火龙果•整理
uml.org.cn

Kettle 培训手册

Etl 介绍

ETL（Extract-Transform-Load 的缩写，即数据抽取、转换、装载的过程），对于金融 IT 来说，经常会遇到大数据量的处理，转换，迁移，所以了解并掌握一种 etl 工具的使用，必不可少。

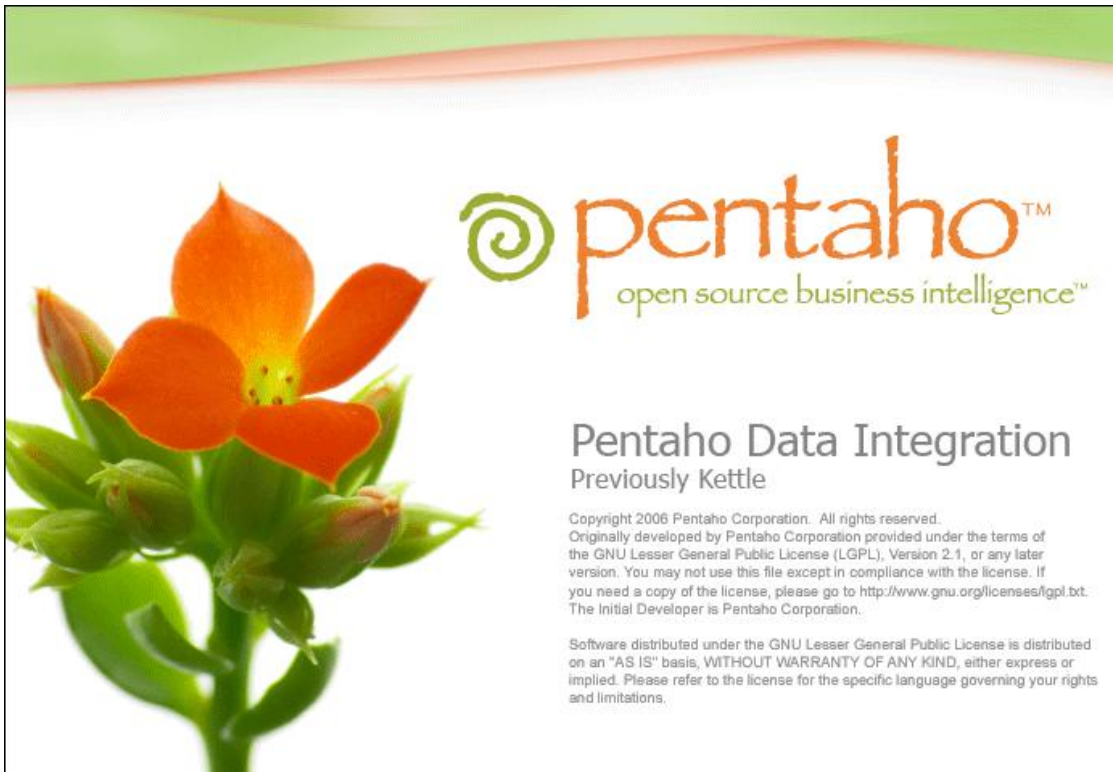
Kettle 是一款国外开源的 etl 工具，纯 java 编写，绿色无需安装，数据抽取高效稳定。Kettle 中有两种脚本文件，transformation 和 job，transformation 完成针对数据的基础转换，job 则完成整个工作流的控制。

kettle 部署运行

将 kettle2.5.1 文件夹拷贝到本地路径，例如 D 盘根目录。

双击运行 kettle 文件夹下的 spoon.bat 文件，出现 kettle 欢迎界面：

火龙果•整理
uml.org.cn



稍等几秒

选择没有资源库，打开 kettle 主界面

火龙果•整理
uml.org.cn

创建 transformation，job

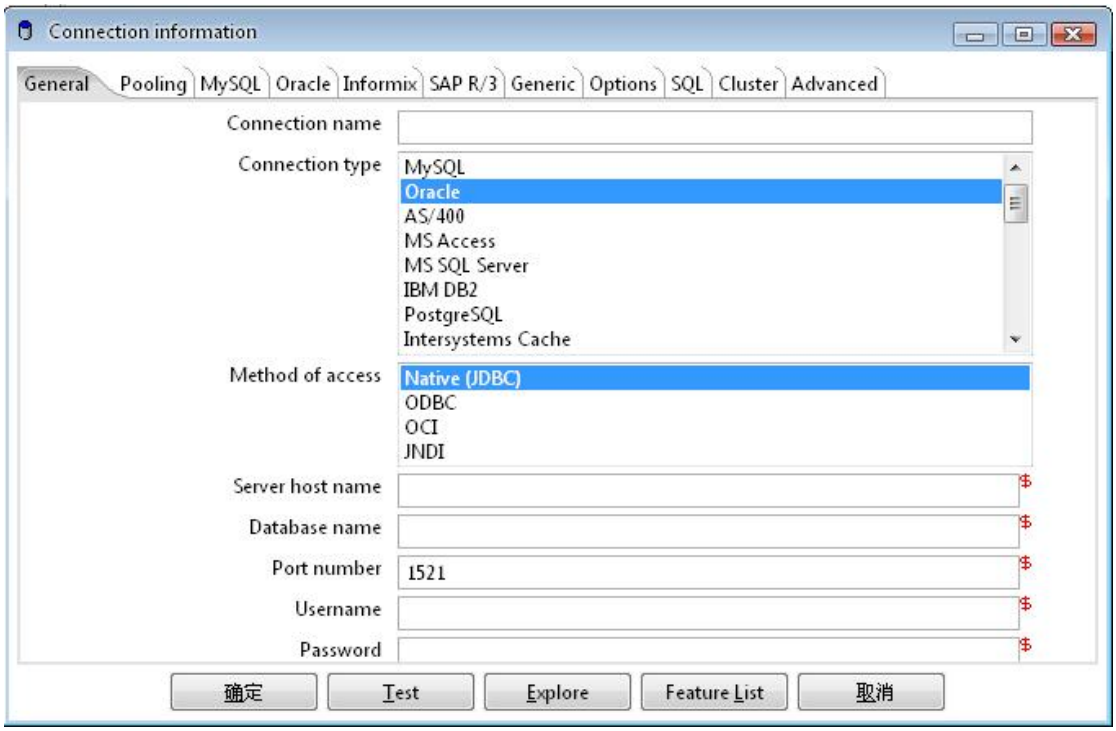
点击页面左上角的创建一个新的 transformation，点击保存到本地路径，例如保存到 D:/etltest 下，保存文件名为 EtltestTrans，kettle 默认 transformation 文件保存后后缀名为 ktr

