

KETTLE SPOON 使用手册

融汇兴业 www.rewin.com.cn 开发部 高健 2010 年 12 月

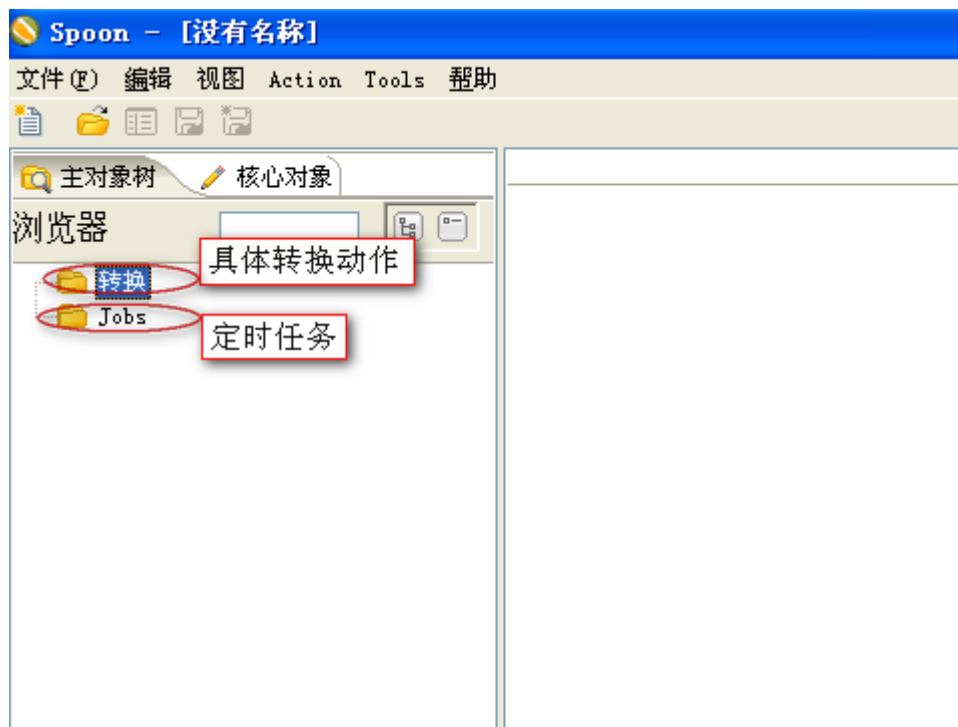
MSN: hhjtu524@sina.com

- 1.到官方网站下载 <http://kettle.pentaho.org/>
- 2.配置 JAVA 环境，解压下载包到英文路径下。如：D:\pdi-ce-4.0.1-stable\data-integration
- 3.文档说明：本文只介绍基本的同步和定时 JOB，其他高级操作请自行拓展。
- 4.工作区介绍：

运行安装目录下 Spoon.bat 欢迎界面：

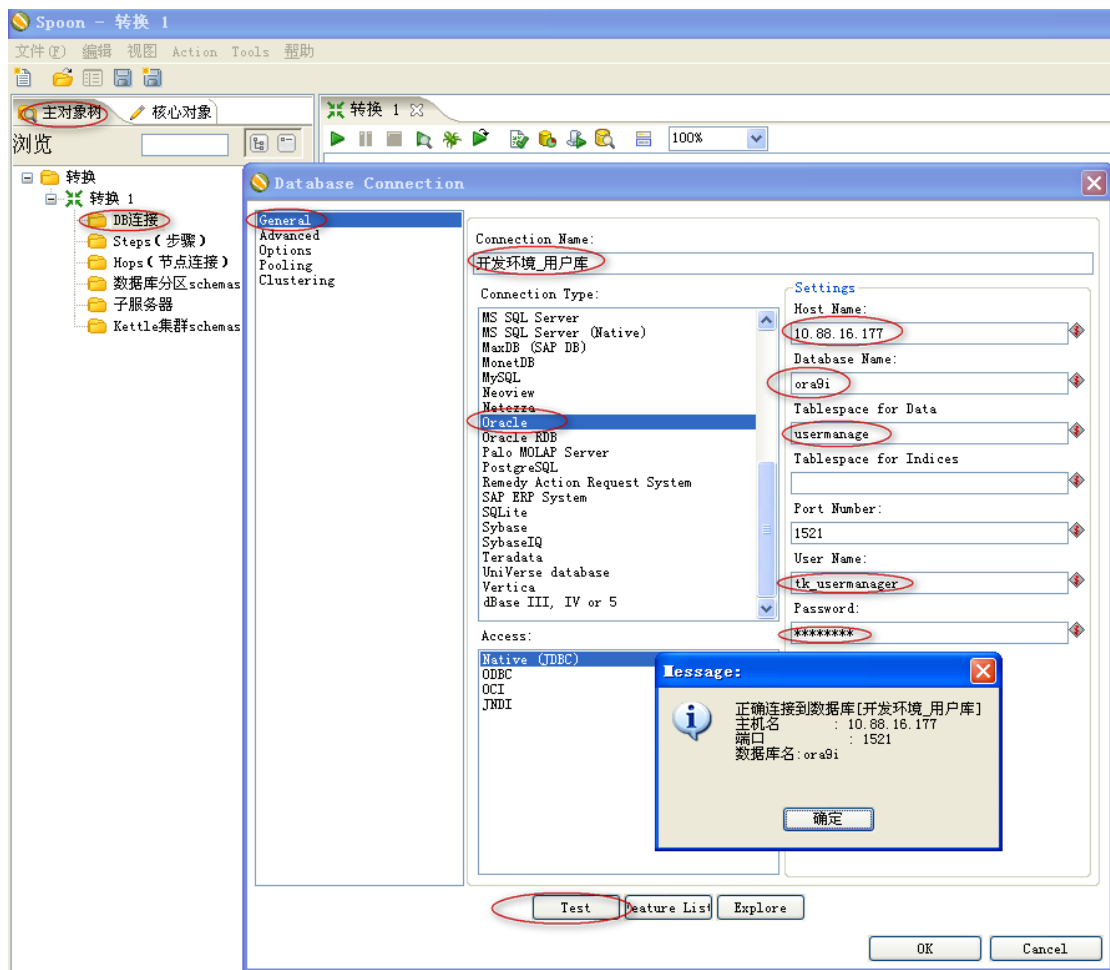
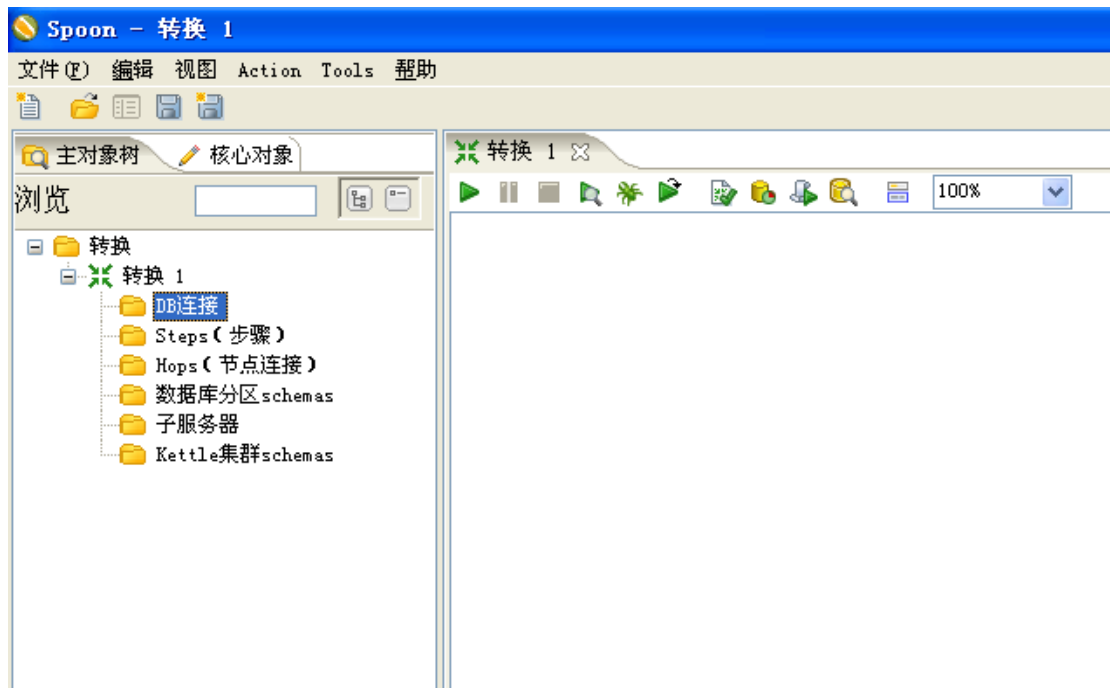


