

技术重塑管理

数据驱动的企业

用友
yonyou

yonyou

UAP 大型企业与组织计算平台

Computing Platform for Large-sized Enterprises and Organizations

| 操作手册 |

UAP DI 操作手册



版权

©2014用友集团版权所有。

未经用友集团的书面许可，本操作手册任何整体或部分的内容不得被复制、复印、翻译或缩减以用于任何目的。本操作手册的内容在未经通知的情形下可能会发生改变，敬请留意。请注意：本操作手册的内容并不代表用友软件所做的承诺。

目录

版权	2
变更记录	9
1. 概述	10
1.1. 名词解释	10
1.2. DI 特性	10
2. 安装部署	11
2.1. 服务器部署	11
2.2. 客户端安装	11
2.2.1. 安装 JDK 和 JRE	11
2.2.2. 运行安装软件	11
3. 主界面介绍	12
3.1. 菜单栏	12
3.2. 工具栏	12
3.3. 项目区	12
3.4. 工作区	13
3.4.1. 转换工作区	13
3.4.2. 作业工作区	13
3.5. 组件区	14
3.5.1. 转换组件	14
3.5.2. 作业组件	16
3.6. 资源管理区	17
3.6.1. AE 元数据	17
3.6.2. 作业	19
3.6.3. 转换	20
3.6.4. 数据库连接	20
3.7. 执行结果监控区	21
3.7.1. 执行历史	21
3.7.2. 日志	22
3.7.3. 作业度量	23
3.7.4. 步骤度量	23
4. 全局应用说明	24
4.1. 搜索元数据	24
4.2. 设置环境变量	25
4.3. 显示环境变量	26
4.4. 显示参数	27
4.5. 设置	28
4.6. 创建数据库连接向导	29
4.7. 复制单表向导	32

4.8.	复制多表向导.....	34
5.	转换步骤介绍	38
5.1.	示例说明.....	38
5.2.	输入组件：文本输入.....	38
5.2.1.	参数说明.....	38
5.2.2.	示例.....	41
5.3.	输入组件：CSV 输入	43
5.3.1.	参数说明.....	43
5.3.2.	示例.....	44
5.4.	输入组件：EXCEL 输入.....	45
5.4.1.	参数说明.....	45
5.4.2.	示例.....	46
5.5.	输入组件：LDAP 输入	46
5.5.1.	参数说明.....	46
5.5.2.	示例.....	47
5.6.	输入组件：META 输入.....	47
5.6.1.	参数说明.....	47
5.6.2.	示例.....	48
5.7.	输入组件：XML 输入	48
5.7.1.	参数说明.....	48
5.7.2.	示例.....	50
5.8.	输入组件：生成记录.....	51
5.8.1.	参数说明.....	51
5.8.2.	示例.....	52
5.9.	输入组件：系统信息.....	53
5.9.1.	参数说明.....	53
5.9.2.	示例.....	53
5.10.	输入组件：表输入.....	54
5.10.1.	参数说明.....	54
5.10.2.	示例.....	54
5.11.	输入组件：配置输入.....	55
5.11.1.	参数说明.....	55
5.11.2.	示例.....	56
5.12.	输出组件：EXCEL 输出.....	57
5.12.1.	参数说明.....	57
5.12.2.	示例.....	58
5.13.	输出组件：META 输出.....	58
5.13.1.	参数说明.....	58
5.13.2.	示例.....	59
5.14.	输出组件：删除.....	59
5.14.1.	参数说明.....	59
5.14.2.	示例.....	59
5.15.	输出组件：插入更新.....	60
5.15.1.	参数说明.....	60

5.15.2.	示例.....	61
5.16.	输出组件：表输出.....	62
5.16.1.	参数说明.....	62
5.16.2.	示例.....	63
5.17.	转换组件：值映射.....	63
5.17.1.	参数说明.....	63
5.17.2.	示例.....	64
5.18.	转换组件：行转列.....	65
5.18.1.	参数说明.....	65
5.18.2.	示例.....	65
5.19.	转换组件：增加常量.....	67
5.19.1.	参数说明.....	67
5.19.2.	示例.....	67
5.20.	转换组件：增加序列.....	68
5.20.1.	参数说明.....	68
5.20.2.	示例.....	69
5.21.	转换组件：字段选择.....	69
5.21.1.	参数说明.....	69
5.21.2.	示例.....	71
5.22.	转换组件：拆分字段.....	72
5.22.1.	参数说明.....	72
5.22.2.	示例.....	73
5.23.	转换组件：排序记录.....	74
5.23.1.	参数说明.....	74
5.23.2.	示例.....	75
5.24.	转换组件：替换 NULL.....	76
5.24.1.	参数说明.....	76
5.24.2.	示例.....	77
5.25.	转换组件：生成 GUID.....	79
5.25.1.	参数说明.....	79
5.25.2.	示例.....	79
5.26.	转换组件：行转列.....	81
5.26.1.	参数说明.....	81
5.26.2.	示例.....	82
5.27.	转换组件：记录去重.....	82
5.27.1.	参数说明.....	82
5.27.2.	示例.....	82
5.28.	转换组件：设置字段.....	84
5.28.1.	参数说明.....	84
5.28.2.	示例.....	84
5.29.	转换组件：连接字段.....	86
5.29.1.	参数说明.....	86
5.29.2.	示例.....	88
5.30.	流程组件：分支.....	88
5.31.	流程组件：空操作.....	91

5.31.1.	参数说明.....	91
5.31.2.	示例.....	91
5.32.	流程组件：过滤记录.....	91
5.32.1.	参数说明.....	91
5.32.2.	示例.....	92
5.33.	脚本组件：SQL 脚本.....	94
5.33.1.	参数说明.....	94
5.33.2.	示例.....	94
5.34.	查询组件：DB 查询.....	95
5.34.1.	参数说明.....	95
5.34.2.	示例.....	95
5.35.	查询组件：META 查询.....	97
5.35.1.	参数说明.....	97
5.35.2.	示例.....	98
5.36.	查询组件：存储过程.....	98
5.36.1.	参数说明.....	98
5.36.2.	示例.....	98
5.37.	查询组件：检查表.....	99
5.37.1.	参数说明.....	99
5.37.2.	示例.....	100
5.38.	查询组件：维度查询.....	100
5.38.1.	参数说明.....	100
5.38.2.	示例.....	101
5.39.	连接组件：合并记录.....	101
5.39.1.	参数说明.....	101
5.39.2.	示例.....	102
5.40.	连接组件：记录连接.....	103
5.40.1.	参数说明.....	103
5.40.2.	示例.....	104
5.41.	质量与统计组件：分组.....	106
5.41.1.	参数说明.....	106
5.41.2.	示例.....	108
5.42.	质量与统计组件：数据检验.....	108
5.42.1.	参数说明.....	108
5.42.2.	示例.....	108
5.43.	作业组件：获取变量.....	110
5.43.1.	参数说明.....	110
5.43.2.	示例.....	111
5.44.	作业组件：设置变量.....	112
5.44.1.	参数说明.....	112
5.44.2.	示例.....	113
5.45.	批量加载组件：ORACLE 加载.....	113
5.45.1.	参数说明.....	113
5.45.2.	示例.....	114
6.	作业项介绍.....	114

6.1.	通用组件：开始.....	114
6.1.1.	参数说明.....	114
6.1.2.	示例.....	114
6.2.	通用组件：空操作.....	116
6.2.1.	参数说明.....	116
6.2.2.	示例.....	116
6.3.	通用组件：作业.....	116
6.3.1.	参数说明.....	116
6.3.2.	示例.....	117
6.4.	通用组件：成功.....	117
6.4.1.	参数说明.....	117
6.4.2.	示例.....	117
6.5.	通用组件：设置变量.....	117
6.5.1.	参数说明.....	117
6.5.2.	示例.....	118
6.6.	通用组件：转换.....	118
6.6.1.	参数说明.....	118
6.6.2.	示例.....	118
6.7.	邮件组件：邮件.....	118
6.7.1.	参数说明.....	118
6.7.2.	示例.....	119
6.8.	邮件组件：UAP 消息.....	119
6.8.1.	参数说明.....	119
6.8.2.	示例.....	119
6.9.	条件组件：检查表.....	120
6.9.1.	参数说明.....	120
6.9.2.	示例.....	120
6.10.	条件组件：检查连接.....	120
6.10.1.	参数说明.....	120
6.10.2.	示例.....	120
6.11.	条件组件：检验值.....	120
6.11.1.	参数说明.....	120
6.11.2.	示例.....	121
6.12.	脚本组件：SQL 脚本.....	121
6.12.1.	参数说明.....	121
6.12.2.	示例.....	121
6.13.	脚本组件：PDM 生成.....	122
6.13.1.	参数说明.....	122
6.13.2.	示例.....	122
6.14.	批量加载组件：ORACLE 到 GBASE.....	123
6.14.1.	参数说明.....	123
6.14.2.	示例.....	123
6.15.	XML 组件：XSD 校验.....	124
6.15.1.	参数说明.....	124
6.15.2.	示例.....	124

6.16.	应用组件：终止作业.....	126
6.16.1.	参数说明.....	126
6.16.2.	示例.....	126
6.17.	应用组件：QDI 调用	126
6.17.1.	参数说明.....	126
6.17.2.	示例.....	127
6.18.	应用组件：表对比.....	127
6.18.1.	参数说明.....	127
6.18.2.	示例.....	128
6.19.	文件传输组件：文件上传.....	129
6.19.1.	参数说明.....	129
6.19.2.	示例.....	129
6.20.	文件传输组件：文件下载.....	129
6.20.1.	参数说明.....	129
6.20.2.	示例.....	130
7.	开始使用 DI.....	130
7.1.	登录服务器.....	130
7.2.	如何新建一个项目.....	131
7.3.	如何导入一个项目.....	131
7.4.	如何导出一个项目.....	133
7.5.	如何打开一个项目.....	134
7.6.	如何创建一个转换.....	135
7.7.	如何创建一个作业.....	135
7.8.	如何嵌套引用作业?	135
7.9.	典型案例.....	136
7.9.1.	案例说明.....	136
7.9.2.	作业调度.....	137
8.	数据库连接管理	143
8.1.	如何新建数据库连接?	143
8.2.	如何编辑/删除数据库连接?	146
8.3.	如何导入数据库对象?	146
9.	调度管理.....	148
9.1.	设置调度策略.....	149
9.2.	查看执行历史.....	150
9.3.	查看所属任务.....	150
9.4.	查看执行日志.....	150
附录	151
作业节点状态说明		151

