# 系统概述

## 目标

学习分析式消息系统kafka数据接入、实时流式计算数据运算，掌握实时流式计算结果数据落地

## 功能

* 开发一个kafka生产者，接入s3的数据
* 开发flink程序对接kafka的队列消息
* 对消息内容按关键字数组进行归类
* 将归类结果写入s3存储

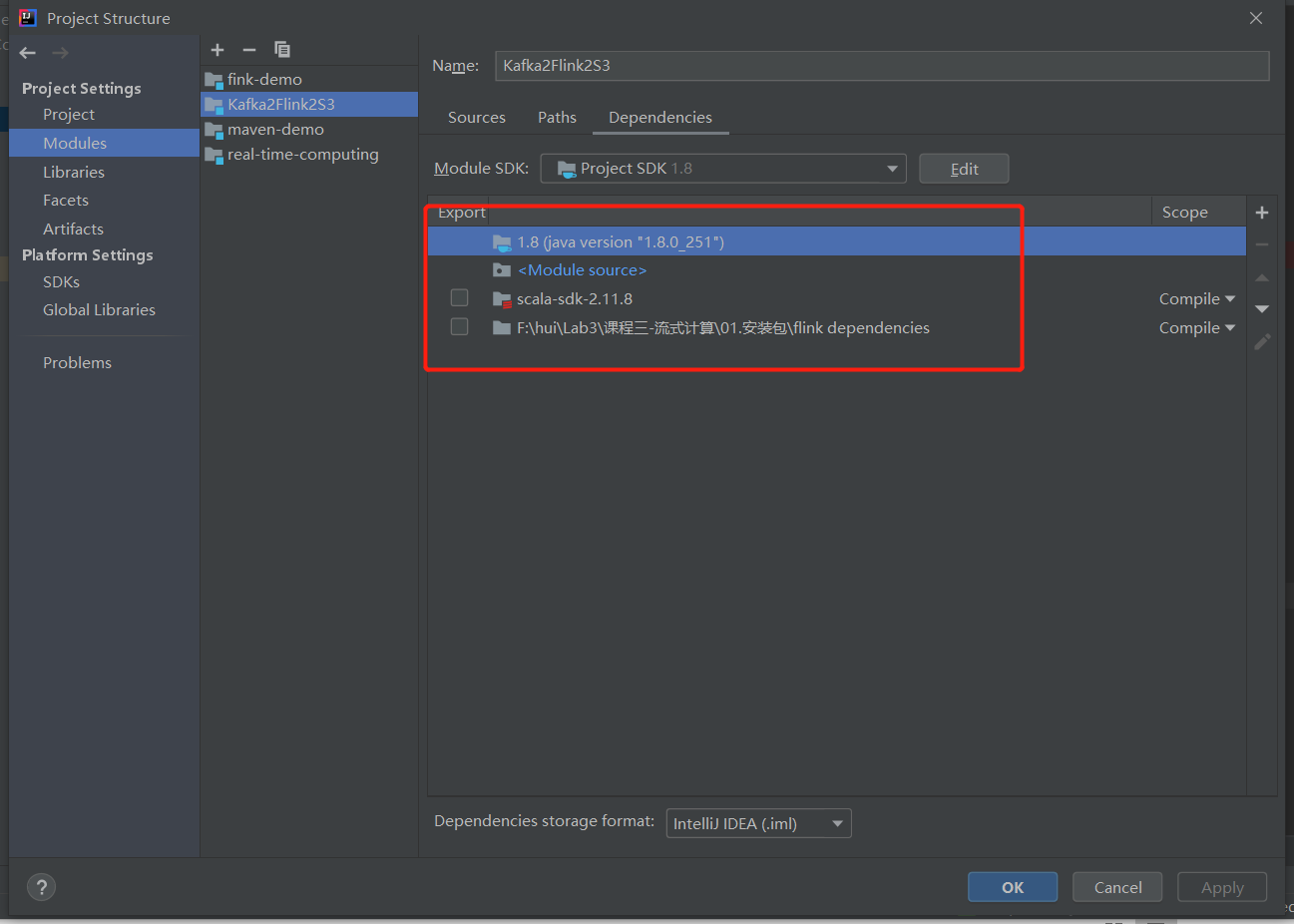
# 运行环境

## 2.1 运行前提

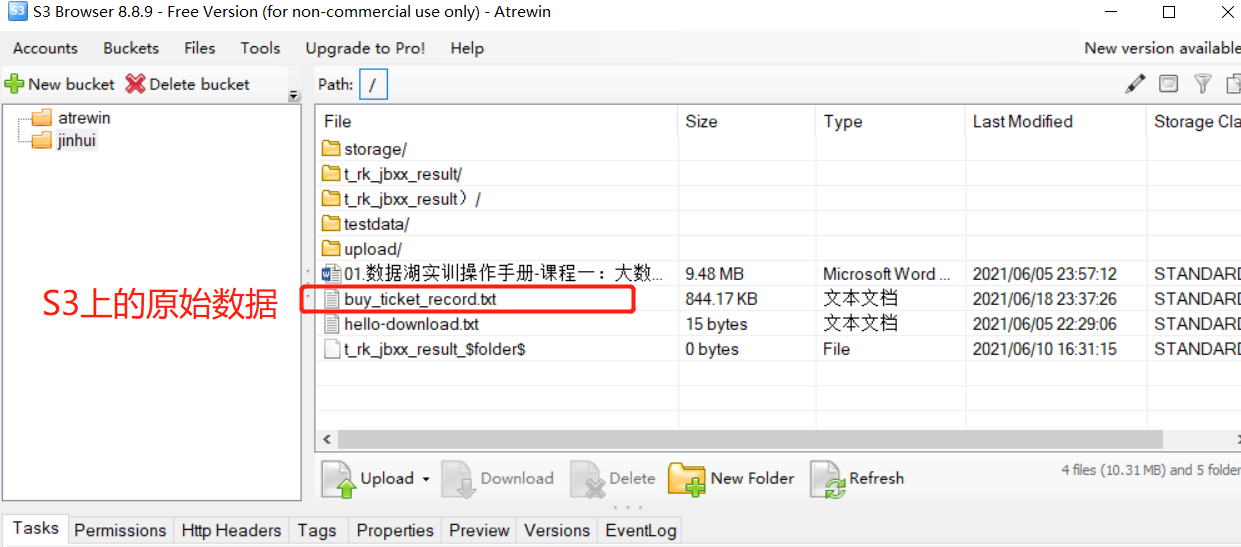
* 安装IntelliJ IDEA
* 安装JDK1.8
* IDEA安装Scala插件
* 安装Scala-2.11.8
* 下载开发包：flink dependencies.zip，