欢迎界面后进入工作平台，一般应用中使用转换和 JOB 即可。

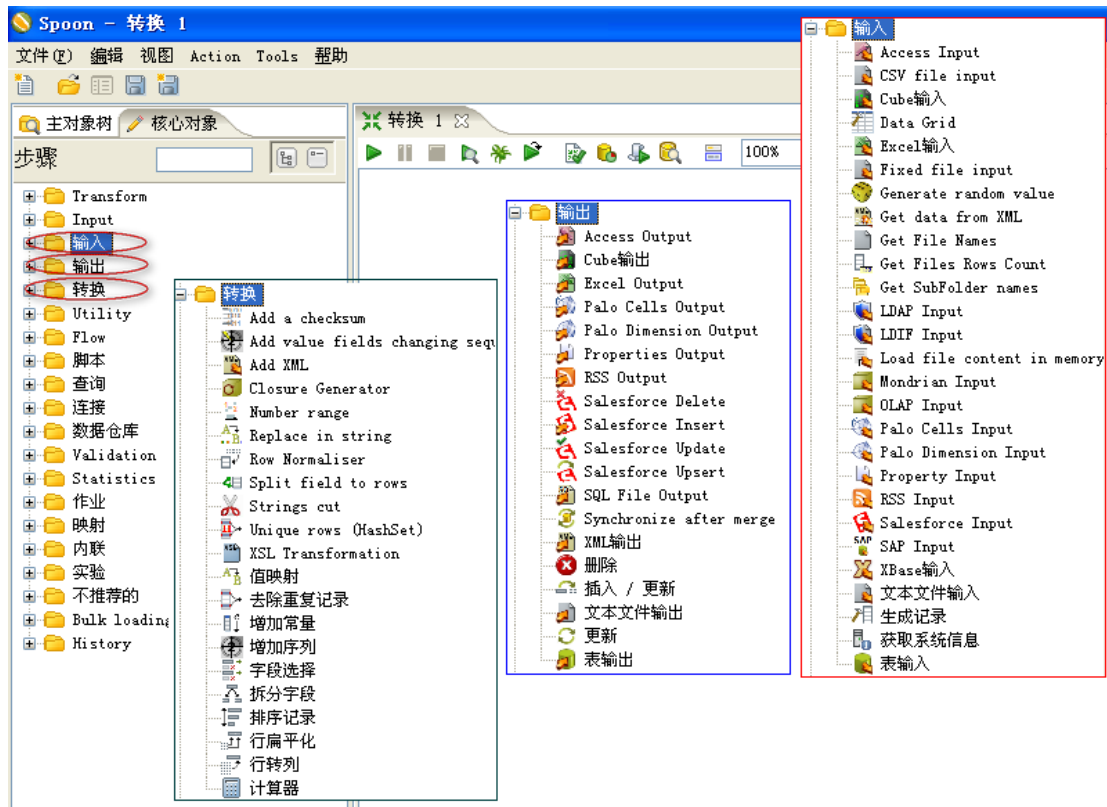


双击转换即新建转换作用。

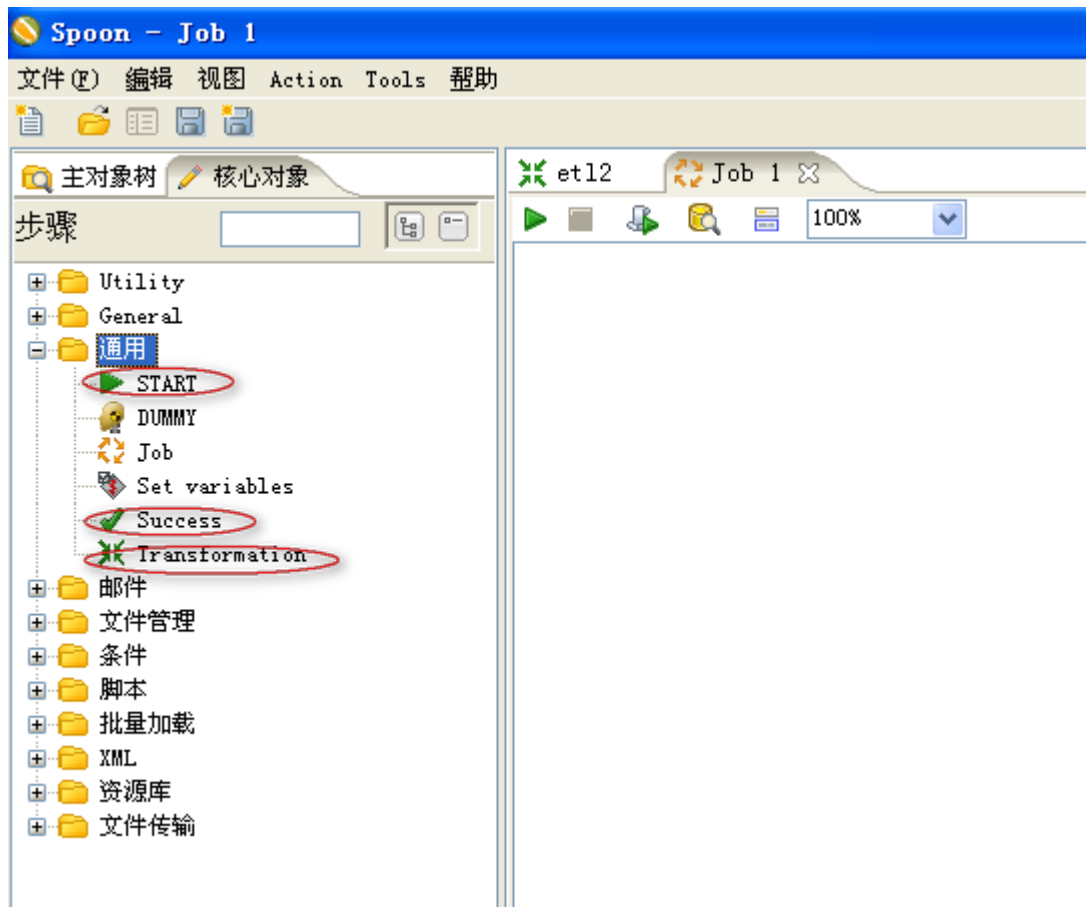
通过主对象树可以配置所要用到的数据库连接。



配置好后点击核心对象。

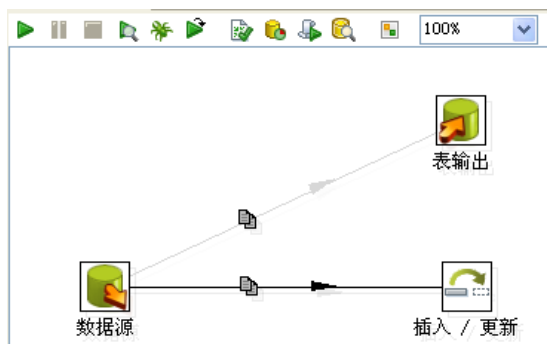


下图是 JOB 用到的基本元素。

