0707日报-第十组

一．任城仪

正在进行的工作：

将本地数据导入mysql

使用sqoop将mysql数据导入hive

使用sqoop将hive数据导入mysql

遇到的问题：

Hadoop报错

下一步要进行的工作：

对导入hive的数据进行处理

1. 宋雨欣

宋雨欣7.7日报

完成的工作：爬虫，运行现有的代码，数据收集，日志数据清洗

计划明天：日志数据清洗

三． 潘炎灵7.7日报

完成的工作

撰写需求分析报告和中期答辩ppt

遇到的问题

有些需求不是很明确

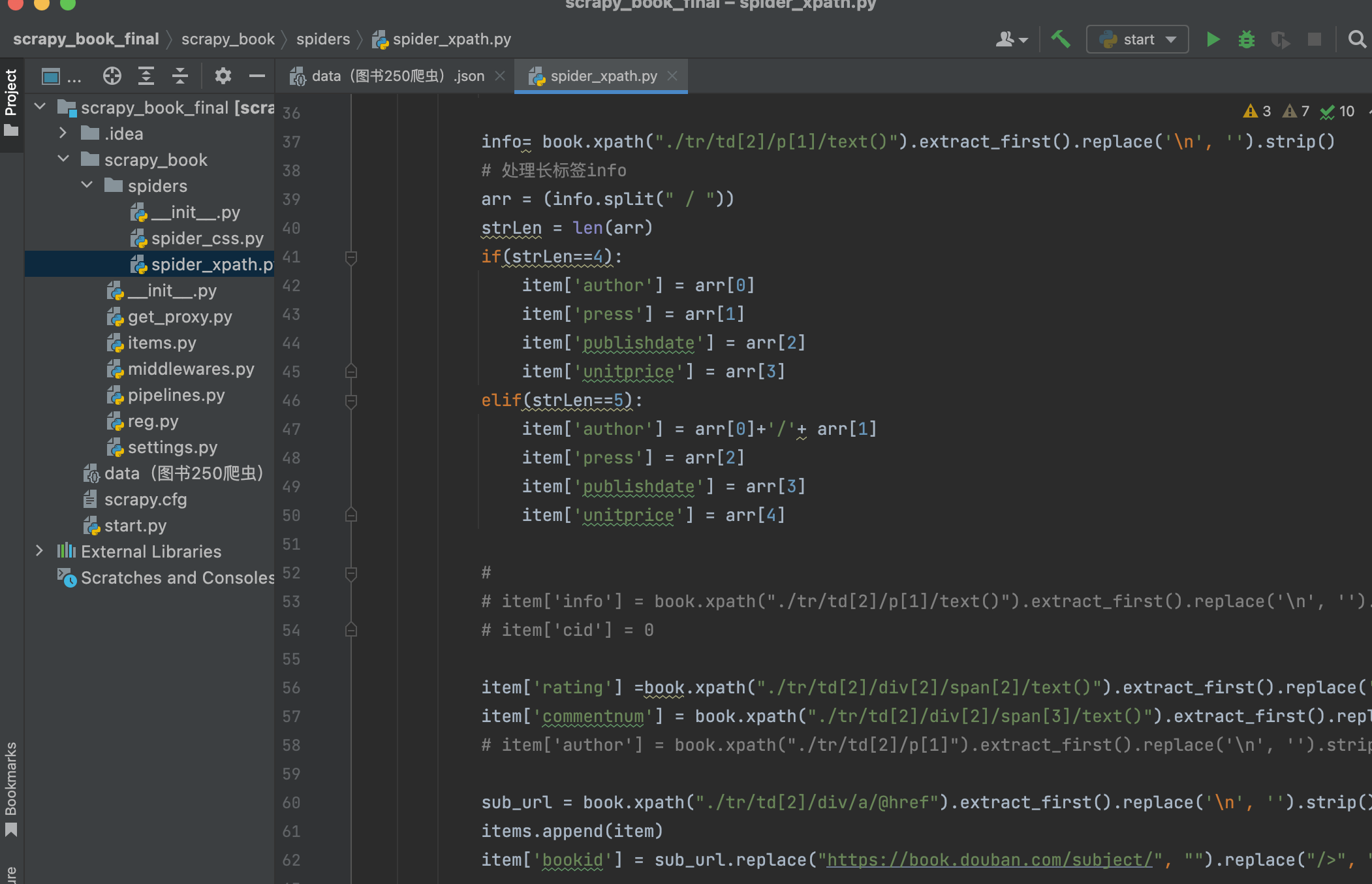
解决方法

进行需求调研

1. 刘诗

完成的工作：

1. Scrapy 爬虫程序的修改，得到与数据库设计相对应的更多字段， 最后爬到豆瓣top250图书信息表



2.得到data.json, 导入MySQL，转化出SQL脚本方便后续使用，

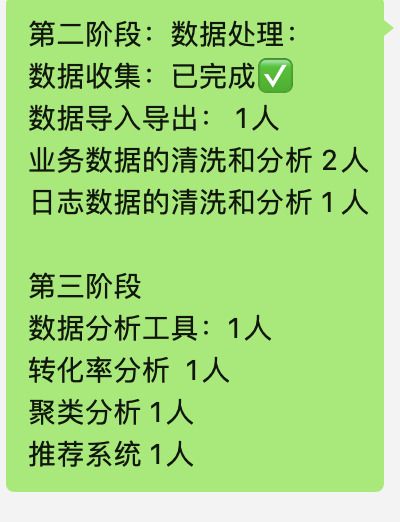


遇到的问题：

爬虫程序访问频繁被限制IP

解决办法：更换UA ，使用VPN走全局代理更换IP。

1. ✅将项目划分了第二、三四阶段， 将第二阶段的任务分工



1. 罗宝泉

运行了业务数据清洗模块的示例代码，阅读并修改相关程序