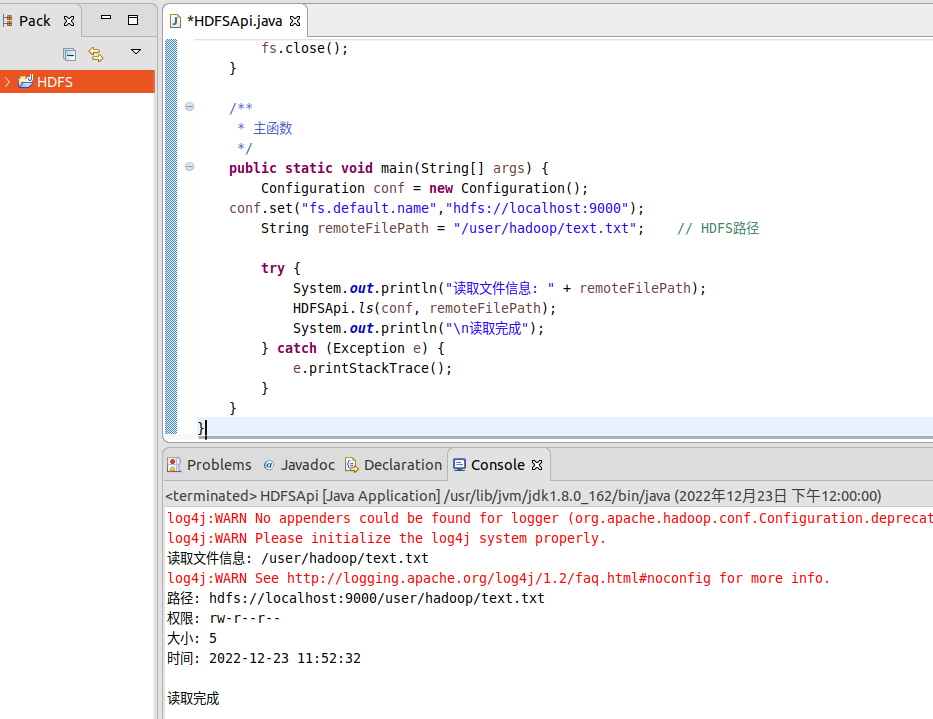
1. 将HDFS中指定文件的内容输出到终端中；



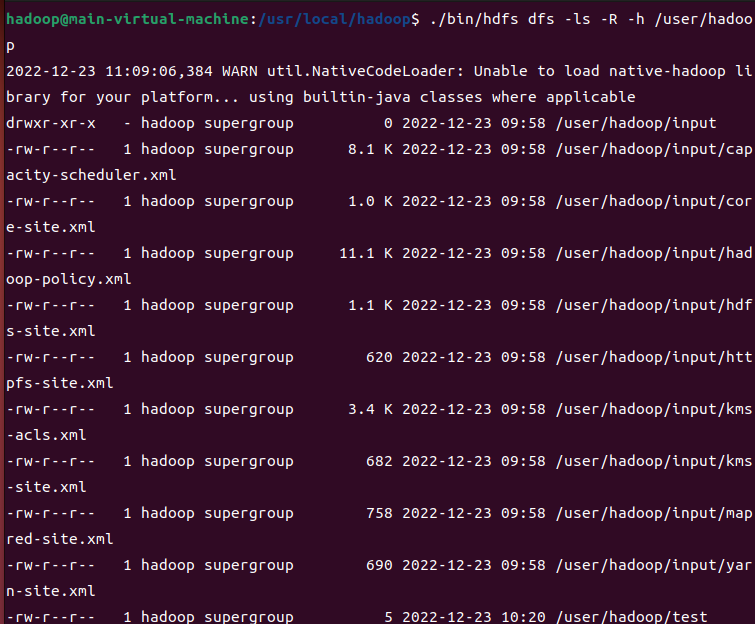


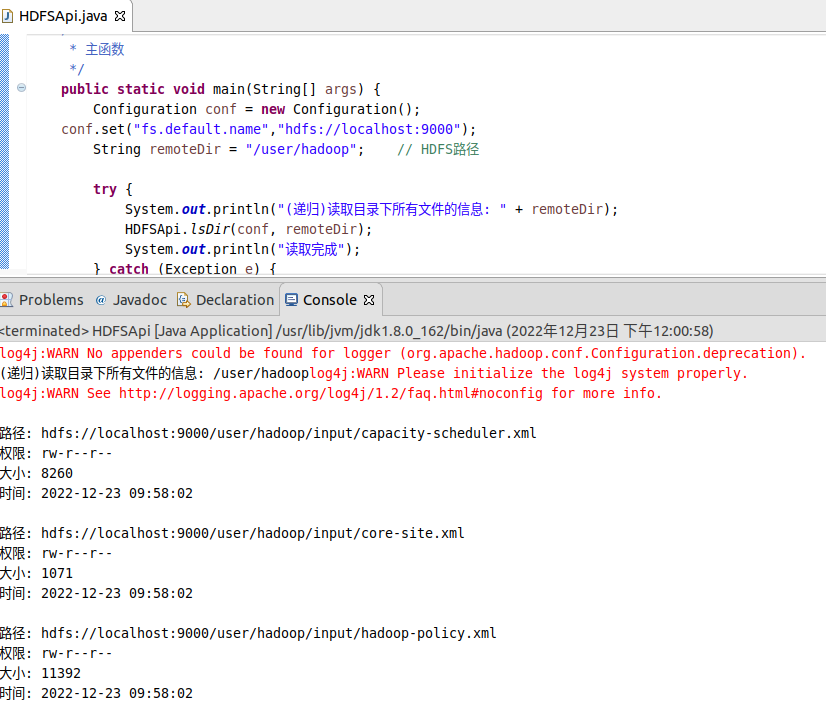
1. 显示HDFS中指定的文件的读写权限、大小、创建时间、路径等信息；

