

说明书及流程图

学号：201830220372

姓名：刘奕鑫

班级：2018级软件1班

功能点

1. 开发一个kafka生产者，接入s3的数据

- 流程图：

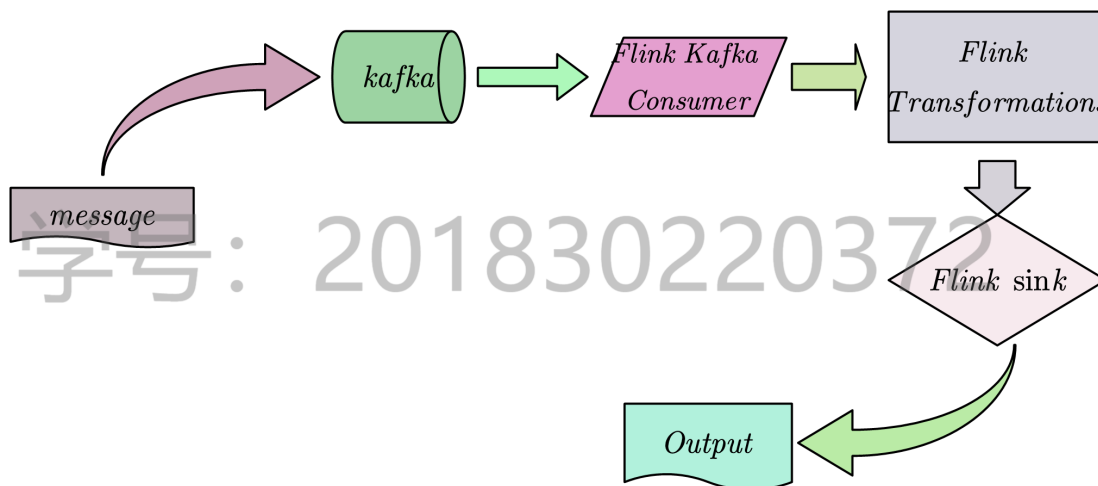


- 主文件：hw3/Main.scala
- 操作说明：程序会将s3的数据读入，然后数据流入kafka。
- 操作结果：

```
开始生产数据: {"buy_time": "2019-04-02 20:55:02.0", "buy_address": "达州市流花路122号", "origin": "达州市", "destination": "宜宾市", "username": "user"}
开始生产数据: {"buy_time": "2019-05-11 20:35:05.0", "buy_address": "汕尾市环市东路326号之一", "origin": "汕尾市", "destination": "达州市", "username": "user"}
开始生产数据: {"buy_time": "2019-07-18 18:24:55.0", "buy_address": "惠州市环市东路326号之一", "origin": "惠州市", "destination": "银川市", "username": "user"}
开始生产数据: {"buy_time": "2019-01-14 15:52:34.0", "buy_address": "郑州市麓景路黄田直街1-2号", "origin": "郑州市", "destination": "永州市", "username": "user"}
开始生产数据: {"buy_time": "2019-07-08 16:53:25.0", "buy_address": "内江市环市东路368号", "origin": "内江市", "destination": "攀枝花市", "username": "user"}
开始生产数据: {"buy_time": "2019-04-06 19:17:29.0", "buy_address": "武汉市沿江东路451号", "origin": "武汉市", "destination": "张家界市", "username": "user"}
开始生产数据: {"buy_time": "2019-08-25 16:18:55.0", "buy_address": "哈尔滨市流花路122号", "origin": "哈尔滨市", "destination": "衡阳市", "username": "user"}
开始生产数据: {"buy_time": "2019-09-13 14:52:20.0", "buy_address": "西宁市先烈南路23号(近区庄立交)", "origin": "西宁市", "destination": "贵阳市", "username": "user"}
开始生产数据: {"buy_time": "2019-10-06 12:33:42.0", "buy_address": "南昌市先烈南路23号", "origin": "南昌市", "destination": "南京市", "username": "user"}
开始生产数据: {"buy_time": "2019-07-10 20:32:50.0", "buy_address": "巴中市麓景路黄田直街1-2号", "origin": "巴中市", "destination": "西宁市", "username": "user"}
开始生产数据: {"buy_time": "2020-03-20 20:34:13.0", "buy_address": "广州市南站", "origin": "广州市", "destination": "潮州市", "username": "吴诗"}
```

