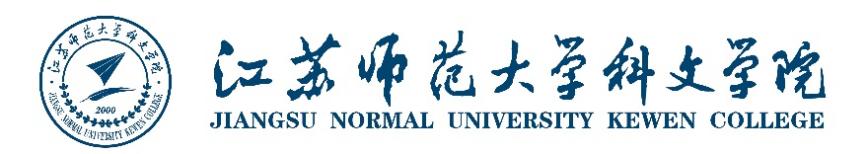
****

**人工智能与软件学院**

实 验 报 告

**课 程 大数据分析与智能决策**

**学 号 7200767121**

**姓 名 彭敏捷**

**班 级 20智科**

**教 师 胡元发**

2023年 4 月至 2023 年 4 月

**一、实验名称**

熟悉常用的HDFS操作

**二、实验目的和要求**

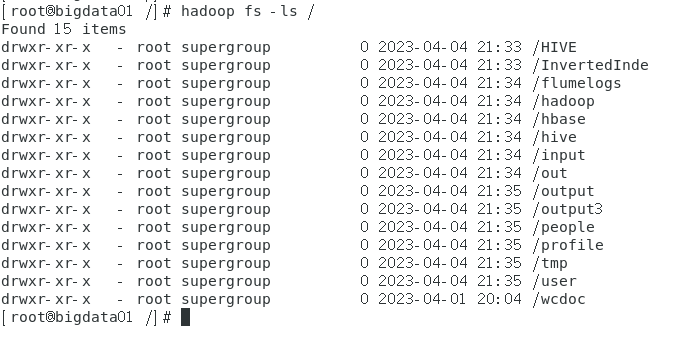
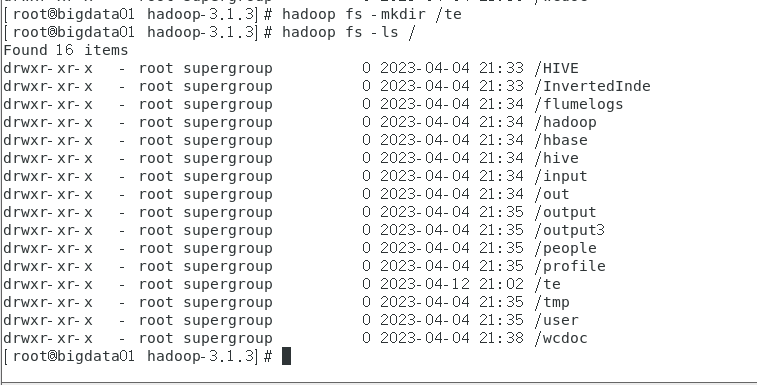
1、利用Hadoop提供的Shell命令完成文件下载、文件输出、追加文件内容等一些列指定任务，掌握分布式文件系统HDFS的基本存取方法；

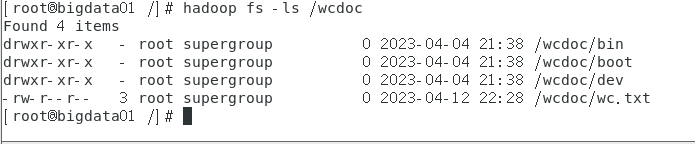
2、利用HDFS提供Java API编程实现文件的读取与管理

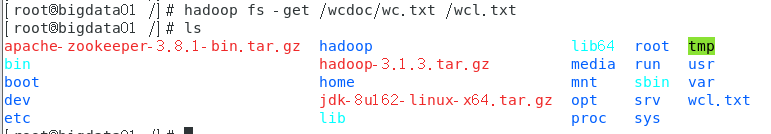
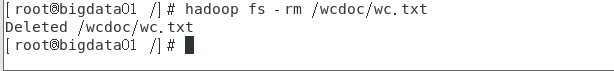
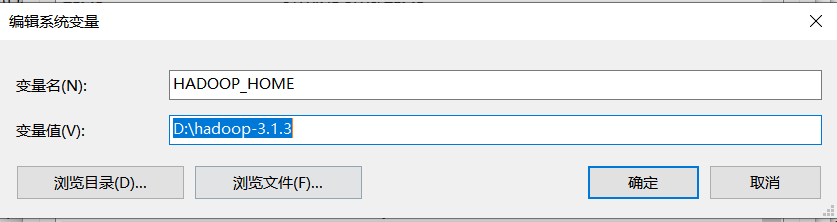
**三、实验环境**