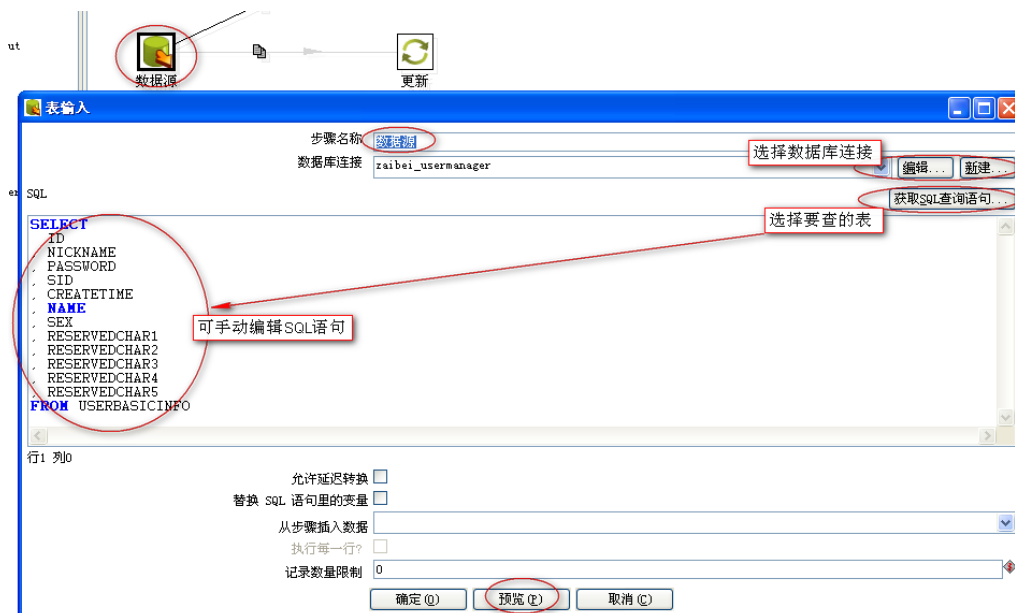


5. 建立转换

选中列表中的“表输入”、“表输出”、“插入/更新”，拖拽到右侧工作区。鼠标选中节点键盘按 **shift** 键，拖拽即可连线。下图是一个转换，有两个动作，一是直接从数据源表同步数据到目标表，一是检查源表的插入和更新，同步到目标表。



黑色线为生效，灰色线为失效，节点和连接可双击进行编辑。



可预览数据是否正确。

预览数据

步骤 数据源的数据 (344 rows)

