实验报告

报告标题：Hadoop环境搭建

学号：21190630

姓名：黄艺杰

日期：2022年11月18日

# 一、实验环境

1. 操作系统：Windows 10、Linux

2. 相关软件（含版本号）：VMWare15pro、FinalShell3.9、Intellij IDEA2019Professional

3. 其它工具：JDK1.8

# 二、实验内容及其完成情况

（针对上述实验内容逐一详述实验过程）

1. Linux（CentOS 7 发行版本）安装及网络配置：

按要求安装虚拟机，安装时确认网络适配器的工作模式为 NAT

使用 ipconfig 命令查看本地 IP 地址

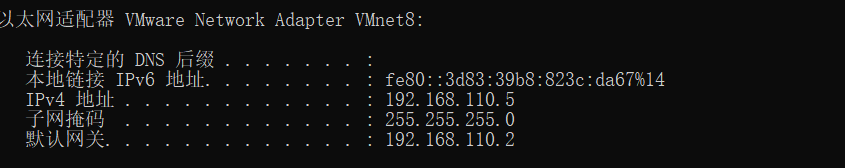


图1.1本地IP地址

在虚拟编辑器中查看VMnet8 的 IP 地址、子网掩码和默认网关，在 NAT 设置中将网关 IP 设置为一个同一子网下不同于主机 IP 的值



图1.2虚拟编辑器VMnet8 的 IP 地址、子网掩码

在网络适配器中更改 VMnet8 的 IP 地址、子网掩码和默认网关。