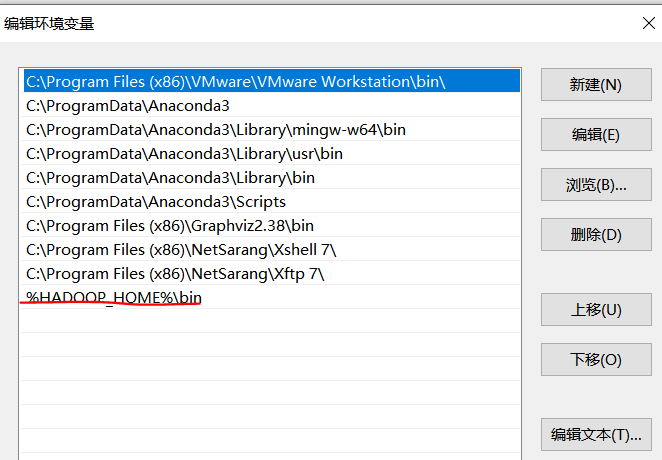
1.VMware Workstation Pro

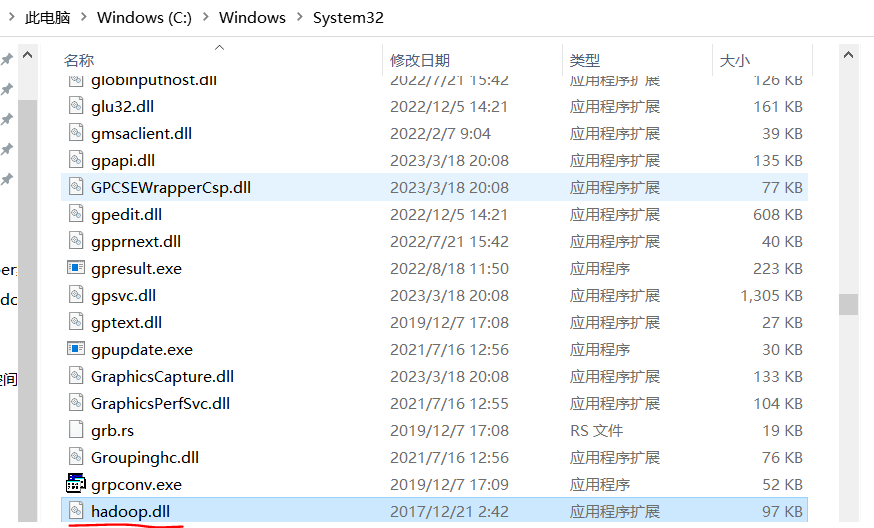
2.Xshell

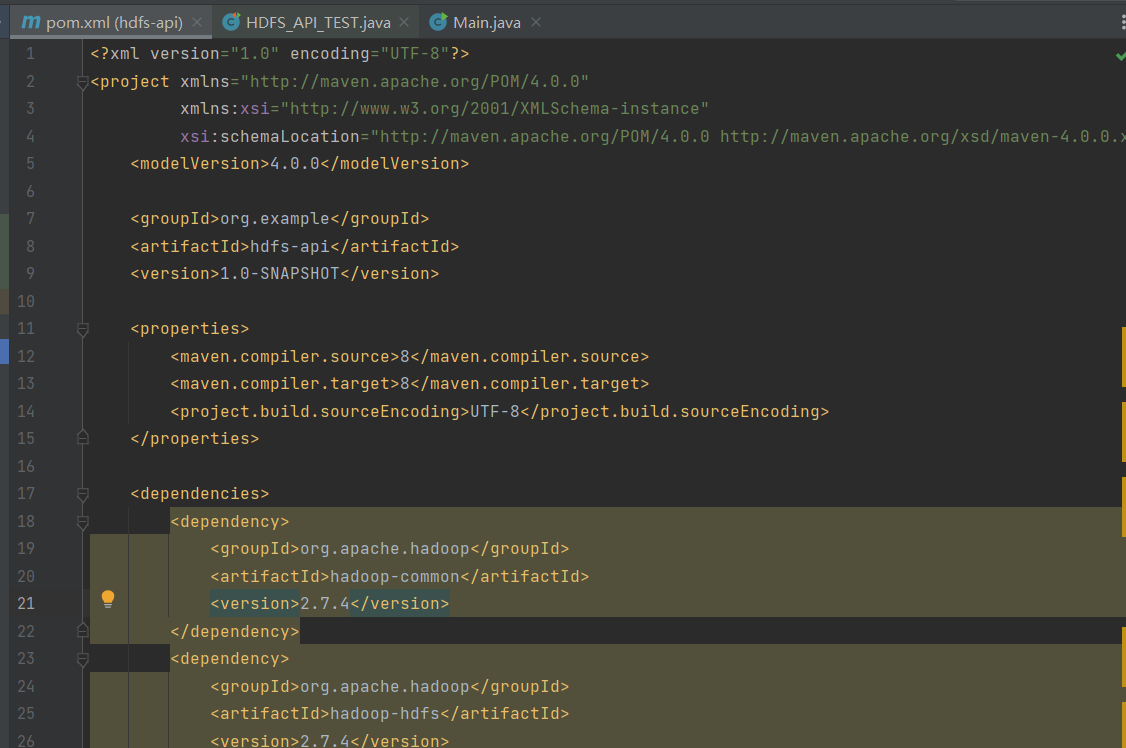
1. **实验步骤**
2. HDFS Shell常用命令
3. 列出文件或文件夹
4. 创建文件夹
5. 上传文件

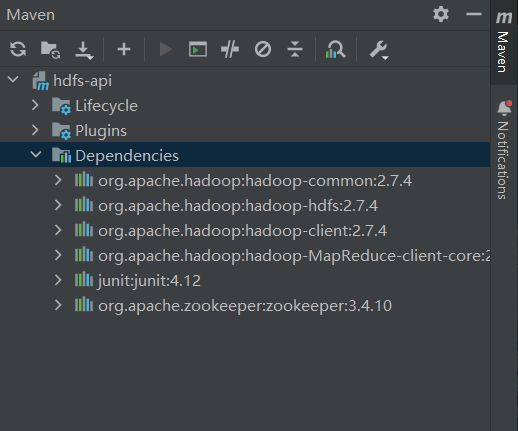


1. 下载文件
2. 查看文件内容5
3. 删除文件夹
4. HDFS的Java API
5. 配置Windows的hadoop环境变量

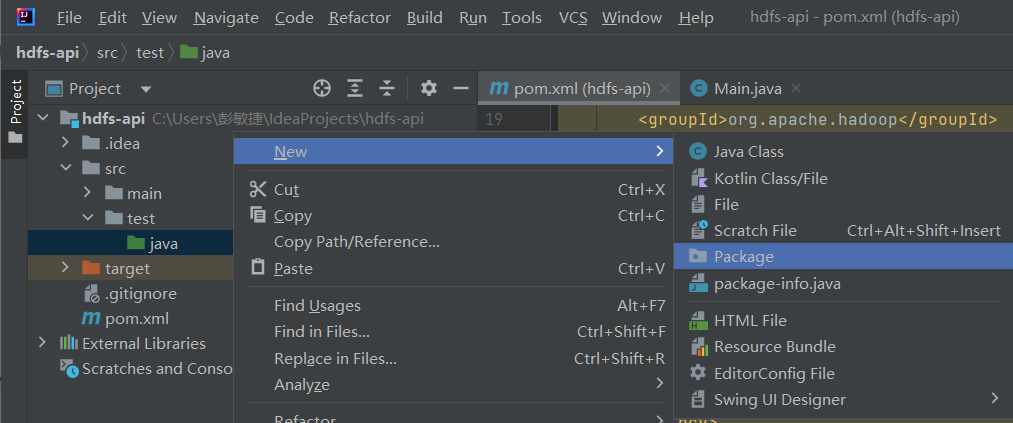


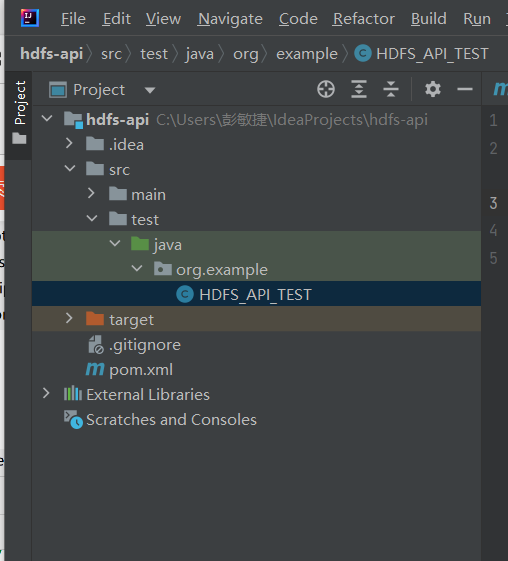
1. 在C:\Windows\System32添加hadoop.dll
2. 创建一个MAVEN项目，在pom.xml配置文件中引入hadoop-common、hadoop-hdfs、haddop-client3中依赖，同时引入junit的[单元测试](https://so.csdn.net/so/search?q=%E5%8D%95%E5%85%83%E6%B5%8B%E8%AF%95&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/weixin_48437766/article/details/_blank)包