变更记录

序号	章节号	章节名称	变更主要内容	备注

1. 概述

1.1. 名词解释

- 数据集成 (Data integration): 数据集成是将不同来源与格式的数据逻辑上或物理上进行集成的过程。
- 作业 (job): 完成整个工作流的控制, 好比工厂里的管理。每个作业中可以包含多个转换, 并设置转换之间的关联关系, 作业支持嵌套。
- 转换 (transformation): 完成针对数据的基础转换, 好比工厂里的流水线, 每个组件相当于一个员工。

1.2. DI 特性

1. 作业, 每个作业中可以包含多个转换, 并设置转换之间的关联关系, 作业支持嵌套。
 - 1) 作业支持调度策略, 包括月, 周, 天, 可以配置周期性调度, 并将调度任务加到现有的调度队列中。
 - 2) 可以完成 Oracle 数据库到 Gbase 数据库的批量加载。
 - 3) 支持文件的上传和下载。
 - 4) 支持原有的 QDI 作业的执行和调度。
 - 5) 支持立即运行, 并对执行结果进行监控。
 - 6) 支持两个库之间的表对比, 并可以将对比 SQL 输出到文件中。
 - 7) 支持将调度执行结果, 通过邮件通知。
2. 转换, 可以支持 46 种转换步骤。
 - 1) 支持数据的批量加载。
 - 2) 支持调用存储过程。
 - 3) 支持立即运行, 并对执行结果进行监控。
 - 4) 支持数据质量控制, 包括字段类型, 数据长度等。
3. 其他特性
 - 1) 支持离线设计。
 - 2) 支持以项目为单位的导入和导出。

- 3) 支持集群调度管理。
- 4) 支持性能和日志监控。
- 5) 全新的用户界面，新的用户界面更加注重用户交互，界面更加美观，操作更加简便灵活。

2. 安装部署

2.1. 服务器部署

DI 的服务器集成在 UAP 平台上，部署了 UAP 平台即部署了 DI 服务器。具体方法参考手册《UAP 平台安装手册》。

2.2. 客户端安装

2.2.1. 安装 JDK 和 JRE

step1. 安装 JDK。

注意：建议安装 1.6 版本，安装路径不要有空格及中文。

step2. 配置环境变量 JAVA_HOME，如下图所示：



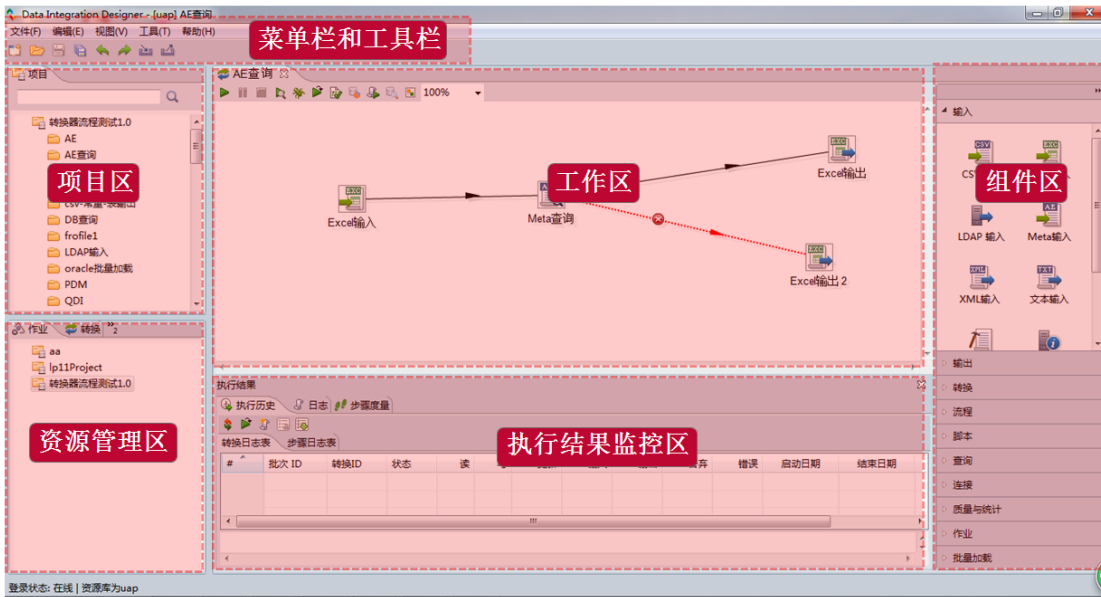
step3. 配置环境变量 Path，在最后面加上，注意分号：

;%JAVA_HOME%\bin;%JAVA_HOME%\jre\bin

2.2.2. 运行安装软件

解压缩安装包 AE-DI-Designer，运行 Data Integration Designer.bat 文件，即可。

3. 主界面介绍



3.1. 菜单栏

DI 主界面的左上角是菜单区。

文件(F) 编辑(E) 视图(V) 工具(T) 帮助(H)

3.2. 工具栏

菜单栏下方是工具栏。



3.3. 项目区

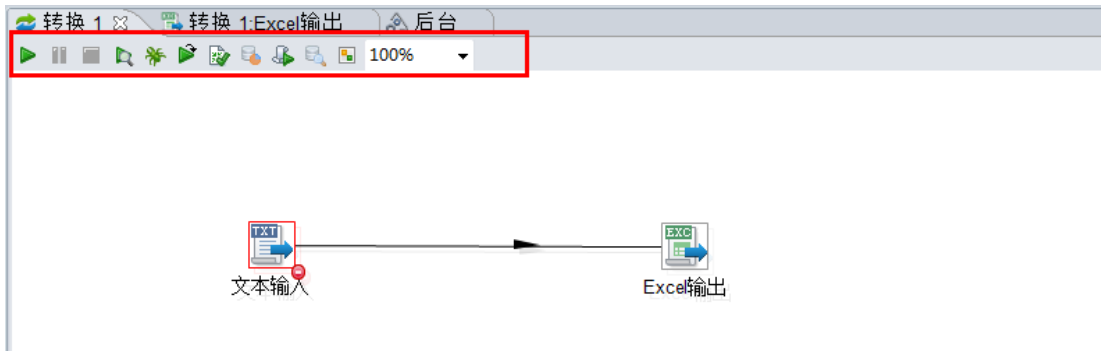
DI 主页面左上角为项目区，即当前用户的工作区，在当前项目下创建的作业和转换都会在此区域列出。

项目区的所有作业和转换都可以通过拖拽的方式，被引用。例如，新建一个作业，名为“文档作业”，选中左侧项目区中的作业 1，拖拽到“文档作业”的设计去，则“文档作业”中，就引用了作业 1，如下图所示：

双击“作业 1”，可以看到这个作业是通过资源库引用指定的，将此作业重命名为“引用作业 1”，如下图所示：

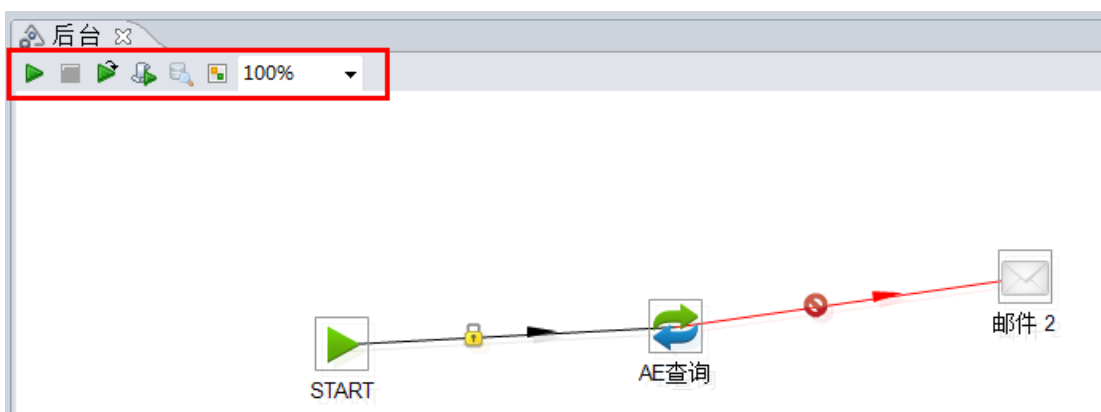
3.4. 工作区

3.4.1. 转换工作区







功能按钮	按钮说明
	运行这个转换
	暂停正在执行的转换
	停止正在执行的转换
	预览这个转换
	调试这个转换
	重放这个转换
	校验这个转换
	分析这个转换在数据库的影响
	产生需要运行这个转换的 SQL
	选择一个可用的数据库连接
	隐藏执行结果面板

3.4.2. 作业工作区



功能按钮	按钮 tips
	运行这个转换
	停止正在执行的转换

	重放这个转换
	产生需要运行这个转换的 SQL
	选择一个可用的数据库连接
	隐藏执行结果面板

3.5. 组件区

3.5.1. 转换组件

组件图标	组件名称	组件类别	备注说明
	CSV 输入	输入	可以从 CSV 文件里面读取数据。
	Excel 输入	输入	可以从 excel 文件里面读取数据。
	LDAP 输入	输入	可用于从 LDAP 服务器中读取有关用户、角色的信息和其他数据。
	Meta 输入	输入	可以从 AE 数据库中读取元数据（语义模型、语义元数据、数据字典）。
	XML 输入	输入	可以从 Designer 支持的系统的 XML 文件里面读取数据。
	文本输入	输入	可以从 Designer 支持的系统的文本文件里面读取数据。
	生成记录	输出	输出一定数量的行，缺省为空。可选包括一定数量的静态字段。
	系统信息	输入	可以获取 DI Designer 环境的系统信息，下表中为系统信息的具体含义。
	表输入	输入	可以通过数据库连接和 SQL 语句读取数据库中的数据，一些基本的 SQL 语句可以自动生成。
	配置输入	输入	从 Java 属性文件中读取。
	Excel 输出	输出	可以写入数据到一个或者多个 Excel 文件中。
	Meta 输出	输出	可以存储信息到 AE 元数据数据库表中
	删除	输出	利用查询关键字在表中搜索行。如果行找到，就删除记录。
	插入更新	输出	利用查询关键字在表中搜索行。如果行没有找到，就插入行。如果能被找到，并且要被更新的字段没有任何改变，就什么也不做。如果有不同，行就会被更新。