## 2.2 环境配置

1. 用IDEA　打开real-time-computing文件
2. 将为Kafka2Flink2S3Model添加依赖



1. 保证自己的S３的账号信息和S3中有待计算的原数据文件



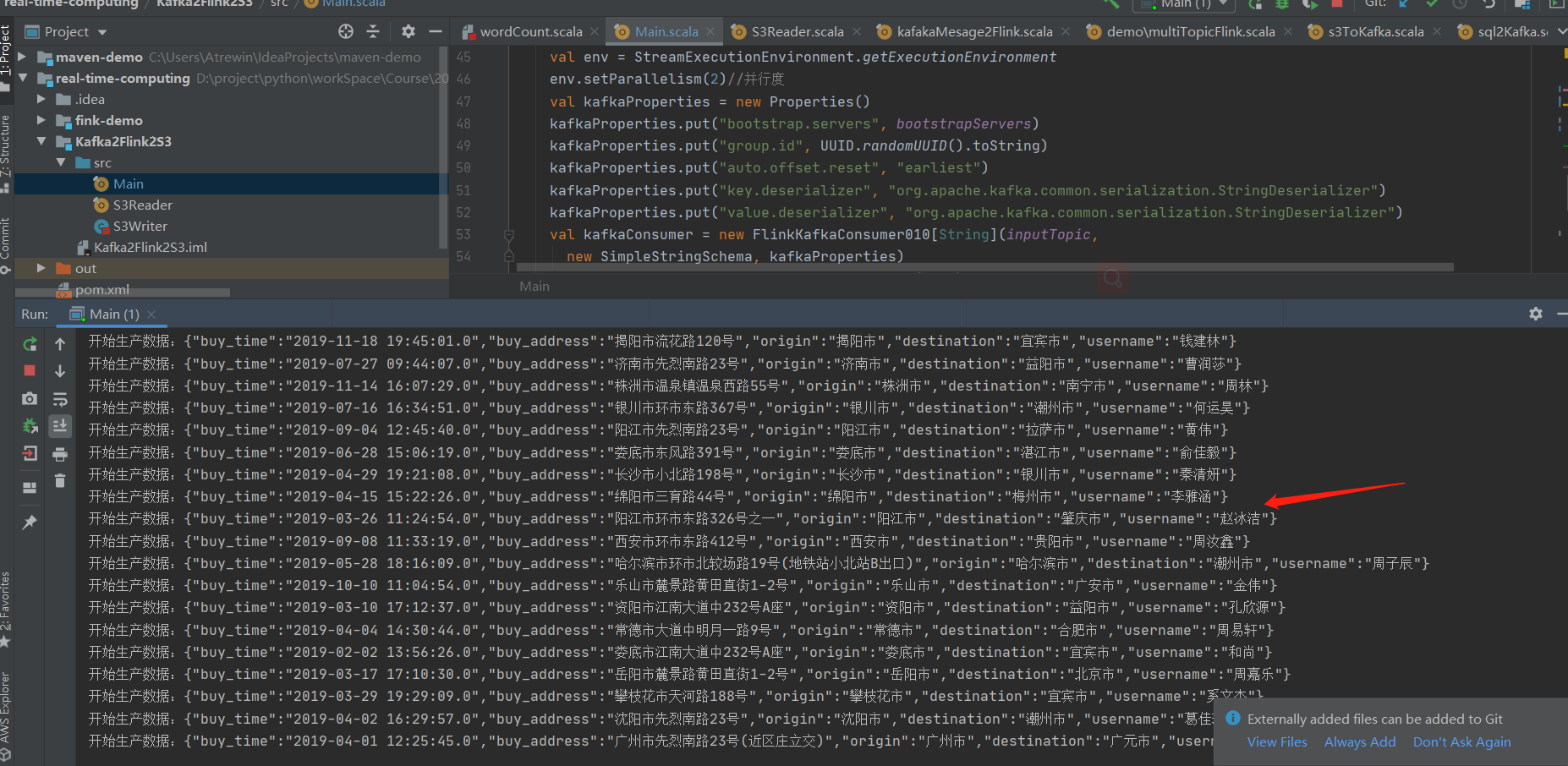
# 使用说明

大作业的工程文件夹是2021-Big-data-development\大数据实时计算\real-time-computing\Kafka2Flink2S3

1. 在Main文件中绑定授权信息
2. val *accessKey* = ""  
   val *secretKey* = ""  
     
   //s3地址  
   val *endpoint* = "scuts3.depts.bingosoft.net:29997"  
   //s3原始数据路径  
   val *key* = "buy\_ticket\_record.txt"  
   //计算经过保存到的桶  
   val *bucket* = "jinhui"  