点击页面左上角的创建一个新的 job，点击保存到本地路径，例如保存到 D:/etltest 下，保存文件名为 EtltestJob，kettle 默认 job 文件保存后后缀名为 kjb

创建数据库连接

火龙果•整理
uml.org.cn

在 transformation 页面下，点击左边的【Main Tree】，双击【DB 连接】，进行数据库连接配置。



Connection name 自命名连接名称

Connection type 选择需要连接的数据库

Method of access 选择连接类型

Server host name 写入数据库服务器的 ip 地址

Database name 写入数据库名

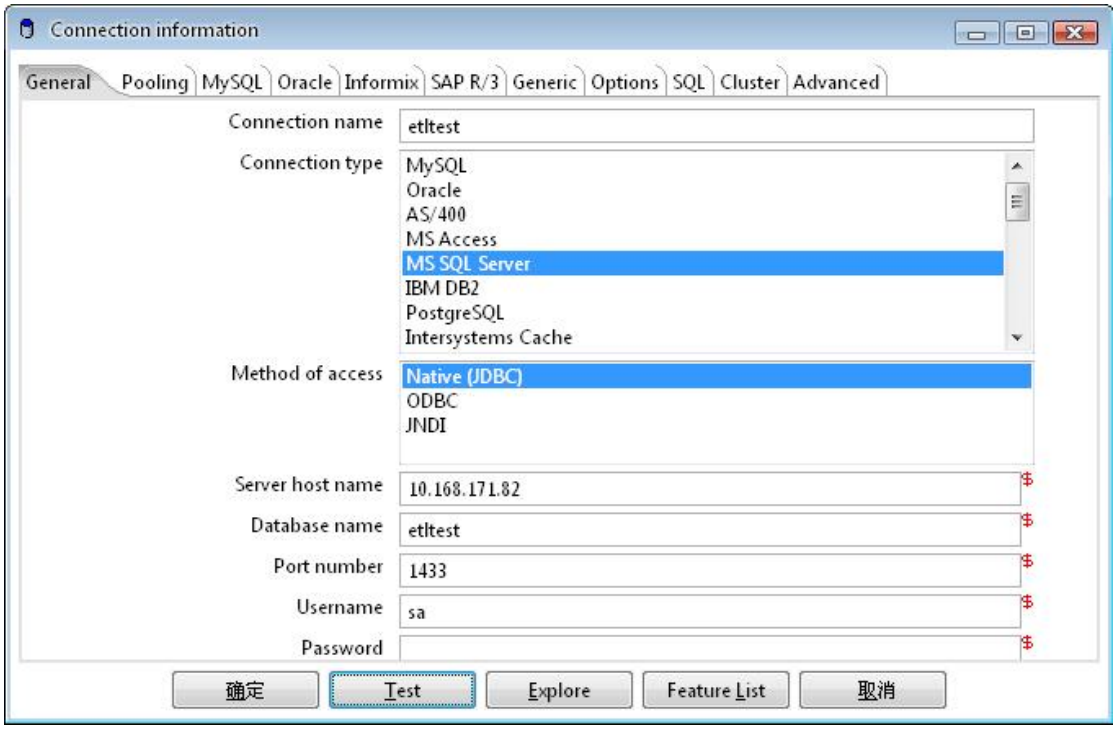
Port number 写入端口号

Username 写入用户名

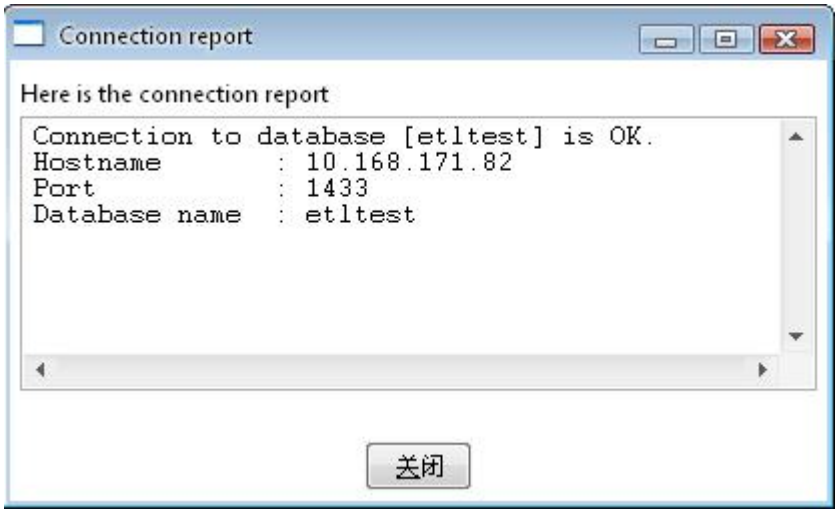
Password 写入密码

例如如下配置：

火龙果•整理
uml.org.cn



点击【test】，如果出现如下提示则说明配置成功



点击关闭，再点击确定保存数据库连接。

火龙果•整理
uml.org.cn

一个简单的 ktr 例子

目标：

从交易表（trade），帐户表（account），客户表（cust）抽数交易相关的所有信息，并判断对公对私分别进行处理，输出到文本文件中。

操作步骤：

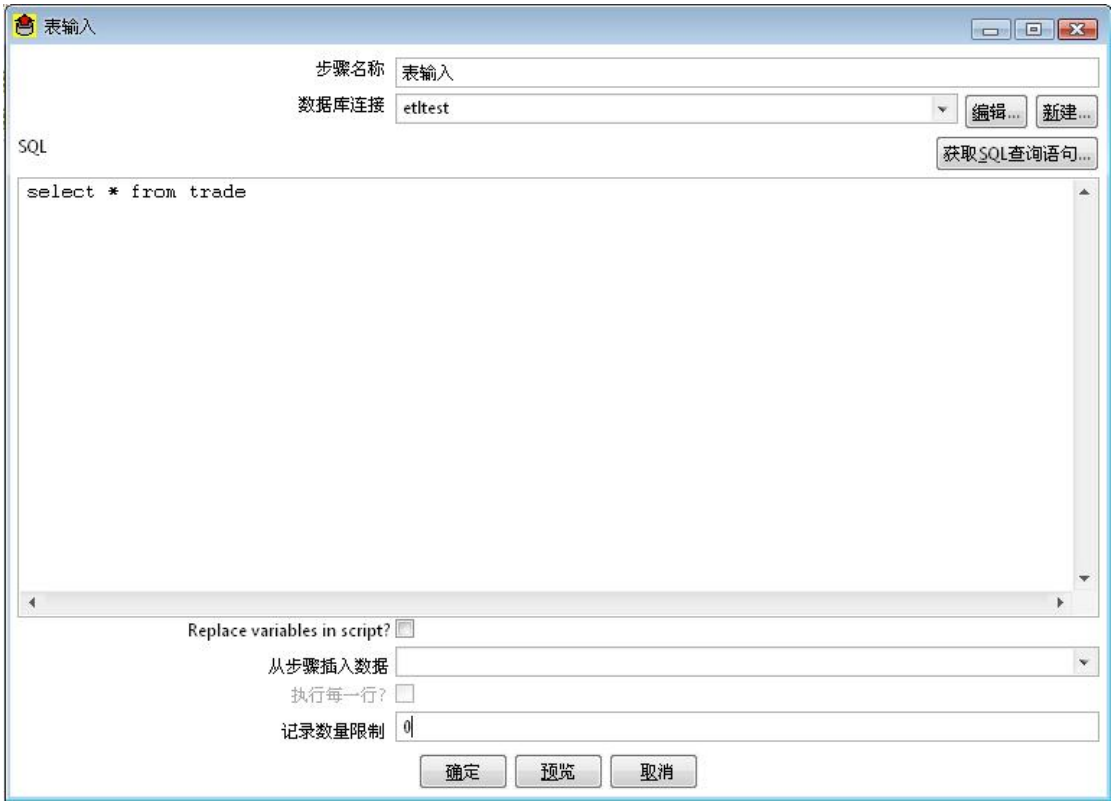
在 EtltestTrans 页面下，点击左侧的【Core Objects】，点击【Input】，选中【表输入】，拖动到主窗口释放鼠标。

