**大数据实时计算使用说明**

## 环境准备

* Windows 环境
* jre 1.8 或以上
* Intellij IDEA
* Flink 依赖
* Scala 依赖

## 执行步骤

### 新建Module

1. **在IDEA中进行Module的创建，依次点击【File->New->Module】，在弹出的对话框左栏选择Java后单击Next，输入Module名称，本实操填写为【dataflow\_opt\_hw】**

电脑萤幕的截图

描述已自动生成

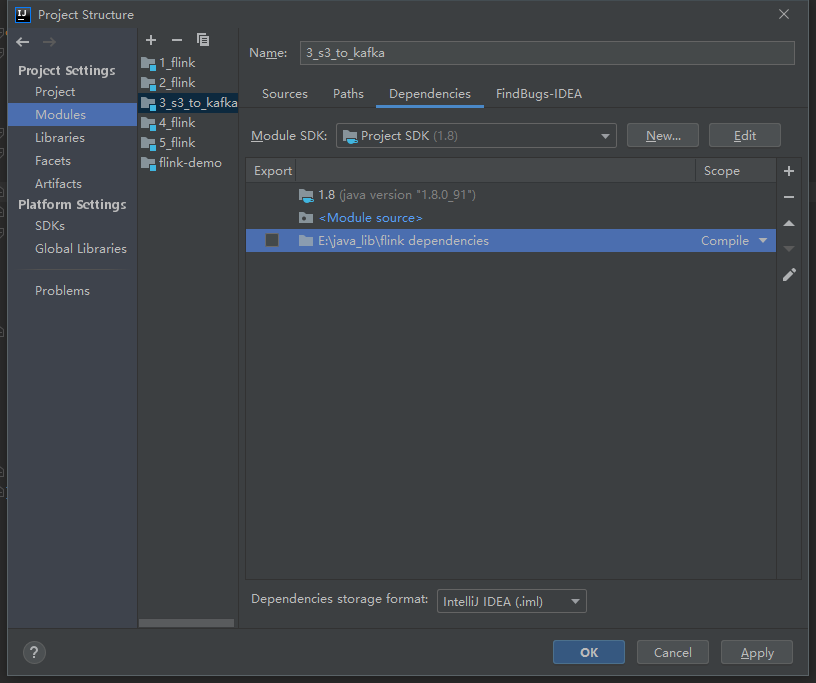
### 依赖引入

1. **右键选择【3\_s3\_to\_kafka module】，点击【Open Module Settings】，在【Dependencies】Tab页中单击“+”按钮，选择 Jars or directories**

截图里有图片

描述已自动生成

1. **在弹出的对话框中，选择前期解压后的【flink dependencies】文件夹路径后，单击OK**



1. **在【Project Structure】中，给module添加Scala SDK，右键选择3\_s3\_to\_kafka module，点击【Open Module Settings】，在左栏选择【Global Libraries】，右键选择【scala-sdk-2.11.8】，选择【Add to modules】**

