# 第三次大作业使用说明书

班级：18级软件工程1班 姓名：许航 学号：201830663360

**一、引言**

本文档用于指导用户了解本次大作业操作流程。

**二、系统配置**

软件：JDK 1.8、IntelliJ IDEA 2020.1.1(Commuity Editon)、flink dependencies.zip、scala-2.11.8

操作系统：windows 10

下载项目所包含的homework3工程导入IntelliJ IDEA中使用。

homework3工程需要在创建maven项目和配置maven环境且项目可以成功bulid后导入使用。

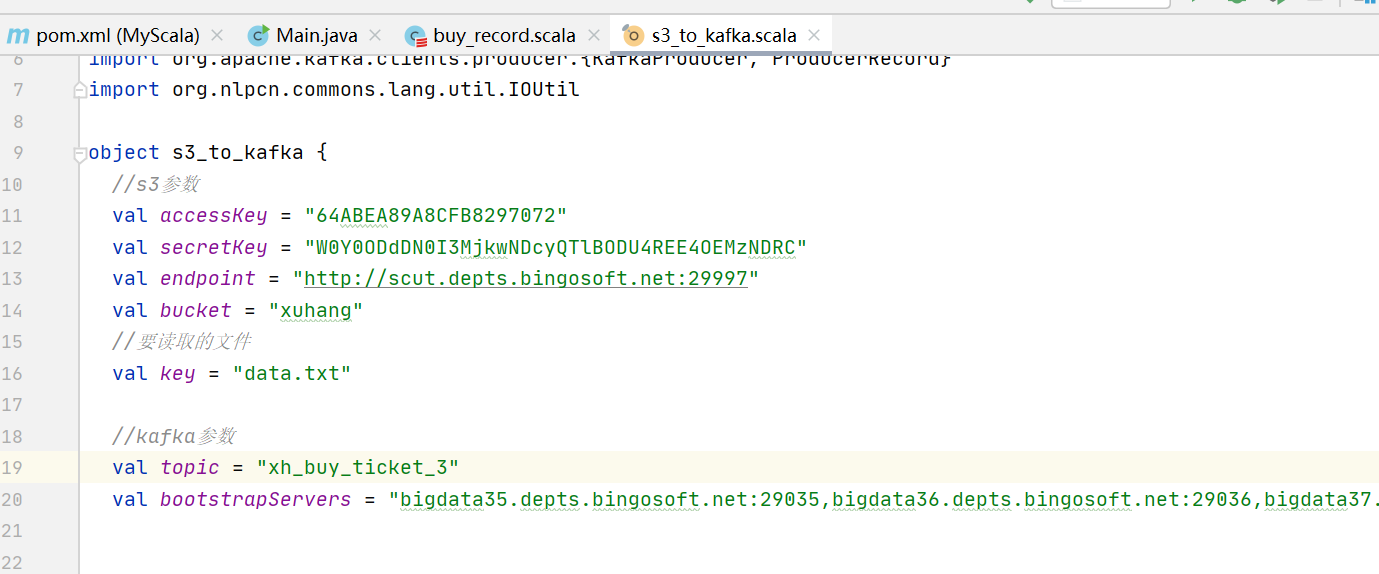
需要导入flink dependencies.zip中的依赖，homework3需要添加scala-sdk-2.11.8。

**三、操作流程**

**\*前置要求：需要配置好maven环境**

1、将s3数据接入到kafka生产者，通过flink程序计算归类后将计算结果保存于本地。

（1）打开maven项目homework3中修改s3\_to\_kafka.scala中的参数



1. //s3参数
2. val accessKey = "64ABEA89A8CFB8297072"
3. val secretKey = "W0Y0ODdDN0I3MjkwNDcyQTlBODU4REE4OEMzNDRC"
4. val endpoint = "http://scut.depts.bingosoft.net:29997"
5. val bucket = "xuhang"
6. //要读取的文件
7. val key = "data.txt"
9. //kafka参数
10. val topic = "xh\_buy\_ticket\_3"