#	ID	NICKNAME	PASSWORD	SID
1	zhaolj01	赵涛	cf3e452b8cda9f05fb5f63c24529a54e	uzhaolj011270201810957
2	wanglingyun	王云	598b4b3095ba32f1cfcff4cd417a4b427	uwanglingyun1185499395657
3	mengqian	孟倩	1265b6f0fcb3c2745278f422a8a11c2a	umengqian1211850748549
4	houm	侯明	6e66f31fce6987801bace9e9b88aa92d	uhoum1283747165145
5	zhouxiaofang	周晓芳	69e0e4795b755623352803624927ea03	uzhouxiaofang1181340506716
6	yangfan04	杨帆	a896f26f6a93ad858140d13ec7fe820	uyangfan041211851423790
7	liuhuan	刘欢	9a80010c9c4d2c90ae626f8149846a	ulihuan1282714205212
8	guoxz	郭晓	189b5cfd1a3ddfb8c3faad11f15940e	uguoxz1163384356159
9	caosl	曹磊	73afc8d3de2961ab1366dff1a1c96dal	ucaosl1174275040195
10	hurigende	胡瑞	1fabdc93b5102776a024e16f24e55a9	uhurigende1196814361332
11	wangzheng01	王正	f89667cc1c882a43849389433155d4ae	uwangzheng011186380644692
12	chenyu03	陈宇	92285db259801a5b5c8508dc7e95e7b7	uchenyu031282548026600
13	chenmj01	陈明	14c775f11476607a2e2ad1abde2c4423	uchenj011282696847722
14	fengtl	冯涛	ac688a10b29a273406c2b5e29526f120	ufengtl1174273440764
15	zhength	郑涛	1dbb6b1094f31e4ee1f508e99abb42ba	uzhength1218088470882
16	dongdw	董冬	8511813fc7084c7fa162cb024799a045	udongdw1218088508264
17	liuchen01	李晨	705a3547a5effaa00db36db407593fea	ulichen011279078382008
18	chenchen	陈晨	9dd40be2c7534bf0496476ca4b07f926	uchench1177486071213
19	linningning	李宁	389eb45909748ec7180b0a34bbe7cbce	uliningning1184641585502
20	pengli	彭利	7b0b673091941734cb3d1ca203723501	upengli1185521639583
21	tianning888	田宁	f443679bf77969e144a65cee0da6f000	utianbing8881185522021138
22	lilan01	李兰	8085b180af62db77436fb0bf1b987845	ulilan011188891688435
23	testgroupleader1	测试组领导1	0235d919acbe2e4d82e04725cd97364e	utestgroupleader11186129848311
24	testgroupleader2	测试组领导2	6a0ab3a30a15a136d4046a374747474	utestgroupleader21186129848311

关闭 显示日志

双击编辑输出动作。

表输出

步骤名称: 表输出

数据库连接: warehouse [编辑...] [新建...]

目标模式: [浏览(B)...]

目标表: USERBASICINFO [浏览(B)...]

提交记录数量: 1000

截剪表: ☐

忽略插入错误: ☐

Specify database fields: ☐

Main options Database fields

Fields to insert:

#	Table field	Stream field
1	ID	ID
2	NICKNAME	NICKNAME
3	PASSWORD	PASSWORD
4	SID	SID
5	CREATETIME	CREATETIME
6	NAME	NAME
7	SEX	SEX
8	RESERVEDCHAR1	RESERVEDCHAR1
9	RESERVEDCHAR2	RESERVEDCHAR2
10	RESERVEDCHAR3	RESERVEDCHAR3
11	RESERVEDCHAR4	RESERVEDCHAR4
12	RESERVEDCHAR5	RESERVEDCHAR5

[Get fields]

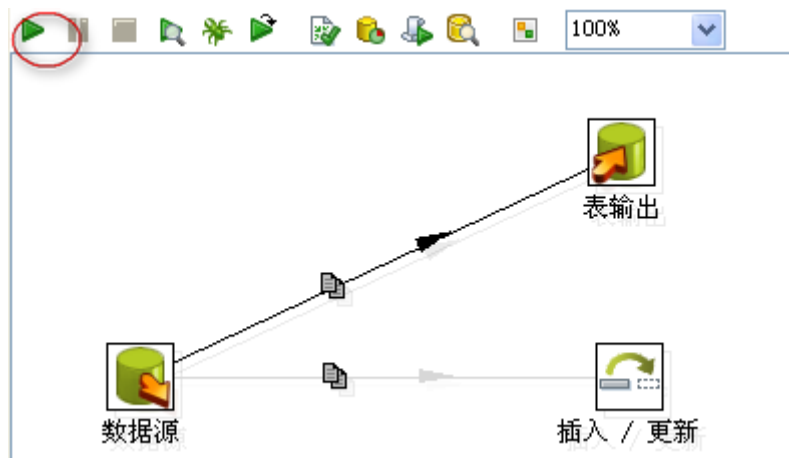
[Enter field mapping]

[确定(O)] [取消(C)] [SQL]

选择要用的字段后点确定，注意字段的约束条件。

说明：现输入和输出表为不同库的两个同构数据表，ID 为主键。

编辑好后点击上面的执行即可。



本地, 远程或集群方式执行

本地执行

远程执行

集群方式执行

远程机器

Pass export to remote server

提交转换

准备执行

开始执行

显示转换

启用安全模式

Clear the log before execution

基本日志

日志级别

重放日期 (yyyy/mm/dd HH:mm:ss)

Parameters

#	Parameter	Value	Default value
1			

参数

#	参数	值
1		

变量

#	变量	值
1	Internal.Job.Filename.Directory	Parent Job File Directory
2	Internal.Job.Filename.Name	Parent Job Filename
3	Internal.Job.Name	Parent Job Name
4	Internal.Job.Repository.Directory	Parent Job Repository Directory

启动 (L)

取消 (C)

