

华东师范大学数据科学与工程学院实验报告

课程名称： 分布式模型与编程

年级： 2016 级

上机实践成绩：

指导教师： 徐辰

姓名： 张宏伟

上机实践名称： Flink 编程

学号： 10165101180

上机实践日期：

上机实践编号： 13

组号：

上机实践时间：

一、实验目的

利用 Flink 实现可扩展的数据分析

二、实验任务

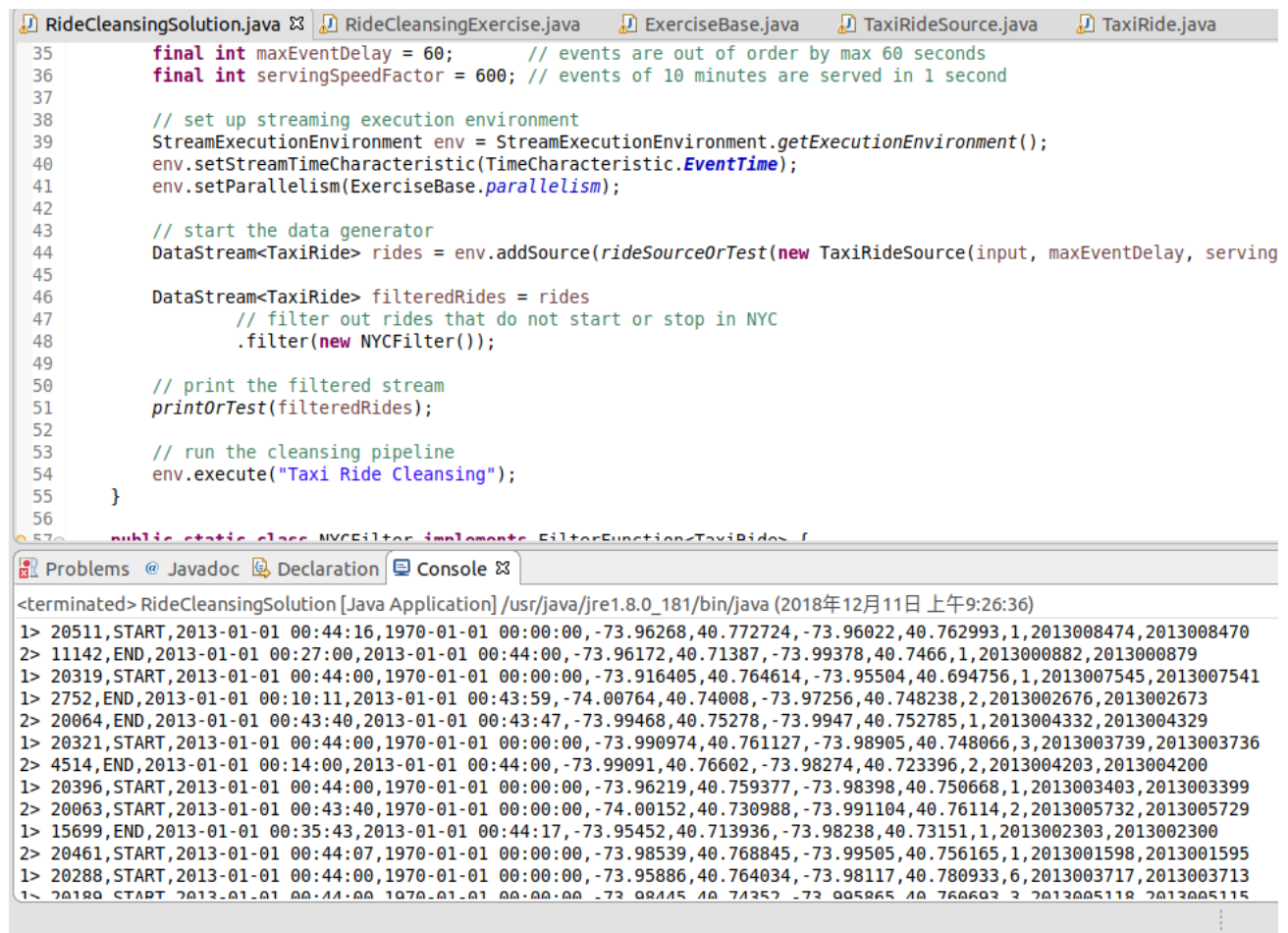
使用官方教程提供的两个数据集， 学习 flink 编程的接口

三、使用环境

Jdk1.8+flink

四、实验过程

运行官方完整实现的 solution 类，筛选出起始点在纽约的行程。



```
RideCleansingSolution.java RideCleansingExercise.java ExerciseBase.java TaxiRideSource.java TaxiRide.java
35     final int maxEventDelay = 60; // events are out of order by max 60 seconds
36     final int servingSpeedFactor = 600; // events of 10 minutes are served in 1 second
37
38     // set up streaming execution environment
39     StreamExecutionEnvironment env = StreamExecutionEnvironment.getExecutionEnvironment();
40     env.setStreamTimeCharacteristic(TimeCharacteristic.EventTime);
41     env.setParallelism(ExerciseBase.parallelism);
42
43     // start the data generator
44     DataStream<TaxiRide> rides = env.addSource(rideSourceOrTest(new TaxiRideSource(input, maxEventDelay, serving
45
46     DataStream<TaxiRide> filteredRides = rides
47         // filter out rides that do not start or stop in NYC
48         .filter(new NYCFilter());
49
50     // print the filtered stream
51     printOrTest(filteredRides);
52
53     // run the cleansing pipeline
54     env.execute("Taxi Ride Cleansing");
55 }
56
57 public static class NYCFilter implements FilterFunction<TaxiRides {
<terminated> RideCleansingSolution [Java Application] /usr/java/jre1.8.0_181/bin/java (2018年12月11日 上午9:26:36)