图1.3虚拟编辑器VMnet8 的 IP 地址、子网掩码和默认网关

在虚拟机中更改 IP 地址、子网掩码和默认网关，DNS。



图1.4虚拟机中IP 地址、子网掩码和默认网关，DNS

使用FinalShell选用SSH连接，连接CentOS 7，虚拟机的IP地址为192.168.110.101，连接成功，完成网络配置

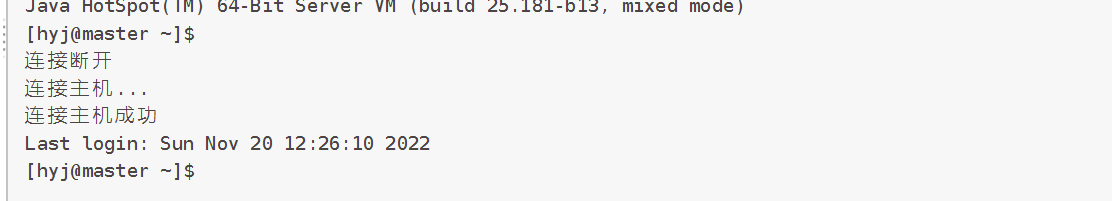


图1.5使用FinalShell连接虚拟机

1. JDK1.8的安装配置

将 JDK 安装包上传至虚拟机中，并使用 tar 命令进行解压缩，将解压后的文件夹放到 /usr/java/jdk1.8 中



图2.1将JDK安装包解压

使用 vim 编辑 .bash\_profile，在 PATH 中加入 JDK 的路径

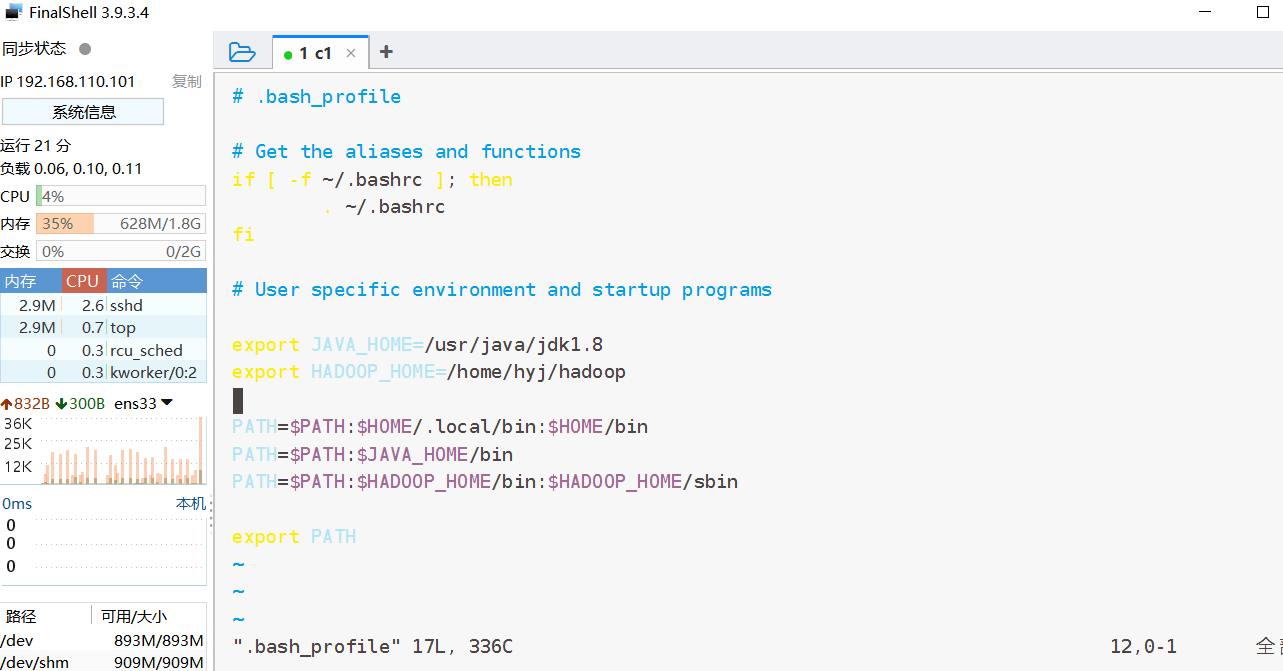


图2.2在.bash\_profile加入JDK路径

使用命令查看java版本

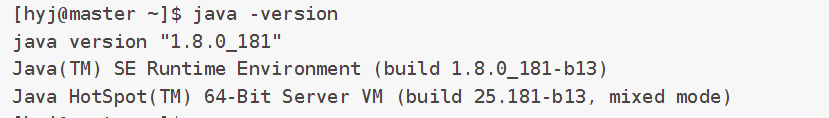


图2.3查看Java版本

1. 单机版Hadoop的安装配置