左侧参数暂不说明，执行完成后可以观看下方的日志。

et12

100%

数据源

表输出

插入 / 更新

执行结果

执行历史

日志

Step Metrics

性能图

#	步骤名称	复制的记录行数	读	写	输入	输出	更新	拒绝	错误	激活
1	数据源	0	0	346	346	0	0	0	0	已完成
2	表输出	0	346	346	0	346	0	0	0	已完成

etl2

100%

数据源

表输出

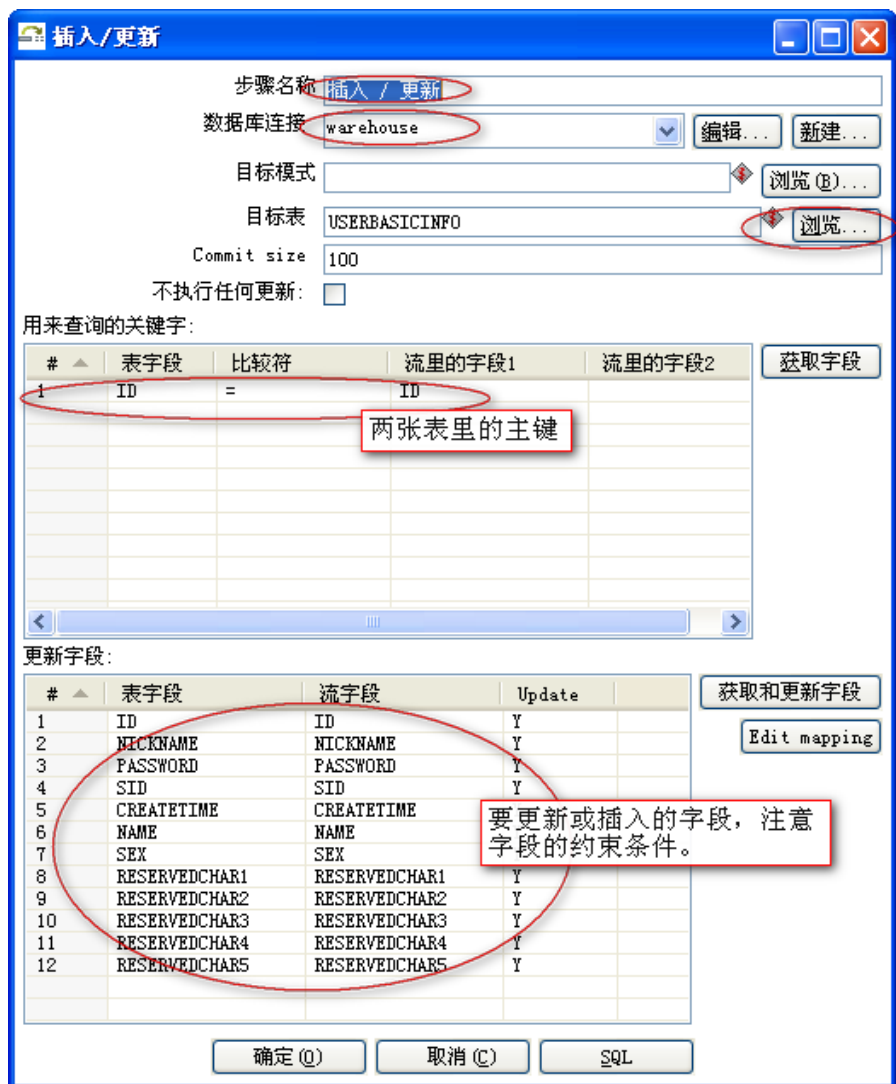
插入 / 更新

执行结果

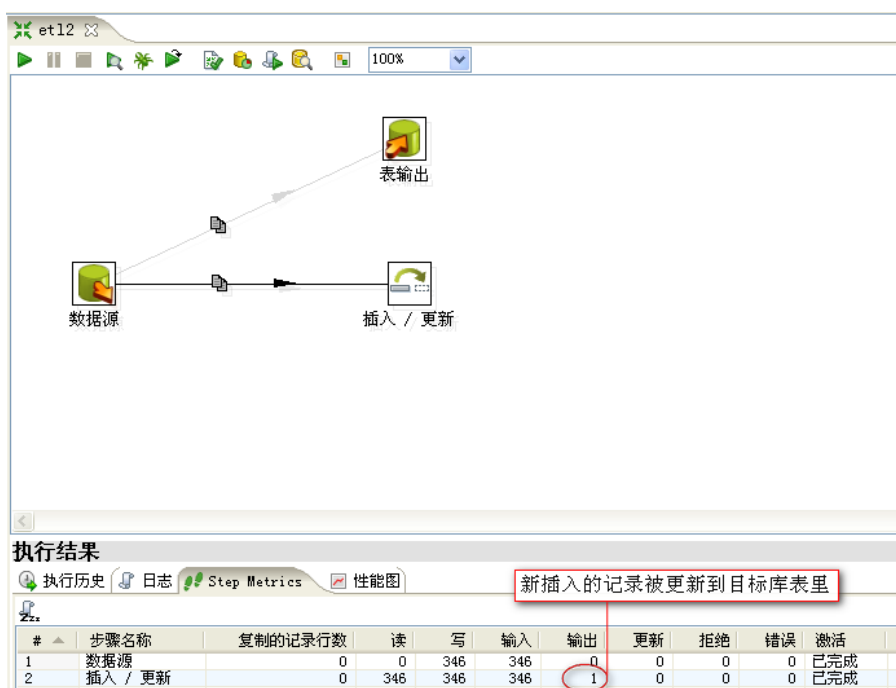
执行历史 日志 Step Metrics 性能图

2010/12/22 13:08:49 - Spoon - 转换已经打开.
2010/12/22 13:08:49 - Spoon - 正在打开转换 [etl2]...
2010/12/22 13:08:49 - Spoon - 开始执行转换.
2010/12/22 13:08:49 - etl2 - 为了转换解除补丁开始 [etl2]
2010/12/22 13:08:49 - Transformation metadata - Natural sort of steps executed in 0ms (3 time p:
2010/12/22 13:08:49 - etl2 - 这项转换可以被回放 回放日期是: 2010/12/22 13:08:49
2010/12/22 13:08:49 - 表输出.0 - Connected to database [warehouse] (commit=1000)
2010/12/22 13:08:50 - 数据源.0 - Finished reading query, closing connection.
2010/12/22 13:08:50 - 数据源.0 - 完成处理 (I=346, O=0, R=0, W=346, U=0, E=0)
2010/12/22 13:08:50 - 表输出.0 - 完成处理 (I=0, O=346, R=346, W=346, U=0, E=0)
2010/12/22 13:08:50 - Spoon - 转换完成!!

双击编辑“插入/更新”动作。



然后点击运行。



执行前，在源库里加条新纪录。确定作业执行正常后可以保存作业，如保存路径到本地磁盘地址 C:\Documents and Settings\w_gaoj\My Documents\etl2.ktr。

6.建立定时执行的 JOB

建立 JOB 的前提是有可以完整执行的作业，上面步骤定制的作业进行保存。

Spoon - etl2

文件(F) 编辑 视图 Action Tools 帮助

新建

打开 CTRL-O

从URL打开文件

Open Recent

关闭 CTRL-F4

Close All

保存 CTRL-S

另存为...

另存为 (VFS)...