双击【表输入】图标

数据库连接选择刚刚创建好的 etltest 数据库连接，在主窗口写入对应的查询语句

```
Select * from trade
```

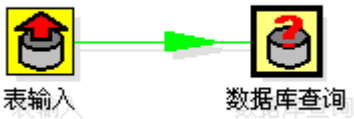
火龙果•整理
uml.org.cn



点击确定完成。

点击左侧的【Lookup】，选中【数据库查询】，拖动到主窗口释放鼠标。

按住 shift 键，用鼠标点中刚才创建的【表输入】，拖动到【数据库查询】上，则建立了两个环节之间的连接。



双击【数据库查询】图标

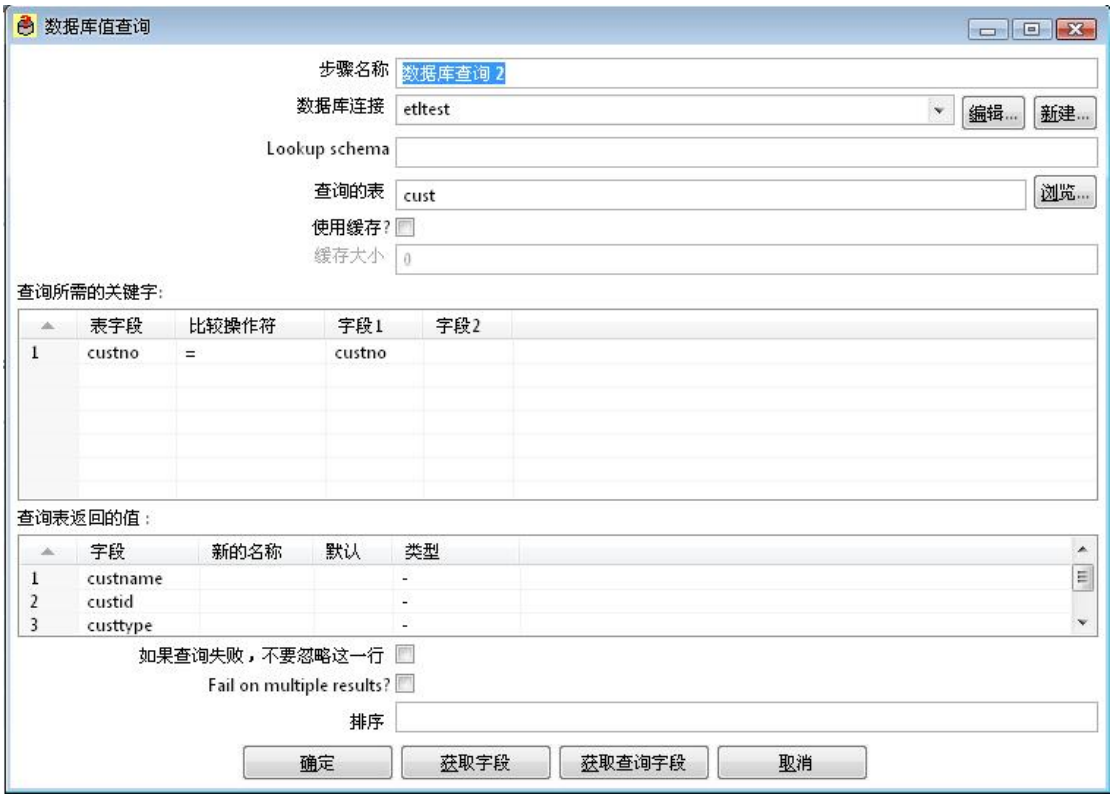
火龙果•整理
uml.org.cn

步骤名称写入 account 表查询，数据库连接选择刚刚创建好的 etltest 数据库连接，查询的表写入 account，查询所需的关键字中，表字段写入 acctno，比较操作符写入“=”，字段 1 写入 acctno。

在查询表返回的值里面写入 custno，确定完成。

同上，再创建一个数据库查询，命名为 cust 表查询，查询的表写入 cust，查询所需的关键字写入 custno=custno，查询表返回的值写入 custname，custid，custtype

火龙果•整理
uml.org.cn



点击左侧的【Transform】，选中【过滤记录】，拖动到主窗口释放鼠标。

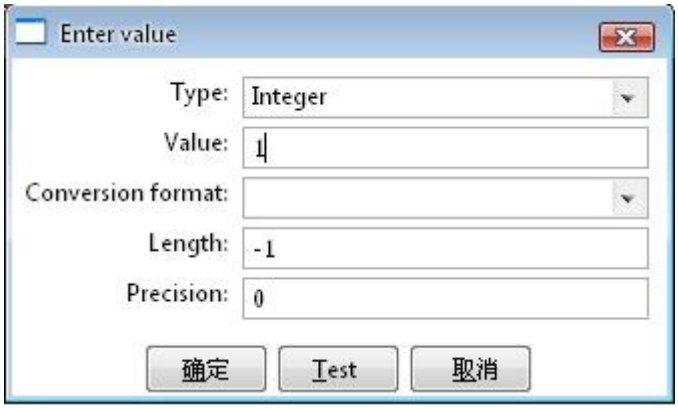
点击左侧的【Scripting】，选中两个【Modified Java Script Value】，拖动到主窗口释放鼠标。分别双击打开，重命名为“对公类型修改”和“对私类型修改”。