1> 20511,START,2013-01-01 00:44:16,1970-01-01 00:00:00,-73.96268,40.772724,-73.96022,40.762993,1,2013008474,2013008470
2> 11142,END,2013-01-01 00:27:00,2013-01-01 00:44:00,-73.96172,40.71387,-73.99378,40.7466,1,2013000882,2013000879
1> 20319,START,2013-01-01 00:44:00,1970-01-01 00:00:00,-73.916405,40.764614,-73.95504,40.694756,1,2013007545,2013007541
1> 2752,END,2013-01-01 00:10:11,2013-01-01 00:43:59,-74.00764,40.74008,-73.97256,40.748238,2,2013002676,2013002673
2> 20064,END,2013-01-01 00:43:40,2013-01-01 00:43:47,-73.99468,40.75278,-73.9947,40.752785,1,2013004332,2013004329
1> 20321,START,2013-01-01 00:44:00,1970-01-01 00:00:00,-73.990974,40.761127,-73.98905,40.748066,3,2013003739,2013003736
2> 4514,END,2013-01-01 00:14:00,2013-01-01 00:44:00,-73.99091,40.76602,-73.98274,40.723396,2,2013004203,2013004200
1> 20396,START,2013-01-01 00:44:00,1970-01-01 00:00:00,-73.96219,40.759377,-73.98398,40.750668,1,2013003403,2013003399
2> 20063,START,2013-01-01 00:43:40,1970-01-01 00:00:00,-74.00152,40.730988,-73.991104,40.76114,2,2013005732,2013005729
1> 15699,END,2013-01-01 00:35:43,2013-01-01 00:44:17,-73.95452,40.713936,-73.98238,40.73151,1,2013002303,2013002300
2> 20461,START,2013-01-01 00:44:07,1970-01-01 00:00:00,-73.98539,40.768845,-73.99505,40.756165,1,2013001598,2013001595
1> 20288,START,2013-01-01 00:44:00,1970-01-01 00:00:00,-73.95886,40.764034,-73.98117,40.780933,6,2013003717,2013003713
1> 20180,START,2013-01-01 00:44:00,1970-01-01 00:00:00,-73.98445,40.74352,-73.995865,40.760603,2,2013005118,2013005115
```

五、总结

Flink 的官方教程做的是真的好…样例多+解释详细，反观 giraph…