首先将hadoop压缩包上传到虚拟机，并解压hadoop-2.6.0-cdh5.7.0.tar

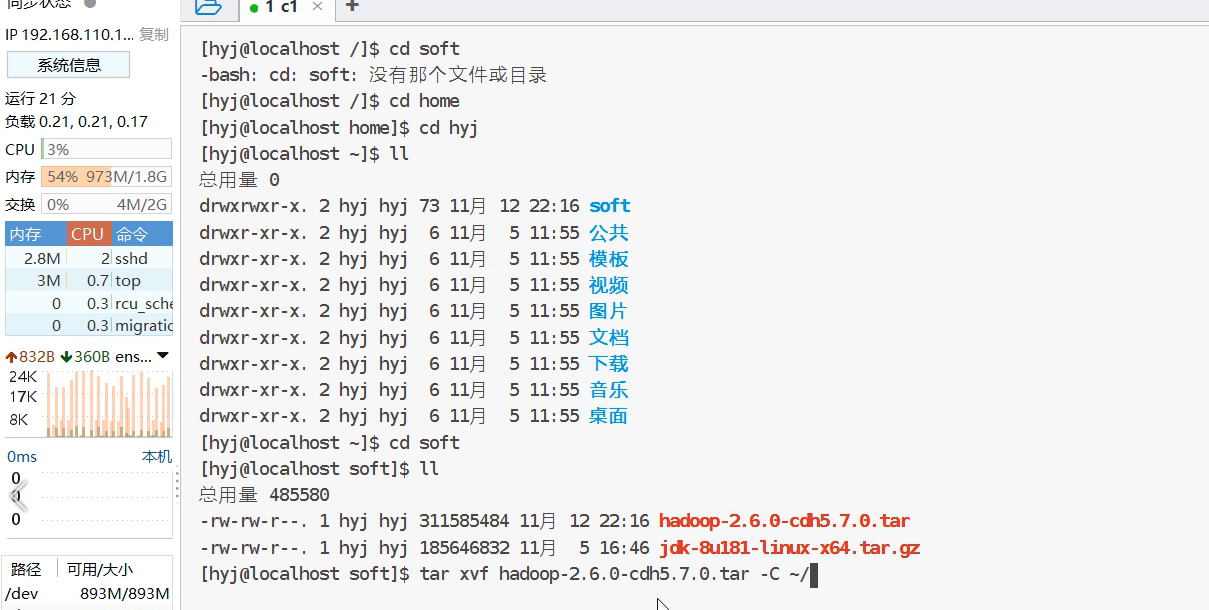


图3.1解压hadoop压缩包

配置 Hadoop 环境变量，使用 vim 编辑 .bash\_profile后执行命令重载

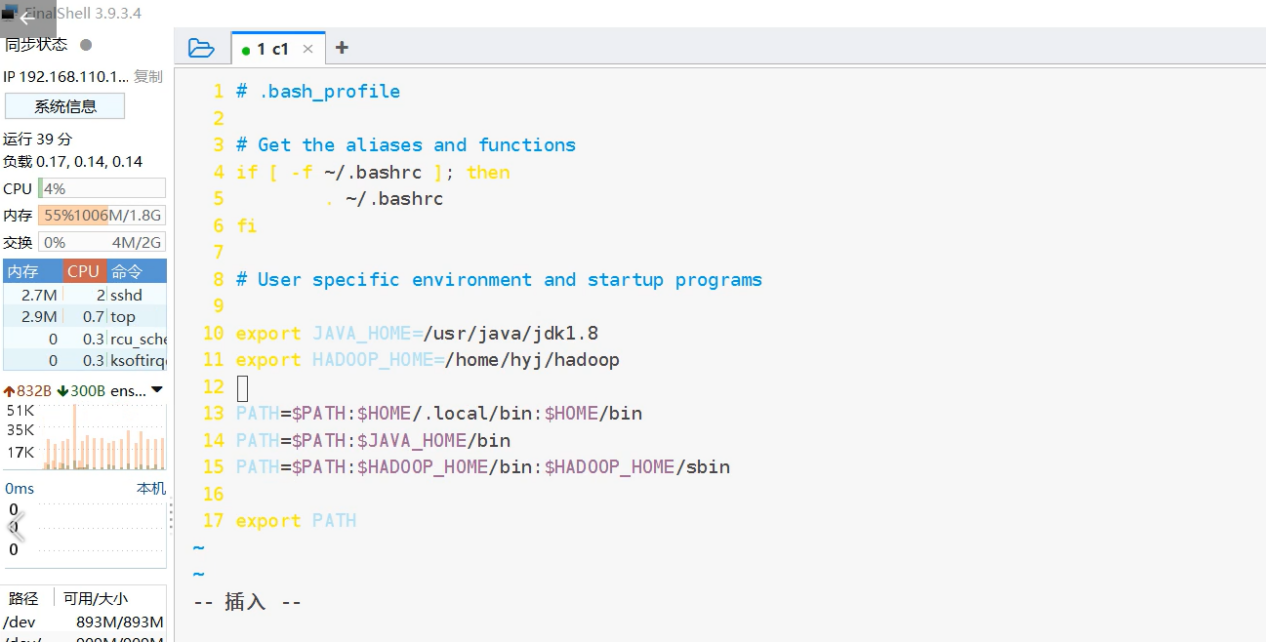


图3.2编辑.bash\_profile

 编辑 Hadoop 环境配置文件hadoop-env.sh

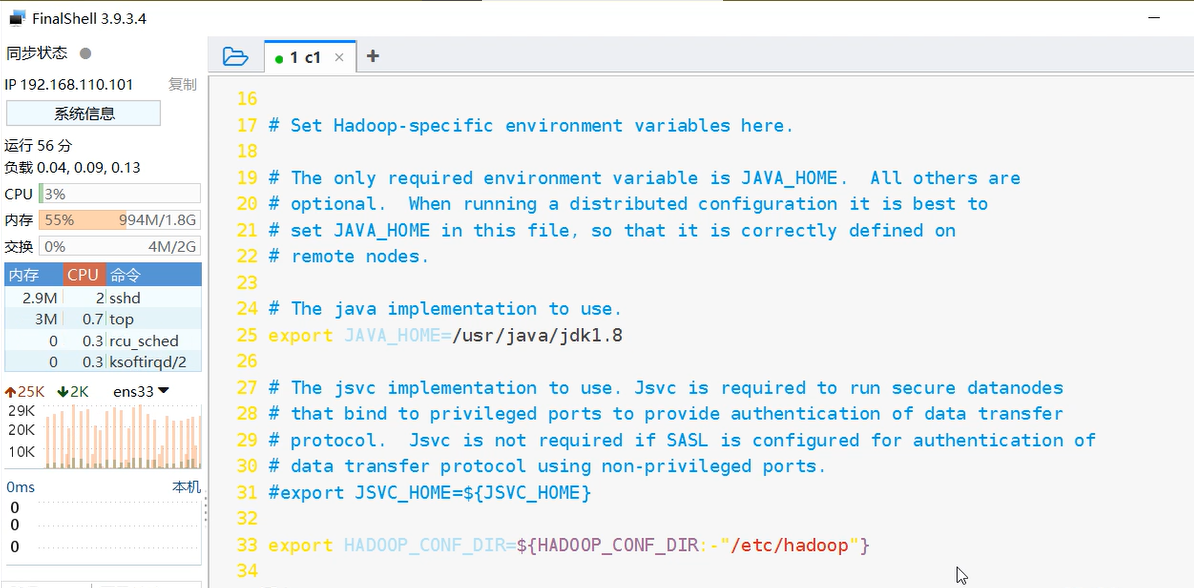


图3.3编辑Hadoop环境配置文件

配置 Hadoop 核心文件



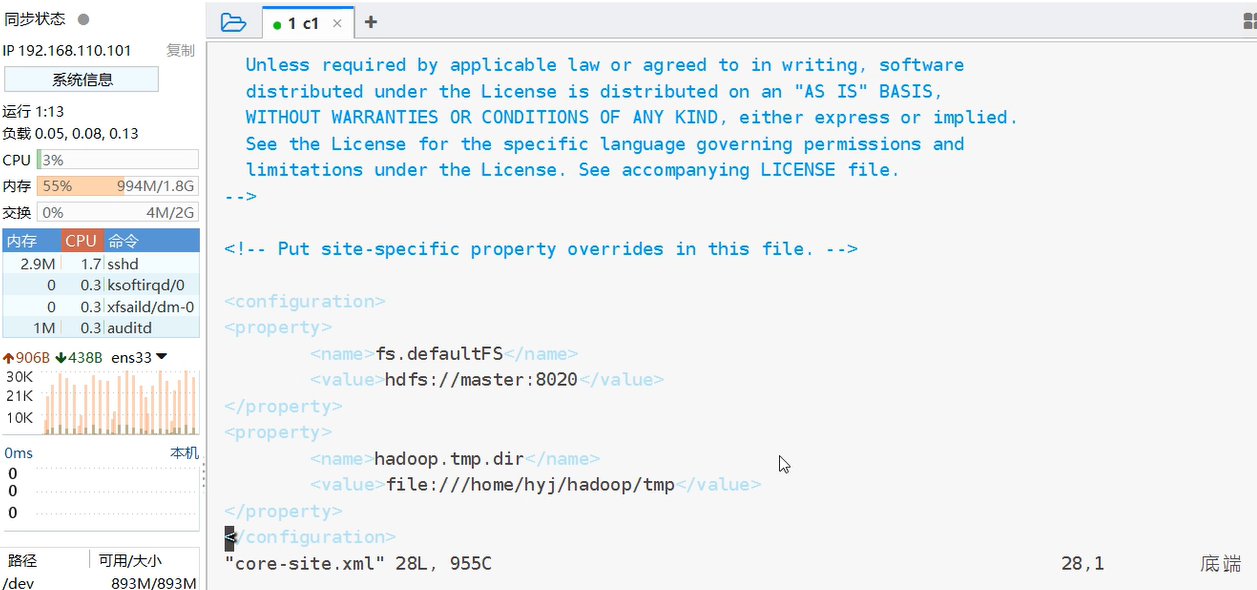


图3.4配置 Hadoop 核心文件

配置 HDFS 文件





图3.5配置 HDFS 文件

执行初始化 NameNode 命令



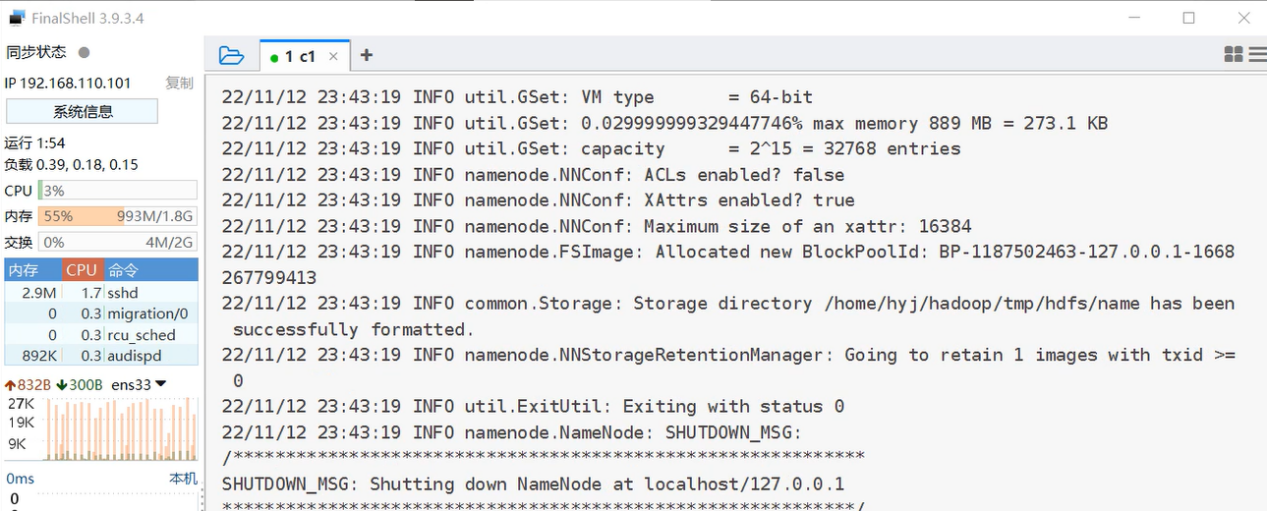