	维度更新	输出	步骤可以存储信息到数据库表中
	表输出	输出	可以存储信息到数据库表中.
	值映射	转换	简单的映射字符串, 从一个值映射到另一个值。
	列转行	转换	通过查询键值对来反向规格化数据。也可以立即转换数据类型。
	增加常量	转换	主要是添加常量到流中。
	增加序列	转换	在流中增加一个序列。
	字段选择	转换	可用于在数据流中选择、删除、重命名、更改数据类型以及配置字段的长度和精度。
	拆分字段	转换	允许你根据分隔符来拆分字段
	排序记录	转换	可根据指定的字段以及是按升序还是降序对行进行排序。
	替换 null	转换	替换 null 值步骤允许根据一定的条件来替换 null 值
	生成 GUID	转换	用来生成指定长度 GUID, 可用来生成数据库表主键
	行转列	转换	将行数据转化为不同列
	记录去重	转换	此步骤用来去掉输入流中的相同记录
	设置字段	转换	这个步骤允许你可以根据需要设置字段值
	连接字段	转换	可以根据需要设置字段值, 用于将多个字段连接为一个目标字段。
	分支	流程	这个步骤节点所做的事情类似编程语言中 Switch/Case 所做的事情, Switch/Case 步骤节点根据不同的值会进行不同的操作
	空操作	流程	空操作不做任何事。它的主要作用是, 在你想测试什么的时候, 充当一个占位符。例如有一个转换, 你至少需要两个彼此连接的步骤。如果你想测试文本文件输入步骤, 你可以将它连接到一个空操作步骤。
	记录过滤	流程	允许你根据条件和比较符来过滤记录。一旦这个步骤连接到先前的步骤中, 你可以简单的单击“”, “=”和 “” 区域来构建条件
	SQL 脚本	脚本	可以执行 SQL 脚本, 或者在转换初始化的时候执行, 或者在步骤的每一个输入行执行。

	DB 查询	查询	允许你在数据库表中查找值
	Meta 查询	查询	允许你在 AE 元数据数据库表中查找值
	存储过程	查询	允许你运行一个数据库存储过程，获取返回结果
	检查表	查询	检查某一表或者某些表在指定数据库、指定模式中是否存在
	维度查询	查询	
	合并记录	连接	合并行可用于比较两行数据流。
	记录连接	连接	
	分组	质量与统计	用于计算限定字段组的值。
	数据检验	质量与统计	用来检查输入的数据质量
	获取变量	作业	允许你从任务中或者虚拟机获取变量
	设置变量	作业	允许你在一个任务中或者虚拟机中设置变量。它仅仅可以用一行数据来设置变量值
	Oracle 加载	批量加载	允许你大批量加载数据到 Oracle 数据库。

3.5.2. 作业组件

组件图标	组件名称	组件类别	备注说明
	开始	通用	作业执行的开始步骤，并且只能有一个。
	DUMMY	通用	此作业项什么都不做，可以用作流程分支。
	作业	通用	可以引用已有的作业并执行
	成功	通用	清除前面步骤的失败状态，并强制转换为成功
	设置变量	通用	根据设计需要用来定义作业和转换中用到的变量
	转换	通用	可以引用已有的转换并执行
	UAP 邮件	邮件	可以发送执行结果到 UAP 消息中心。
	邮件	邮件	可以用于发送邮件

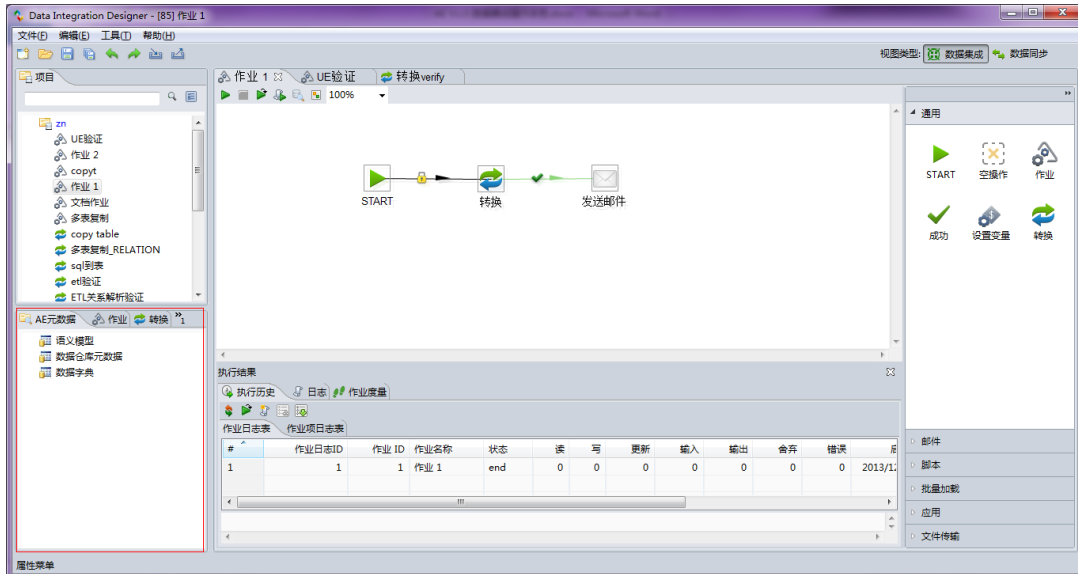
	检查表	条件	可以用于检查表是否存在。
	检查连接	条件	可以用于检查某一数据库连接是否有效可用
	检验值	条件	可以用于检验结果流里字段或变量的值。
	PDM 生产	脚本	可以通过 PDM 文件生成数据库对象。当物理模型与目标库的数据类型一致时支持 Oracle, gbase, sqlserver, db2. 当物理模型与目标库的数据类型不一致时只支持物理模型为 Sqlserver2005
	SQL 脚本	脚本	可以用执行 SQL 语句
	Oracle 到 GBase	批量加载	可以用于从 Oracle 批量载入数据到 GBase
	XSD 校验	XML	验证 Xml 的结构
	QDI 调用	应用	可以用于执行 QDI 作业。
	中止作业	应用	可以用于中止正在执行的作业
	表对比	应用	可以用于执行全表对比
	文件上传	文件传输	可以用于执行文件上传到文件服务器
	文件下载	文件传输	可以用于执行从文件服务器下载文件到本地

3.6. 资源管理区

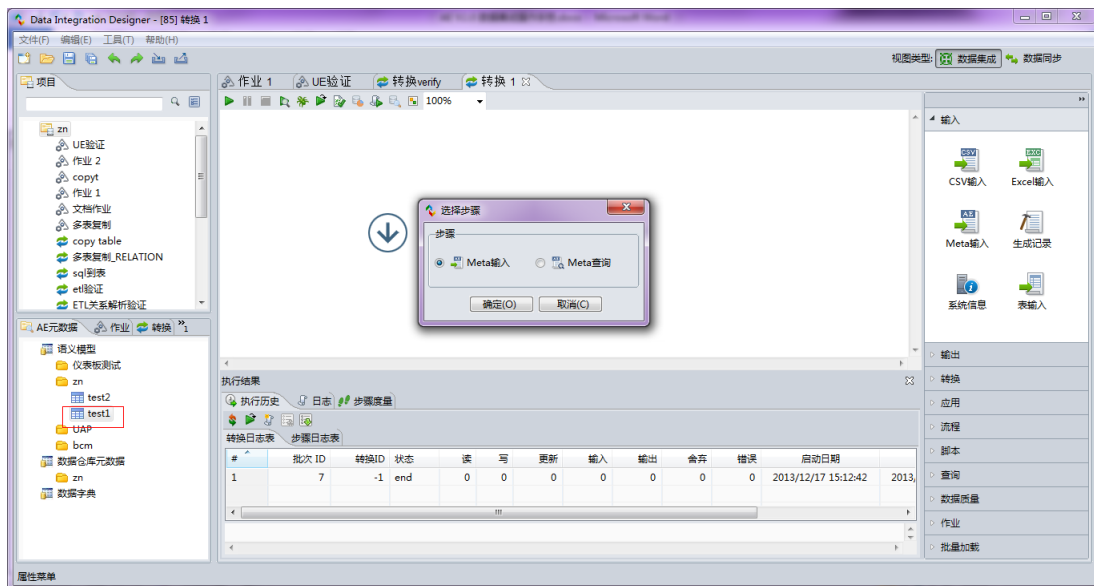
资源管理区，顾名思义此区域完成对 DI 资源的管理，其中包括 AE 元数据，作业，转换和数据库连接，接下来我们对每种资源进行说明。

