第一步:对于Mven工程第一步需要通过maven install将本地jar包导入本地仓库或者私服，

命令如下：mvn install:install-file -Dfile=E:\mvn-jar\DataShuffle-2.0-SNAPSHOT.jar -DgroupId=rzx -DartifactId=DataShuffle -Dversion=2.0 -Dpackaging=jar

在pom文件中添加对应的依赖。

第二步:新建ShuffleDomain对象，并传入两个参数arg1：要接收数据的topic名称。arg2：最后要发送的topic名称。

第三部：新建JobInfo对象。

并设置初始化参数，可供设置的参数有多个：

Streaming\_Seconds(age:Int)：设置流的间隔时间（必填）

Kafka\_List(arg：String)：设置kafka集群的list，多个机器的话用”，”号隔开。（必填）

Data\_Type(arg:String)：设置每个字段的类型，每个类型必须与文件中字段严格对应。并以”，”号隔开（必填）

Data\_Size(arg:Int)：设置文件中一行数据总字段数量了，必须与文件严格一致（必填）

Data\_Regs(arg:String)：设置最后输出数据的分隔符（必填）

Date\_Format\_=("yyyy-MM-dd HH:mm:ss")：如果有日期类型需要设置此项，用来对日期格式进行初始化，需要注意的是只支持以冒号，横线，斜线，空格分隔的日期格式，并且不能出现单个的如y，m，d，H，m，s。

Data\_IsDel(arg:Boolean)：设置是否删除有缺失值或类型异常的数据（可选）

第四步：调用shuffle方法进行数据清洗：

Shuffle方法有两个。一个必须传入JobInfo对象，另一个则需要传入JobInfo对象和默认替换的值：

Shuffle（arg：JobInfo）：该方法默认删除出现缺失或异常的数据。要替换的值。

Shuffle（arg1：JobInfo，arg2：String）：该方法默认出现缺失值或者异常值时用传入的参数进行替换。注意传入的值必须与文件字段相互对应。并且以逗号隔开

下面一段代码是不含日期类型时的一个例子，并且对与异常值进行替换的例子：

job.Kafka\_List\_=("t004:9092,t002:9092,t003:9092")  
job.Stream\_Seconds\_=(3)  
job.Data\_Type\_=("String,String,Int")  
job.Data\_Size\_=(3)  
//job.Data\_Celue\_=("")  
job.Data\_Regx\_=("|")  
shuffle.shuffle(job,"a,a,0"