同时，分别创建【过滤记录】和【对公类型修改】,【对私类型修改】的连接。

双击【规律记录】打开。

第一个<field>里面选择 custtype，点击<value>，在 Enter value 里面写入 1，确定

火龙果•整理
uml.org.cn



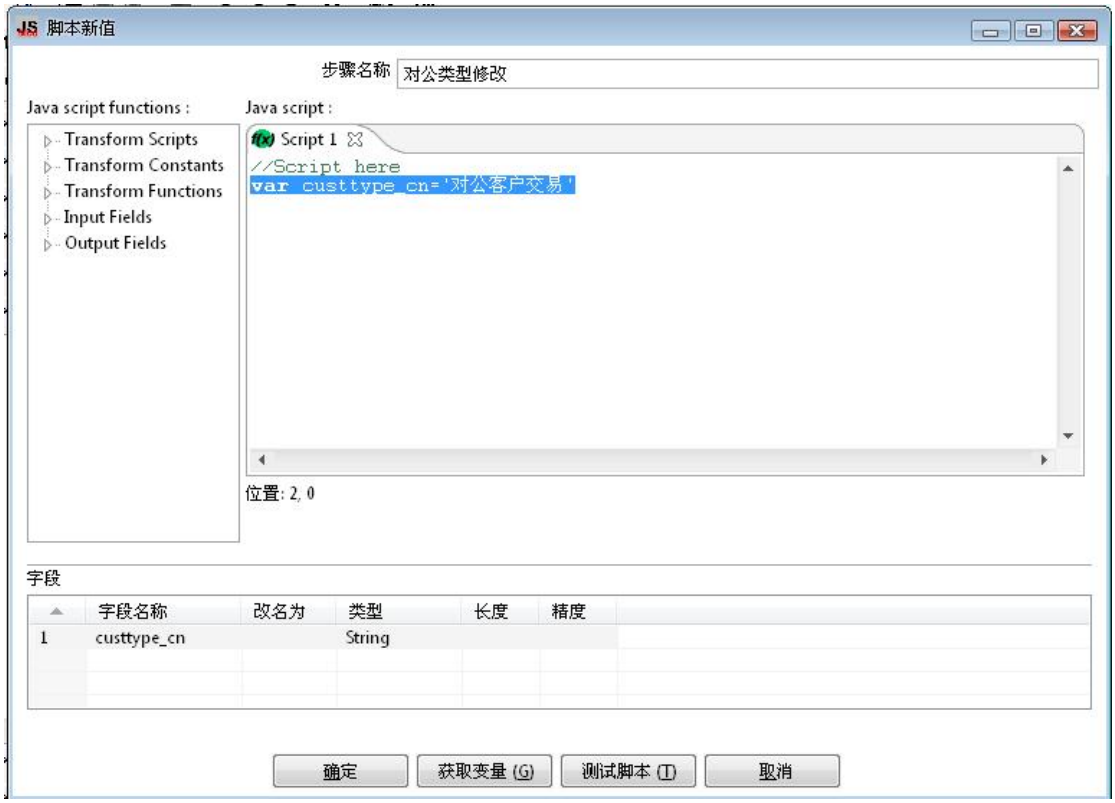
在发送 true 数据给步骤里，选择【对私类型修改】，在发送 false 数据给步骤里，选择【对公类型修改】，确定保存。

双击【对公类型修改】，在里面写入 javascript 脚本语句

```
var custtype_cn='对公客户交易'
```

火龙果•整理
uml.org.cn

在字段中写入 custtype_cn，类型选为 string。确定。



同理，在【对私类型修改】中，在里面写入 javascript 脚本语句

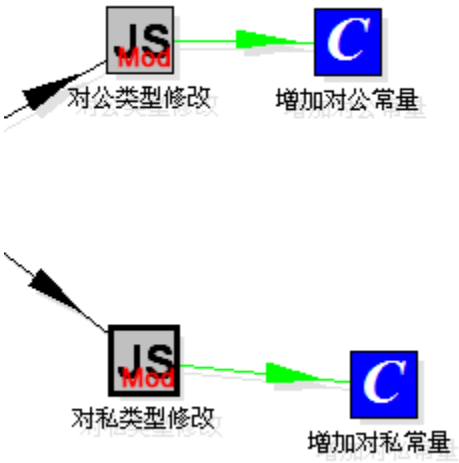
```
var custtype_cn='对私客户交易'
```

在字段中写入 custtype_cn，类型选为 string。确定。

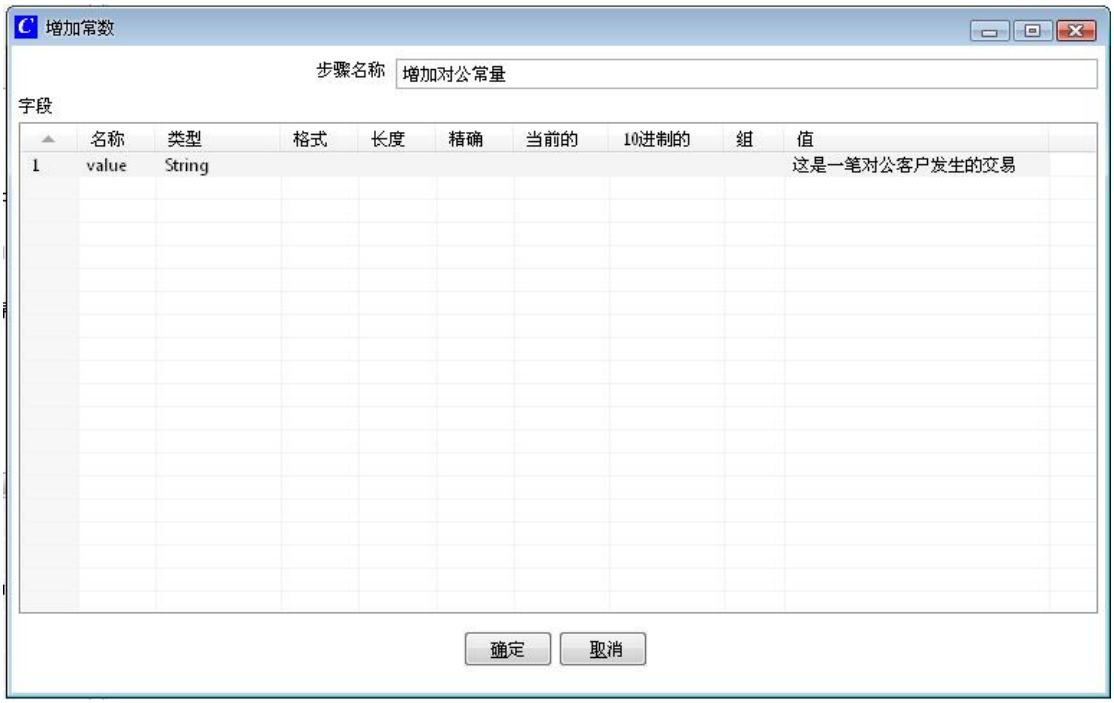
点击左侧的【Transform】，选中两个【增加常量】，拖动到主窗口释放鼠标。分别双击打开，重命名为“增加对公常量”和“增加对私常量”。