3.6.1. AE 元数据

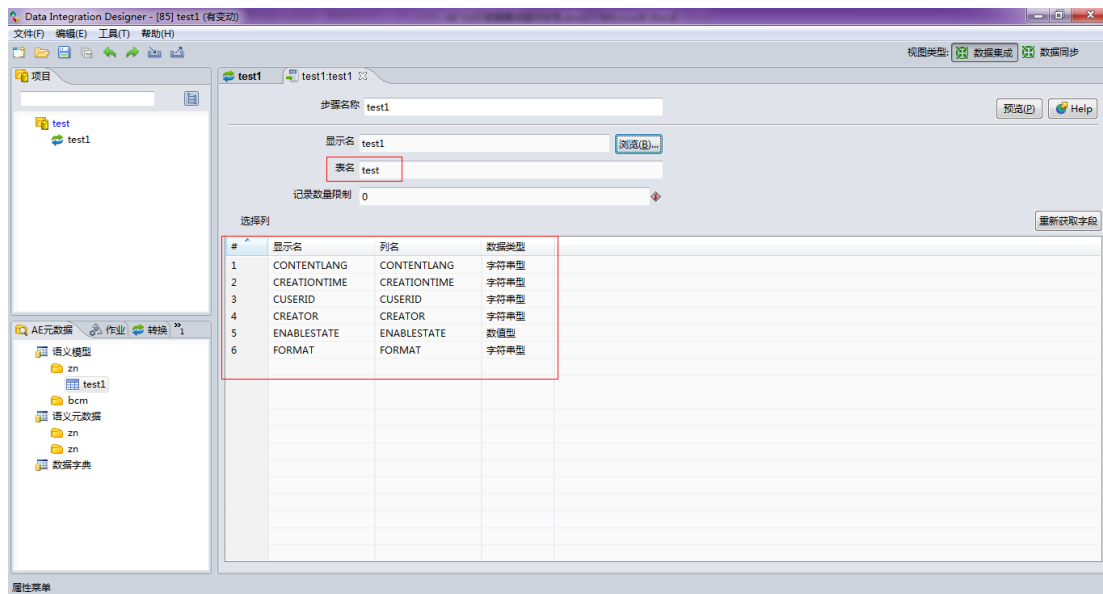
AE 元数据，将语义层的模型，表，数据字典都直接取出来作为资源供 DI 直接使用，AE 元数据共分为三类，语义模型，语义元数据和数据字典。各个类目下分别列出了了 DI 所连的服务上的所有语义模型，语义元数据的表和数据字典，如下图所示：



AE 元数据下所包含的模型和表，都可以作为资源，拖拽到右侧使用，例如将语义模型下面，zn 分类下，名字为“test1”的语义模型拖到右侧转换中，会弹出提示框，将此语义模型作为输入还是作为查询，如下图所示：



选择“Meta 输入”，点击确定，然后双击 test1 的转换步骤，可以看到所有的数据库的设置以及字段信息都已经获取出来了，如下图：

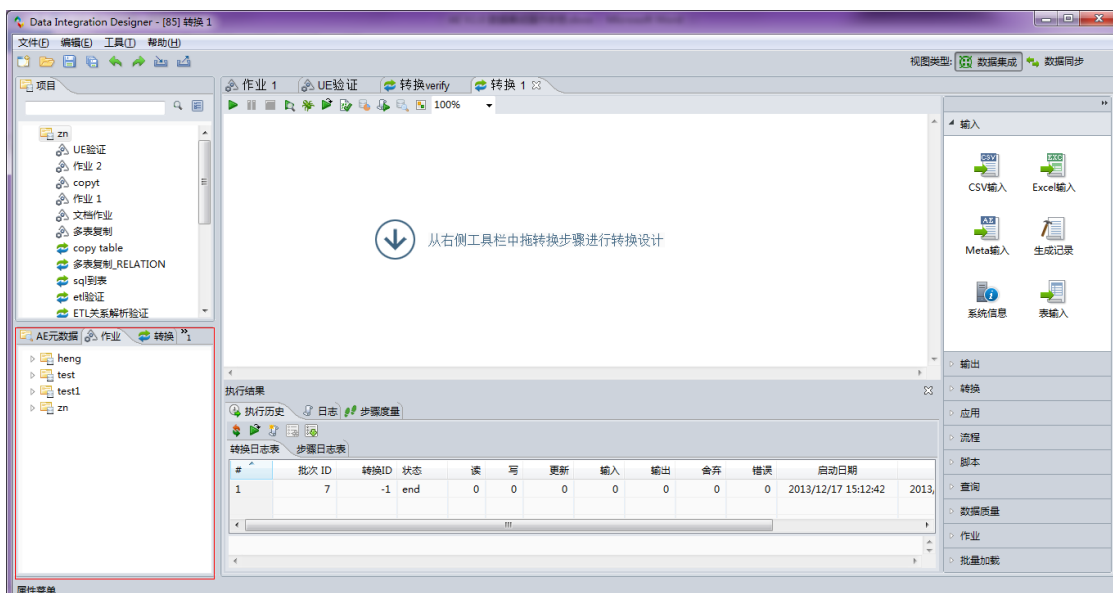


上述的拖入操作，相当于从“工具箱”拖入一个“AE 元数据输入”的转换器，然后双击该转换器，配置完相应的设置，可以看出通过左侧资源管理处的拖拽，可以减少很多操作步骤。

同样，我们还可以直接拖拽“语义元数据”和“数据字典”中的表，完成相应的功能。

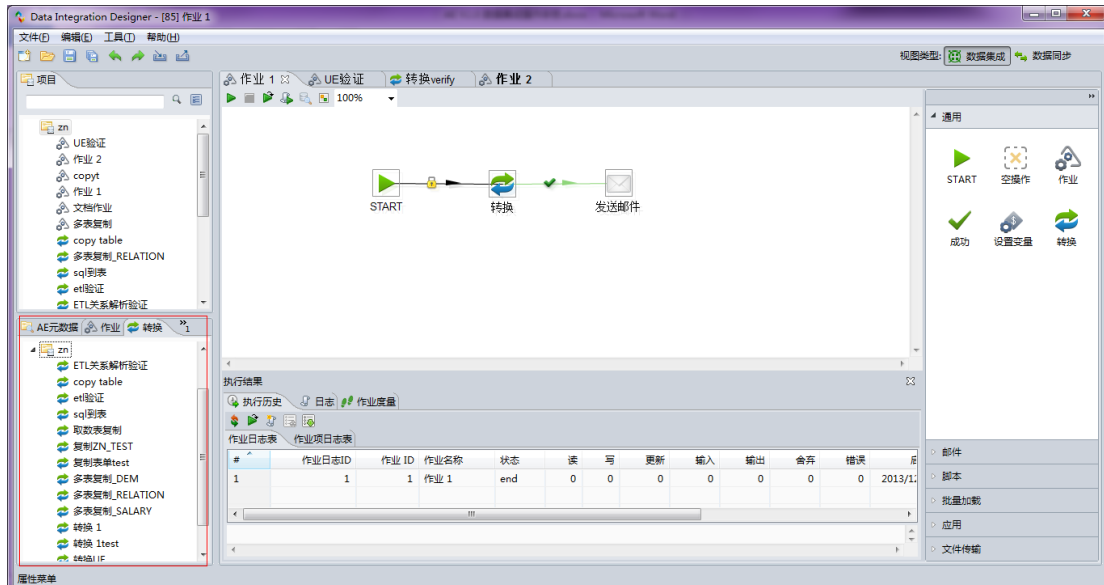
3.6.2. 作业

单击资源管理中的“作业”页签，在此页签中，列出了服务端所有项目包含的作业，作业按照对应的项目分类。如下图所示：



3.6.3. 转换

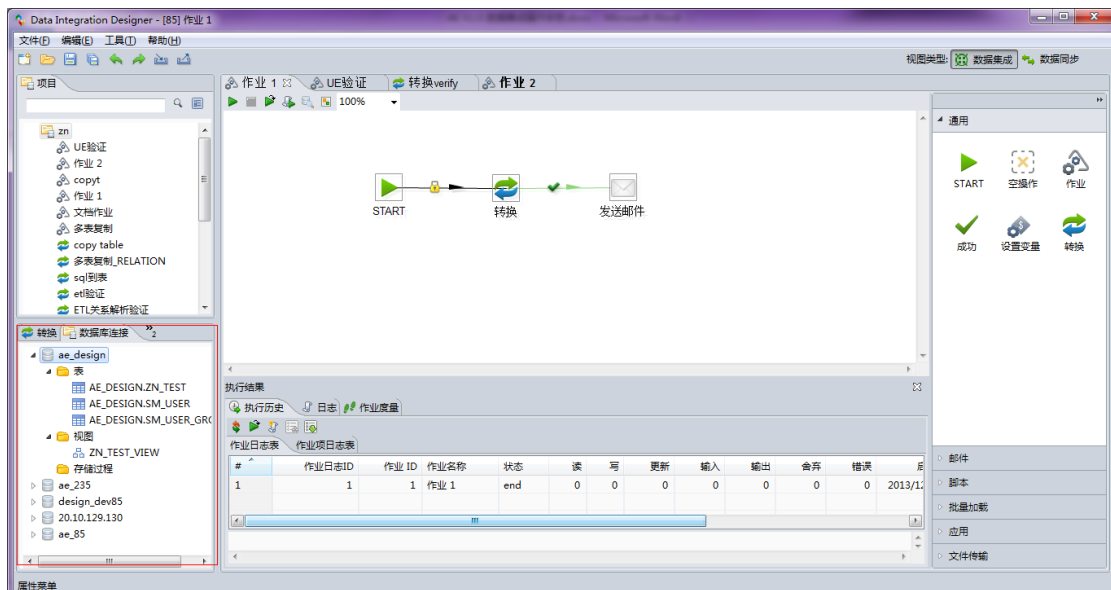
单击资源管理中的“转换”页签，在此页签中，列出了服务端的所有项目包含的转换，转换按照对应的项目分类。如下图所示：