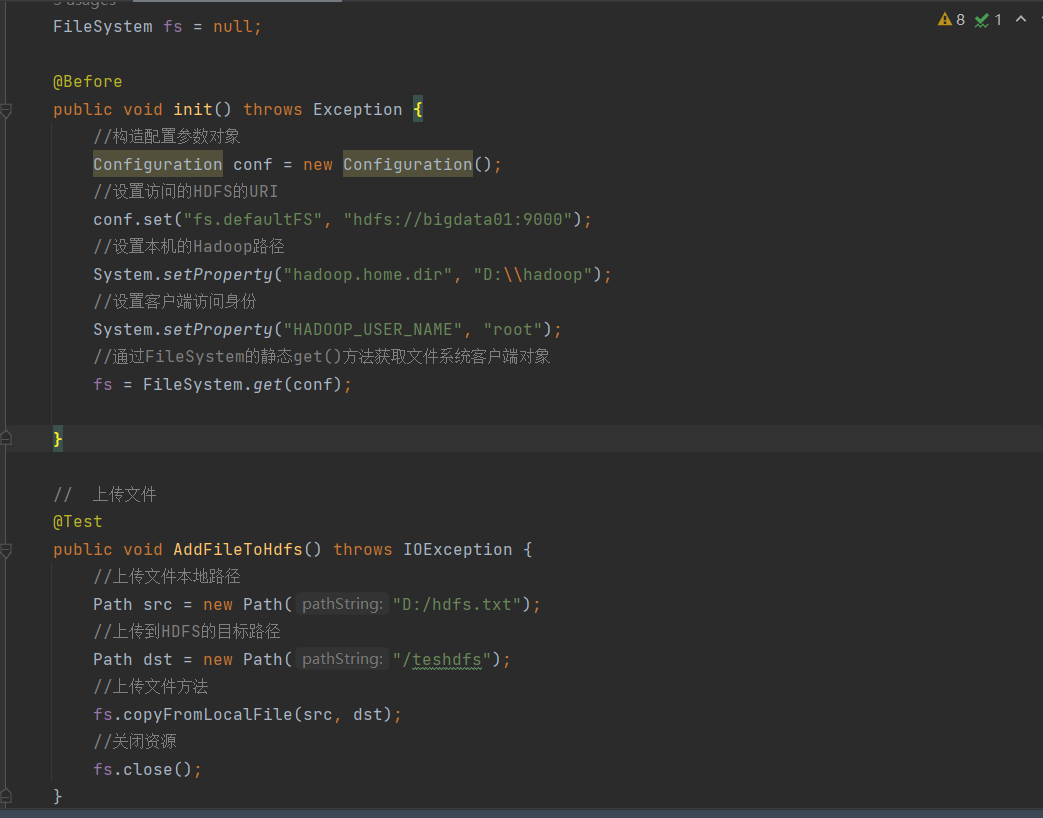


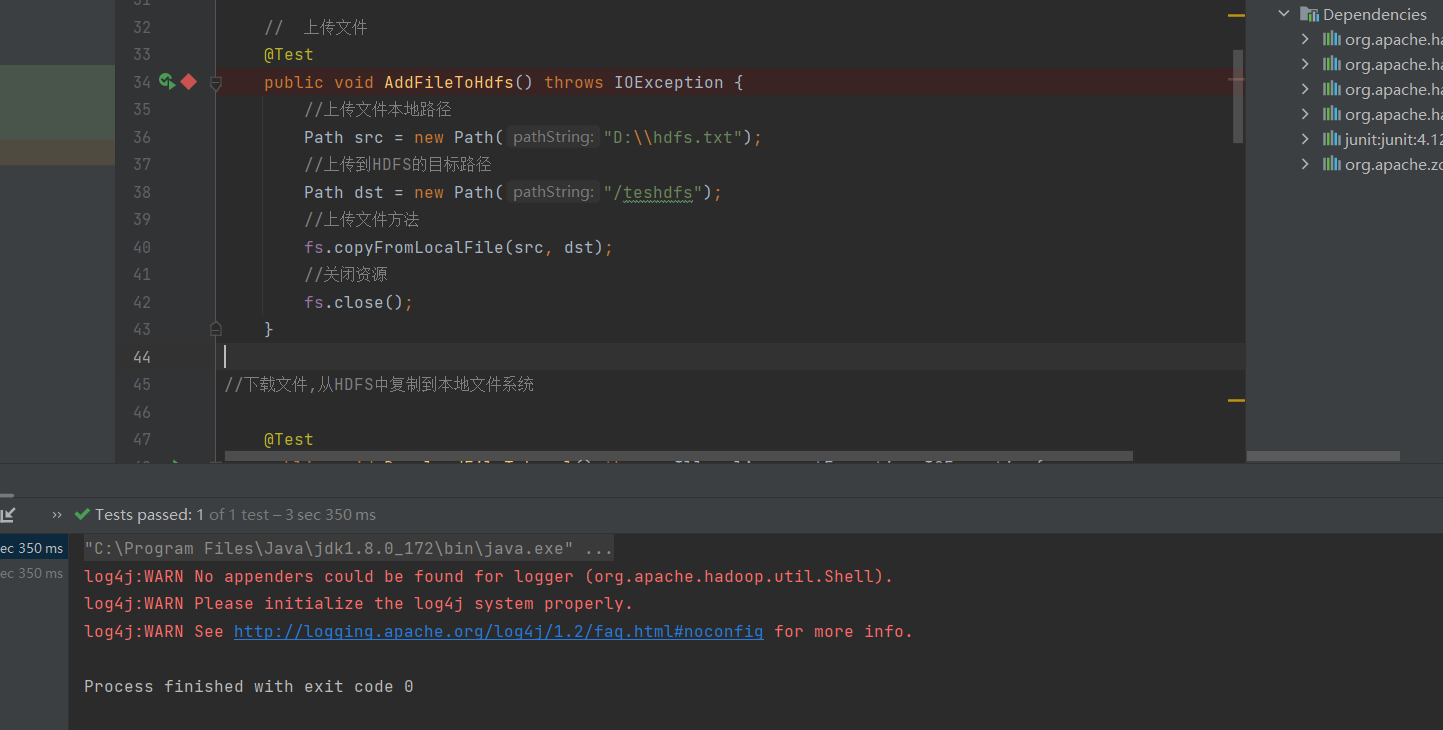


