0706日报-第十组

一．刘诗

工作总结

1. git报错

原因：系统升级之后Xcode shell 需要重新下载

1. 之前的爬虫程序报错

原因：报错信息为

ModuleNotFoundError: No module named 'douban.settings'

Settings。py 里的pipeline name设置出错，

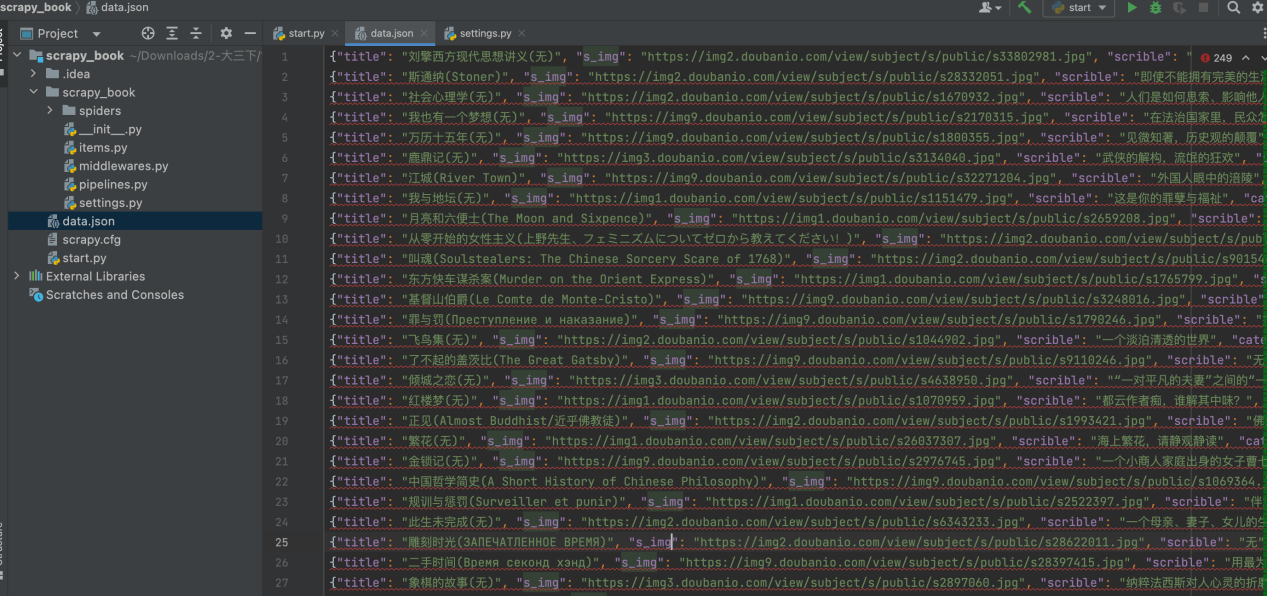
ITEM\_PIPELINES = {

'douban.pipelines.DoubanPipeline': 300}