分别建立【对公类型修改】和【对私类型修改】与【增加对公常量】和【增加对私常量】的连接

火龙果•整理
uml.org.cn



双击【增加对公常量】，名称写入 value，类型选择 string，值写入“这是一笔对公客户发生的交易”，确定保存。

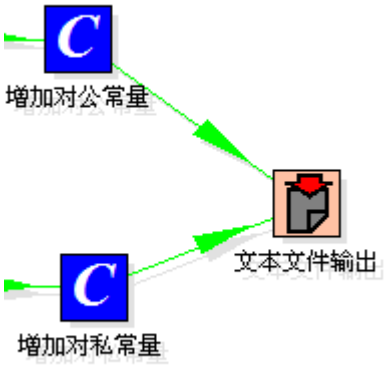


同理，双击【增加对私常量】，名称写入 value，类型选择 string，值写入“这是一笔对私客户发生的交易”，确定保存。

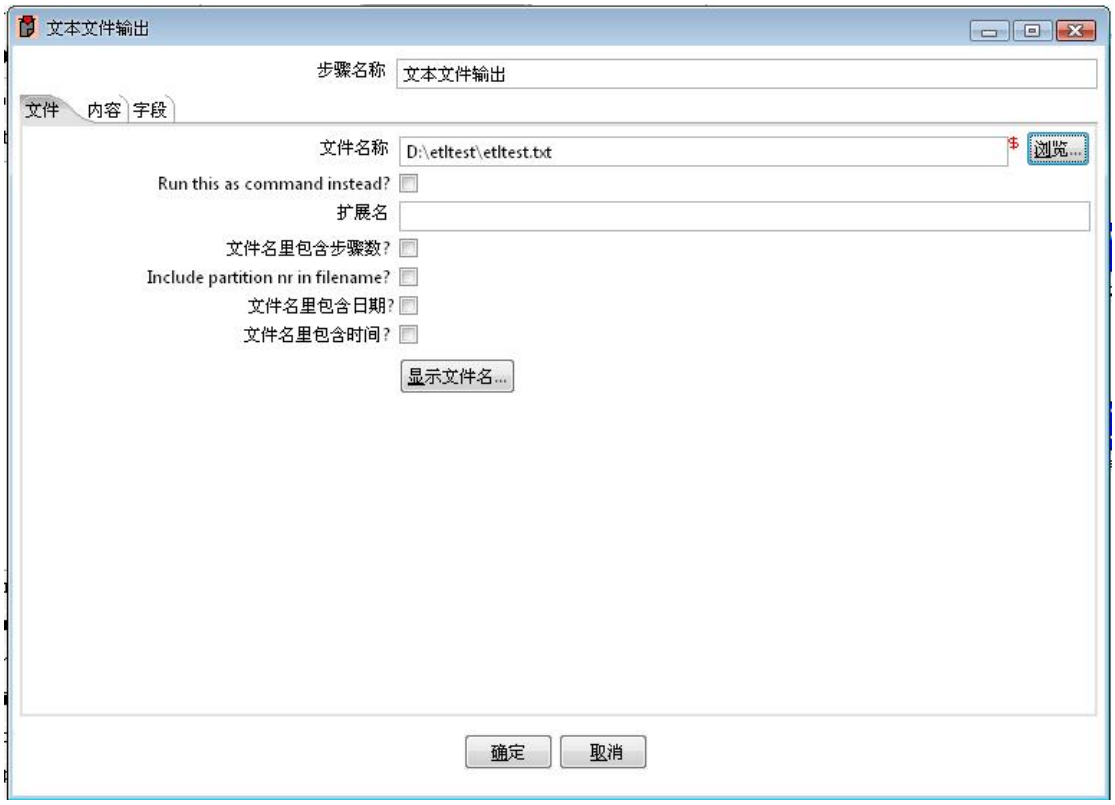
点击左侧的【Output】，选中【文本文件输出】，拖动到主窗口释放鼠标。

火龙果•整理
uml.org.cn

建立【增加对公常量】,【增加对私常量】和【文本文件输出】的连接。



双击打开【文本文件输出】，文件名称写入 D:\etltest\etltest.txt



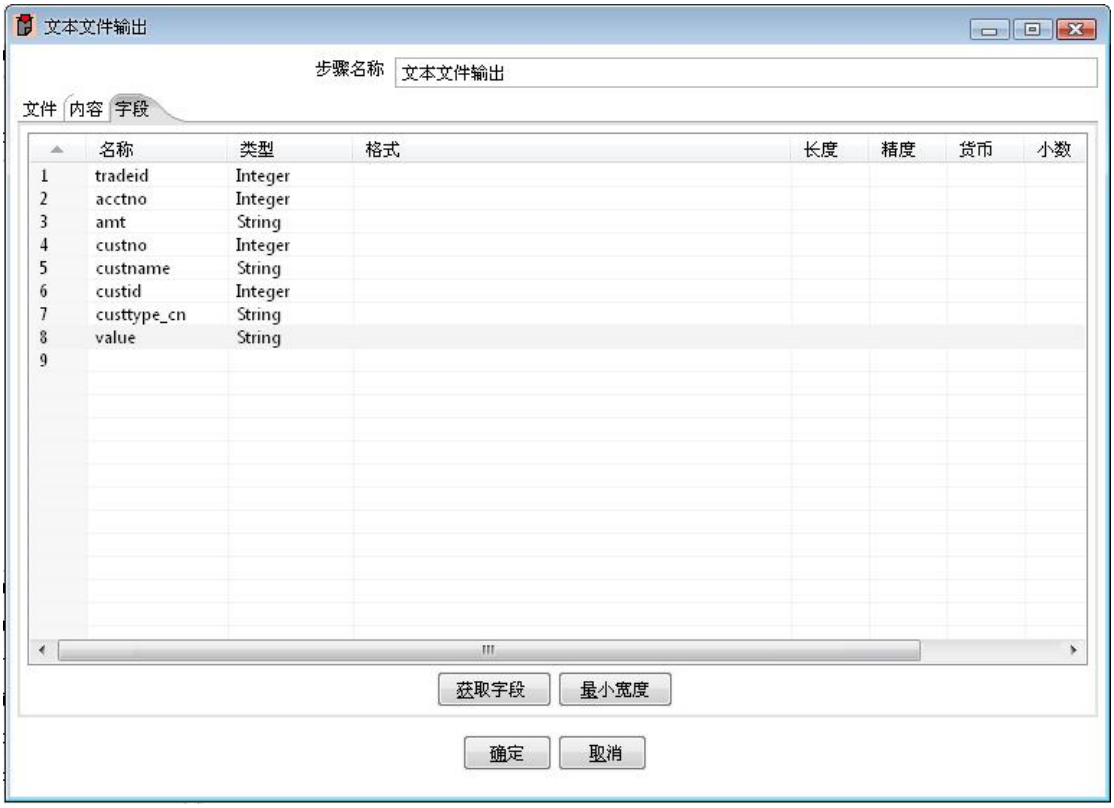
点击内容标签，根据情况进行修改，例如

火龙果•整理
uml.org.cn

点击字段标签

名称依次写入 tradeid，acctno，amt，custno，custname，custid，custtype_cn，value，类型根据各个字段实际类型进行选择

