文本

描述已自动生成

# 基于kafka的流式计算开发使用说明

姓名：李琰朕

学号：201830661250

班级：软件三班

年级：2018级

指导教师：汤峰

报告日期：2021年6月20日

# 环境配置：

* Java==1.8
* Flink依赖
* Aws开发依赖
* Scala开发依赖

# 软件配置：

1. 配置个人topic：

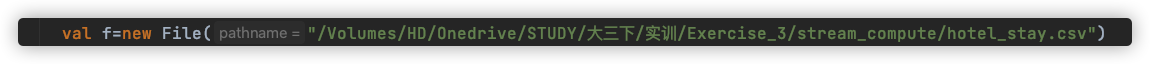
图形用户界面, 文本, 应用程序

描述已自动生成如上图所示，需要将producer和consumer中的topic改为个人topic

1. 文本

   描述已自动生成配置个人S3存储设置：

如上图所示，需要将producer中的S3设置改为自己的密钥和和账户以及存储桶的名称和要读取的文件

1. 配置从consumer中读取的消息在本地的保存路径：

# 软件使用：

1. 先运行producer对象，获得如下输出：

文本

描述已自动生成

代表kafka生产者已经将S3中的数据发送到Broker

1. 随后运行Consumer对象，获得如下输出：

文本

描述已自动生成

随后我们可以打开目录查看result.csv文件中是否写入了数据：

表格

描述已自动生成

1. 运行main函数，会调用Filter和Uploader两个程序来完成最后的数据分片和上传：

图形用户界面, 文本

描述已自动生成

可以看到程序已经将数据按照月份进行分片并上传到S3的Upload文件夹下：

图形用户界面, 应用程序, Word

描述已自动生成