数据迁移数据量大的时候，验证起来非常费劲，设计到大量数据，2个不同的数据库和表的验证，在凌云项目期间做了个数据的验证工具，附上说明：

1. 运行方式：

指定单个或则多个规则运行：java -jar ./dbfair.jar -Dfile=test.xml -Dfile=test2.xml

指定一个csv，所有的规则在csv中：java -jar ./dbfair.jar -Dfile=test.csv

运行结果是一个xml，在该xml中可以看到成果和失败的用例。

1. 数据比较规则：

规则是一个xml，由2部分组成：dbSet部分和Flows部分，其中dbSet是数据库的别名、链接方式、用户名和密码构成，Flows定义了一系列的比较规则，由至少1个flow组成。flow一般由描述部分、本次比较的类别（目前只有**COUNT和FILEDHANDLE**），是否有公共字段、进行比较结果的groovy的文件夹地址和groovy文件、sql语句以及填充sql的value值。

一个典型的dbSet如下：



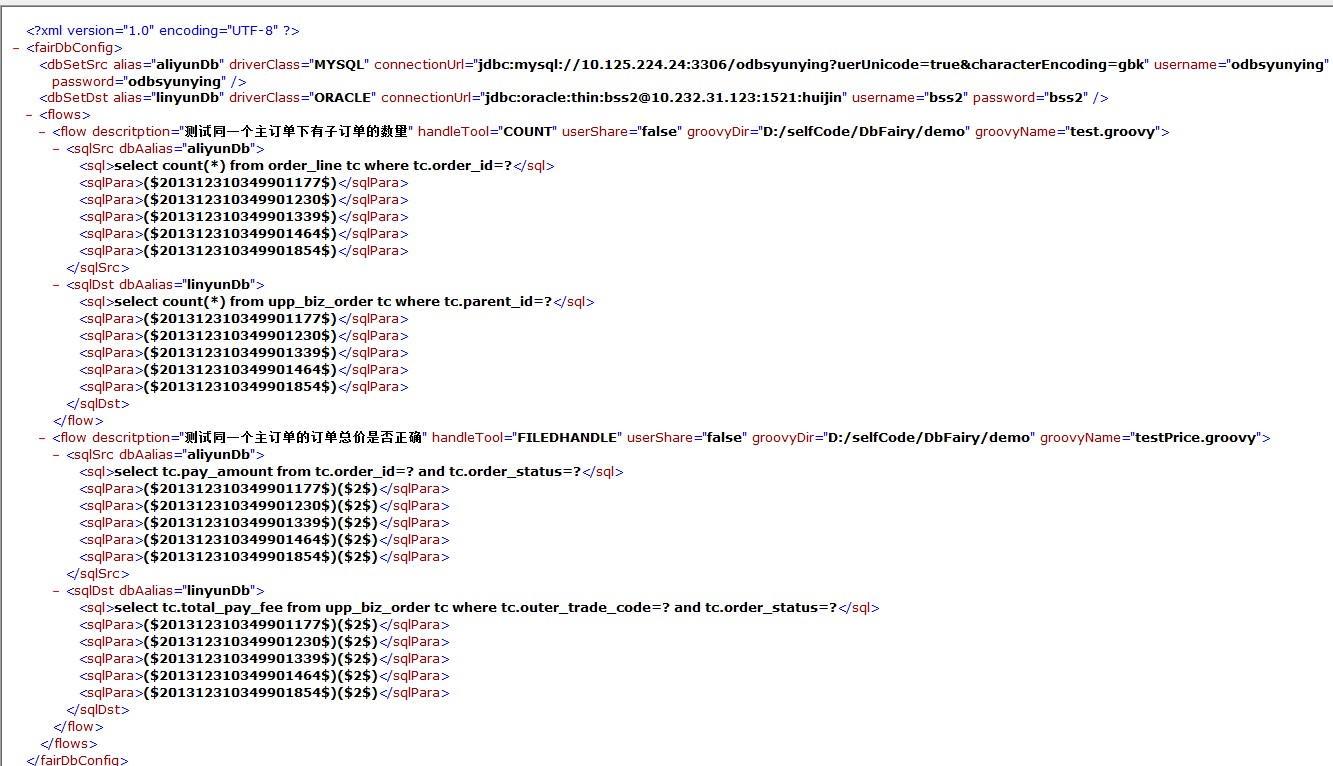
这个dbSet设置了2个数据库的的别名，驱动方式、url、用户名和密码

一个flow典型的flow如下：



以上表明，这个用例是测试订单量的，每次sql取出value值，将aliyunDb的结果和linyunDb的结果进行比较，将2次sql查询的结果使用test.groovy进行比对。

一个较为完整的xml如下：



有的时候，2个比较的关键字段值是一样的，而且都是从某个sql结果中取得的，那么我们就可以使用userShare，将userShare设置成true，同时设置成sqlShare，例如



以上get(order)语句中的值，由sqlShared的查询结果取出，每次取出一次，这样如果查询出1000个orderId，则sqlsrc和sqlDst可以进行1000次比较。

注意事项：driverClass 指明了数据的类型，目前支持：MYSQL和ORACLE

Alias和dbAalias不能设置错误，sql是根据别名进行数据库的链接。

handleTool：比较的类别，目前支持：**COUNT（数量）、FILEDCOMPARE（字段）**

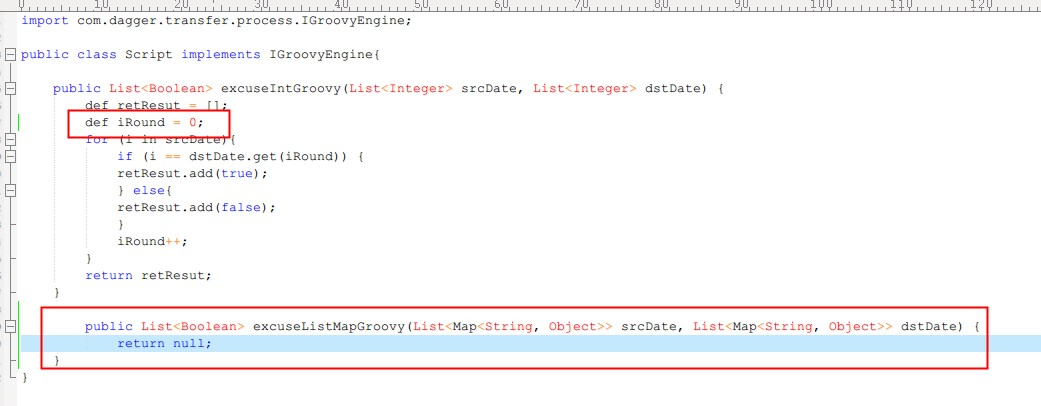
**由于比较的类别千差万别，所以由groovy进行具体的内容最终比较：**

**这个groovy实现IGroovyEngine的接口；**



**一个数量比较的demo为：**

Script.groovy：



最后的结果输出到dbFairy.html