图3.6执行初始化 NameNode 命令，出现successfully formatted

添加 hostname为master

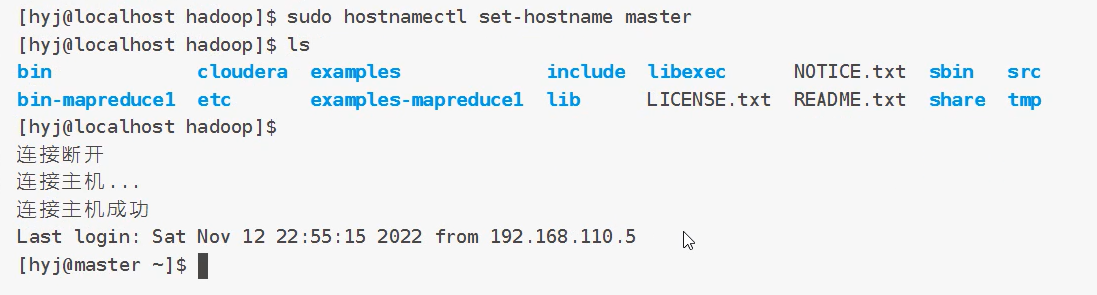


图3.7成功添加hostname

启动 Hadoop，执行 start-dfs.sh 命令，之后执行 jps，可以看到 NameNode，则运行成功。



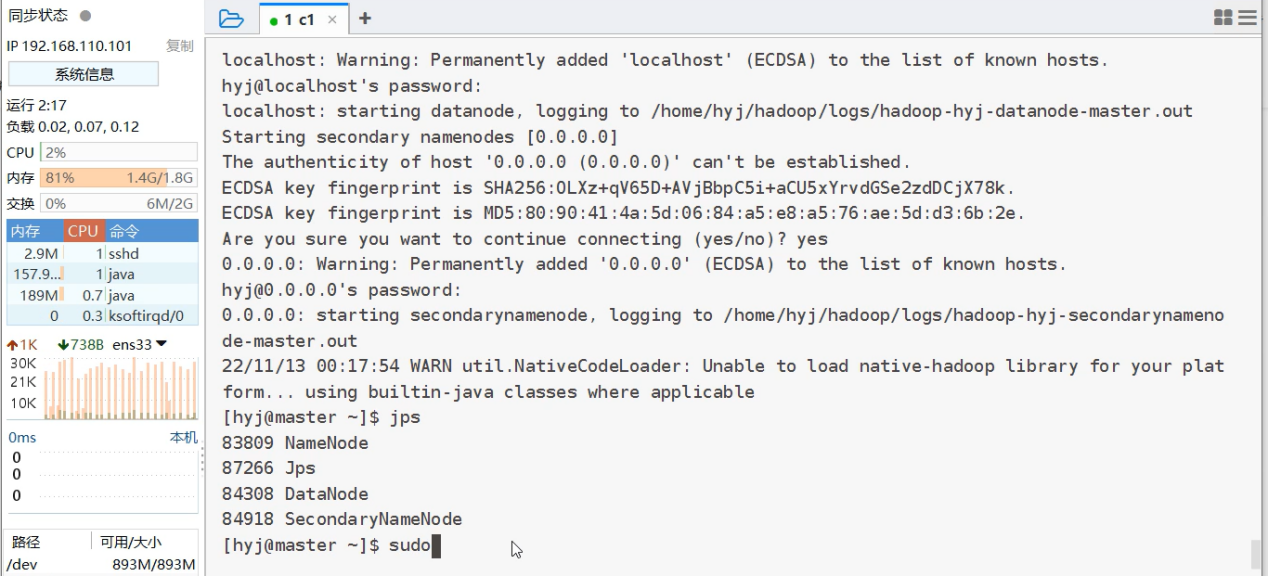


图3.8成功启动Hadoop

设置防火墙允许访问 50070 端口，通过浏览器查看CentOS 7的50070端口

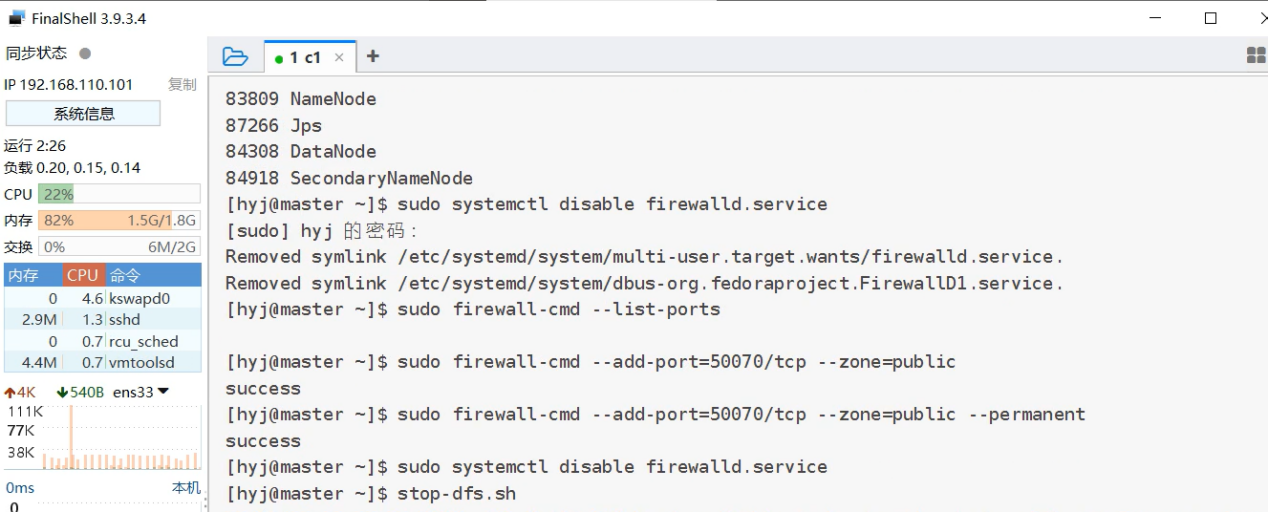


图3.9设置防火墙允许访问 50070端口



图3.10通过浏览器查看CentOS 7的50070端口

接下来因在启动过程中需要多次输入用户密码，可以配置 SSH 免输入密码

先产生密钥，在复制一份公钥，将密钥下载并设置使用





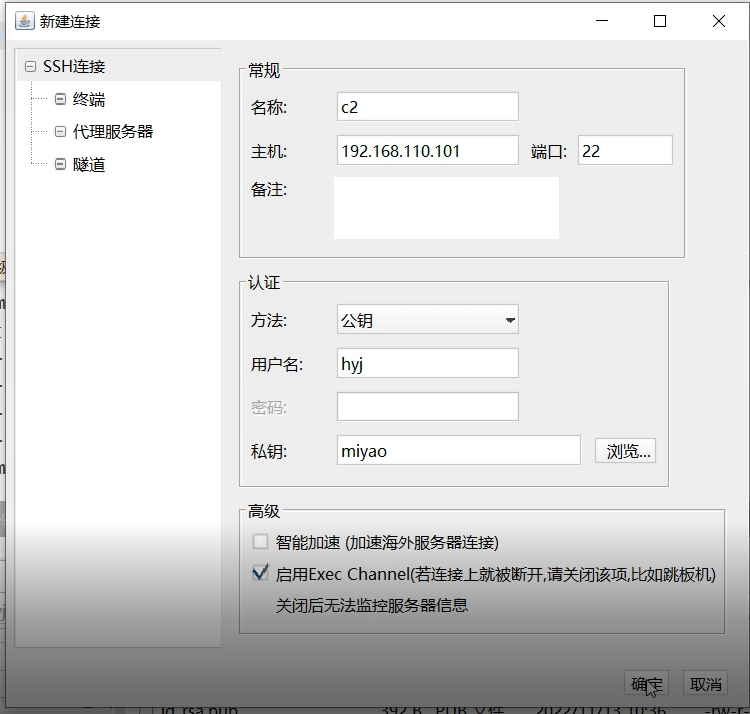
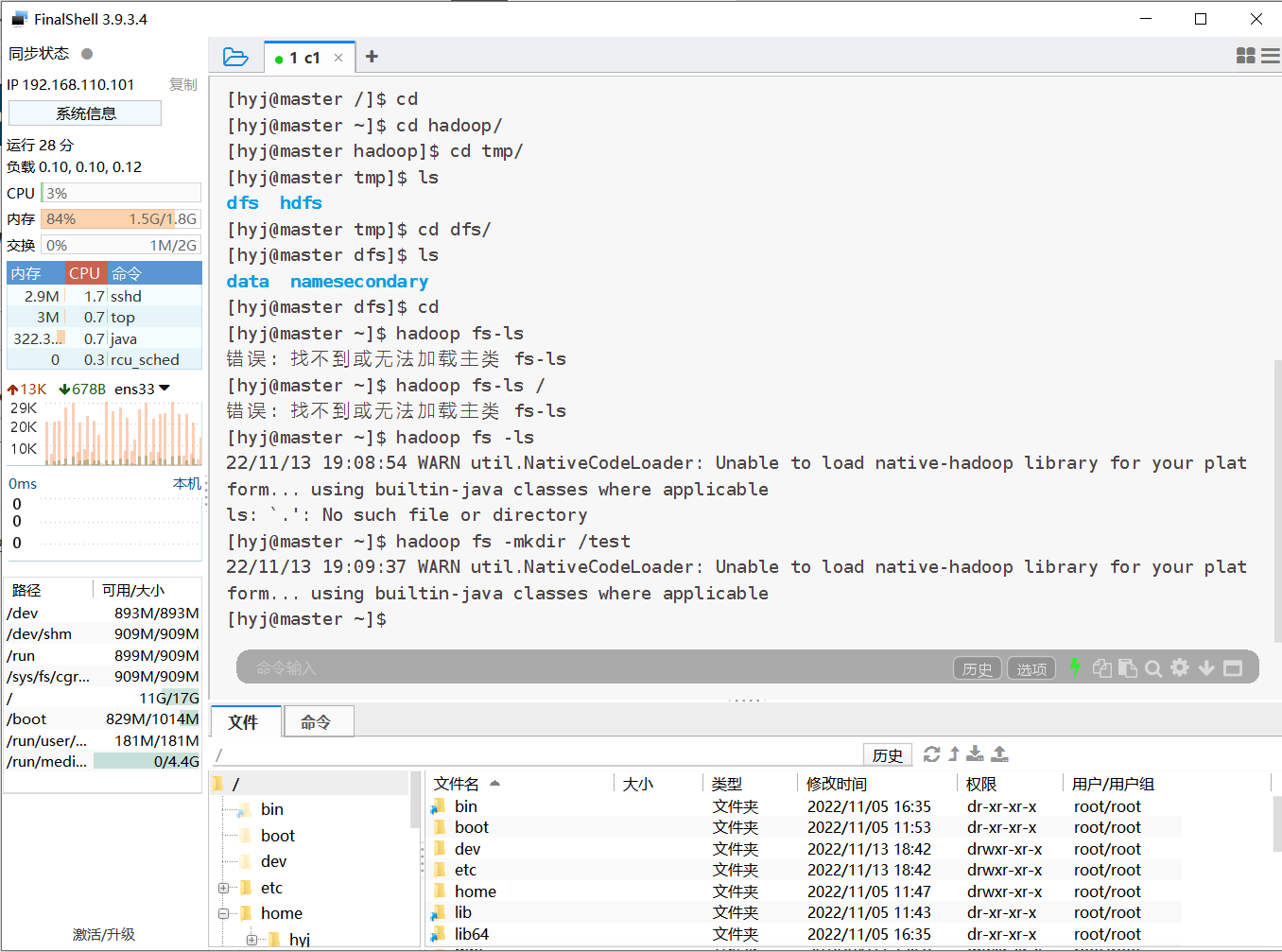