打印

从XML文件导入 CTRL-I

Export

退出

转换 CTRL-N

作业(J) CTRL-ALT-N

数据库连接

Validation

Statistics

作业

映射

内联

实验

不推荐的

Bulk loading

History

数据源

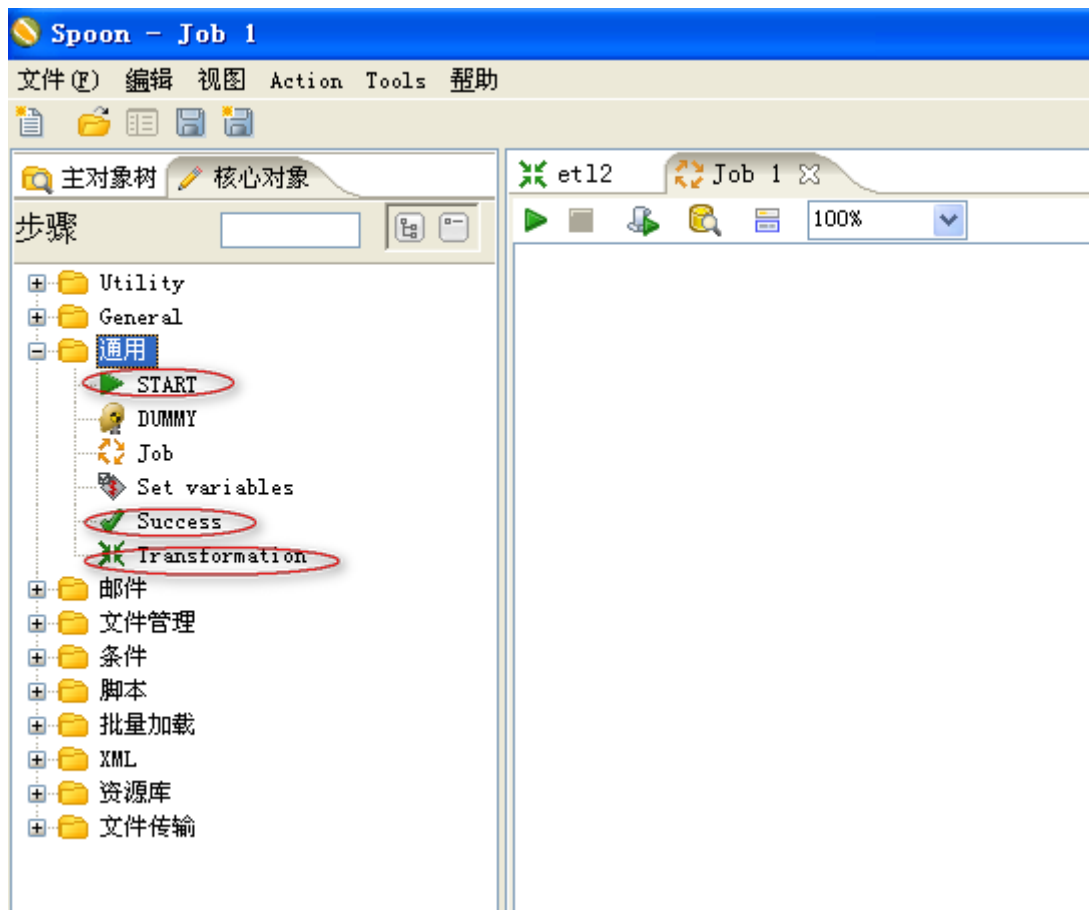
表输出

插入 / 更新

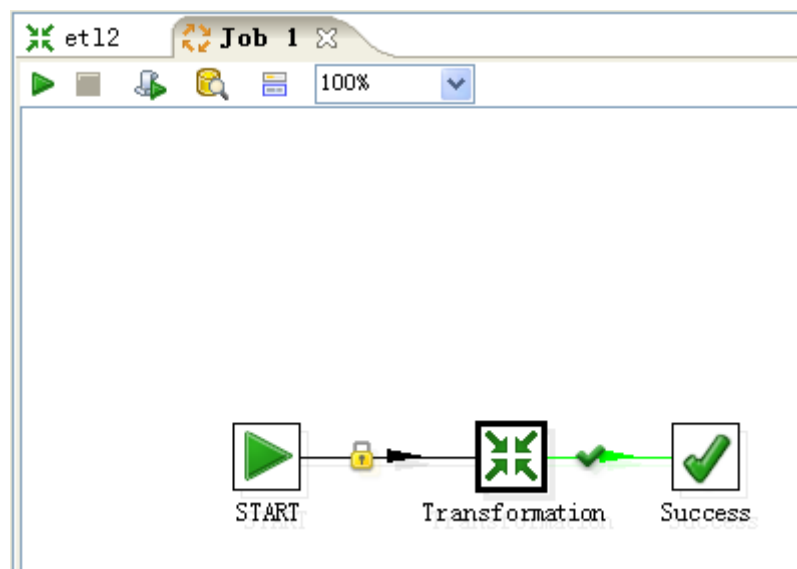
