**华东师范大学数据科学与工程学院实验报告**

|  |  |  |
| --- | --- | --- |
| **课程名称**：分布式编程模型与系统 | **年级**：2018 | **上机实践成绩**： |
| **指导教师**：徐辰 | **姓名**： | **学号**： |
| **上机实践名称**： Hadoop部署与编程 **上机实践日期：** | | 2021.3.11-4.1 |
| **上机实践编号**：实验1 | **组号**： | **上机实践时间**：**15:00-16:30** |

1. **实验目的**

* **学习编写简单的基于Java API操作HDFS的程序。**
* **掌握在IDEA中调试HDFS相关程序，以及在单机为分布、分布式模式下提交运行HDFS相关程序的方法。**
* **理解HDFS中块的概念**

1. **实验任务**

* 完成基于JavaAPI编写的“将本地文件上传到HDFS中”的程序。
* 在IDEA中调试以上程序，在单机伪分布、分布式模式下提交该程序。

1. **实验环境**

* **操作系统：Ubuntu 18.04**
* **JDK版本:1.8**
* **Hadoop版本：2.10.1**
* **IDEA版本: 2020.2.3(Ultimate 版)**

1. **实验步骤**

**5.4.1 编写HDFS应用程序**

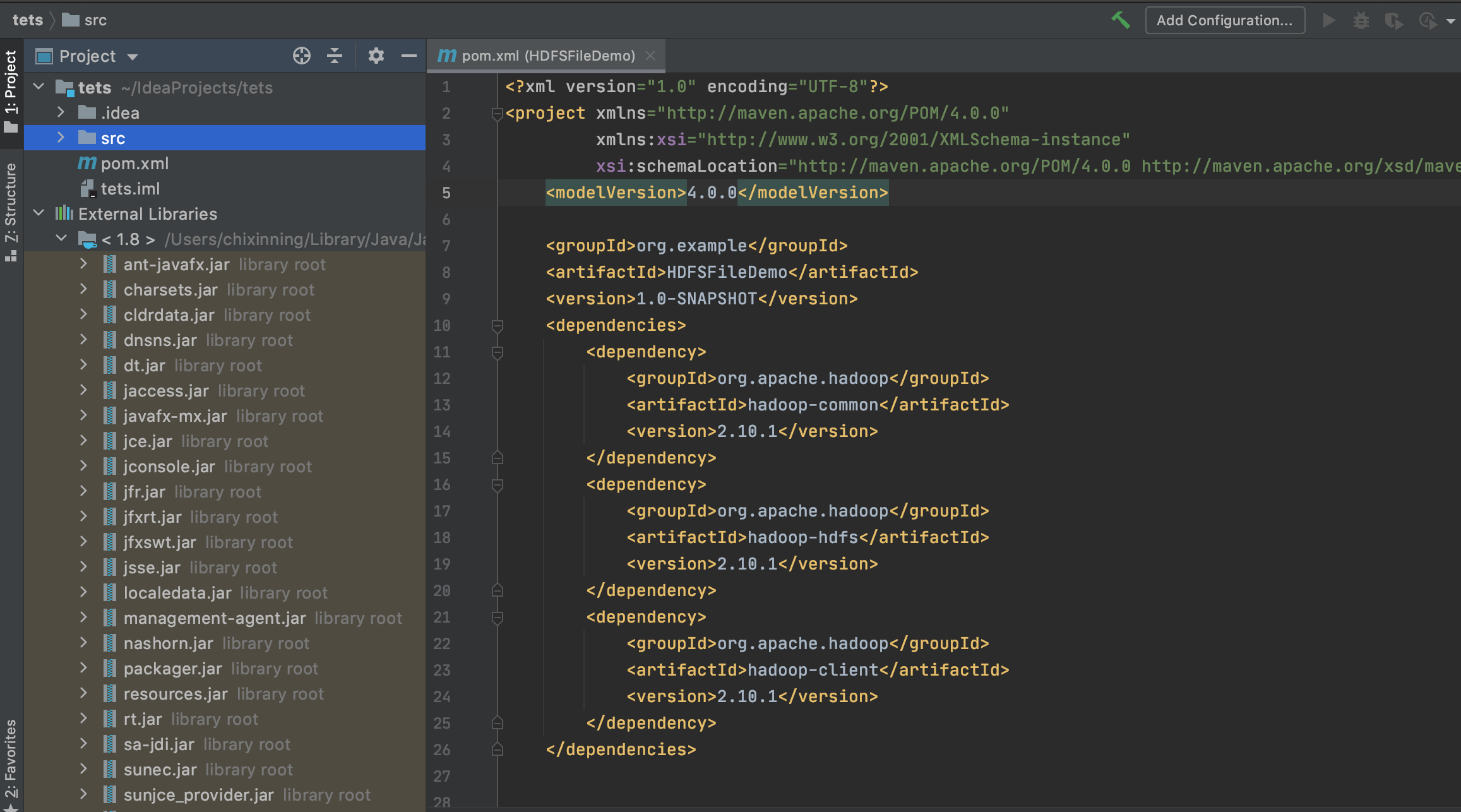
1. **新建Maven项目**

实验5是我第一次接触Maven项目的创建和pom.xml的dependency的添加，按照以前的Maven项目的创作，会有一定的插件安装的报错。

最后询问了助教学长，更改了Maven项目的创建方式(quick-start)得到问题的解决。

1. **编写Java程序**

**编写一个Java程序，实现将本地文件上传到HDFS中，即HDFS Shell中的put命令。**

****

**5.4.2 调试HDFS应用程序**

**(1)配置运行环境**

**(2)开启HDFS服务**

**(3)调试应用程序**

**5.4.3 运行HDFS应用程序**

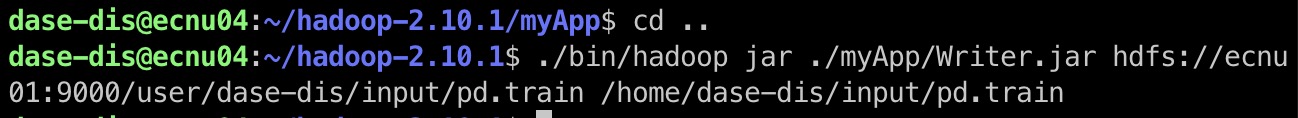
1. **在本地使用IDEA打Jar包：**

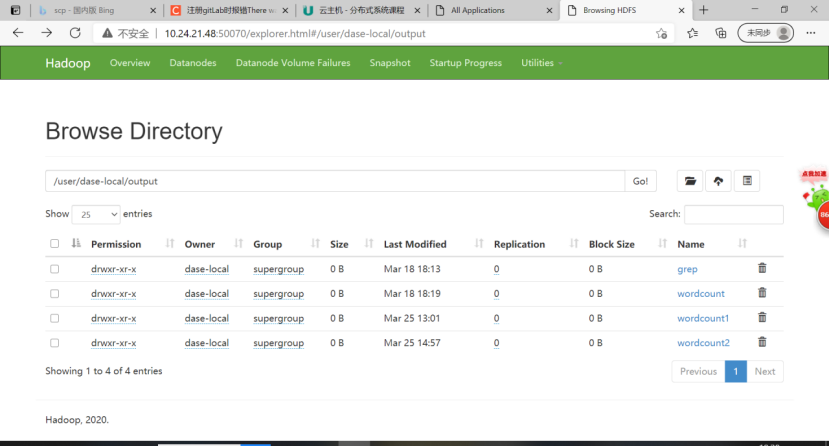
**在这次实验中第一次打Jar包，不论是实验5还是实验6后续的实验都需要打jar包。**