火龙果•整理
uml.org.cn

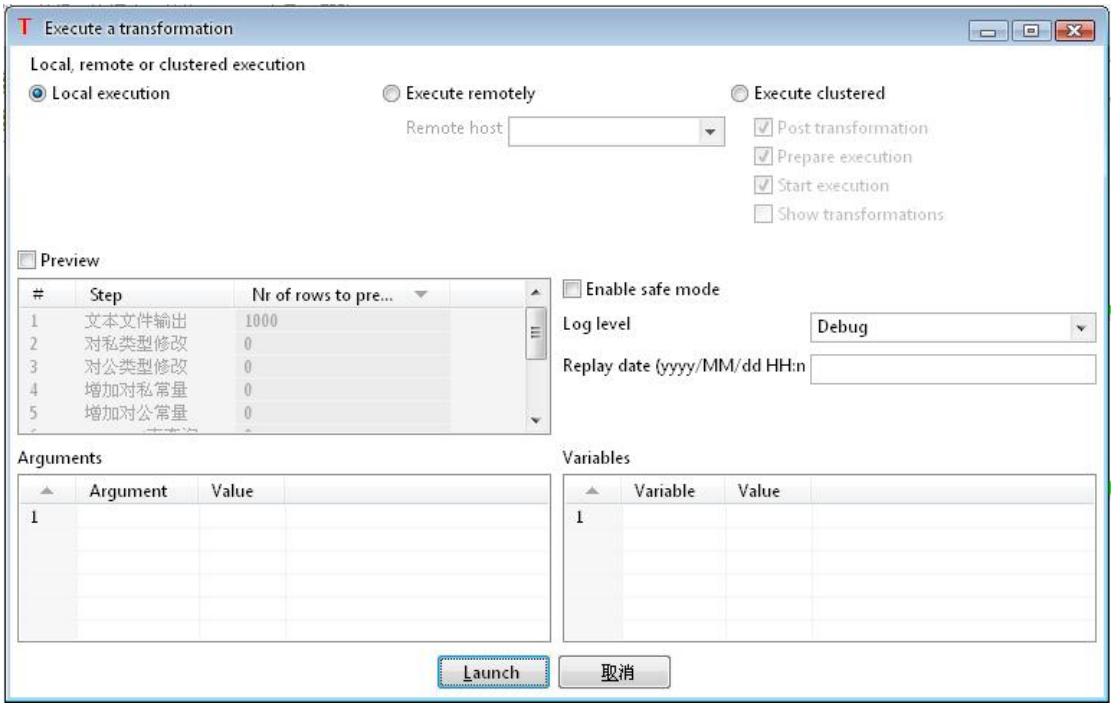


确定保存

点击保存创建好的 transformation。

点击运行这个转换。

火龙果•整理
uml.org.cn



点击 launch，开始运行

当所有状态都变成已完成时，则转换完成。

#	步骤名称	复制的记录行数	读	写	输入	输出	更新	Rejected	错误	激活	时间
1	表输入	0	0	16	16	0	0	0	0	已完成	0.7
2	account表查询	0	16	16	0	0	0	0	0	已完成	1.0
3	cust表查询	0	16	16	0	0	0	0	0	已完成	1.5
4	过滤记录	0	16	16	0	0	0	0	0	已完成	1.5
5	对私类型修改	0	12	12	0	0	0	0	0	已完成	2.0
6	对公类型修改	0	4	4	0	0	0	0	0	已完成	2.1
7	增加对私常量	0	12	12	0	0	0	0	0	已完成	2.1
8	增加对公常量	0	4	4	0	0	0	0	0	已完成	2.1
9	文本文件输出	0	16	0	0	16	0	0	0	已完成	2.5

