Flink入门-1

今天开始玩Flink了。首先按照官方文档编写测试程序。

首先每一步用mvn生成代码模板。

mvn archetype:generate -DarchetypeGroupId=org.apache.flink -DarchetypeArtifactId=flink-quickstart-java -DarchetypeVersion=1.7.0

直接用cmd来建千万不要使用powershell。别问我为什么。建好的结构如下所示：

Folder PATH listing

Volume serial number is 9270-7902

E:.

└───flink-test

│ pom.xml

│

└───src

└───main

├───java

│ └───com

│ └───zl

│ └───test

│ └───flink

│ BatchJob.java

│ StreamingJob.java

│

└───resources

log4j.properties

不废话，先编一下看看效果。flink-test>mvn clean package。看到下面的提示，觉得心情还不错。

[INFO] BUILD SUCCESS

我们稍微改一下pox文件。因为默认是stream，但是我觉得还是batch比较适合入门。

pom.xml中改一下下面这句：

<mainClass>com.zl.test.flink.StreamingJob</mainClass>

**<mainClass>com.zl.test.flink.BatchJob</mainClass>**

然后再编一下。

然后先运行一下。

./bin/flink run flink-test-1.0-SNAPSHOT.jar

然后，报错了。错误信息如下，原因是模板不能直接运行。

Caused by: java.lang.RuntimeException: No data sinks have been created yet. A program needs at least one sink that consumes data. Examples are writing the data set or printing it.

那我们稍做改造吧。

Flink的程序思路就是读、算、写。这里面的算比较麻烦。我们先去掉。修改一下源码如下：

final ExecutionEnvironment env = ExecutionEnvironment.getExecutionEnvironment();

//读数据

DataSet<String> text = env.readTextFile("/home/test/flink-1.6.0/README.txt");

//写结果

text.writeAsText("/home/test/flink-1.6.0/README1.txt");

// execute program

env.execute("Flink Batch Java API Skeleton");

行了。再执行一下。得到下面的信息。就证明执行成功了。

[test@TOGGSERAP01 flink-1.6.0]$ ./bin/flink run flink-test-1.0-SNAPSHOT.jar

Starting execution of program

Program execution finished

Job with JobID 8e8143c962913a389cd0e06580fafb99 has finished.

Job Runtime: 177 ms

看一下结果。README1文件也生成了。好了。现到这吧。下次再接着写。希望这个能对大家有点帮助。刚刚开始记录学习。文档写得不好。谢谢。