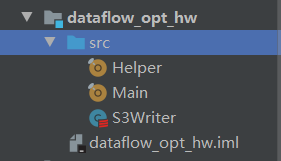
**社交网站的屏幕截图

描述已自动生成**

1. **在弹出的对话框中，选中dataflow\_opt\_hw后确定即可**
2. **至此，依赖引入完毕**

### 放入源文件

**A. 将三个源文件【Main.scala】【S3Writer.scala】【Helper.scala】放入 【dataflow\_opt\_hw\src】目录下，如下图**



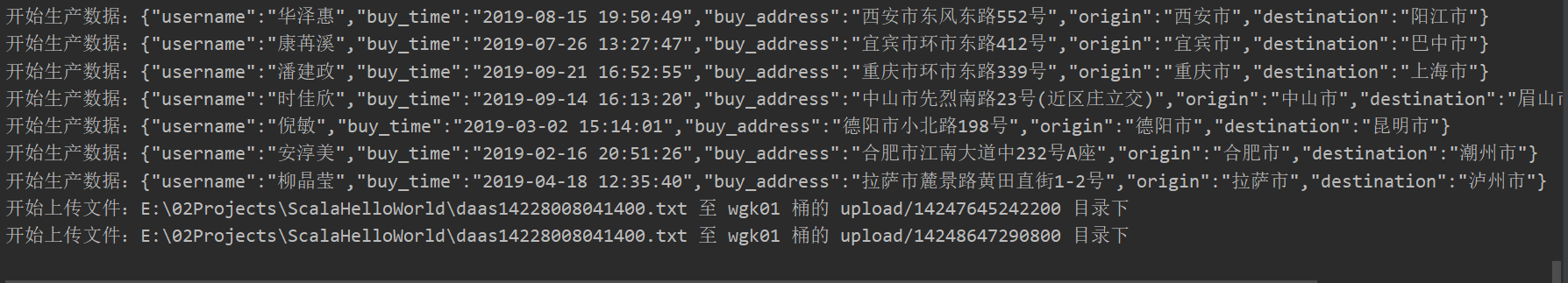
**B. 执行 Main 文件即可**

## 执行结果：

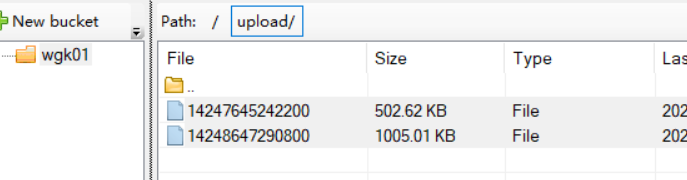
首先，程序会从 S3 读取 daas.txt 并传入 Kafka 的消息队列中。

然后，Flink 会从 Kafka 中作为消费者取出流数据

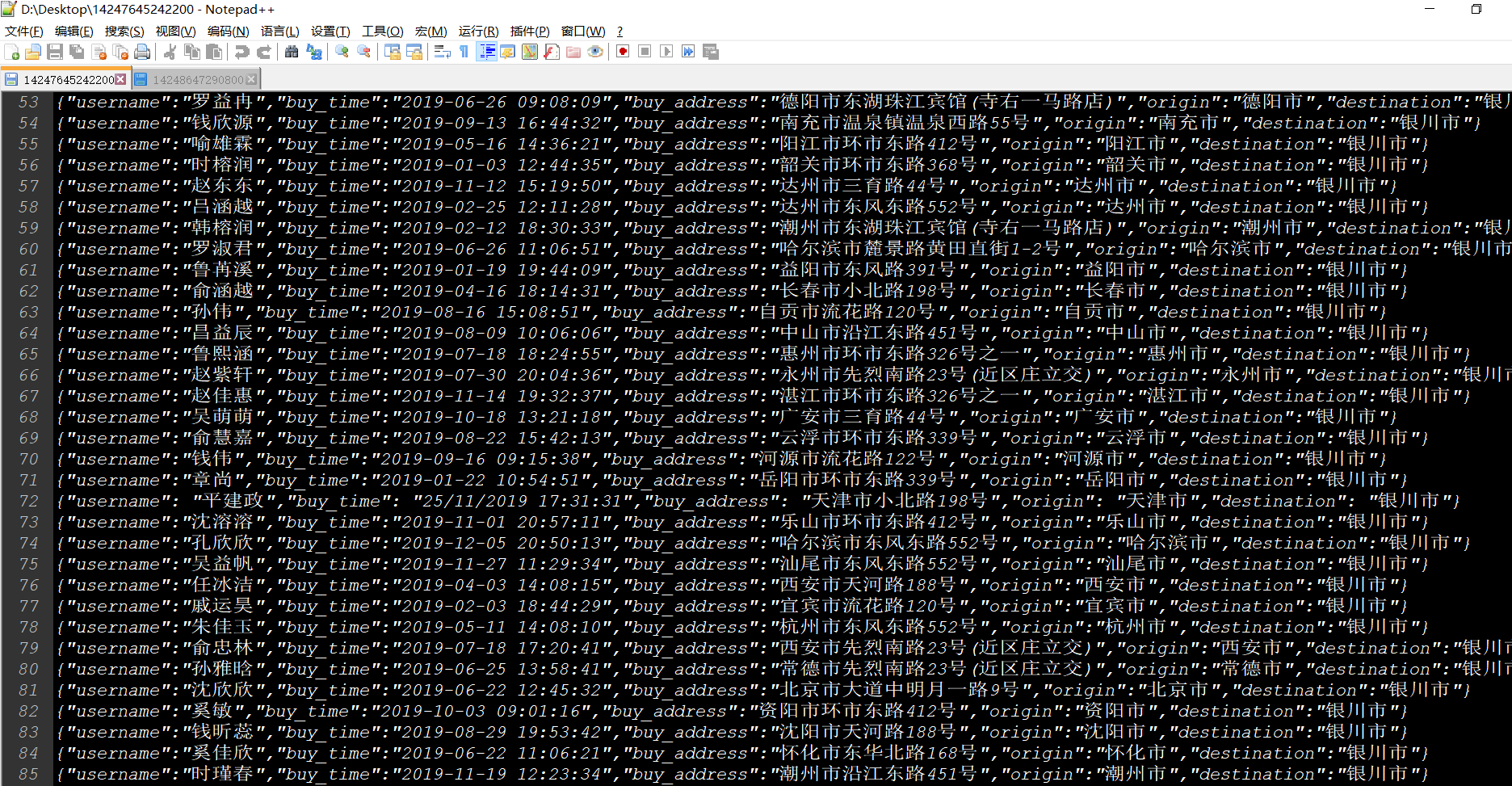
流数据最后在 S3Writer 文件中根据 destination 划分，并上传到 S3 的 upload/ 目录下。