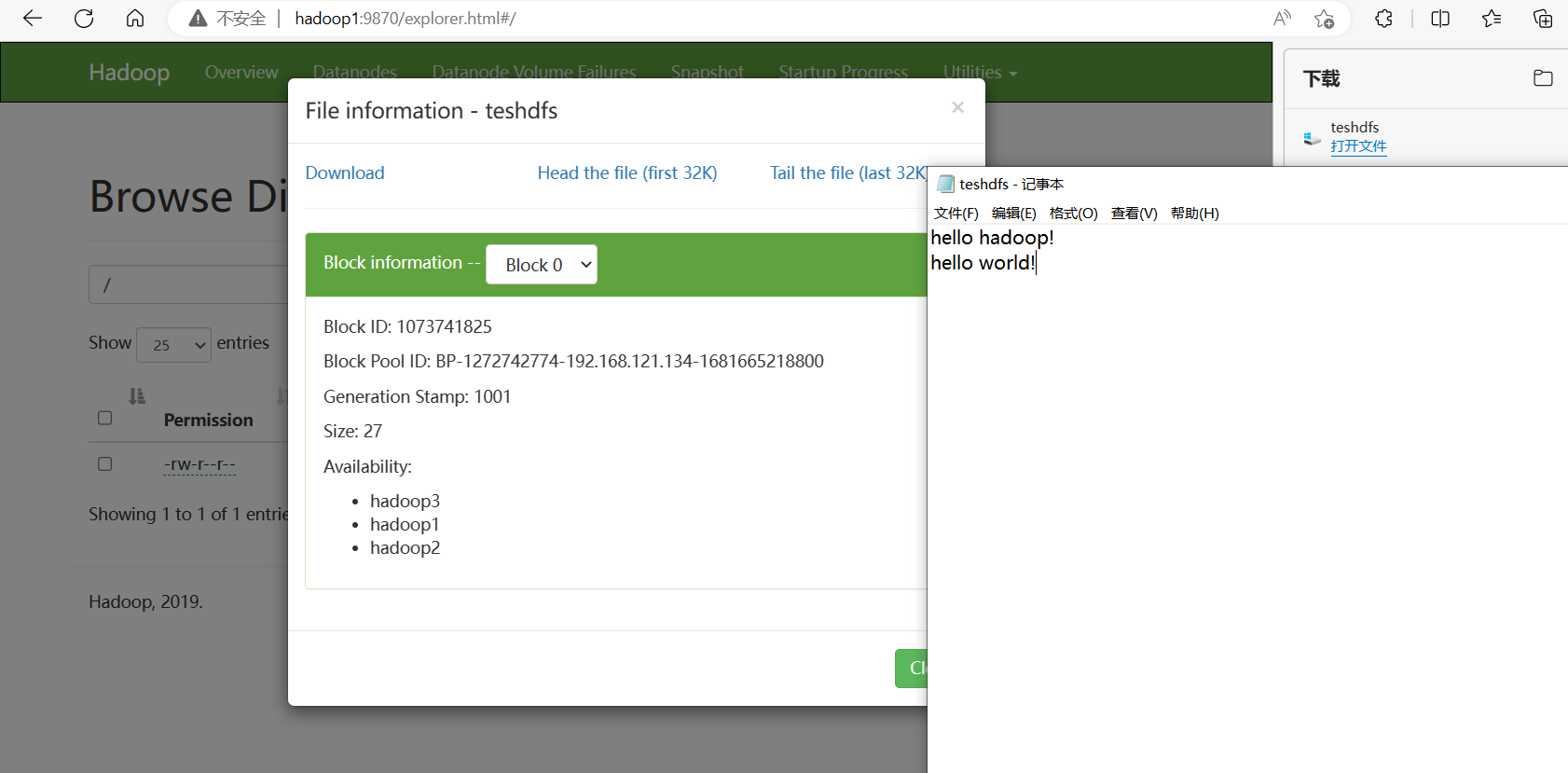
（4）在src\test\java目录下新建package，包名为org.example，新建HDFS\_API\_TEST类