图3.11配置SSH

最后是利用hadoop创建文件夹，在HDFS中可查看



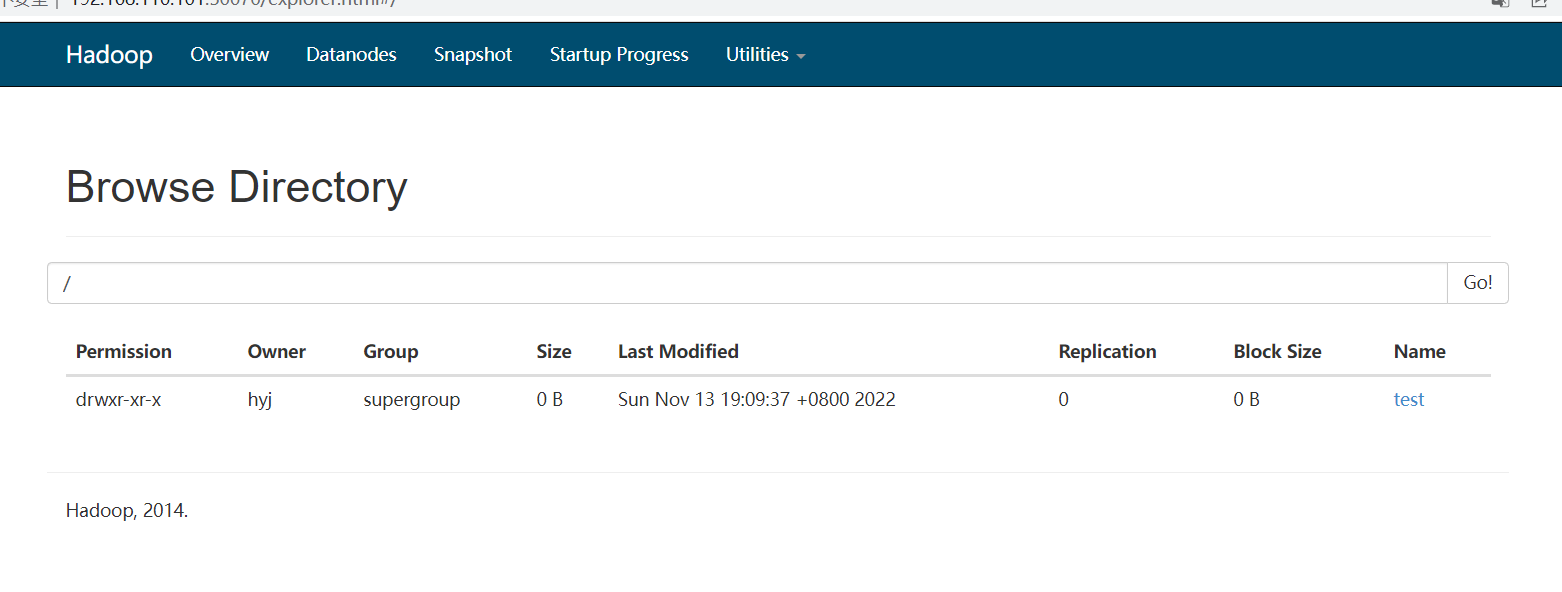


图3.12利用hadoop创建文件夹，在HDFS中查看

1. Hadoop的Java编程

在下载下来的setting.xml文件中修改镜像为阿里云镜像

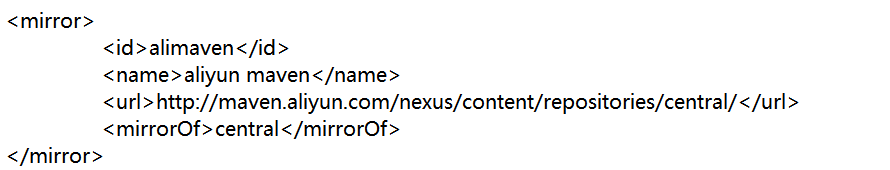


图4.1修改setting.xml为阿里云镜像

在初次打开IDEA时，配置环境，选择通过 maven 环境管理创建,选择Project SDK为JDK1.8，勾选Create from archetype框并选择maven-archetype-quickstart