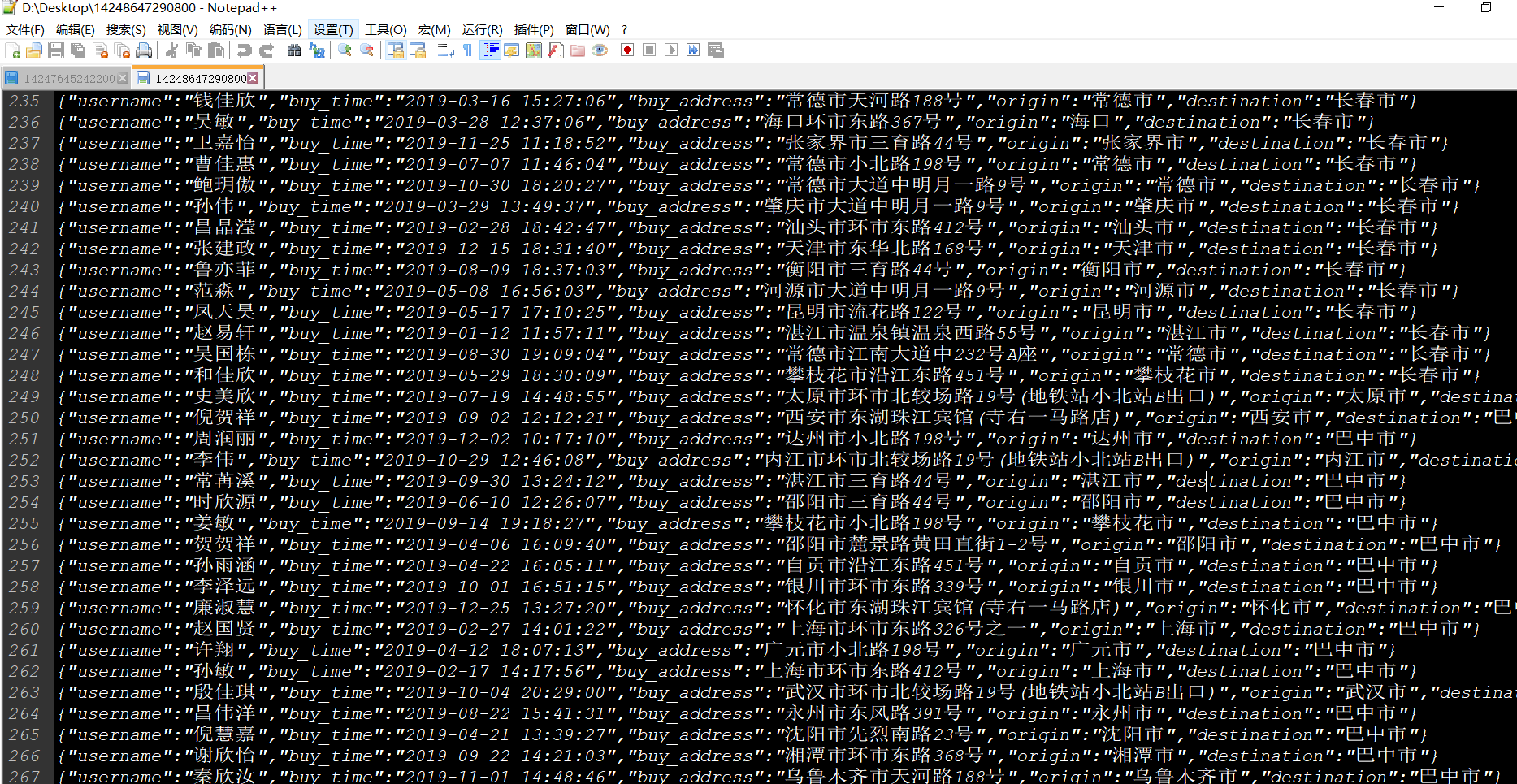
执行输出



上传到 S3 的文件



从 S3 下载的第一份文件



从 S3 下载的第二份文件