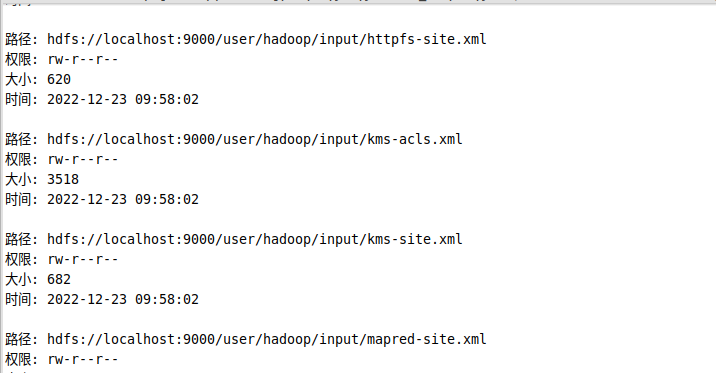




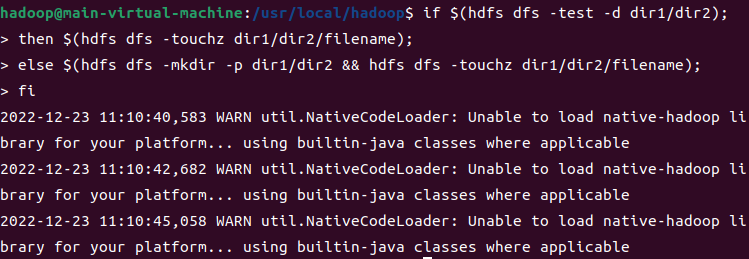
1. 给定HDFS中某一个目录，输出该目录下的所有文件的读写权限、大小、创建时间、路径等信息，如果该文件是目录，则递归输出该目录下所有文件相关信息；