在 maven settings 中修改的配置

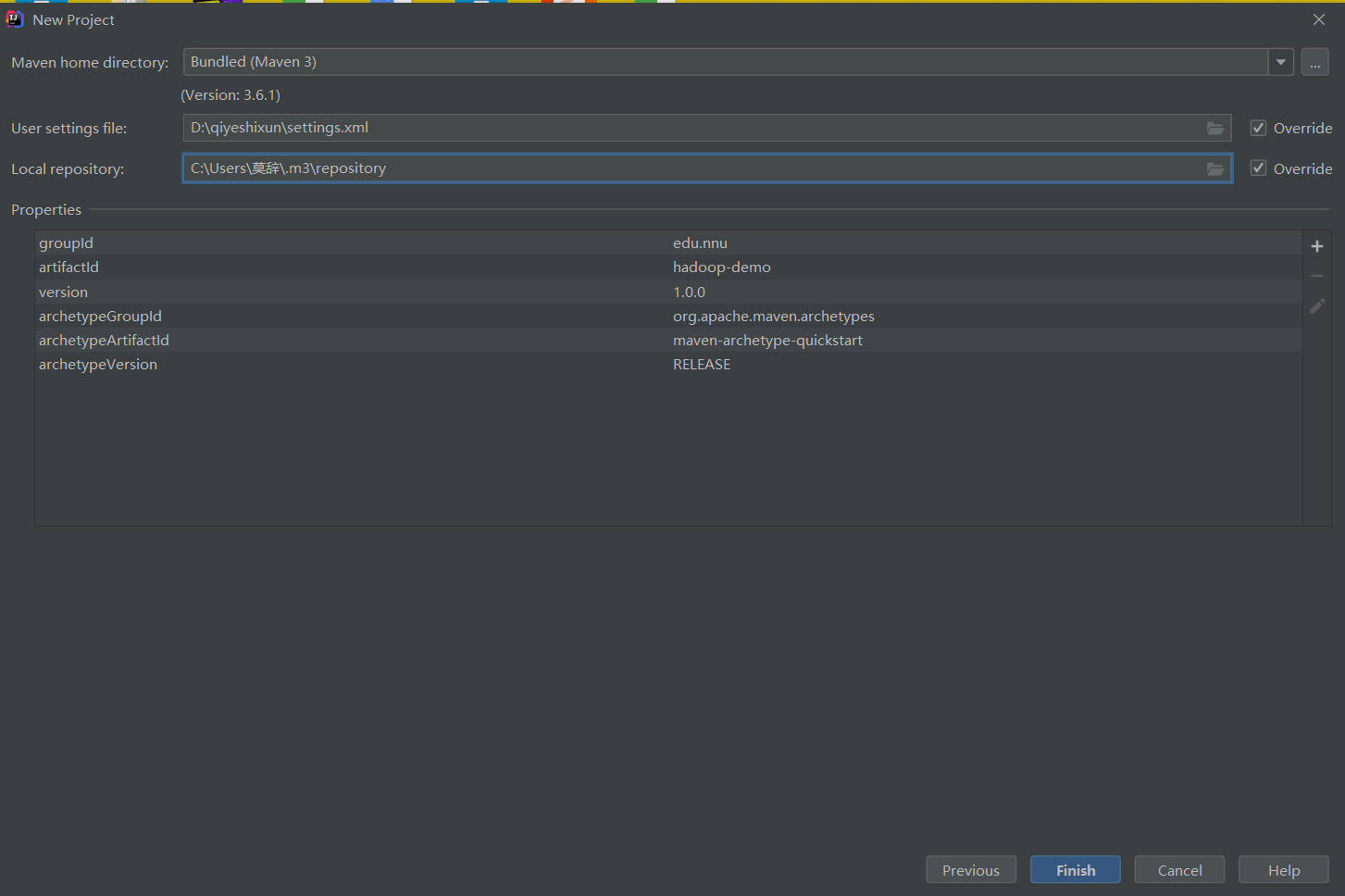


图4.2配置IEDA环境

创建项目后，在pom.xml添加hadoop版本、hadoop-common依赖、hadoop-hdfs依赖

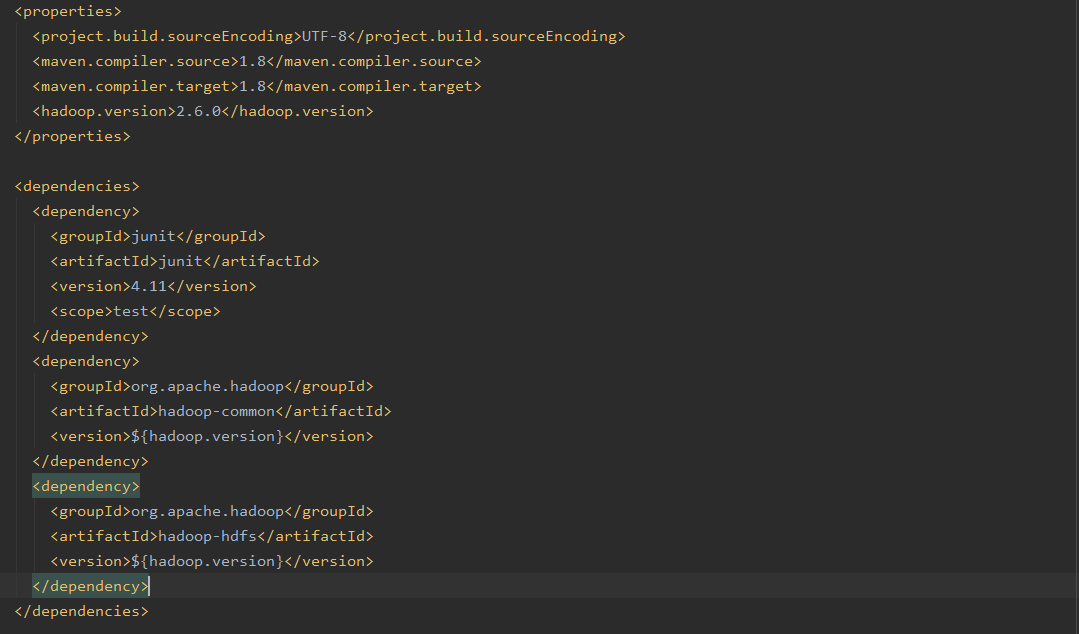


图4.3修改pom.xml中的配置

删除test文件夹中的文件，在App.java中完成一些基本链接和config配置

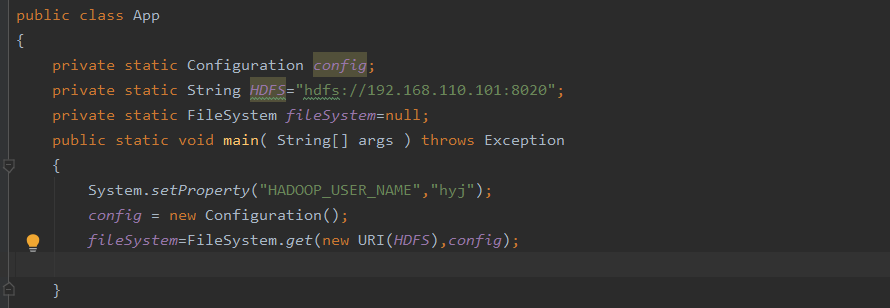
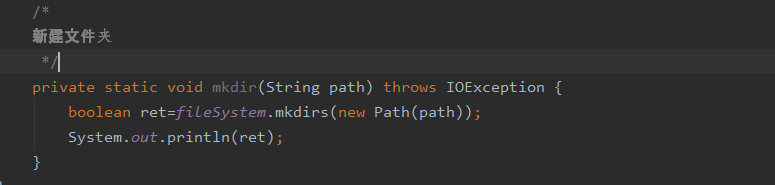
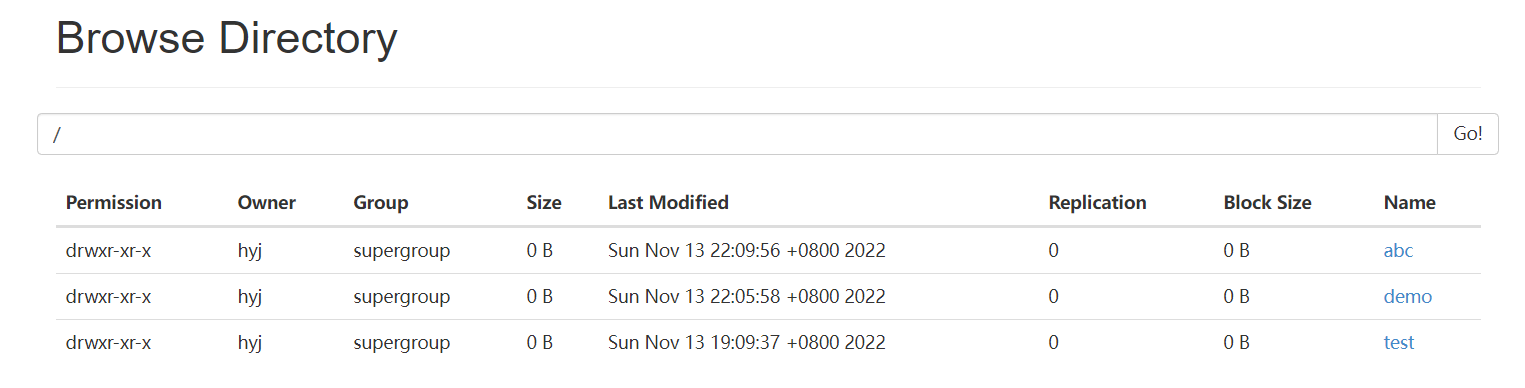


