使用说明书

1. **运行环境**

* **安装S3 Browser**
* **Scala sdk-2.11.8**
* **Flink dependencies**
* **Kafka bootstrapServers , topics**

1. **运行前的准备**

先将代码中的S3登录信息以及kafkaserver配置信息补充完整（accessKey、secretKey、endpoint、bootstrapServers）。选择合适的正确的topic或者上传文件到S3中后进行生产数据到kafka中。

1. **运行时**

Flink接入kafka消息队列后，可以自动实时接收kafka的数据，进行实时流式计算数据运算，对消息内容进行按关键字归类，时间任务会确保在一段时间内没有新数据到来时会将文件流关闭进行保存归类结果，在有新数据到来后，程序对其进行分类，会记录下其归类结果所对应的文件名，之后再将对应的被修改后的结果文件同步写入S3中存储落地，避免了将其他未修改的文件也上传，增加了时延。

1. **运行结束后**

可以在本地项目文件夹data下看到分类保存后的结果。在S3桶save中查看结果。

1. **流程图**