每个转换都可以通过拖拽的方式，完成转换的引用，类似项目区中转换的引用，由于操作步骤跟作业的引用方式相同，在此就不再举例。

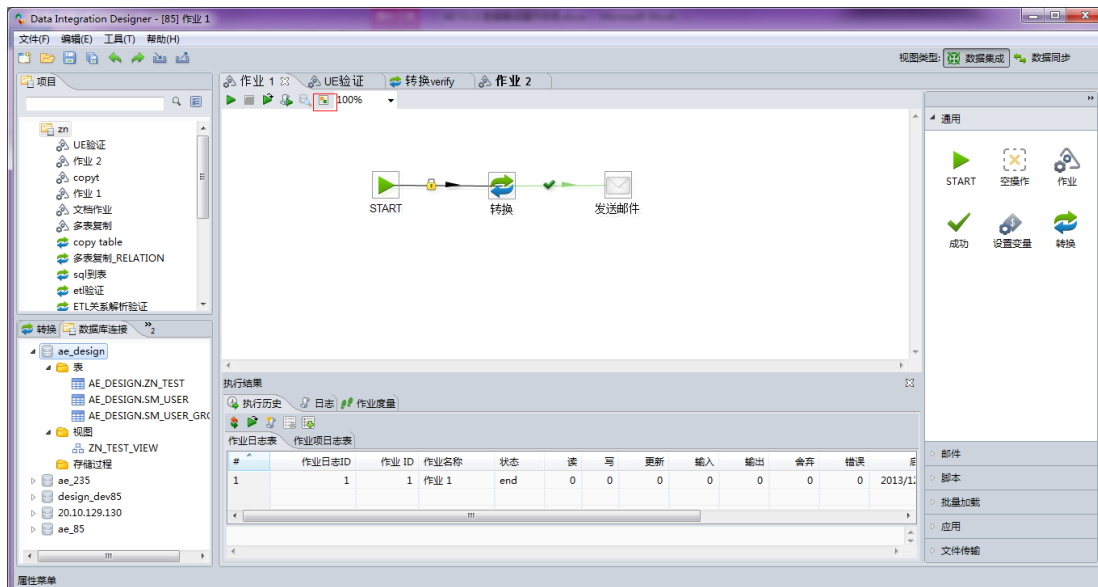
3.6.4. 数据库连接

单击资源管理中的“数据库连接”页签，在此页签中，列出了服务端的所配置的数据库连接，即服务端配置的数据源。默认情况下，数据库连接下没有任何对象，用户可以根据自己的需要，将所需的对象导入到资源管理区，数据库连接下的对象包括：表，视图和存储过程，如下图所示：



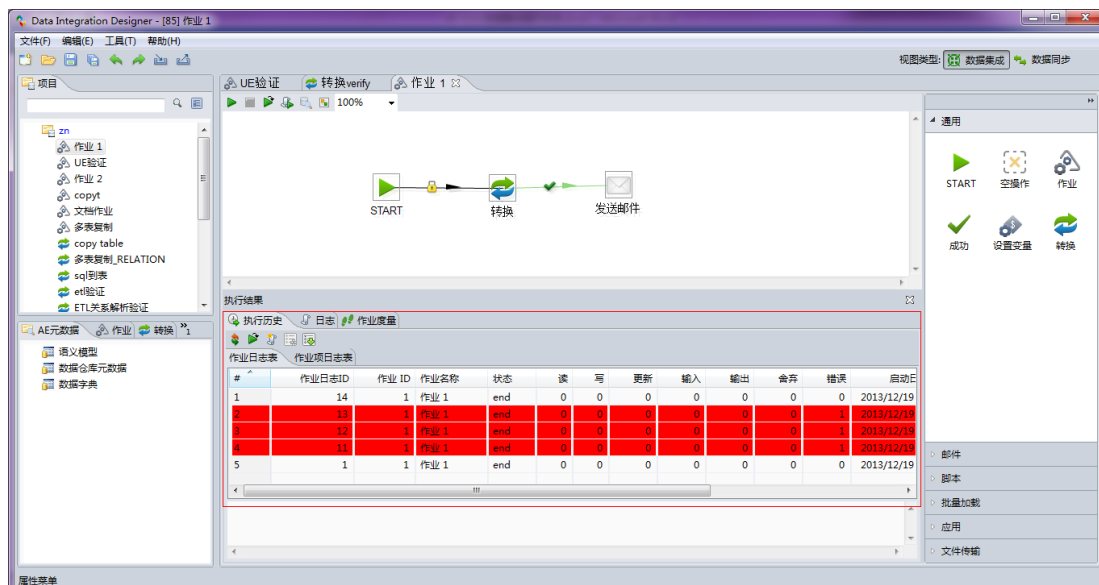
3.7. 执行结果监控区

执行结果监控区，位于 DI 主界面的正下方，完成对执行结果的监控，包括查看执行历史，执行日志和作业度量。可以通过单击工具栏上的按钮，显示和隐藏执行结果监控区，如下图所示：



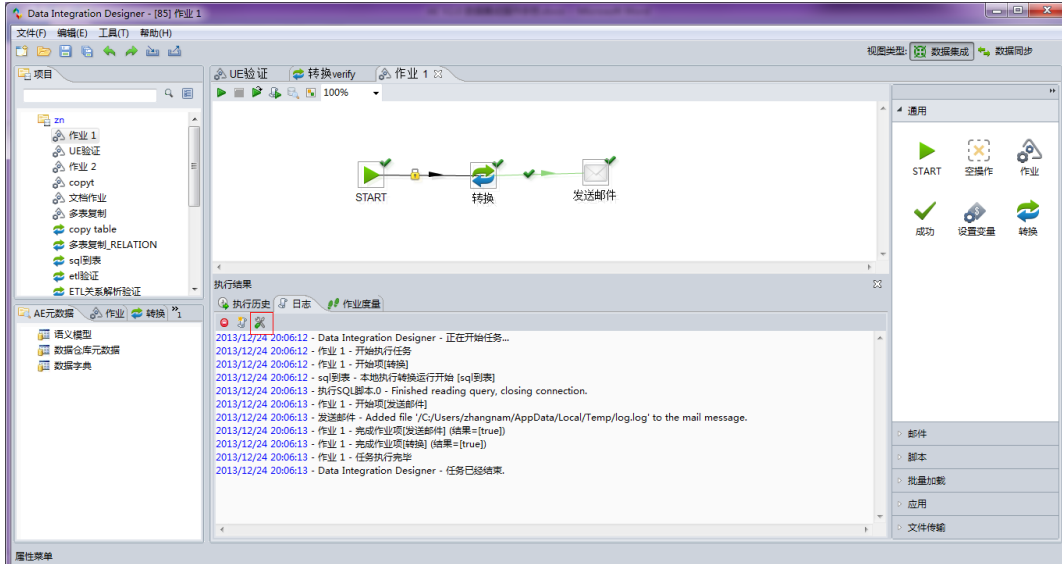
3.7.1. 执行历史

打开作业 1，单击工具栏上的“运行”按钮，运行当前作业，执行历史页签，列出了当前作业的执行历史，我们可以通过工具栏上的<回放>按钮，完成回放功能，我们还可以查看所属当前作业的作业项的执行情况，如下图所示：

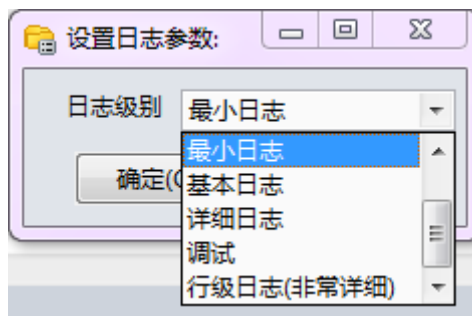


3.7.2. 日志

在<日志>页签，记录了当前作业执行的日志，错误的日志会使用红色的文字标记。
单击工具栏上方的<日志设置>按钮，可以对输出日志的级别进行设置，如下图所示：



DI 提供了 7 中输出日志的级别，分别为没有日志，错误日志，最小日志，基本日志，详细日志，调试和行级日志。如下图：



没有日志：不显示任何输出。

错误日志：仅显示错误信息。

最小日志：小时最小的日志，只输出作业的开始和结束。

基本日志：默认的日志级别。

详细日志：在基本日志的基础上，更加详细，包含了每个步骤的详细处理过程。

调试：调试目的，调试输出。

行级日志：打印出每一行记录的信息。

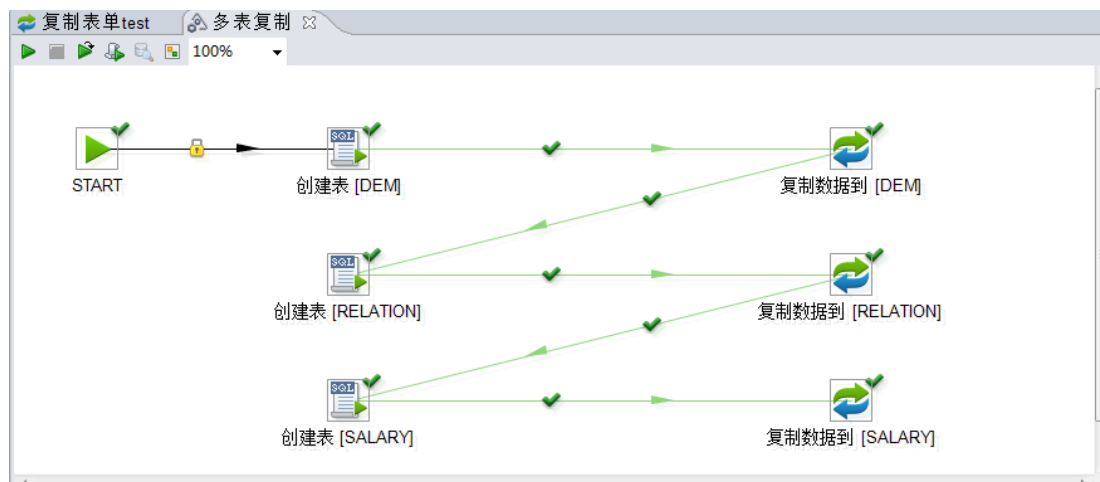
从日志中我们可以分析出作业和转换的运行状态和结果。作业运行状态有两种：**true**（成功）和 **false**（失败），注意的是运行结果失败，并不代表运行异常和出错，有的时候只是用来控制流程的一种决策和手段。对于转换的日志输出中会有几个关键字，说明如下：

l: 当前步骤生成的记录数（从表输入或文件读入）

- O: 当前步骤输出的记录数（输出到表或文件）。
- R: 当前步骤从前一个步骤读取的记录数。
- W: 当前步骤向后面步骤抛出的记录数。
- U: 当前步骤更新过的记录数。
- E: 当前步骤处理出错的记录数。