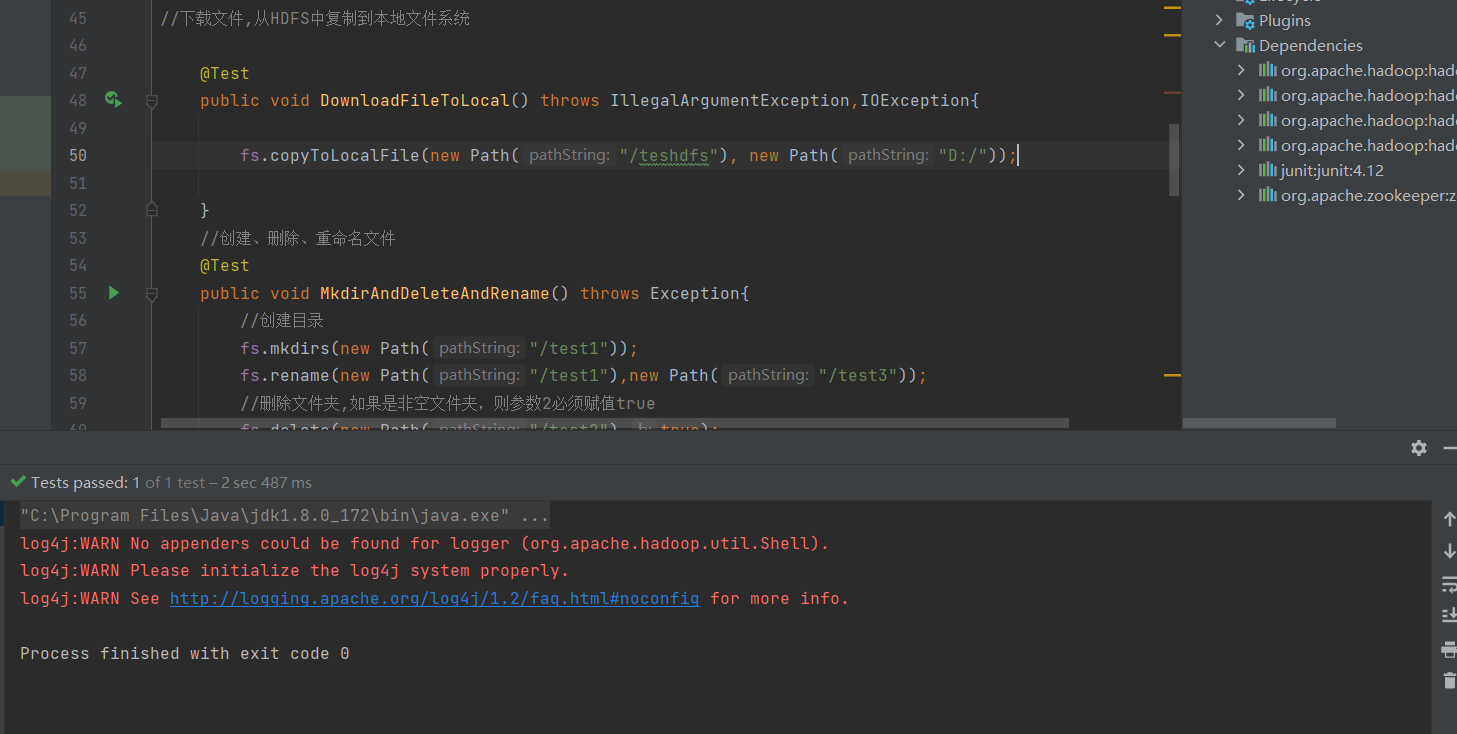


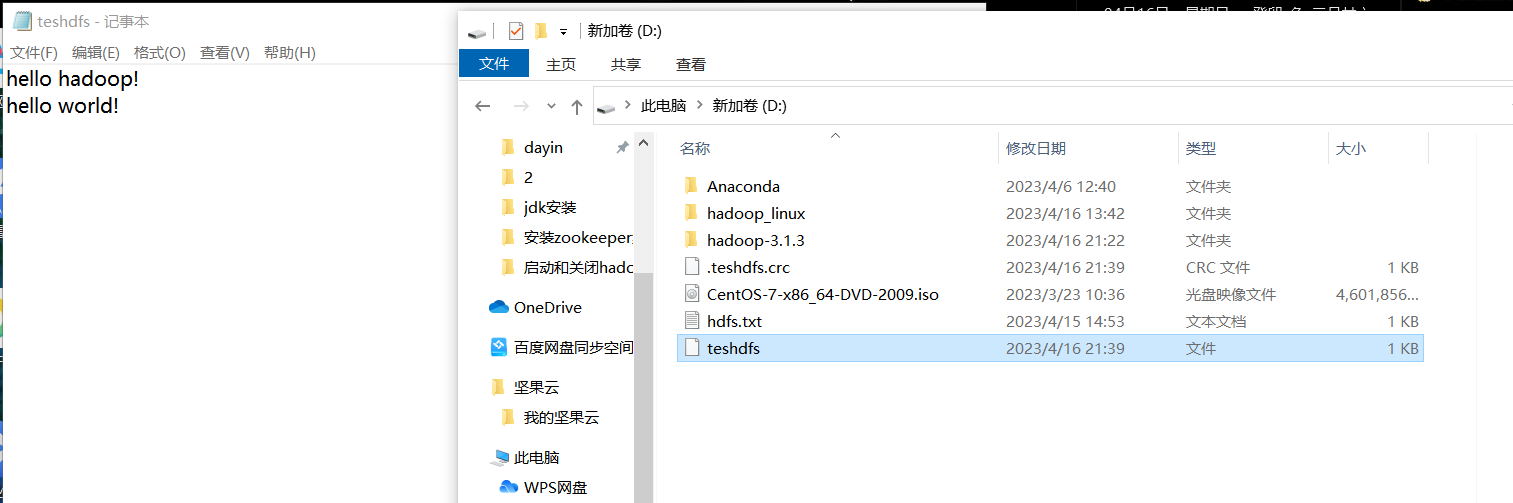


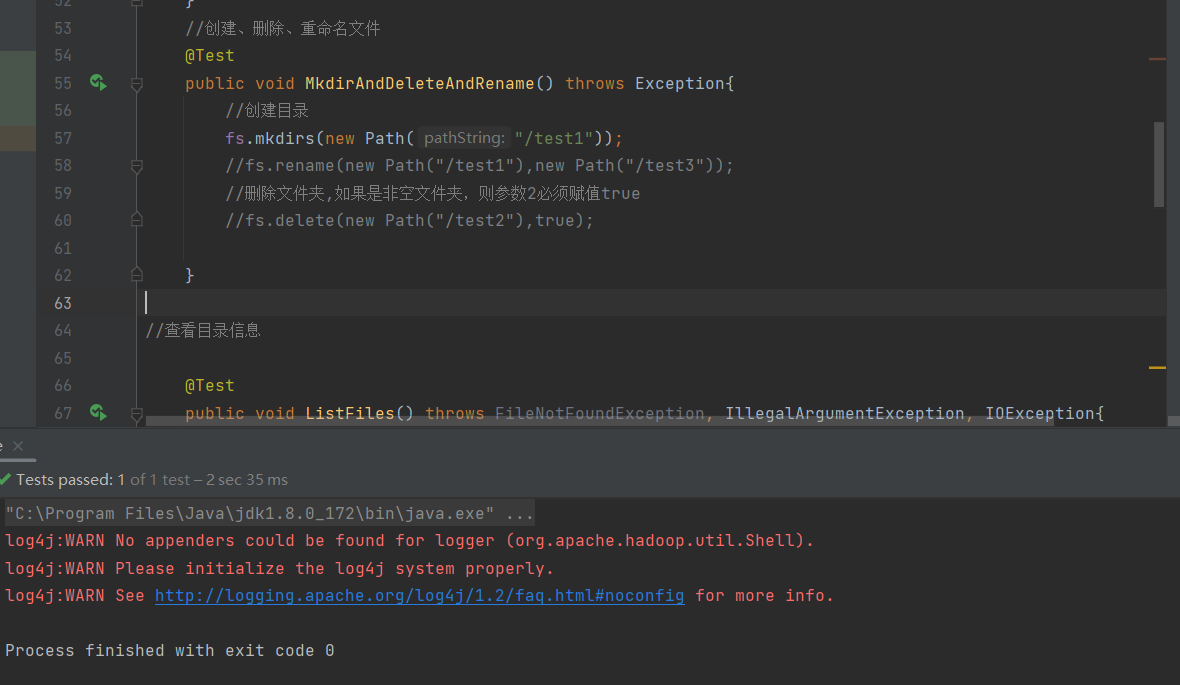
1. 编写Java类HDFS\_API\_TEST,初始化客户端对象，并运行测试。