图4.4完成HDFS和config的配置

测试新建文件夹，分别是demo和多级文件夹abc/efg/hjk





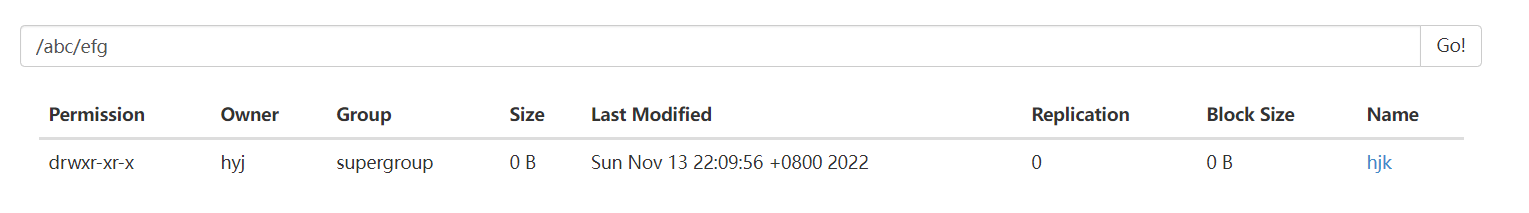
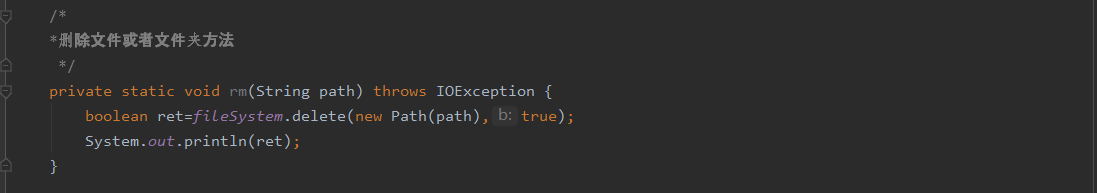
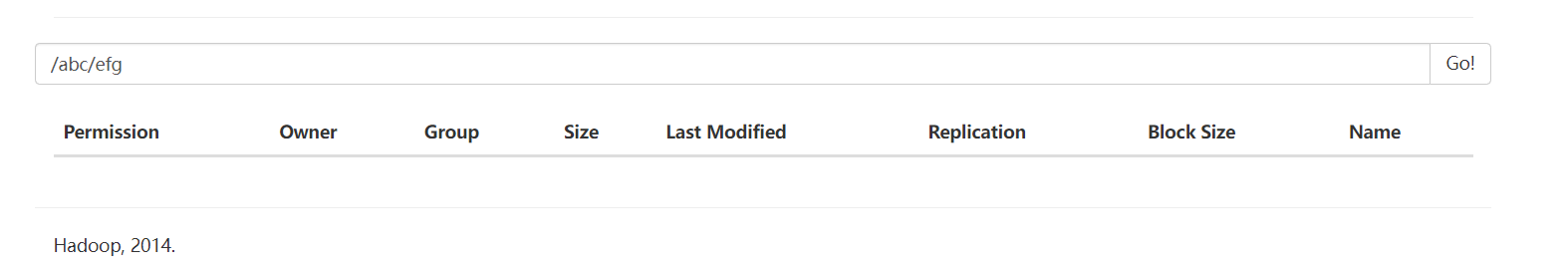


图4.5新建文件夹并查看运行结果

分别删除abc/efg/hjk，删除abc，在HDFS的Browse上分别看到abc下的efg中为空，根目录下没有了abc。





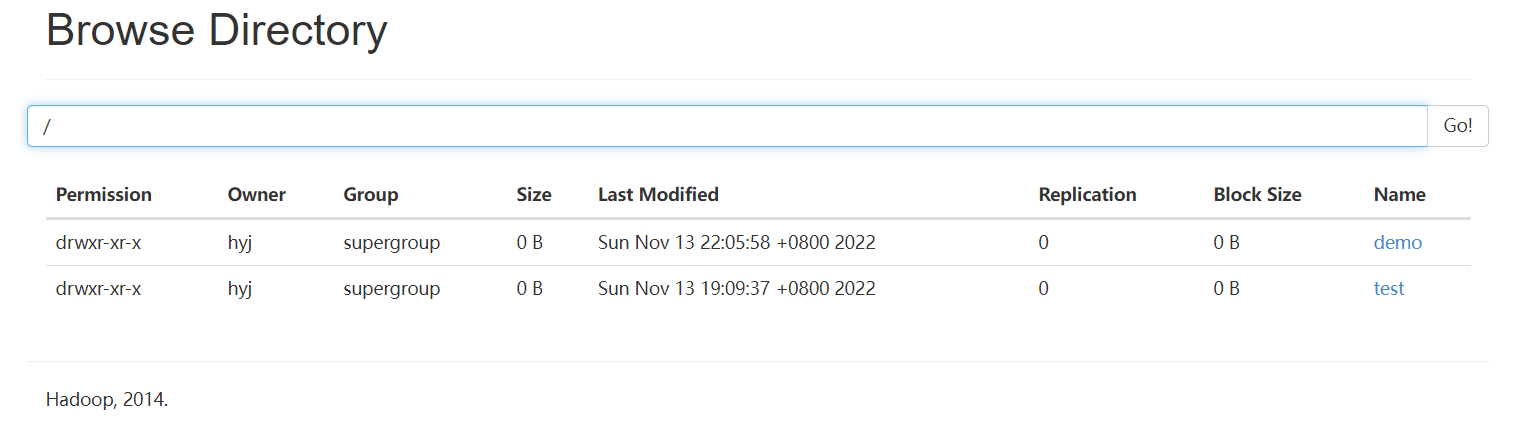
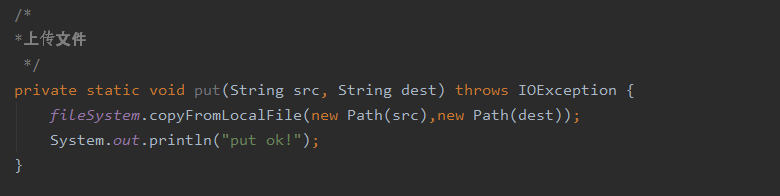


图4.6删除文件夹并查看运行结果

用put上传同一个本地文件1.txt，一个不改名，一个改为a.txt，可以看到它们除了姓名信息不同，其他信息是一样的



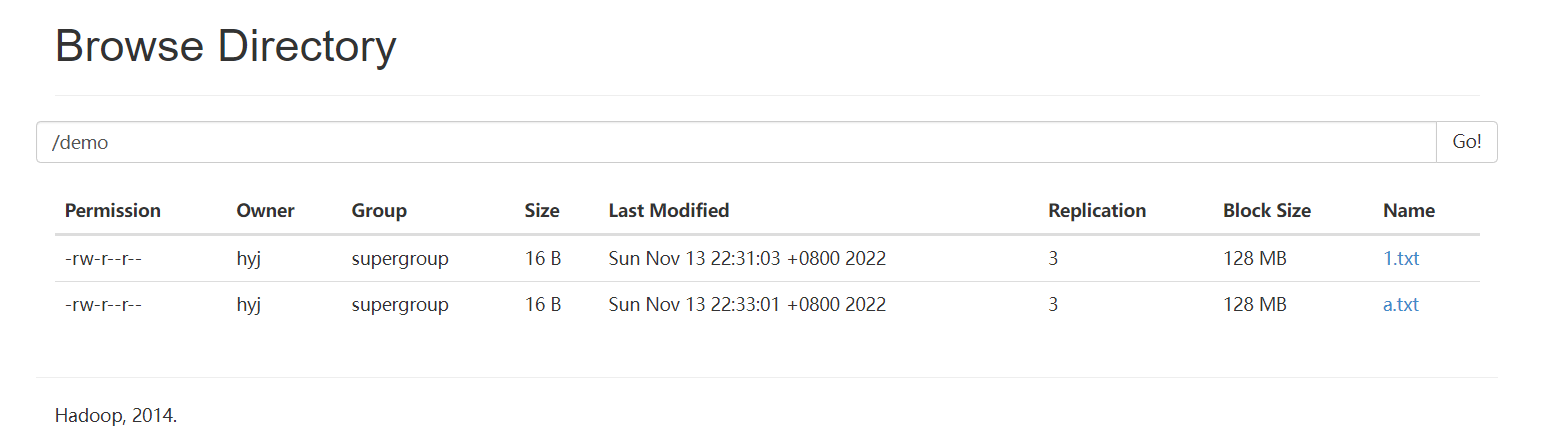
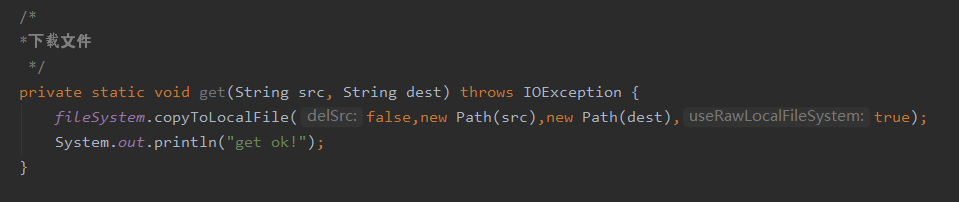