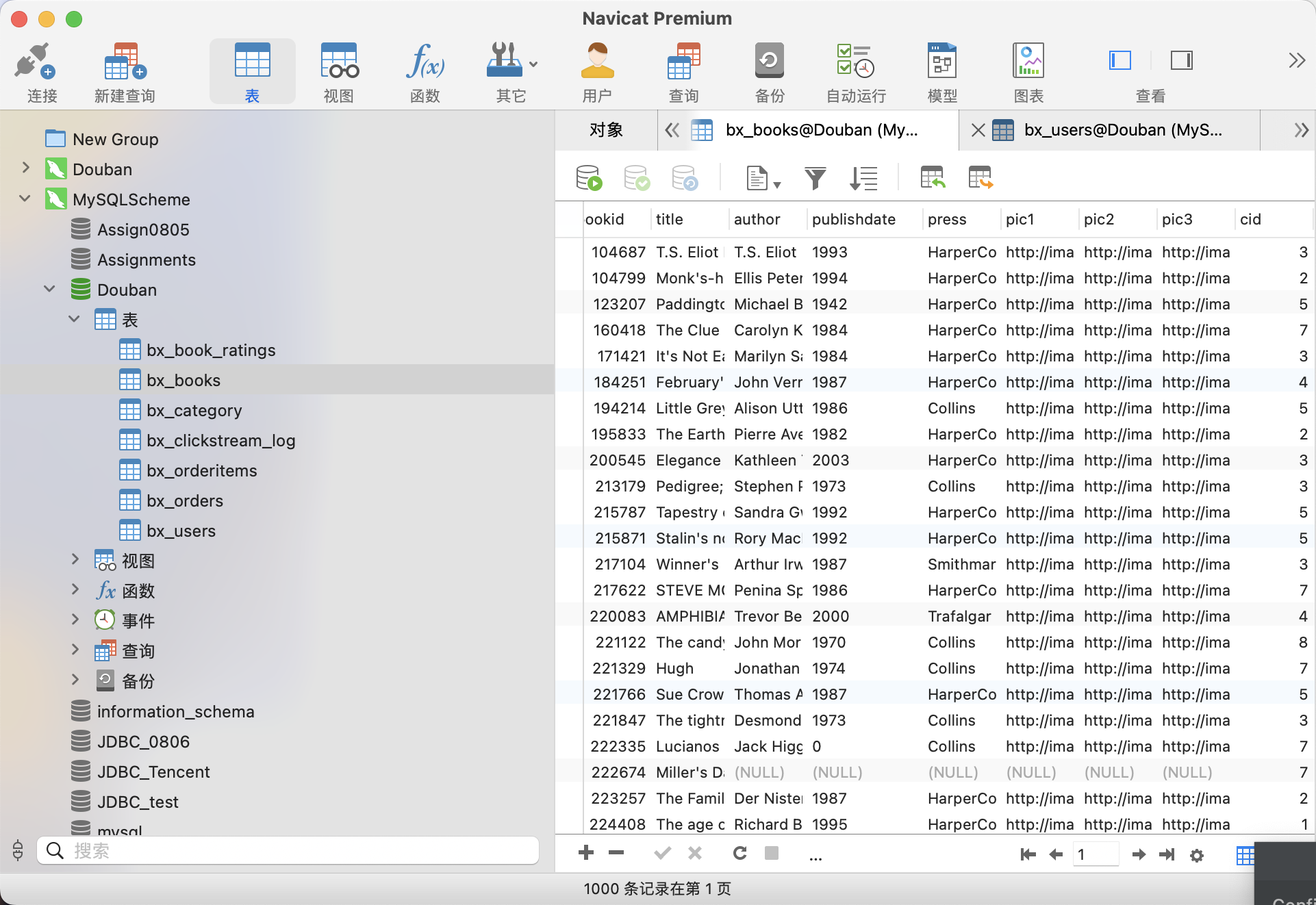
应该为

ITEM\_PIPELINES = {  
 'scrapy\_book.pipelines.ScrapyBookPipeline': 300,  
}

爬虫运行结果->

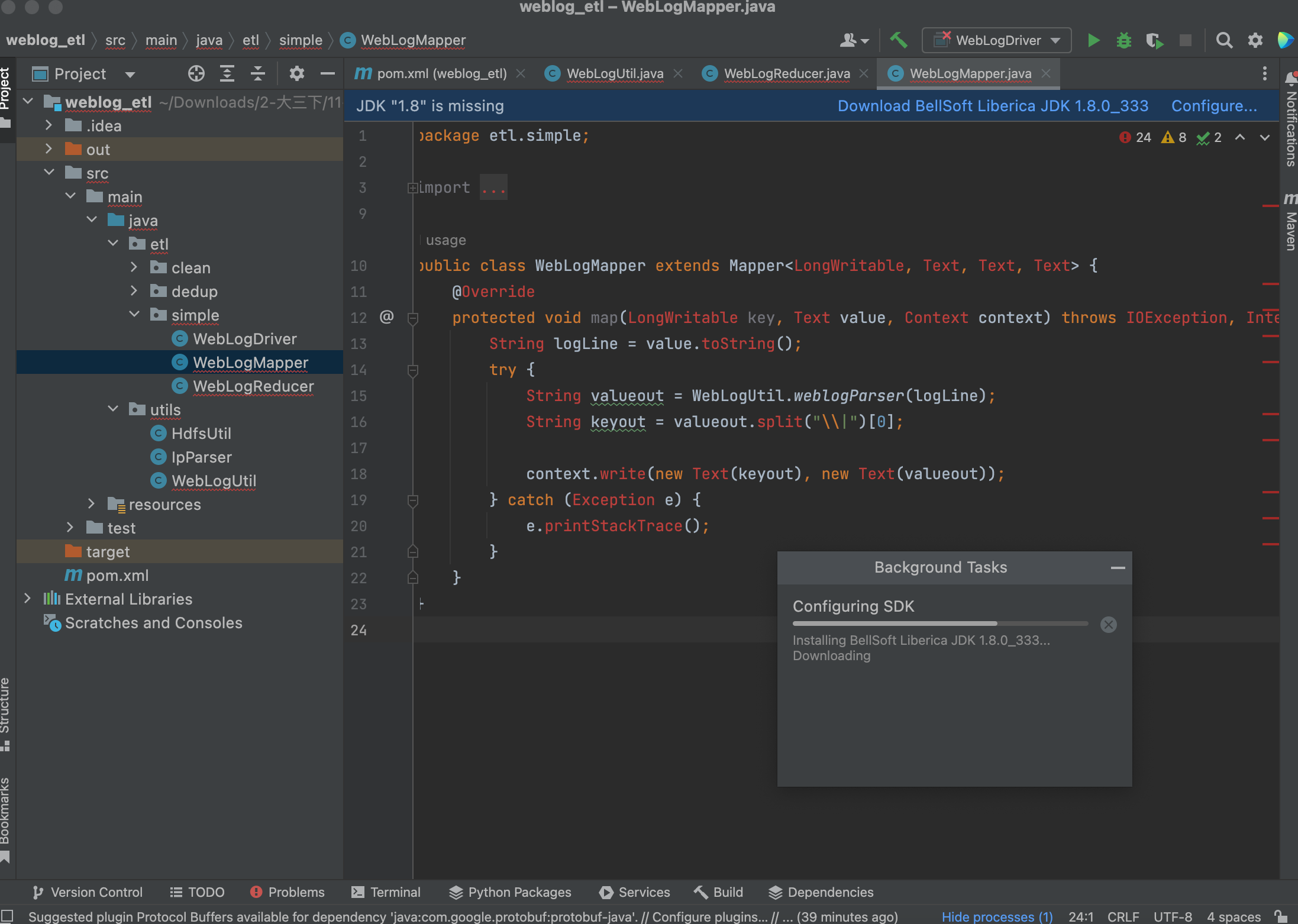


1. 将示例数据导入Mysql数据库



遇到的问题：csv文件没有表头字段名，第一行会丢失

1. 下载后续程序需要的JDK



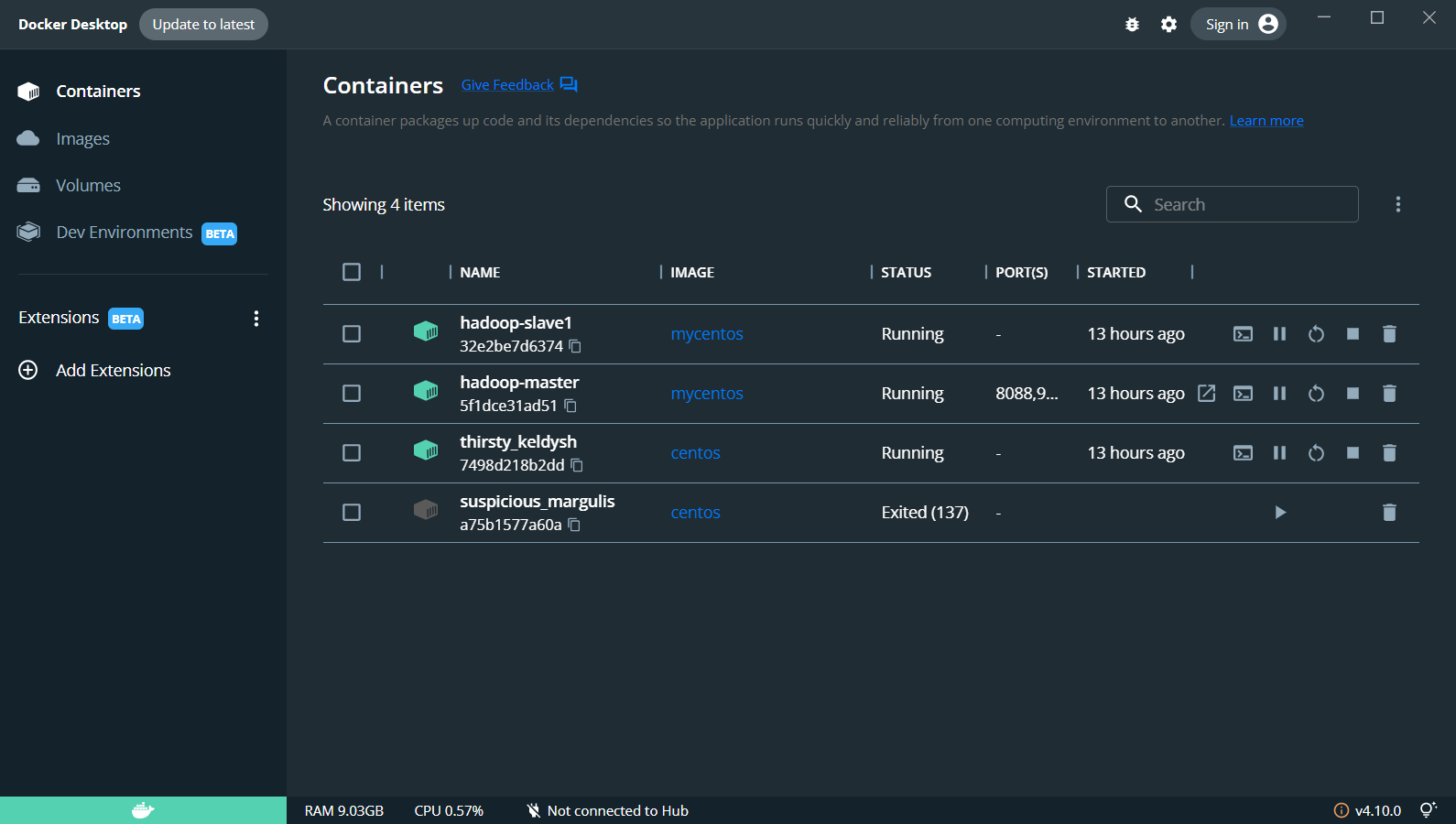