s3参数修改为用户自己的授权信息

bucket为s3桶中需要读取的数据所存放桶名称

topic为想要生产数据的kafka topic（确保该topic无其他的内容）

data.txt为事先上传于s3桶中的文件（已将文件备份保存与该文档同一目录下）

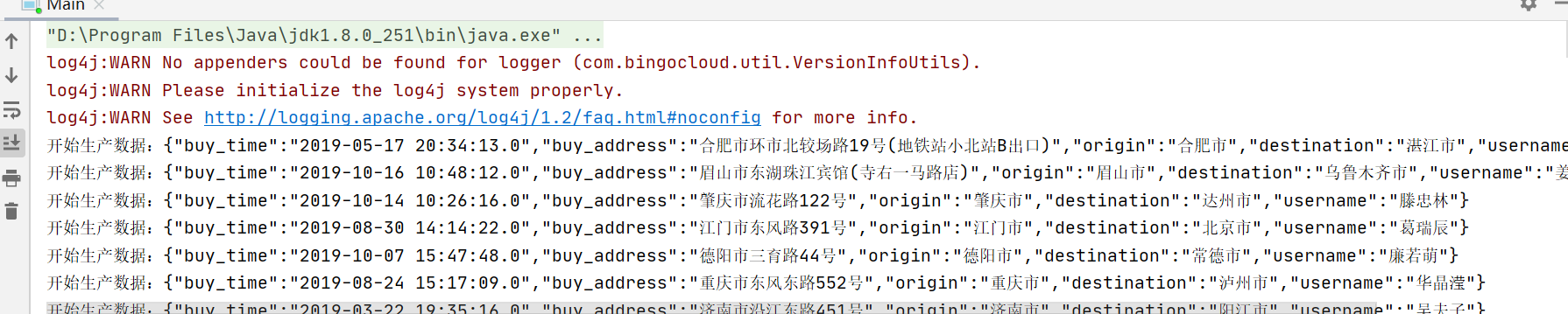
本使用说明所使用的data数据集为如下格式：



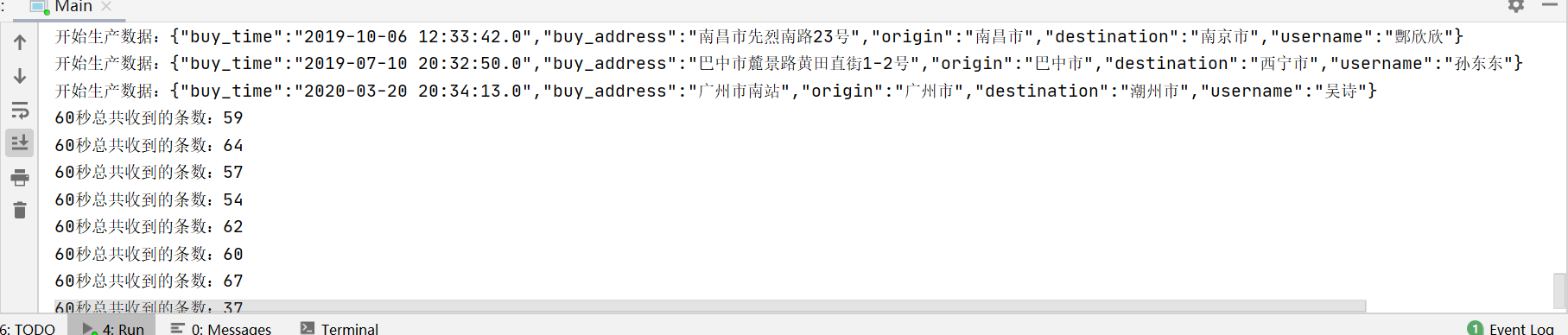
（2）修改Main.java的topic，修改为和上一步一致的topic，以确保消费的数据确实是s3导入的数据。