执行结果

执行历史 日志 Step Metrics 性能图

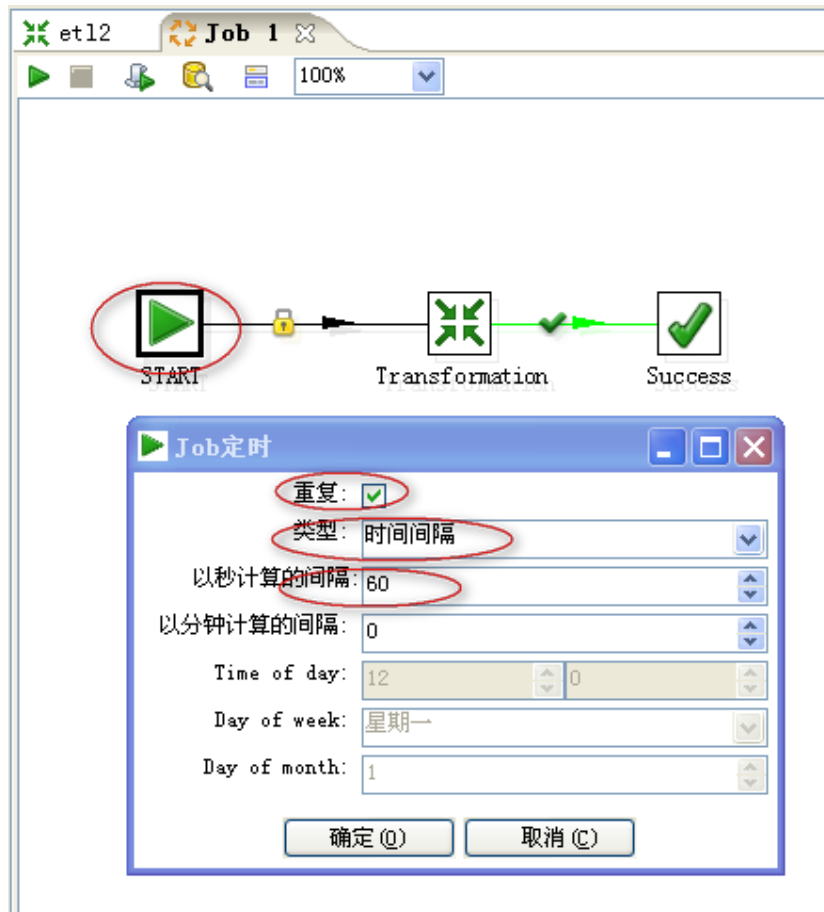
#	步骤名称	复制的记录行数	读	写	输
1	数据源	0	0	346	3
2	表输出	0	346	346	



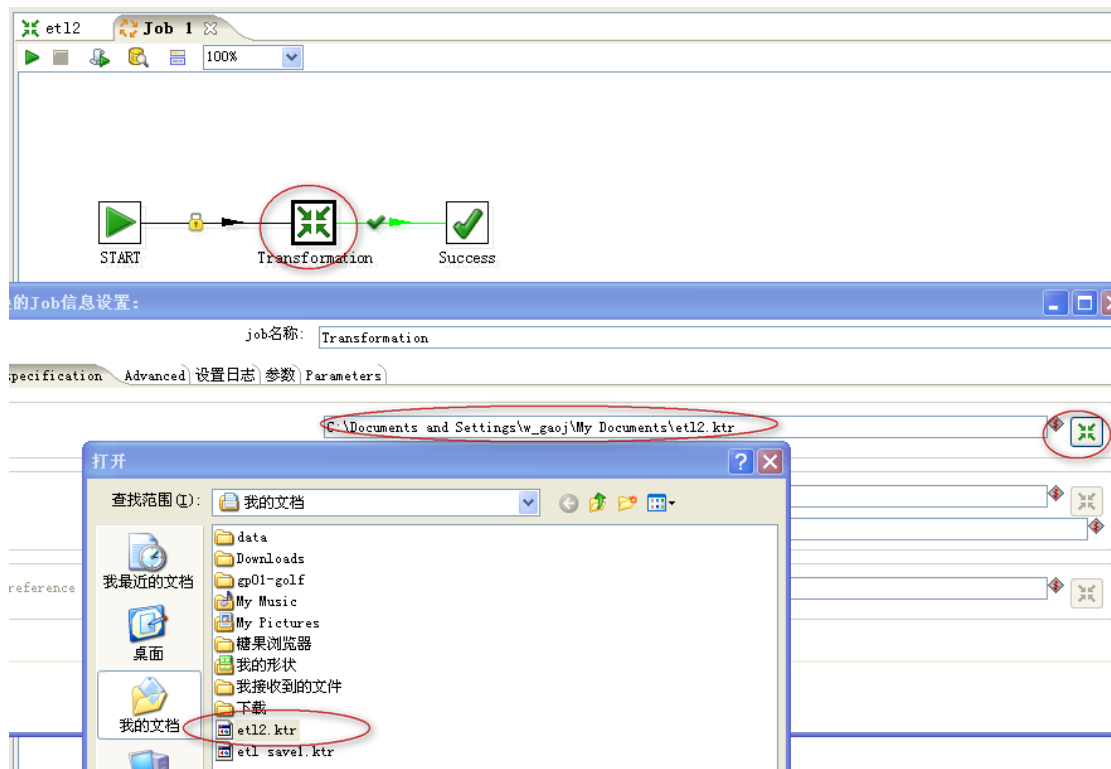
从左侧列表拖拽“START”，“Success”，“Transformation”到工作区并连接，如下图。