另一个简单的 ktr 例子

目的：

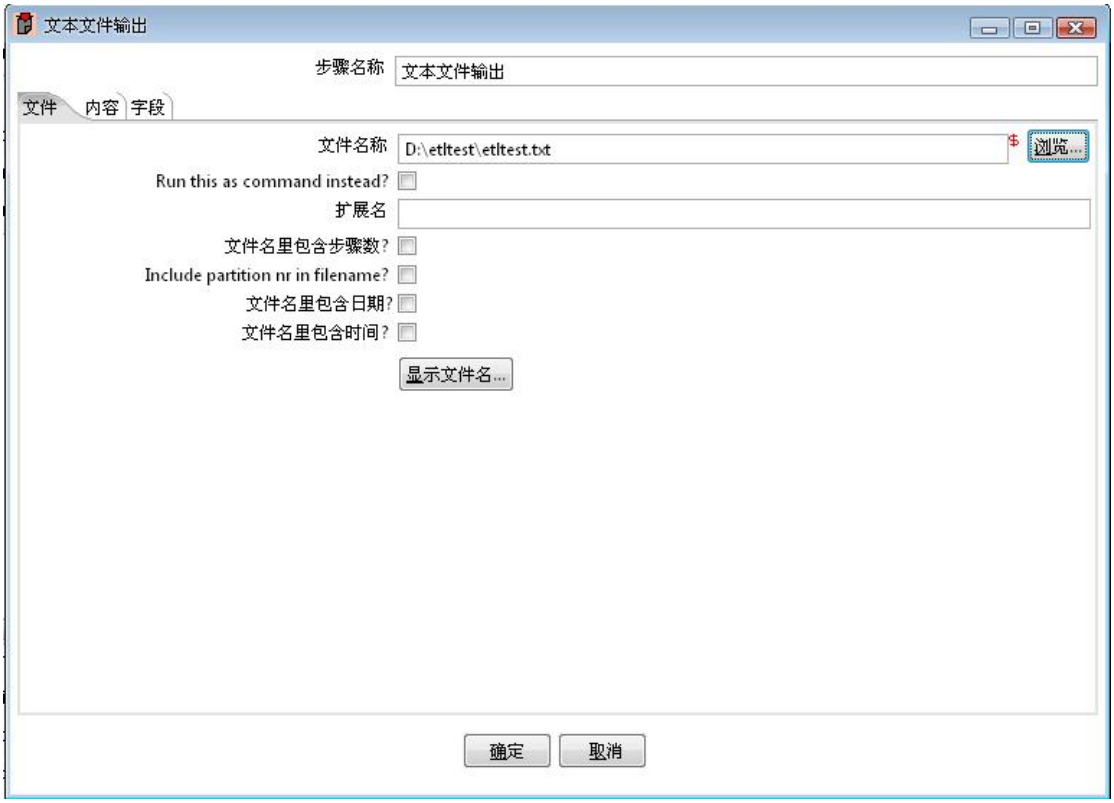
将上一个 ktr 生成的文本导入到数据库中。

操作步骤：

火龙果•整理
uml.org.cn

创建一个 transformation，命名为 EtltestTransfile2db.ktr，创建数据库连接 etltest，点击【Input】，选中【文本文件输入】，拖到主窗口，释放鼠标，双击打开

文件名称里面写入 D:\etltest\etltest.txt

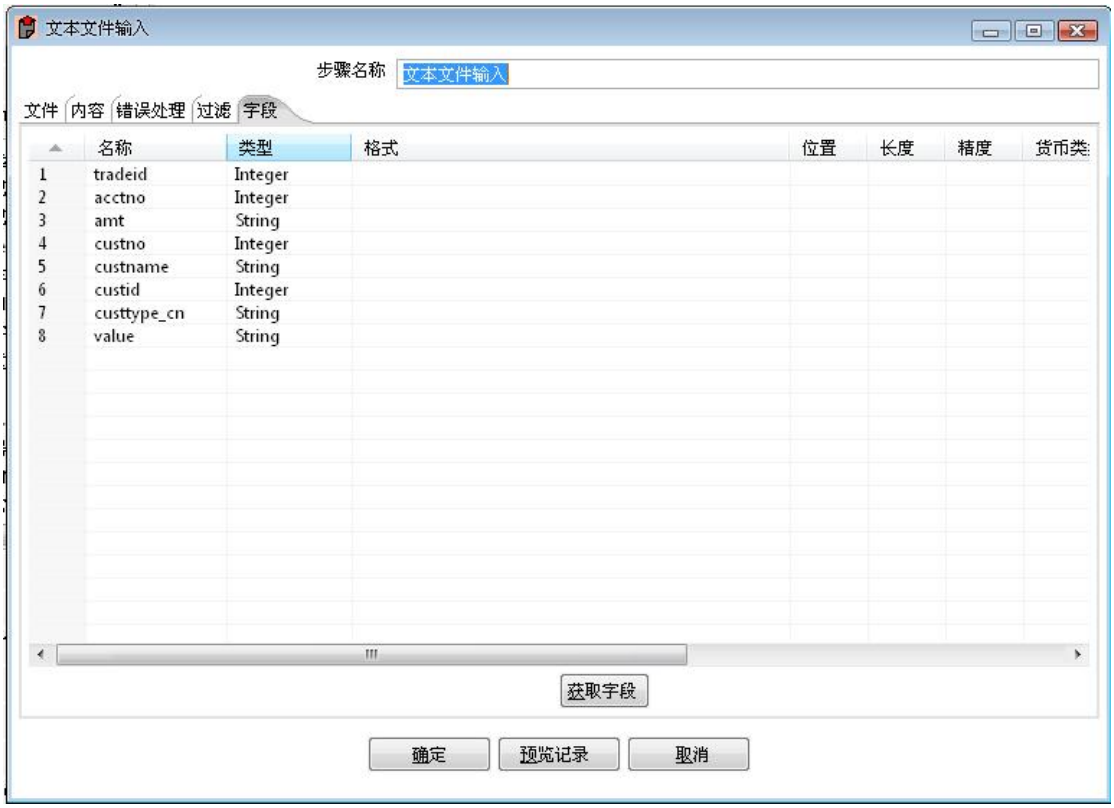


点击内容标签，分隔符写入@@，将头部的钩去掉

火龙果•整理
uml.org.cn

点击字段标签，名称依次写入 tradeid，acctno，amt，custno，custname，custid，custtype_cn， value，类型根据各个字段实际类型进行选择