   val *keyPrefix* = "upload/"  
   //上传文件名  
   val *filename* = "result"+System.*nanoTime*() + ".txt"  
   //上传数据间隔 单位毫秒  
   val *period* = 30000  
   //输入的kafka主题名称  
   val *inputTopic* = "mn\_buy\_ticket\_1\_jinhui"  
   //kafka地址  
   val *bootstrapServers* = "bigdata35.depts.bingosoft.net:29035,bigdata36.depts.bingosoft.net:29036,bigdata37.depts.bingosoft.net:29037"
3. 运行代码Main文件

4.经过

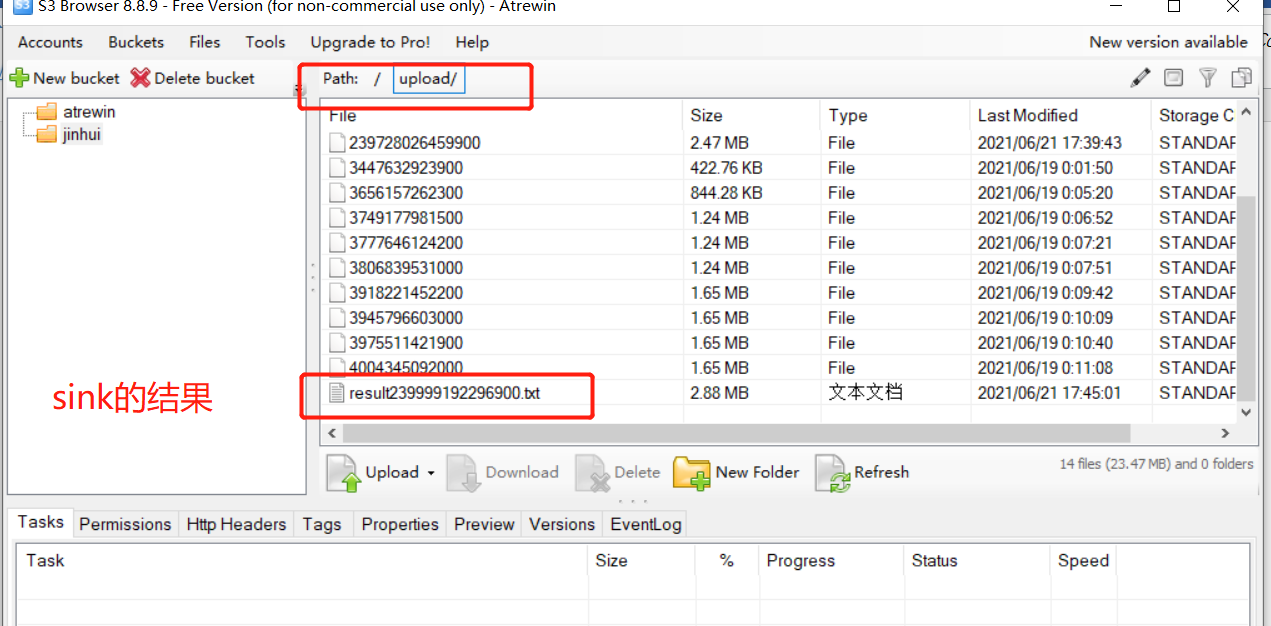
A：S3 的数据流入到Kafka的对应主题下。



B:flink 消费Kafka数据后经过沉淀到S3



C：查看分类经过



D：下载到本地查看