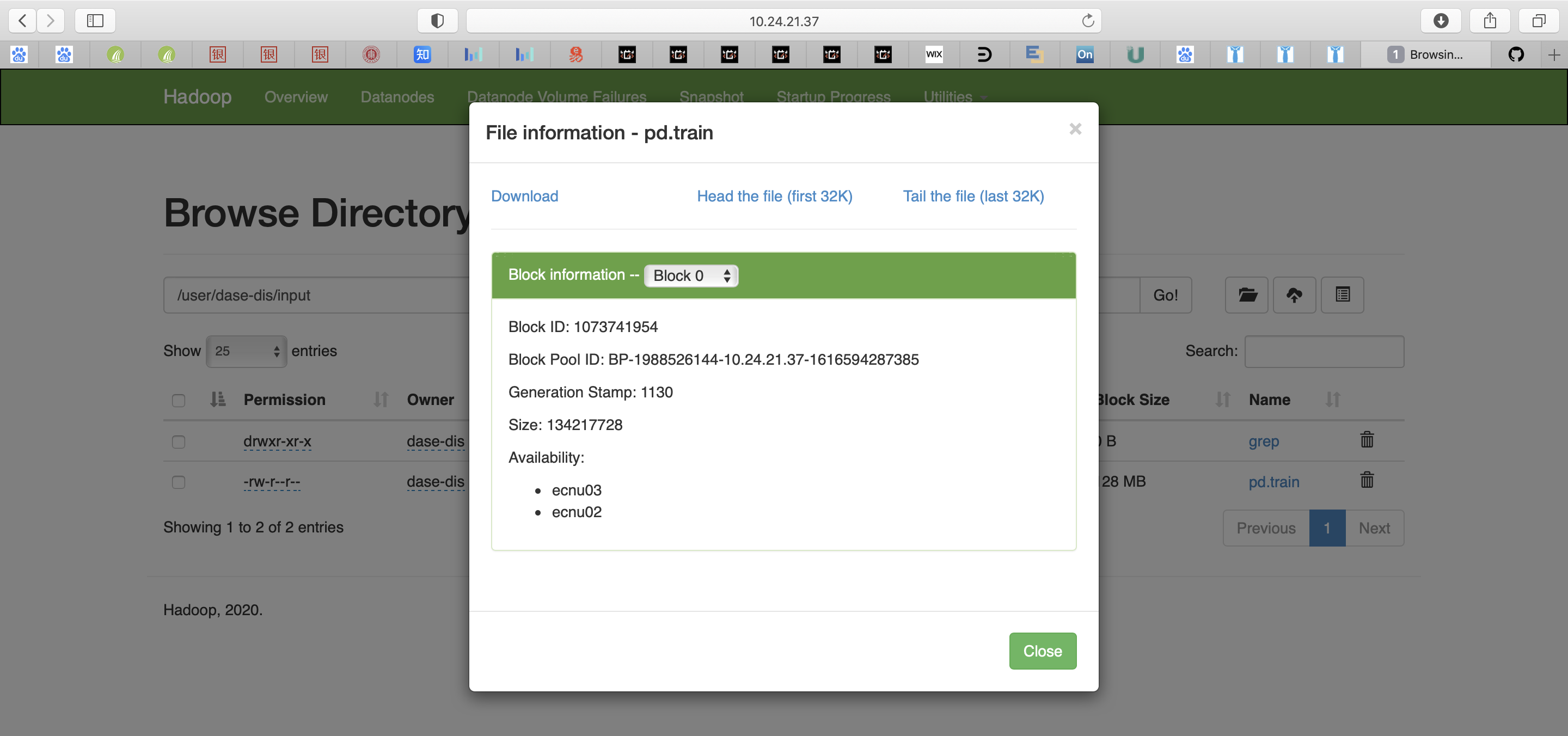
**在第一次打Jar包时，会出现main-class not found的报错，**

**(2)单机伪分布：**

**(3)分布式:**







1. **思考题**
2. **HDFS上每个文件块多大？请结合WebUI简要说明。**
3. **HDFS上每个文件的副本数如何制定?**
4. **总结**
5. **实验目的**

1. **实验任务**

1. **实验环境**

1. **实验步骤**

1. **思考题**
2. **总结**
3. **实验目的**

1. **实验任务**

1. **实验环境**

1. **实验步骤**

1. **思考题**
2. **总结**
3. **实验目的**

1. **实验任务**

1. **实验环境**

1. **实验步骤**

1. **思考题**
2. **总结**