3.7.3. 作业度量

通过作业度量窗口，我们可以看到当前作业下的每个作业项的运行结果和相关信息。例如，作业“多表复制”完成了三个表的复制，流程如下图所示，点击运行之后，在作业度量窗口，可以查看每个作业项的执行结果，如下图所示：

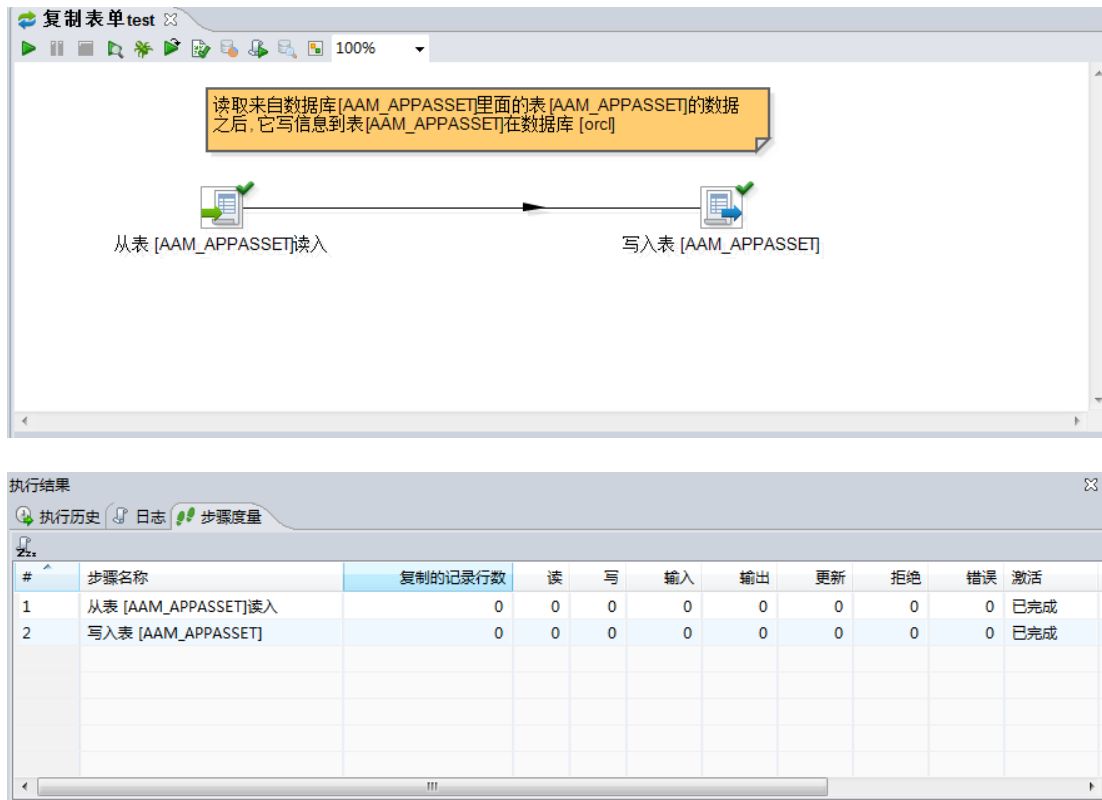


任务 / 任务条目	注释	结果	原因	文件名
START	开始执行任务		开始	
START	任务执行完毕	成功		
创建表 [DEM]	开始执行任务		Followed无条件的链接	
创建表 [DEM]	任务执行完毕	成功		
复制数据到 [DEM]	开始执行任务		Followed link after success	
复制数据到 [DEM]	任务执行完毕	成功		
创建表 [RELATION]	开始执行任务		Followed link after success	
创建表 [RELATION]	任务执行完毕	成功		
复制数据到 [RELATION]	开始执行任务		Followed link after success	
复制数据到 [RELATION]	任务执行完毕	成功		
创建表 [SALARY]	开始执行任务		Followed link after success	
创建表 [SALARY]	任务执行完毕	成功		
复制数据到 [SALARY]	开始执行任务		Followed link after success	
复制数据到 [SALARY]	任务执行完毕	成功		
任务: 多表复制	任务执行完毕	成功	完成	

3.7.4. 步骤度量

通过步骤度量窗口，我们可以看到当前转换下的每个步骤的运行结果，此结果跟上面介绍转换的日志输出结果一致，包括输入，输出，读写等。例如，转换“复制表单 test”完成的是单表表的复制，包含两

个步骤，流程如下图所示，点击运行之后，在步骤度量窗口，可以查看每个步骤的的执行结果，如下图所示：

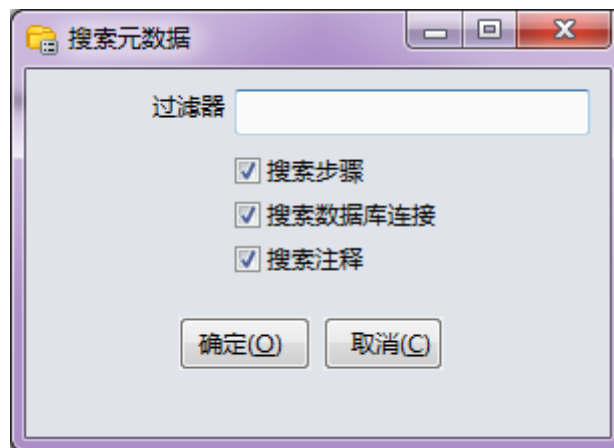


4. 全局应用说明

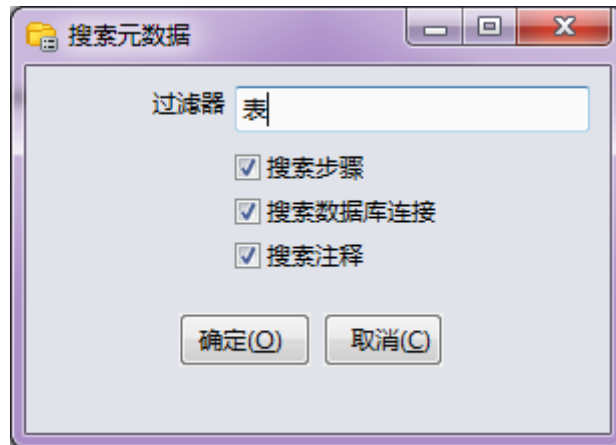
4.1. 搜索元数据

搜索某个转换/作业下的元数据信息。

选择<编辑>菜单下的<搜索元数据>菜单项，弹出“搜索元数据”对话框，如下图所示：



在弹出框中，勾选搜索范围，填写过滤条件。例如：勾选<搜索步骤>复选框，在过滤器中输入“表”，如下图所示：



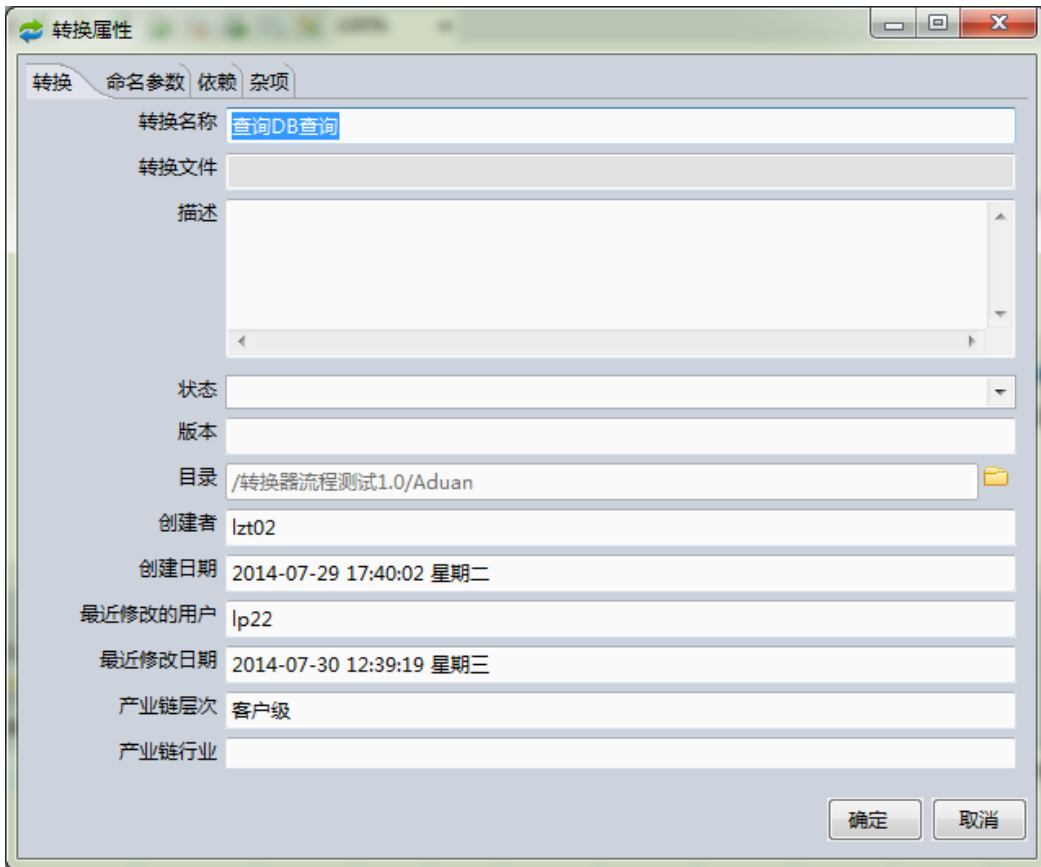
点击<确定>按钮，则搜索出步骤中，带有“表”字样的元数据信息，如图所示：

#	转换/作业	步骤/数据库连接/注释	字符串	位置
1	/zn/转换 1	检查表是否存在	检查表是否存在	Step name
2	/zn/转换 1	检查表是否存在	result	resultfieldname
3	/zn/转换 1	表输出	表输出	Step name
4	/zn/转换 1	表输出	1000	commitSize
5	/zn/转换 1	表输出	false	truncateTable (Boolean)
6	/zn/转换 1	表输出	false	ignoreErrors (Boolean)
7	/zn/转换 1	表输出	false	useBatchUpdate (Boolean)
8	/zn/转换 1	表输出	true	partitioningEnabled (Boolean)
9	/zn/转换 1	表输出	false	partitioningDaily (Boolean)
10	/zn/转换 1	表输出	true	partitioningMonthly (Boolean)
11	/zn/转换 1	表输出	false	tableNameInField (Boolean)