图4.7上传文件夹并查看运行结果

从demo中使用get()下载1.txt到本地指定路径。



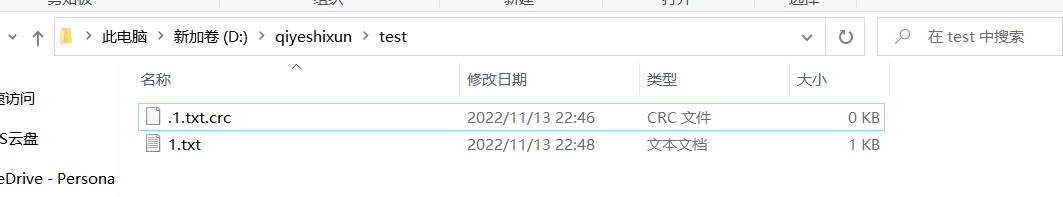


图4.8下载文件并查看运行结果

用list()方法查看根目录和demo文件夹中详细信息。

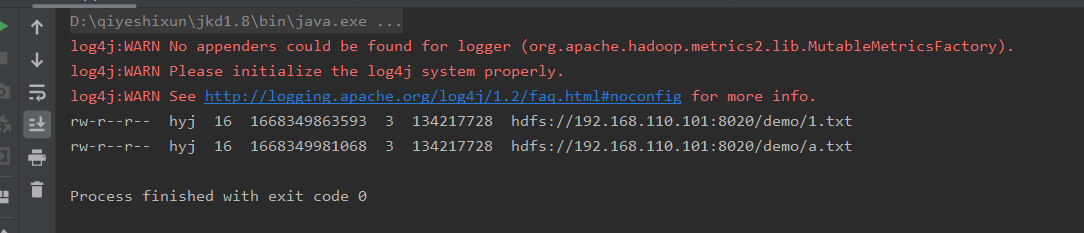
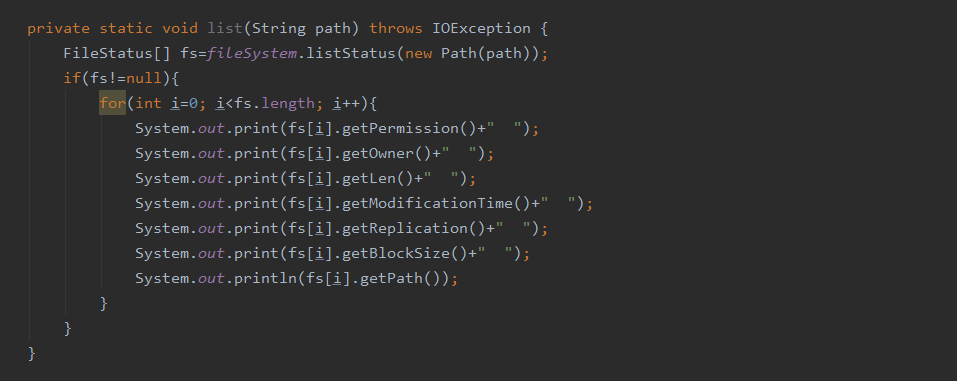


图4.9查看根目录和demo文件夹中详细信息

最后用put()函数上传四大名著文件到HDFS的/book目录中。

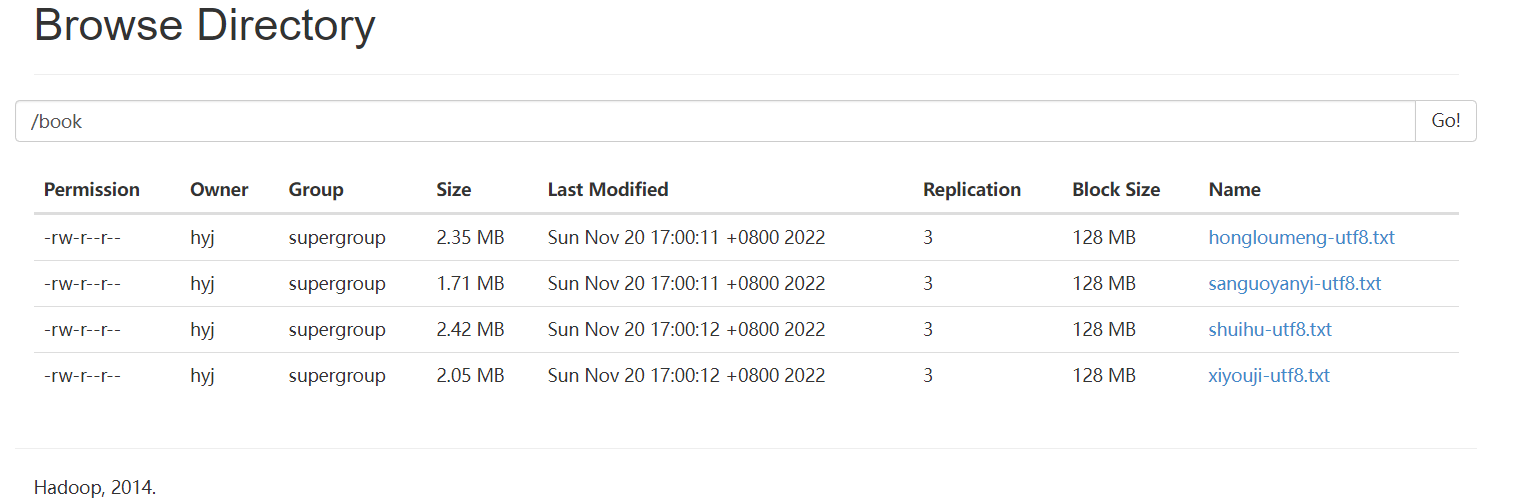


图4.10上传四大名著到HDFS

# 三、实验总结

（可以总结实验中出现的问题以及解决的思路，也可以列出没有解决的问题）

1. 在用浏览器打开HDFS页面前，用以下指令在虚拟机中关闭防火墙。

关闭防火墙：sudo systemctl stop firewalld

打开 hdfs 服务：start-dfs.sh

1. 对于某些文件，有些是root下的文件，并不属于所有者，这时候对文件进行操作时记得要用sudo提高权限。
2. 在IDEA上设定不同路径时，Windows 宿主机文件目录路径中的‘/’要使用 ‘\’来进行引导。