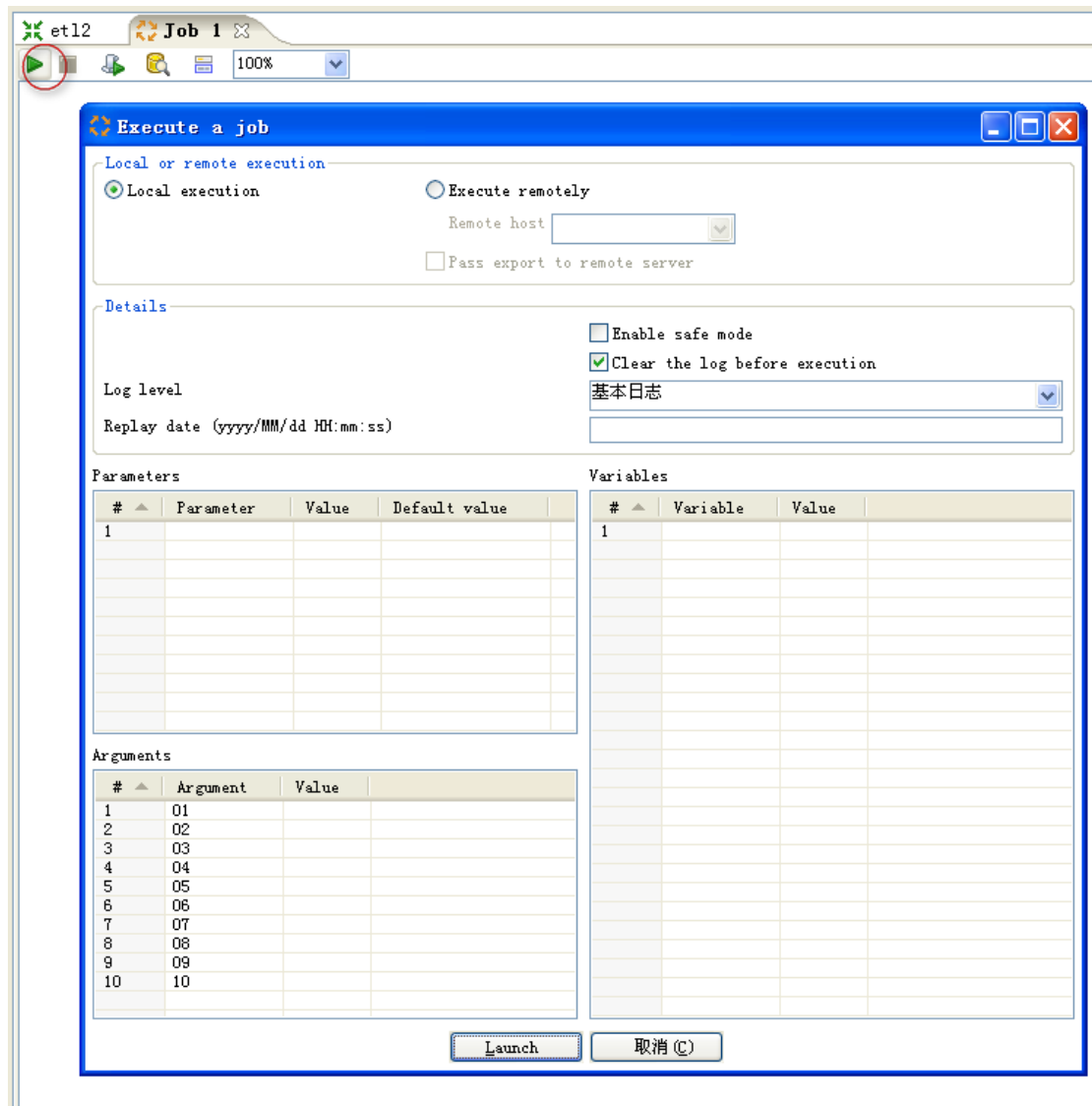
然后双击“START”动作进行编辑。



然后双击编辑“Transformation”活动。



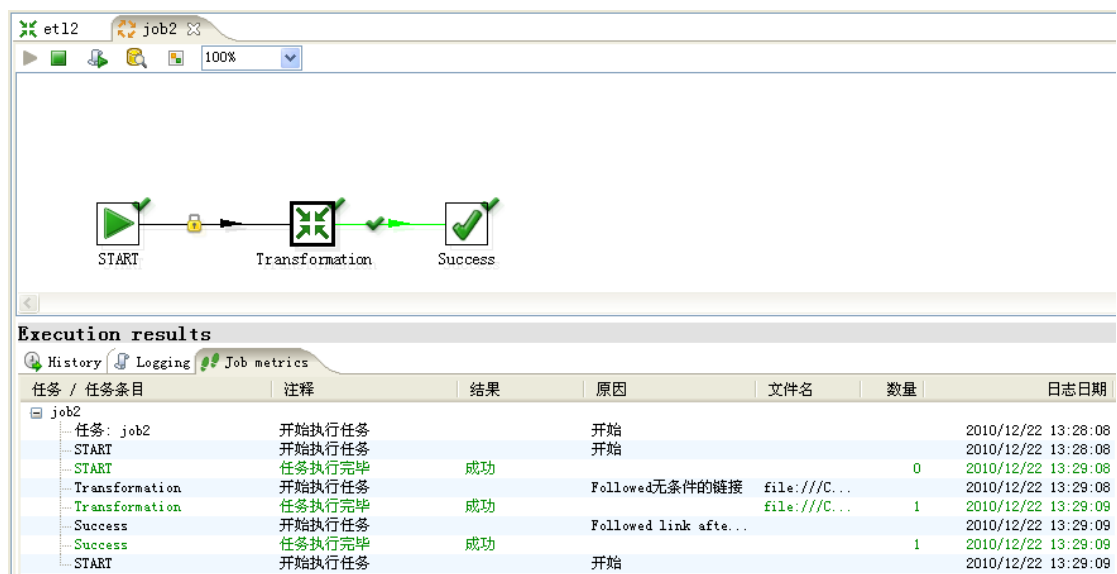
编辑完成后直接点击运行即可。



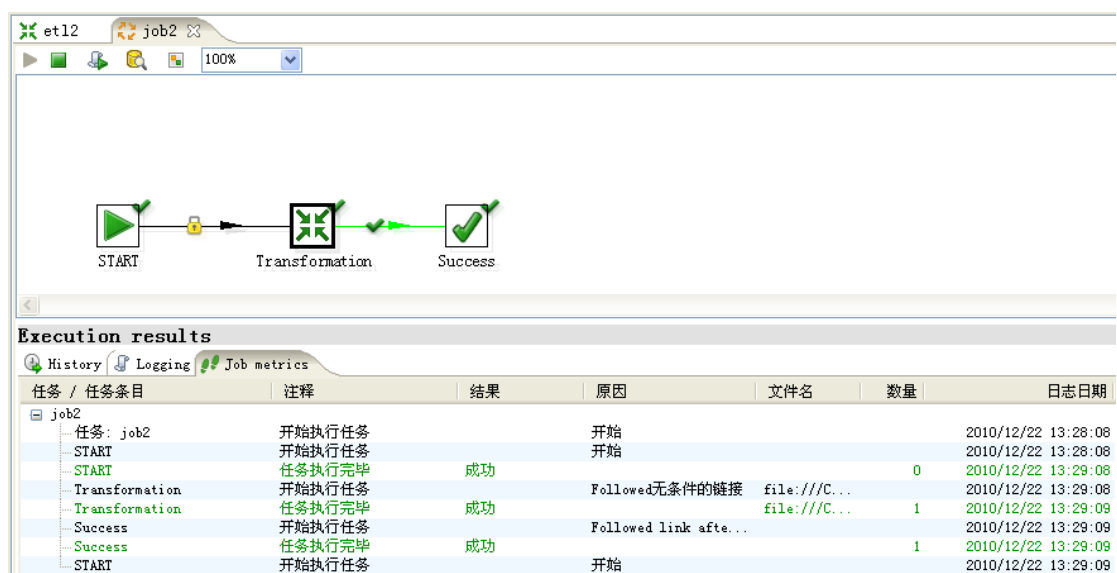
运行后查看日志。



当任务执行以后可以查看相对应的日志，也可自行配置日志。



多次执行的记录结果。



可以返回工作区，查看相应的资源，如下图。