问题：速度挺慢的

1. 宋雨欣

今日总结：完成1-5实验，正在运行爬虫和数据处理。

遇到问题：打开docker Hadoop-slaver2容器消失，解决办法：重启电脑

明天计划。实现6-8实验



1. 罗宝泉
2. 完成的工作

本日解决了实验六的剩余问题及研究并开始实验七

1. 遇到的问题及解决方法

爬虫程序运行出现问题，在组长的指导下将ITEM\_PIPELINE中的内容修改后解决。

1. 明日工作

完成实验七及实验八

1. 任城仪

完成的工作：创建Scrapy爬虫项目，爬取豆瓣图书信息。

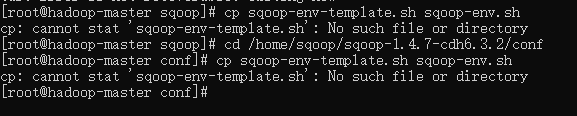
遇到的问题：访问太频繁IP地址被禁用。

接下来的计划：分析需求，爬取其他需要的数据。

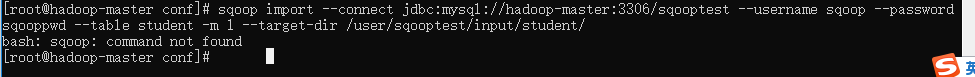
1. 潘炎灵
2. 完成的工作

实验3,4,5

1. 遇到的问题
   1. 找不到***sqoop-env-template.sh***



* 1. Sqoop安装失败



1. 解决方法
   1. 自己创建sqoop-env.sh
   2. 配置文件损坏，重新解压