4.2. 设置环境变量

选择<编辑>菜单下的<设置环境变量>菜单项，弹出“设置变量”对话框，如下图所示：

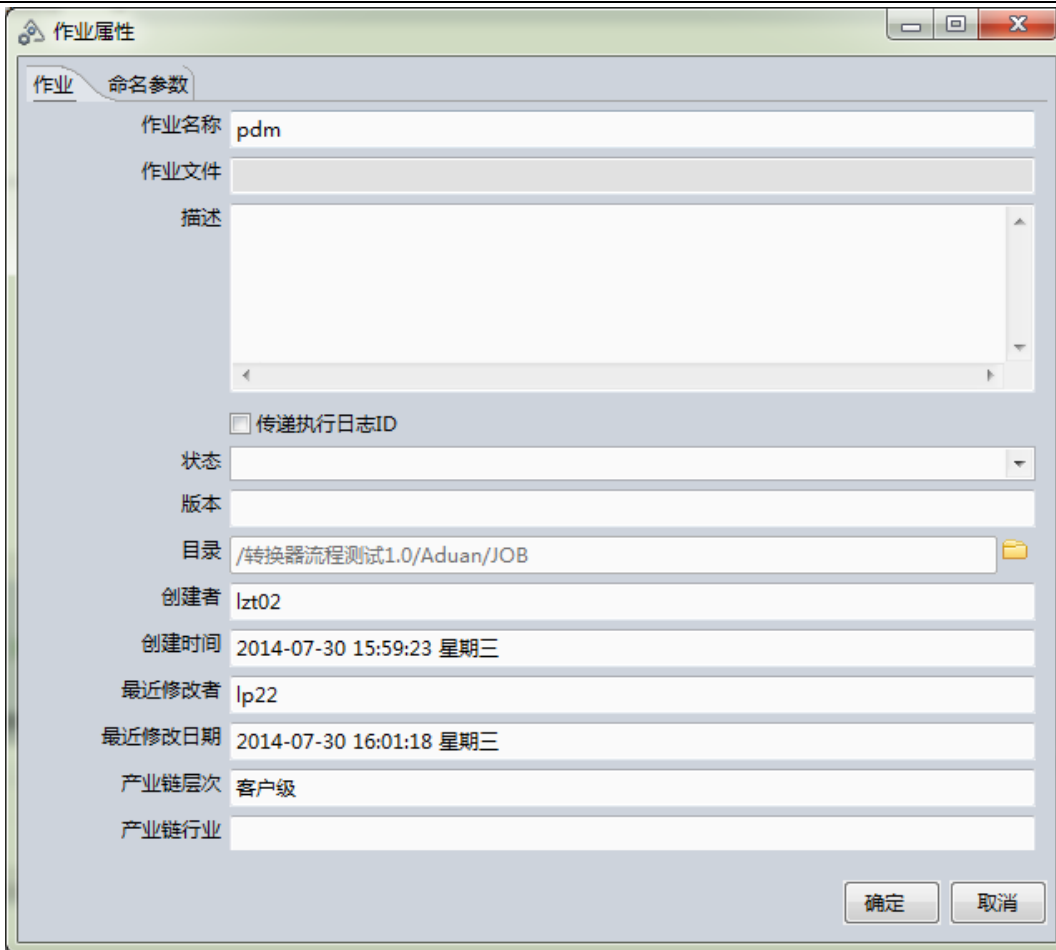
4.5. 设置



The screenshot shows a window titled "转换属性" (Conversion Properties) with a tabbed interface. The "转换" (Conversion) tab is selected. The fields are as follows:

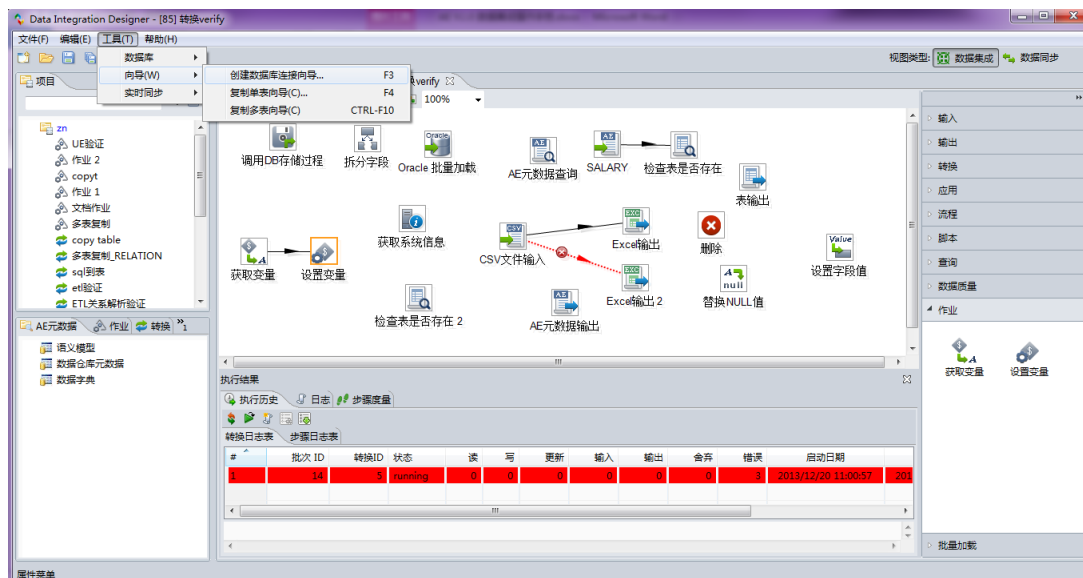
Field Name	Value
转换名称 (Conversion Name)	查询DB查询
转换文件 (Conversion File)	
描述 (Description)	
状态 (Status)	
版本 (Version)	
目录 (Directory)	/转换器流程测试1.0/Aduan
创建者 (Creator)	lzt02
创建日期 (Creation Date)	2014-07-29 17:40:02 星期二
最近修改的用户 (Last Modified User)	lp22
最近修改日期 (Last Modified Date)	2014-07-30 12:39:19 星期三
产业链层次 (Industry Chain Level)	客户级
产业链行业 (Industry Chain Industry)	

Buttons: 确定 (OK), 取消 (Cancel)