2. 开发flink程序对接kafka的队列消息

- 流程图：



- 主文件：hw3/Main2.scala
- 操作说明：根据kafka地址获取文件内容

- 操作结果：

```
{ "buy_time": "2019-06-13 19:01:10.0", "buy_address": "西安市环市东路368号", "origin": "西安市", "destination": "张家界市", "username": "李佳琪" }
{ "buy_time": "2019-03-01 11:46:05.0", "buy_address": "长沙市沿江东路451号", "origin": "长沙市", "destination": "达州市", "username": "孙欣怡" }
{ "buy_time": "2019-06-09 19:24:33.0", "buy_address": "广元市环市东路326号之一", "origin": "广元市", "destination": "潮州市", "username": "李禹辰" }
{ "buy_time": "2019-04-06 19:42:37.0", "buy_address": "绵阳市麓景路黄田直街1-2号", "origin": "绵阳市", "destination": "长春市", "username": "葛欣" }
{ "buy_time": "2019-05-20 19:58:55.0", "buy_address": "佛山市东湖珠江宾馆(亭右一马路店)", "origin": "佛山市", "destination": "银川市", "username": "李俊" }
{ "buy_time": "2019-01-28 20:49:38.0", "buy_address": "揭阳市环市东路367号", "origin": "揭阳市", "destination": "益阳市", "username": "钱佳欣" }
{ "buy_time": "2019-11-24 14:52:04.0", "buy_address": "兰州市三育路44号", "origin": "兰州市", "destination": "乐山市", "username": "钱东东" }
{ "buy_time": "2019-04-02 20:55:02.0", "buy_address": "达州市流花路122号", "origin": "达州市", "destination": "宜宾市", "username": "周天昊" }
{ "buy_time": "2019-07-18 18:24:55.0", "buy_address": "惠州市环市东路326号之一", "origin": "惠州市", "destination": "银川市", "username": "鲁熙涵" }
{ "buy_time": "2019-07-08 16:53:25.0", "buy_address": "内江市环市东路368号", "origin": "内江市", "destination": "攀枝花市", "username": "卞国栋" }
{ "buy_time": "2019-08-25 16:18:55.0", "buy_address": "哈尔滨市流花路122号", "origin": "哈尔滨市", "destination": "衡阳市", "username": "孙建政" }
{ "buy_time": "2019-10-06 12:33:42.0", "buy_address": "南昌市先烈南路23号", "origin": "南昌市", "destination": "南京市", "username": "鄢欣欣" }
{ "buy_time": "2020-03-20 20:34:13.0", "buy_address": "广州市南站", "origin": "广州市", "destination": "潮州市", "username": "吴诗" }
```

3. 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

- 主文件: hw3/Main3.scala
- 操作说明: [origin="济南市"]进行归类分组
- 操作结果：

```
{ "value": { "buy_time": "2019-10-05 12:00:08.0", "buy_address": "济南市天河路188号", "origin": "济南市", "destination": "珠海市", "username": "李俊" } }
{ "value": { "buy_time": "2019-09-06 17:49:11.0", "buy_address": "济南市江南大道中232号A座", "origin": "济南市", "destination": "茂名市", "username": "李俊" } }
{ "value": { "buy_time": "2019-12-09 19:34:46.0", "buy_address": "济南市东风路391号", "origin": "济南市", "destination": "乌鲁木齐市", "username": "李俊" } }
{ "value": { "buy_time": "2019-05-01 11:46:04.0", "buy_address": "济南市环市东路326号之一", "origin": "济南市", "destination": "益阳市", "username": "李俊" } }
{ "value": { "buy_time": "2019-07-05 16:37:21.0", "buy_address": "济南市沿江东路451号", "origin": "济南市", "destination": "攀枝花市", "username": "李俊" } }
{ "value": { "buy_time": "2019-08-30 13:12:34.0", "buy_address": "济南市环市东路368号", "origin": "济南市", "destination": "东莞市", "username": "李俊" } }
{ "value": { "buy_time": "2019-01-10 17:36:45.0", "buy_address": "济南市环市北路较场路19号(地铁站小北站B出口)", "origin": "济南市", "destination": "济南市", "username": "李俊" } }
{ "value": { "buy_time": "2019-08-30 13:11:16.0", "buy_address": "济南市环市东路326号之一", "origin": "济南市", "destination": "湘潭市", "username": "李俊" } }
{ "value": { "buy_time": "2019-09-10 19:45:50.0", "buy_address": "济南市环市北路较场路19号(地铁站小北站B出口)", "origin": "济南市", "destination": "济南市", "username": "李俊" } }
{ "value": { "buy_time": "2019-04-02 15:12:26.0", "buy_address": "济南市麓景路黄田直街1-2号", "origin": "济南市", "destination": "郴州市", "username": "李俊" } }
{ "value": { "buy_time": "2019-08-26 10:46:51.0", "buy_address": "济南市环市东路367号", "origin": "济南市", "destination": "绵阳市", "username": "李俊" } }
{ "value": { "buy_time": "2019-10-20 13:34:36.0", "buy_address": "济南市温泉镇温泉西路55号", "origin": "济南市", "destination": "内江市", "username": "李俊" } }
```