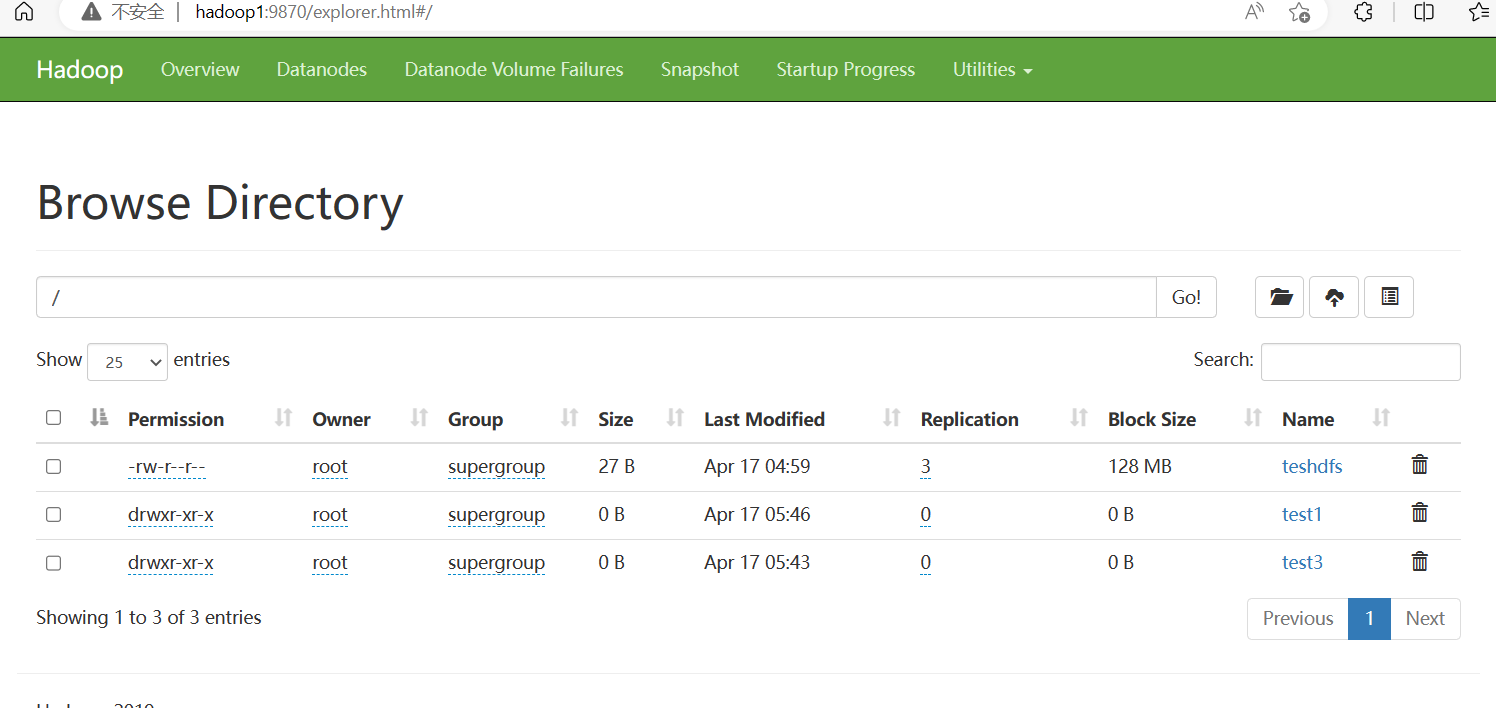
①上传文件

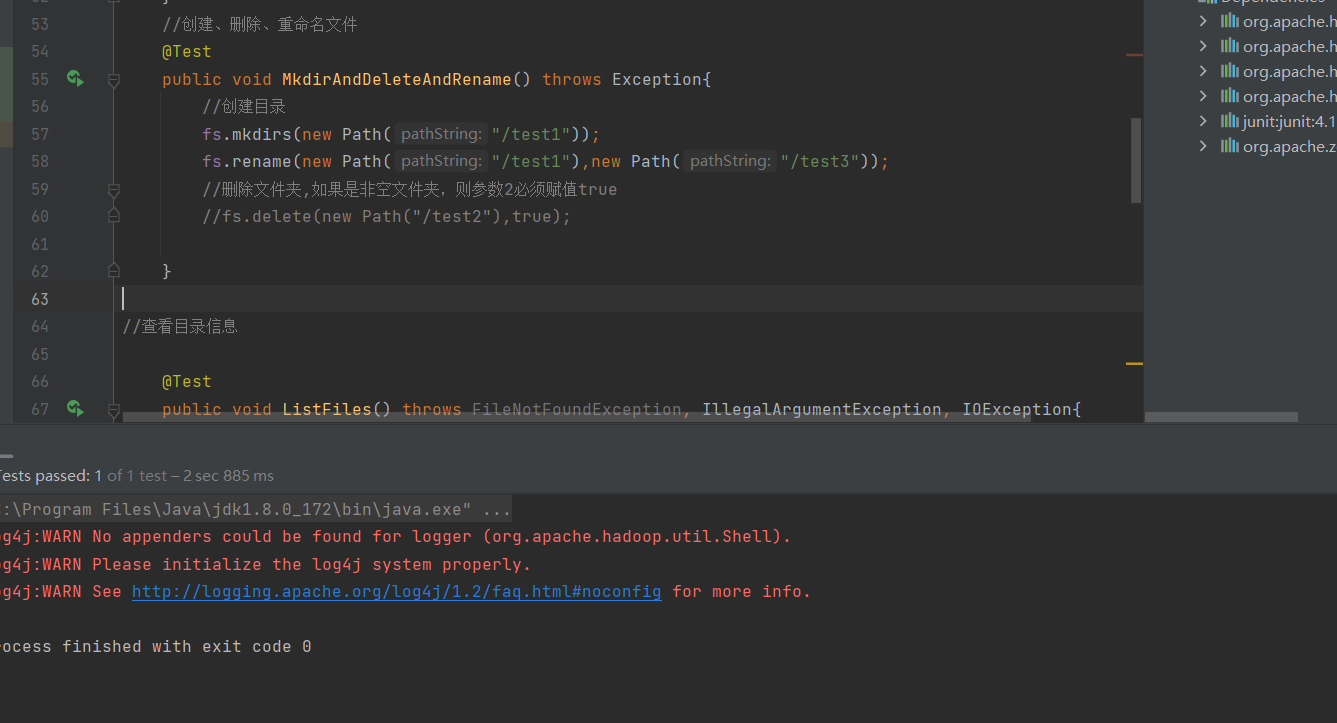


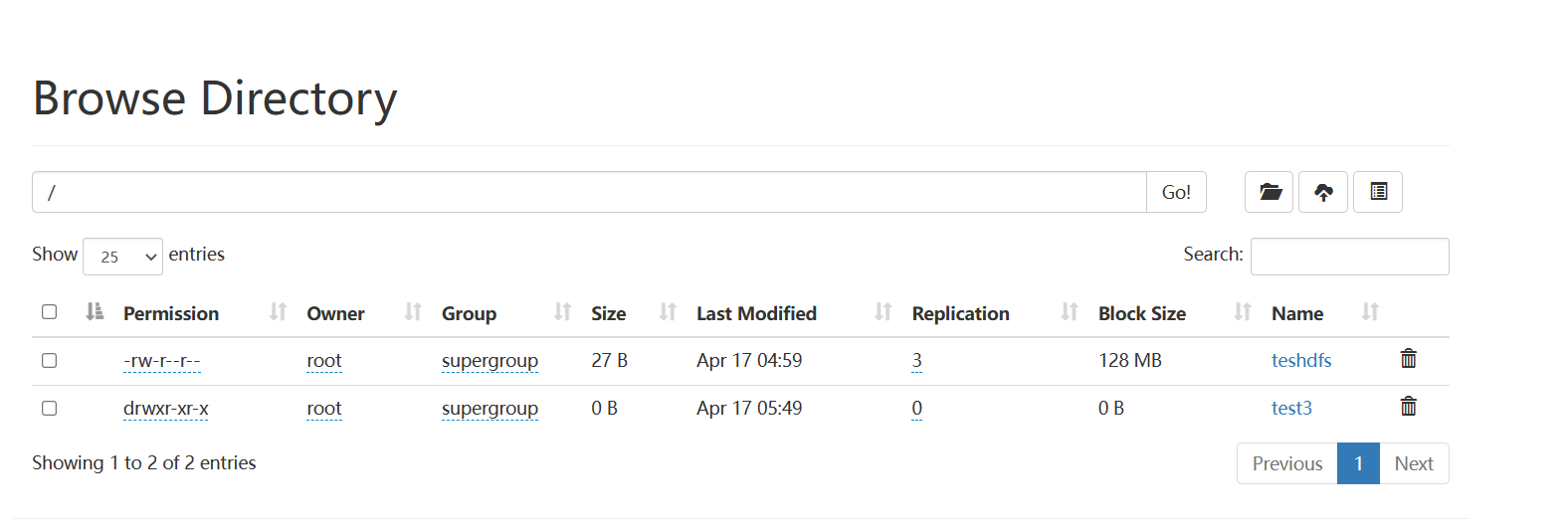
②复制文件到本地文件系统



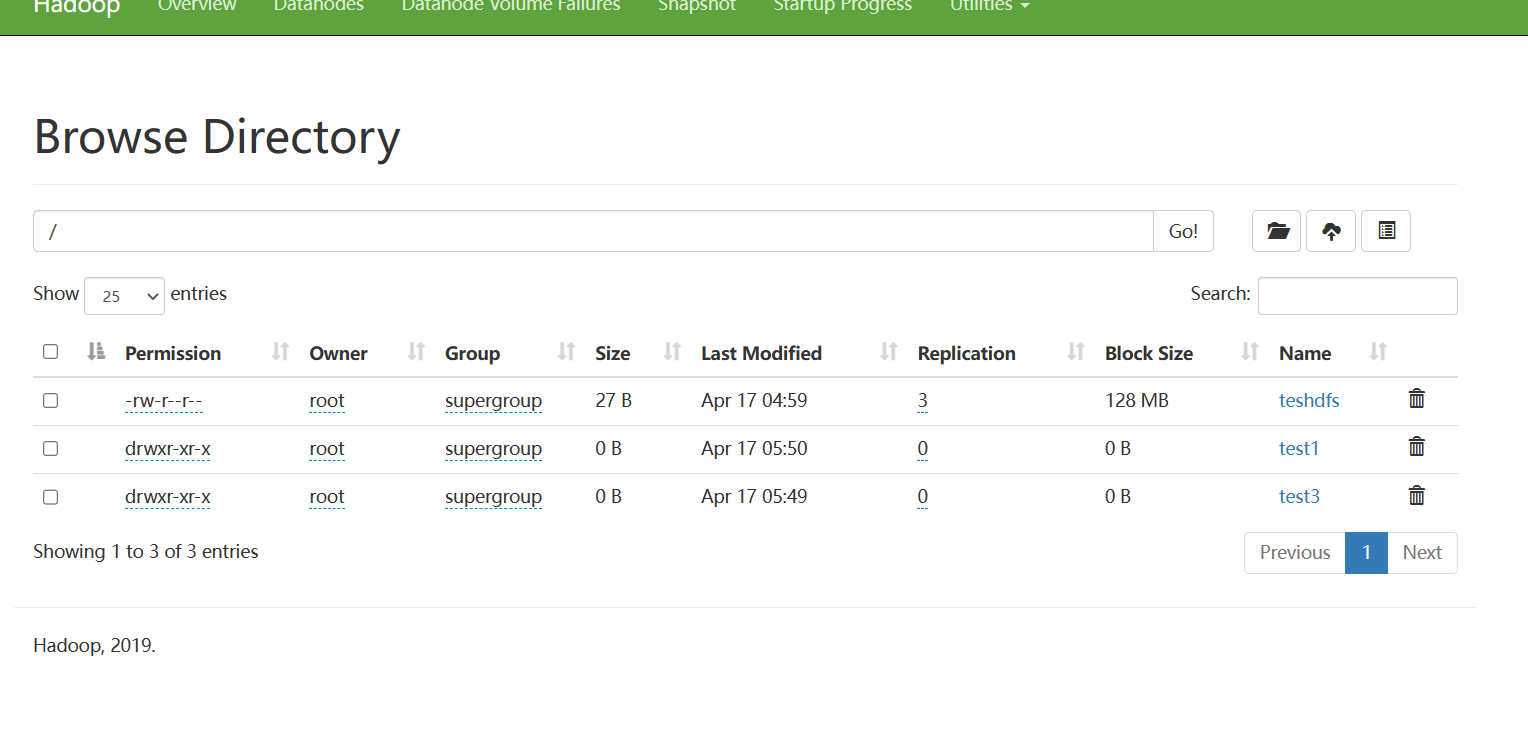
③创建、删除、重命名文件