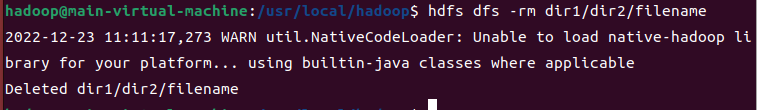


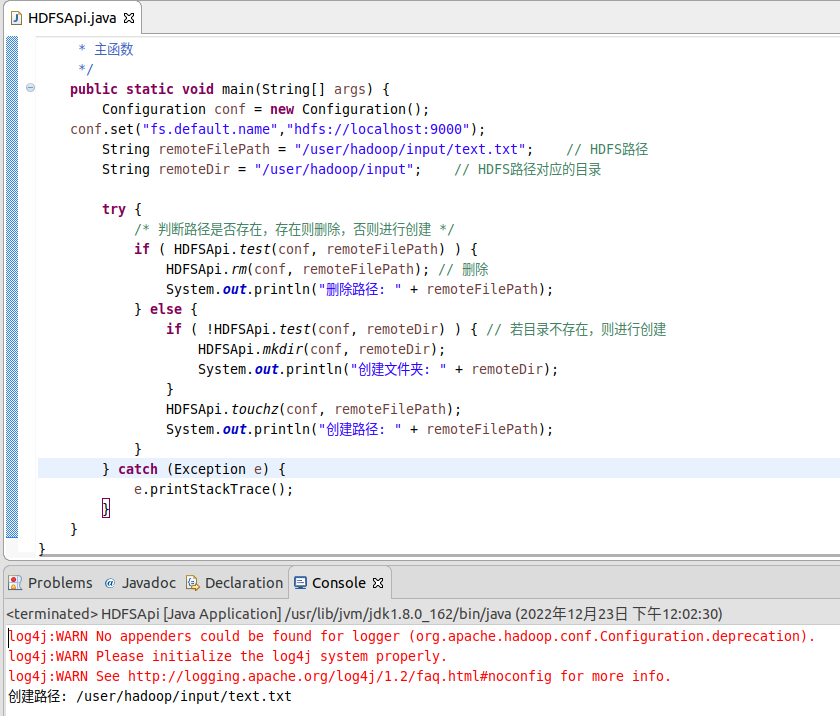


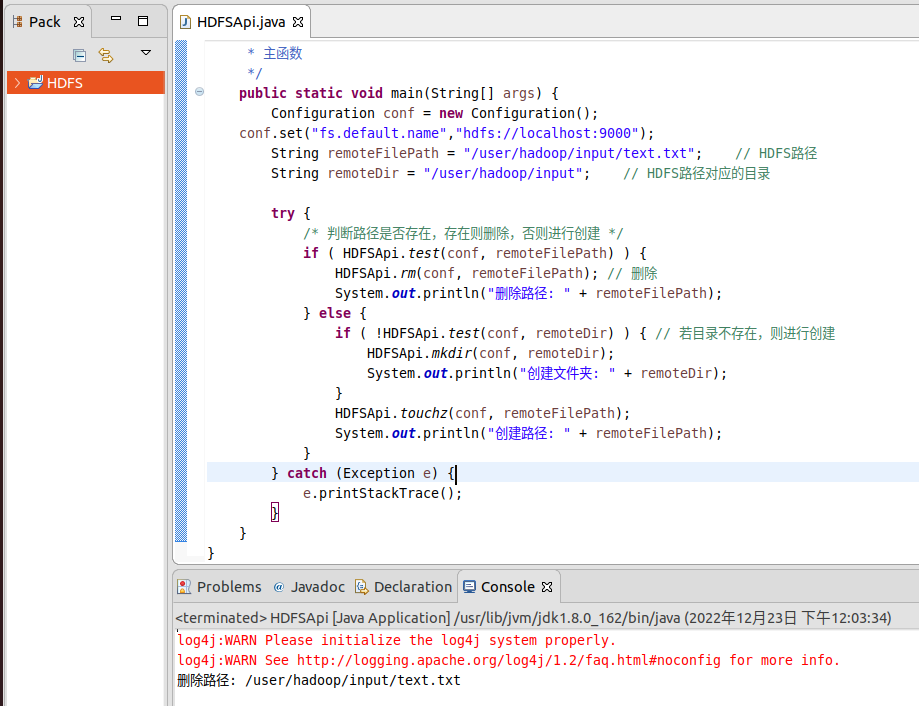


1. 提供一个HDFS内的文件的路径，对该文件进行创建和删除操作。如果文件所在目录不存在，则自动创建目录；

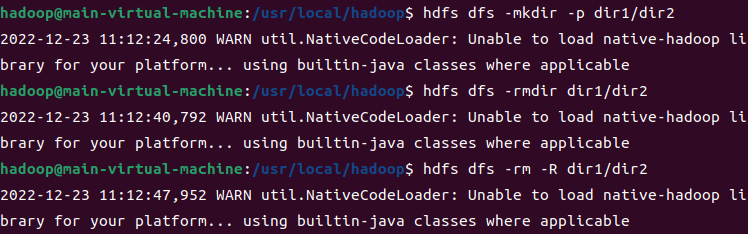






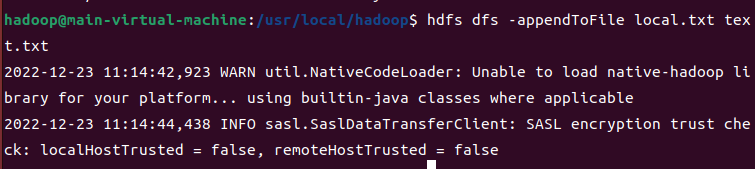