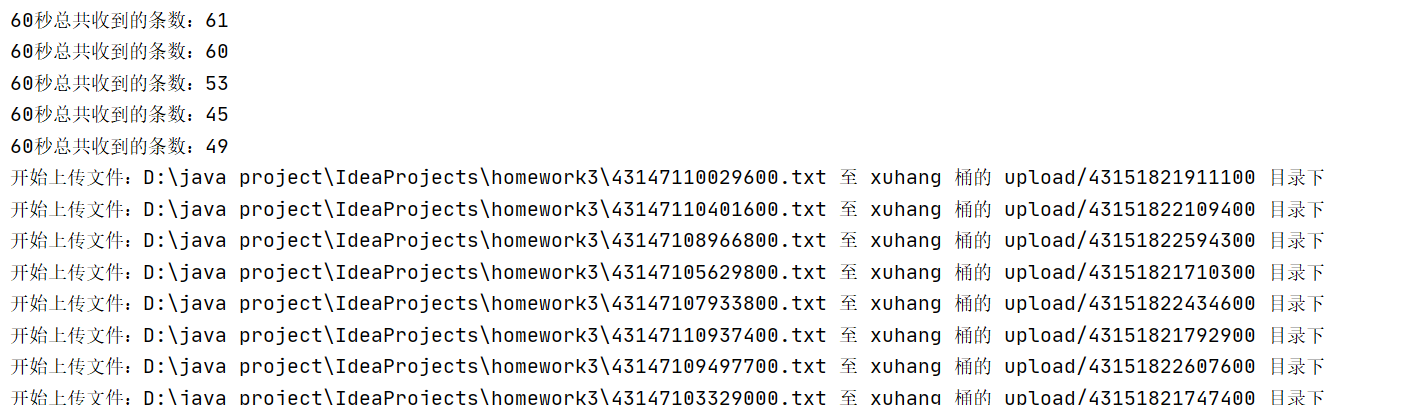
（3）运行Main.java



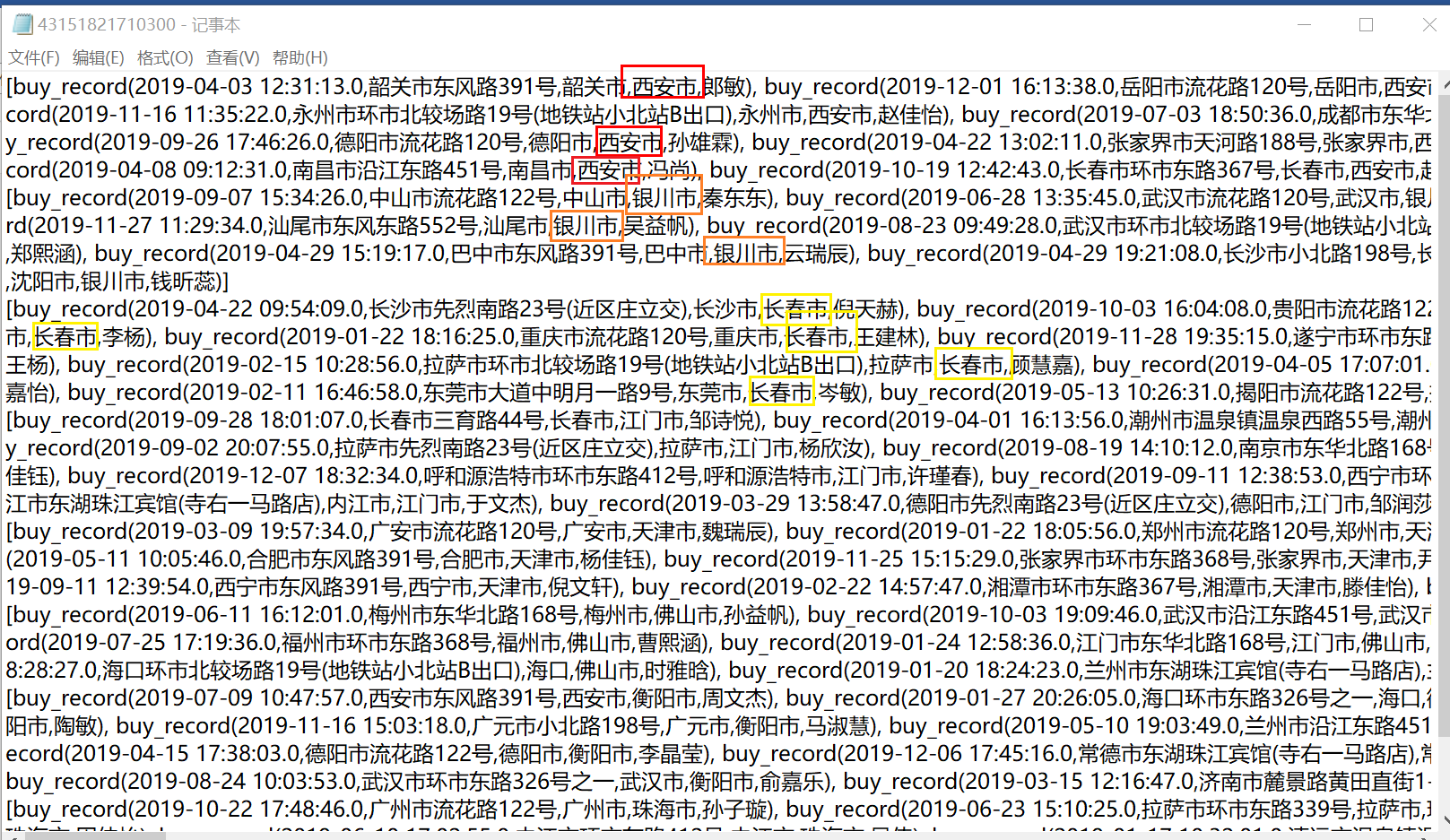
可以看到程序先将数据生产到相应kafka topic中。



然后flink开始分组归类后写入s3桶中。



到s3下载上传的文件检查分组的结果：



可以看到相同destination的购买记录被分在了同一个[]中。

输出8个文件的原因是flink根据keyby之后自动进行的8个分区为了方便其多线程进行运算操作。