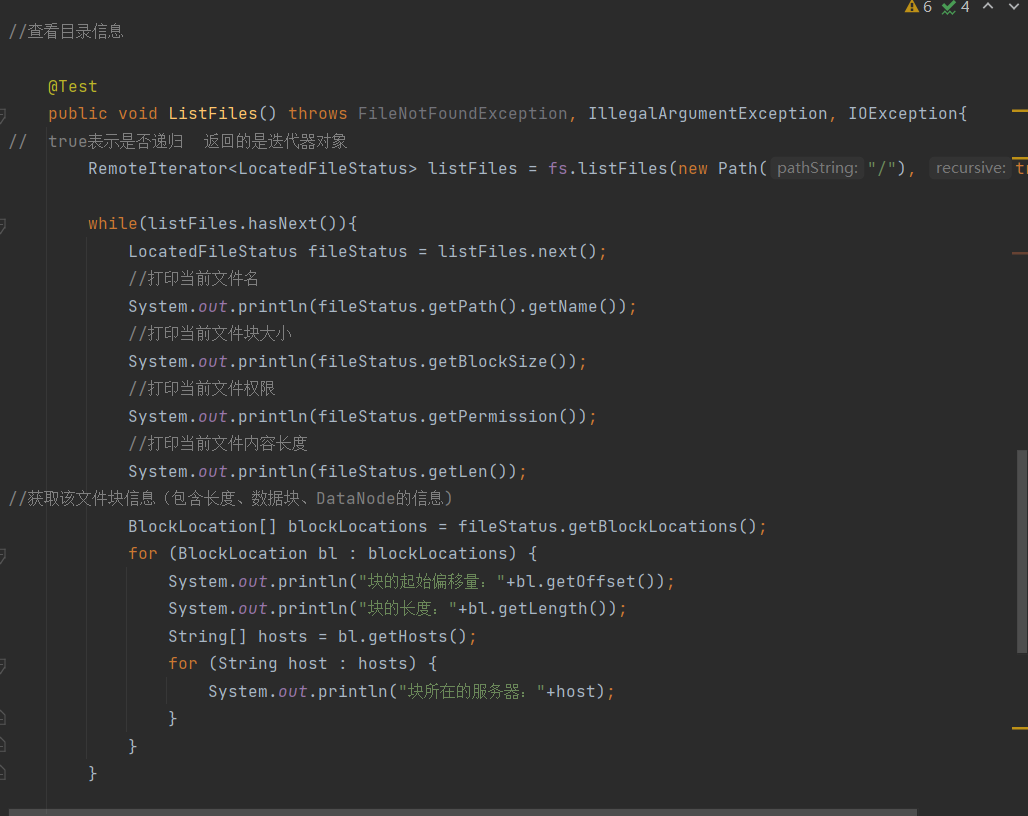


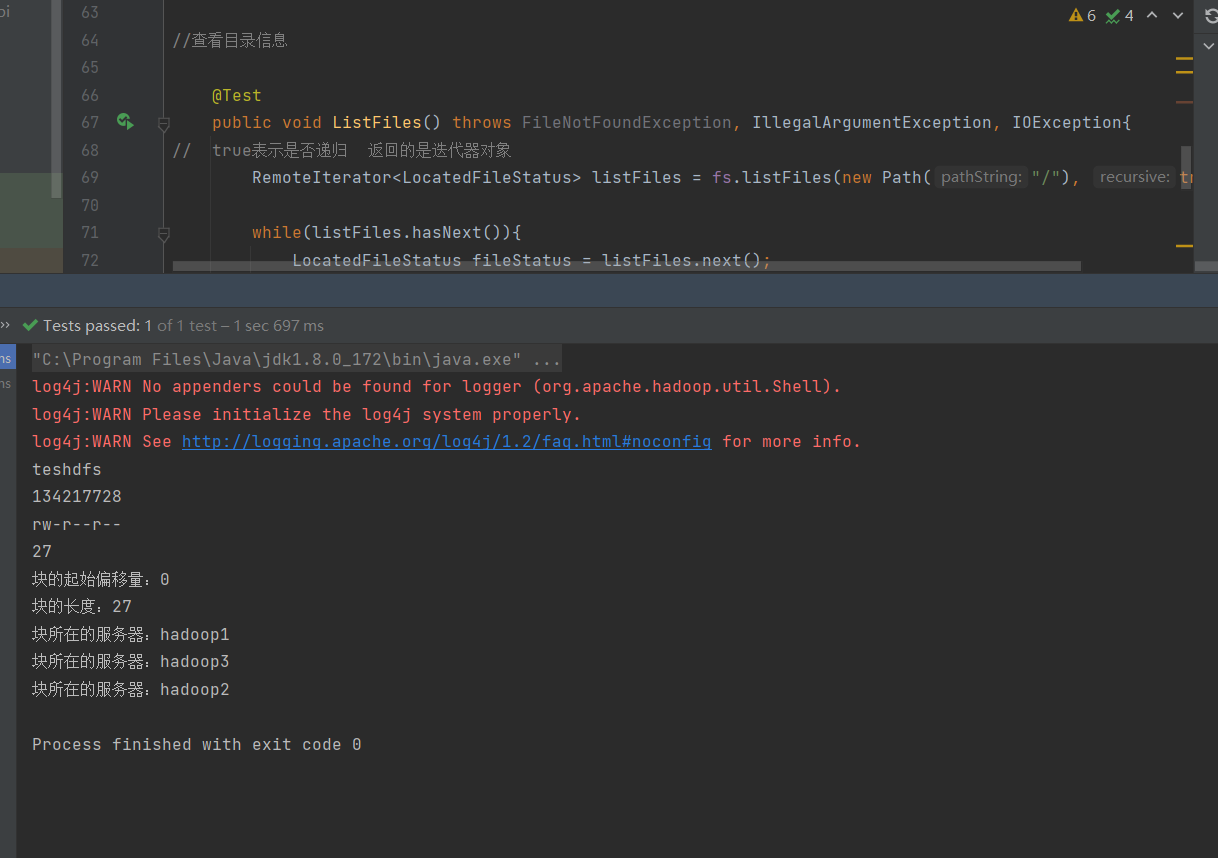


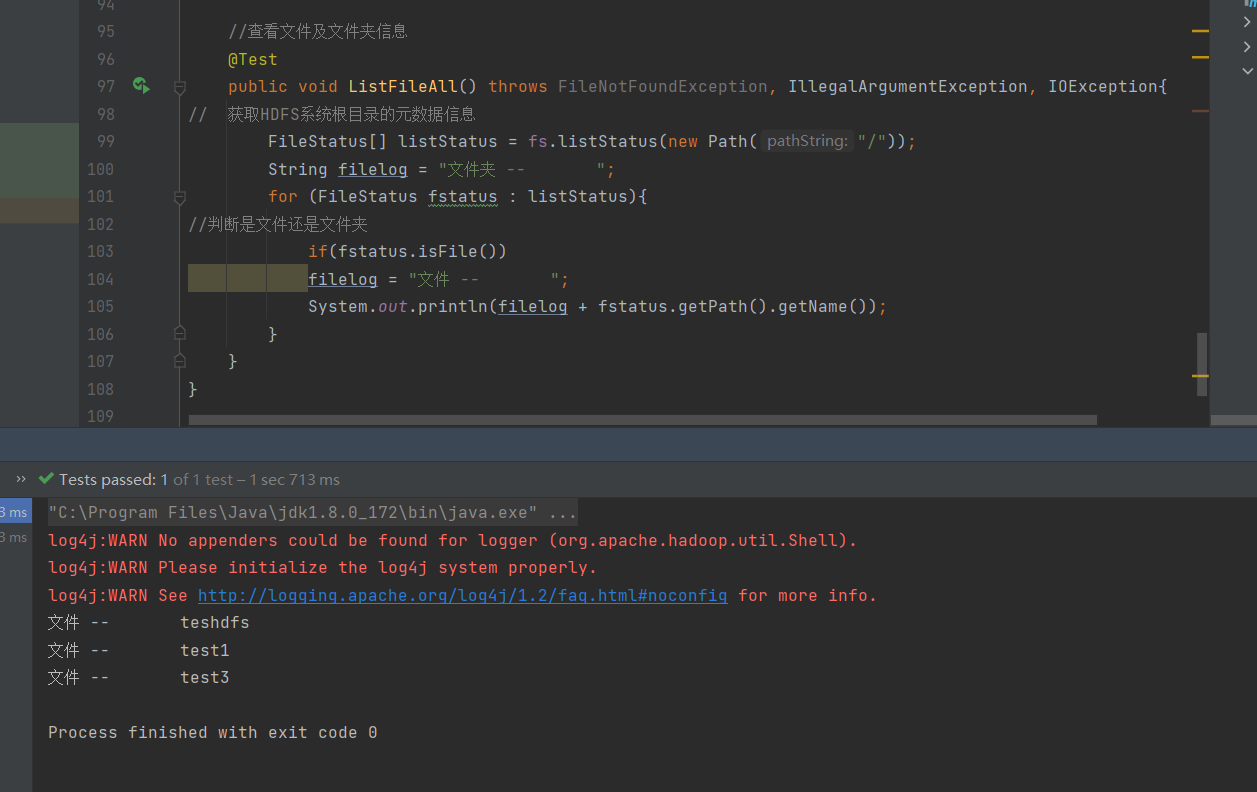






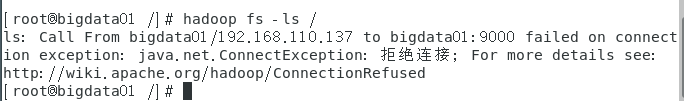
④查看目录信息



⑤查看文件及文件夹信息

1. **实验总结**

1.没有启动HDFS，解决方法：sbin/start-dfs.sh



2.我在进行HDFS的Java API实验时，发现原虚拟机存在问题，于是重新创建了虚拟机，名称不同于原虚拟机名称。

3.我开始并没有配置Windows的hadoop环境变量，也没在C:\Windows\System32添加hadoop.dll所以在运行代码时报错。hadoop.dill通用，不受版本限制。

4.上传文件操作要新建一个文件。

5.创建、删除、重命名文件操作要逐个运行，否则看不到文件变化。

6.不要连续格式化。