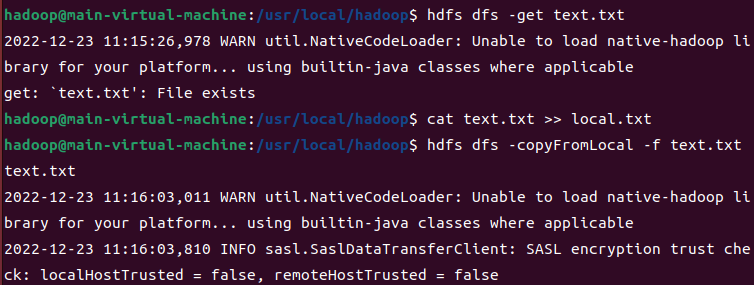
1. 提供一个HDFS的目录的路径，对该目录进行创建和删除操作。创建目录时，如果目录文件所在目录不存在，则自动创建相应目录；删除目录时，由用户指定当该目录不为空时是否还删除该目录；

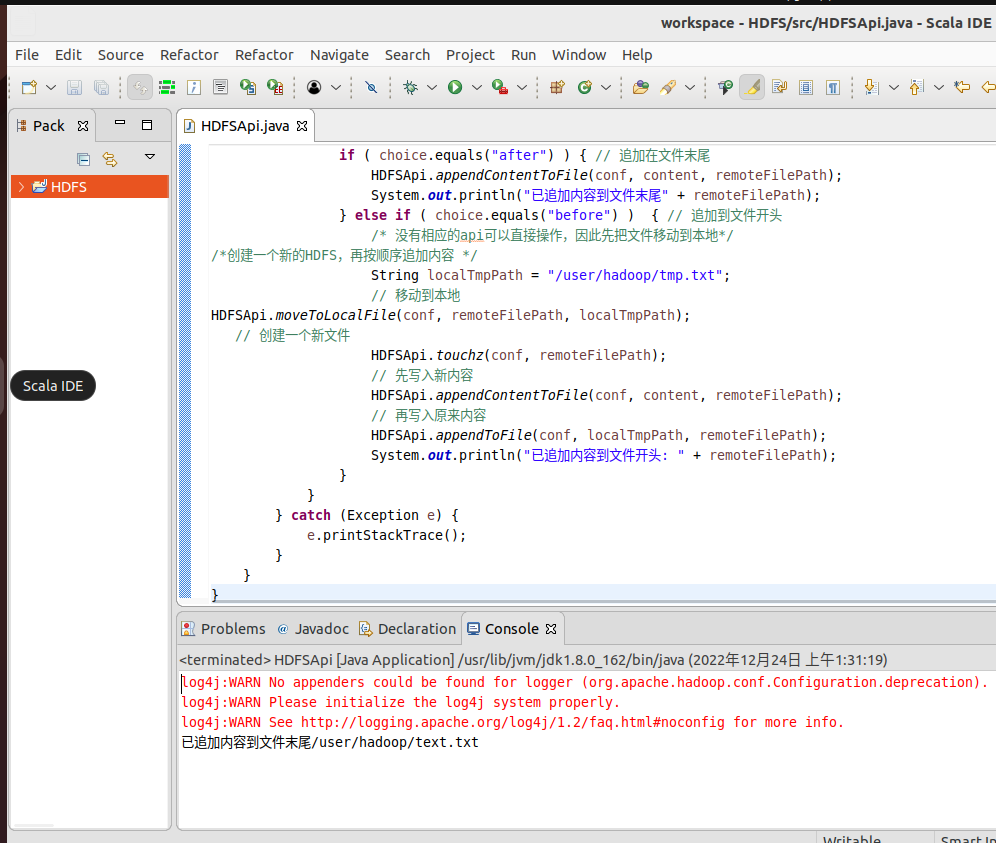




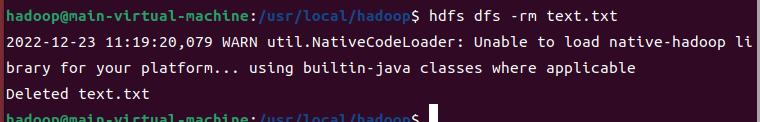
1. 向HDFS中指定的文件追加内容，由用户指定内容追加到原有文件的开头或结尾；