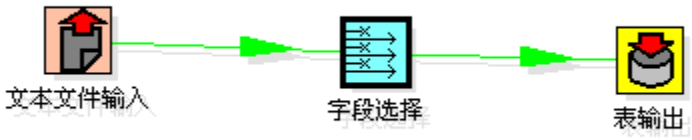
火龙果•整理
uml.org.cn



点击【Transform】，选中【字段选择】，拖到主窗口，释放鼠标

点击【Output】，选中【表输出】，拖到主窗口，释放鼠标

建立【文本文件输入】和【字段选择】与【字段选择】和【表输出】的连接

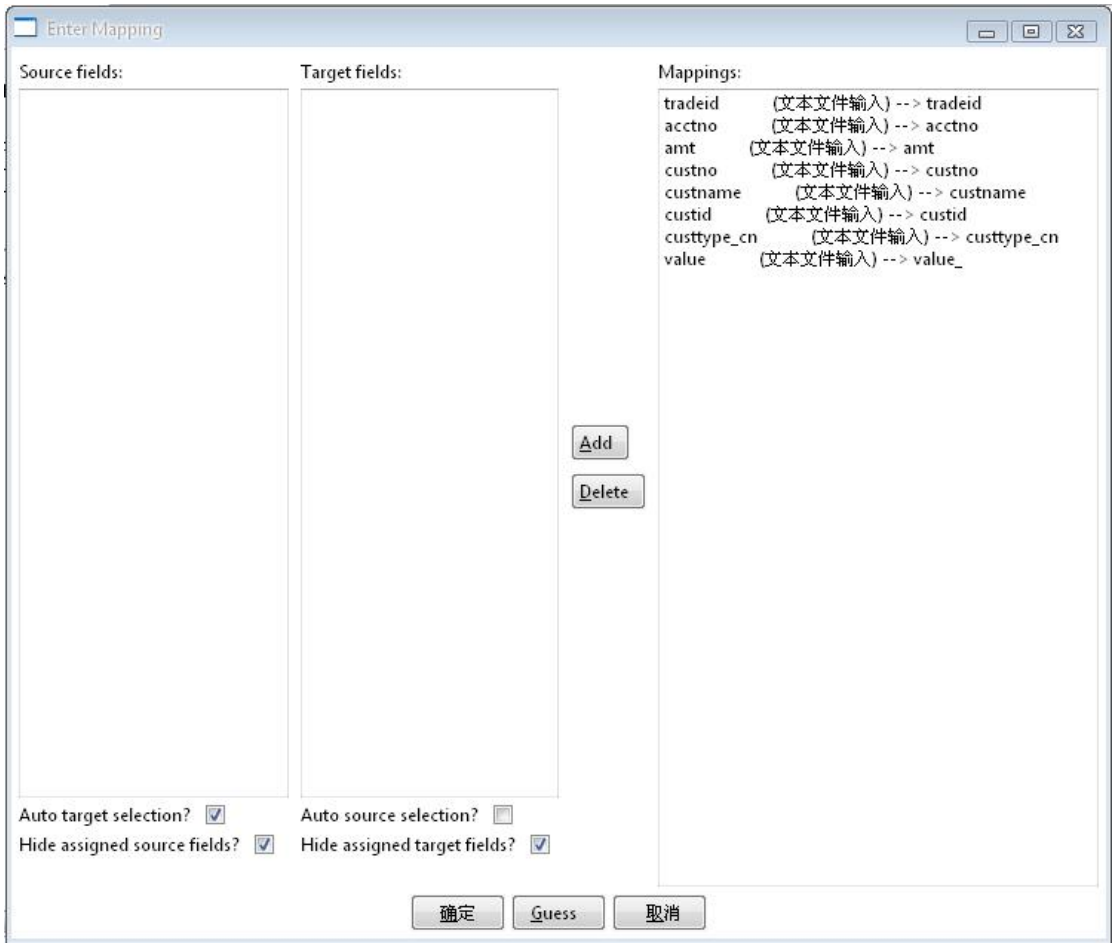


双击【表输出】，目标表中写入 trade_all，提交记录数量写成 1，确定保存。

火龙果•整理
uml.org.cn

双击【字段选择】，点击 获取选择的字段，再点击 Edlt Mapping，点击 OK 确定，编辑所有字段对应关系，点确定。

火龙果•整理
uml.org.cn



点击运行这个转换。，则将上一个 ktr 中生成的文本，导入到数据库当中。

一个简单的 kjb 例子

目的：

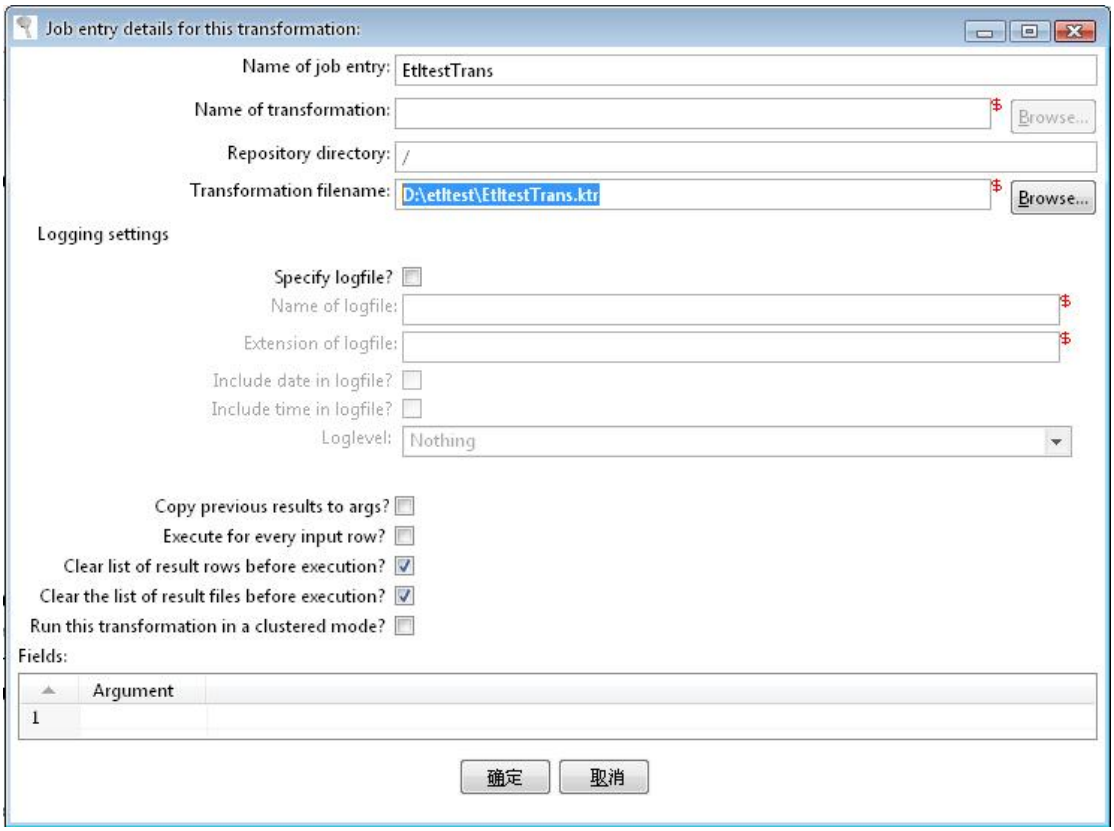
将上两个 transformation 一起在同一个 job 里面调用执行。

操作步骤：

在 EtlscriptJob 页面，点击【Core Objects】，点击【Job entries】，选中【START】拖动到主窗口释放鼠标，再选中两个【Transformation】，拖动到主窗口释放鼠标，建立【START】和【Transformation】与【Transformation】之间的连接。

火龙果•整理
uml.org.cn

双击第一个【Transformation】，在 Transformation filename 中写入 D:\etltest\EtltestTrans.ktr，确定保存。



同时将另外一个【Transformation】，路径指向 D:\etltest\EtltestTransfile2db.ktr，保存。

点击保存创建好的 job。

点击运行这个转换。

待所有任务都显示成功，则为 job 调用 transformation 运行成功。

火龙果•整理
uml.org.cn

EtltestJob			
任务: EtltestJob	Start of job execution		start
Start	Job entry started		start
Start	Job entry ended	成功	
EtltestTrans	Job entry started		Followed unconditional link
EtltestTrans	Job entry ended	成功	
EtltestTransfile2db	Job entry started		Followed link after success!
EtltestTransfile2db	Job entry ended	成功	
任务: EtltestJob	Job execution ended	成功	end