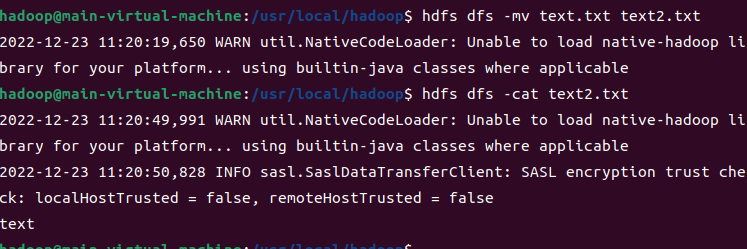


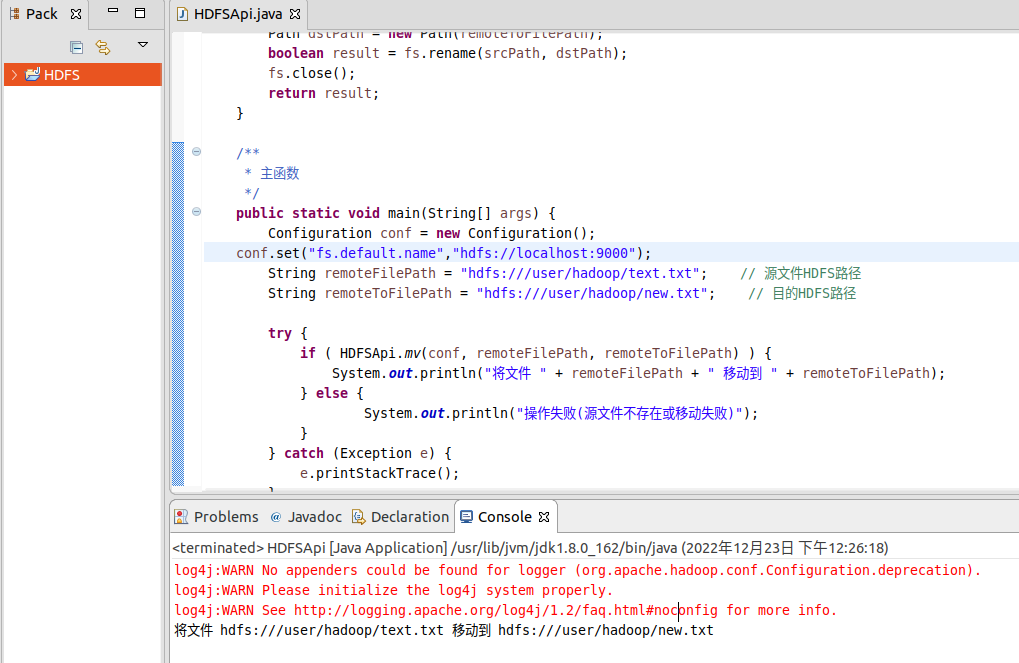
1. 删除HDFS中指定的文件；





1. 在HDFS中，将文件从源路径移动到目的路径。



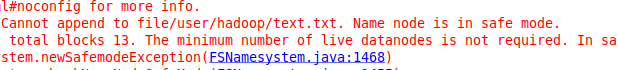


四、实验总结

1. 心得体会

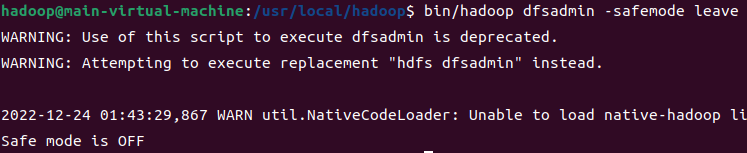
这次实验主要是使用hadoop的shell命令以及java程序对文件进行存储等操作，在本次实验中，我遇到了很多困难，比如说追加数据时关闭safe mode等以及写入数据时权限不足等，这些问题无不困扰我许久，但是在经过这次实验以后，我linux和hadoop的编程能力也得到了大幅提高，总的来说这次实验收获还是比较大的。

1. 遇到的问题及解决办法
2. 在使用java程序向文件中追加指定内容时出现报错 Name node is in safe mode



解决办法:在经过网上查阅资料以后得知，需要关闭hadoop的安全模式，

在使用bin/hadoop dfsadmin -safemode leave以后就可以关闭hadoop的安全模式。



1. 在使用java程序向文件中最佳指定内容时报错 Failed to replace a bad datanode on the existing pipeline due to no more good datanodes being available to try

解决办法:在经过网上查阅资料以后得知，需要在 $HADOOP\_HOME/etc/hadoop 中的hdfs-site.xml追加这一段内容，原因是:默认情况下副本数为3个，在执行写入到HDFS的操作时，当我的一台Datenode写入失败时，它要保持副本数为3，它就会寻找一个可用的DateNode节点来写入.