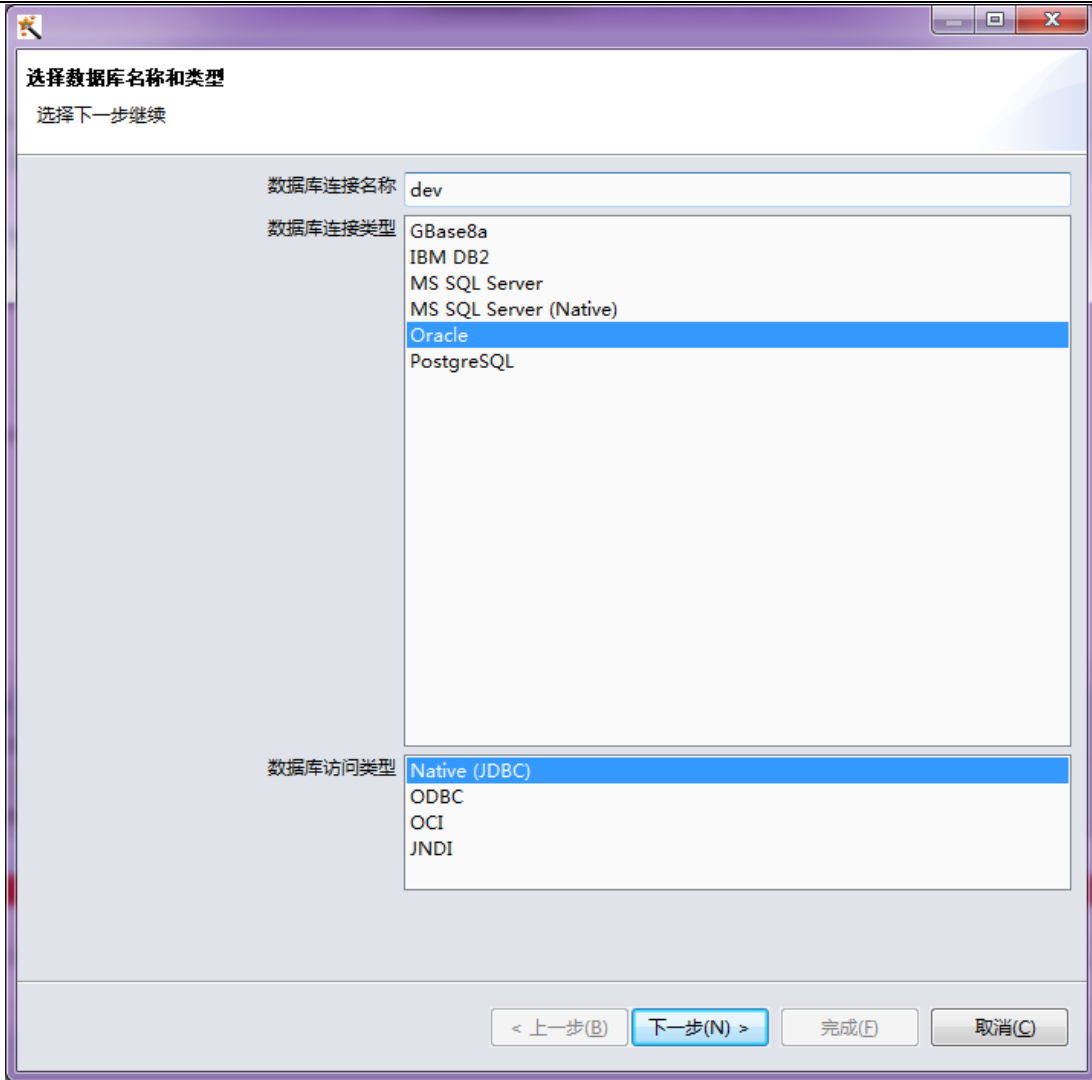


4.6. 创建数据库连接向导

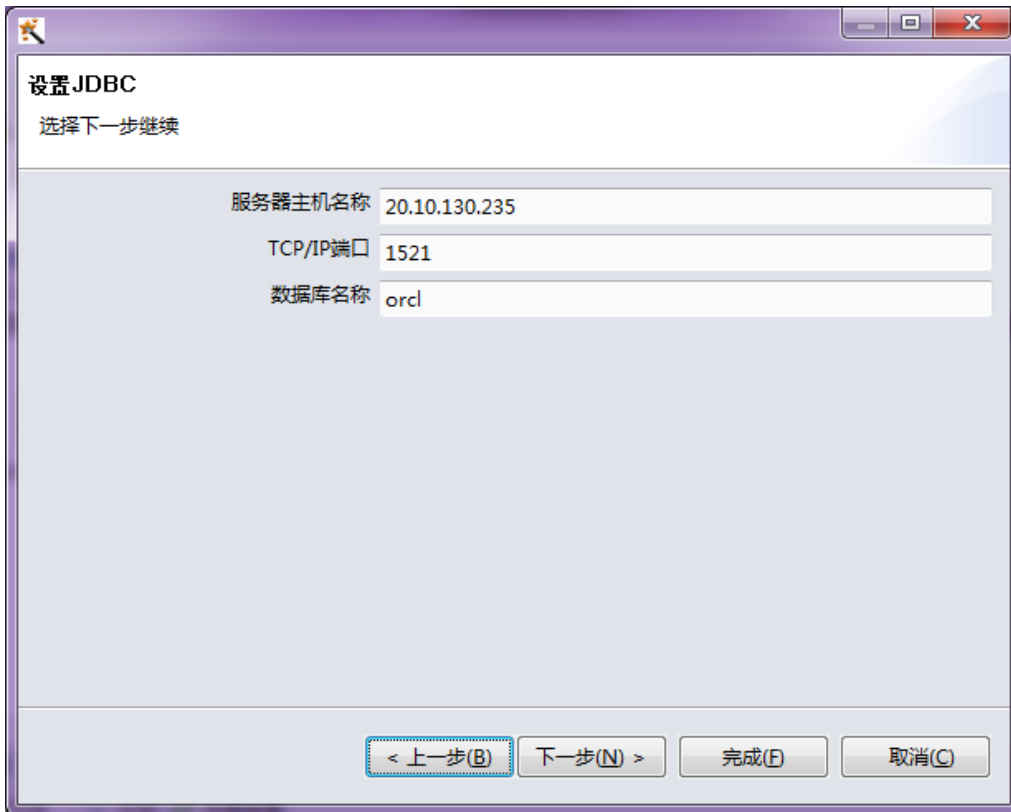
选择<工具>菜单下的，<向导>子菜单下的<创建数据库连接向导>菜单项，如下图所示：



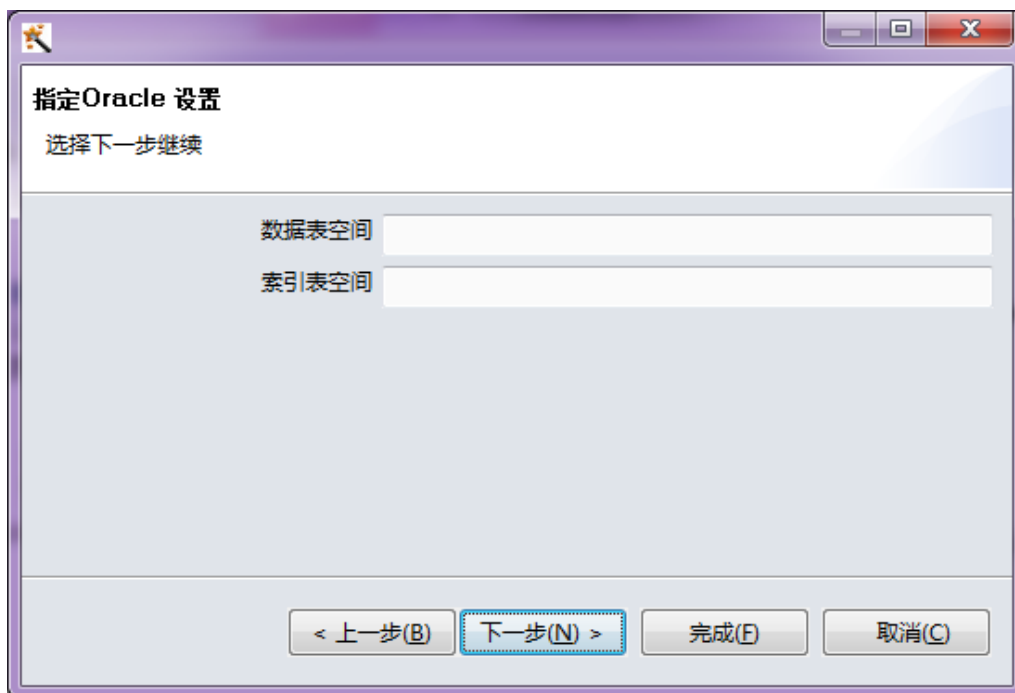
在弹出的向导对话框中，输入数据库连接的名称，选择数据库类型和访问类型，如下图所示：



点击<下一步>，输入服务器主机名称，即数据库所在的服务器，端口号和数据库名称，如下图所示：

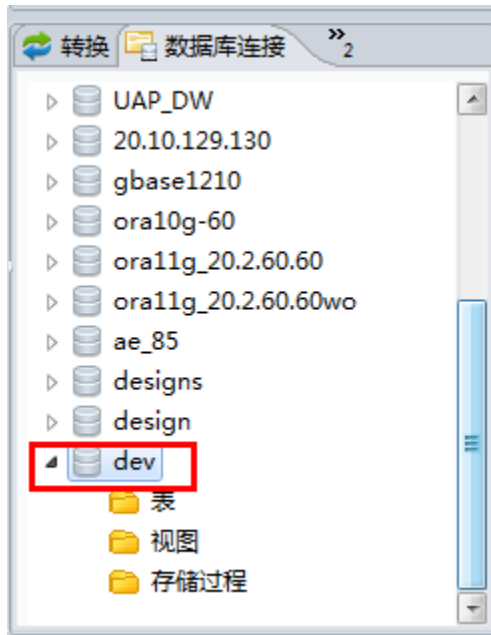


点击<下一步>，设置数据库表空间和索引表空间，这里我们不指定，如下图：



点击<下一步>，设置用户名和密码，设置完成后，可以点击<测试数据库连接>按钮，来测试刚才我们所建立的数据库连接是否连通。点击<完成>按钮，完成对数据库连接的创建。

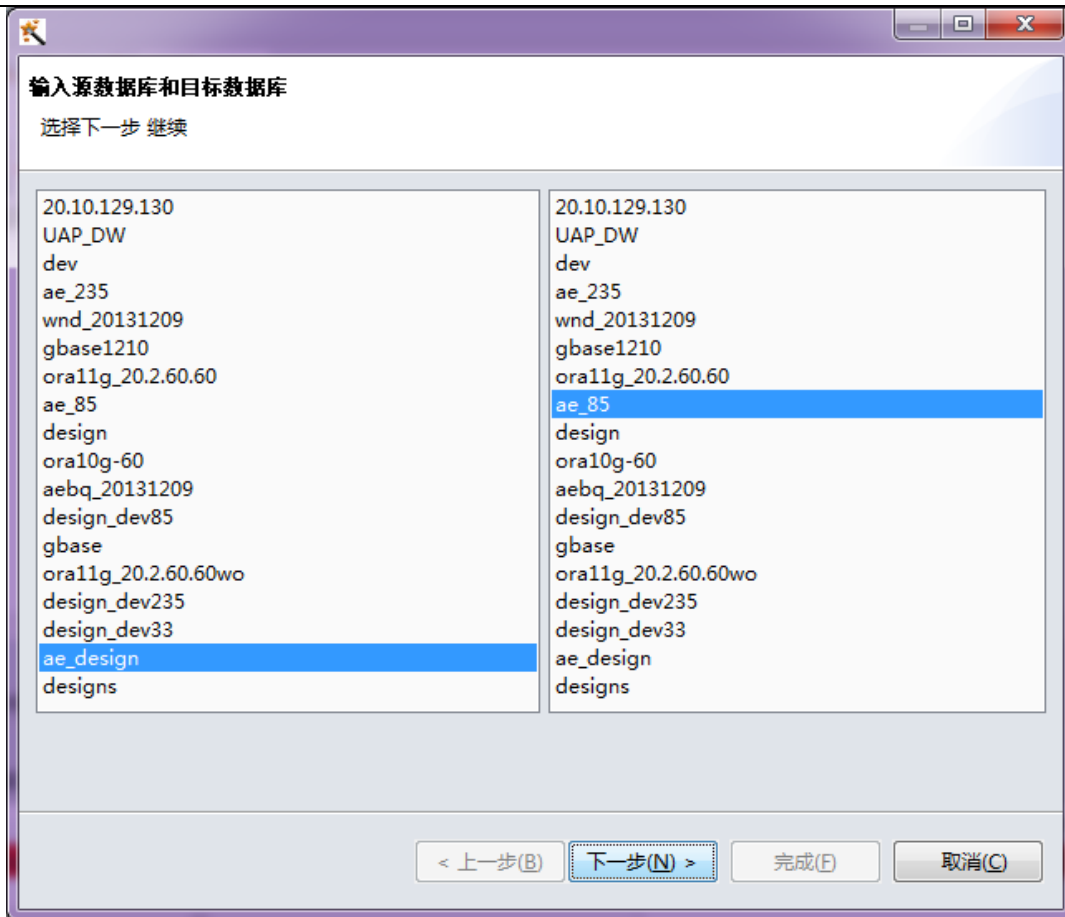
此时在资源管理区，可以看到刚才新建的数据库连接，如下图所示：



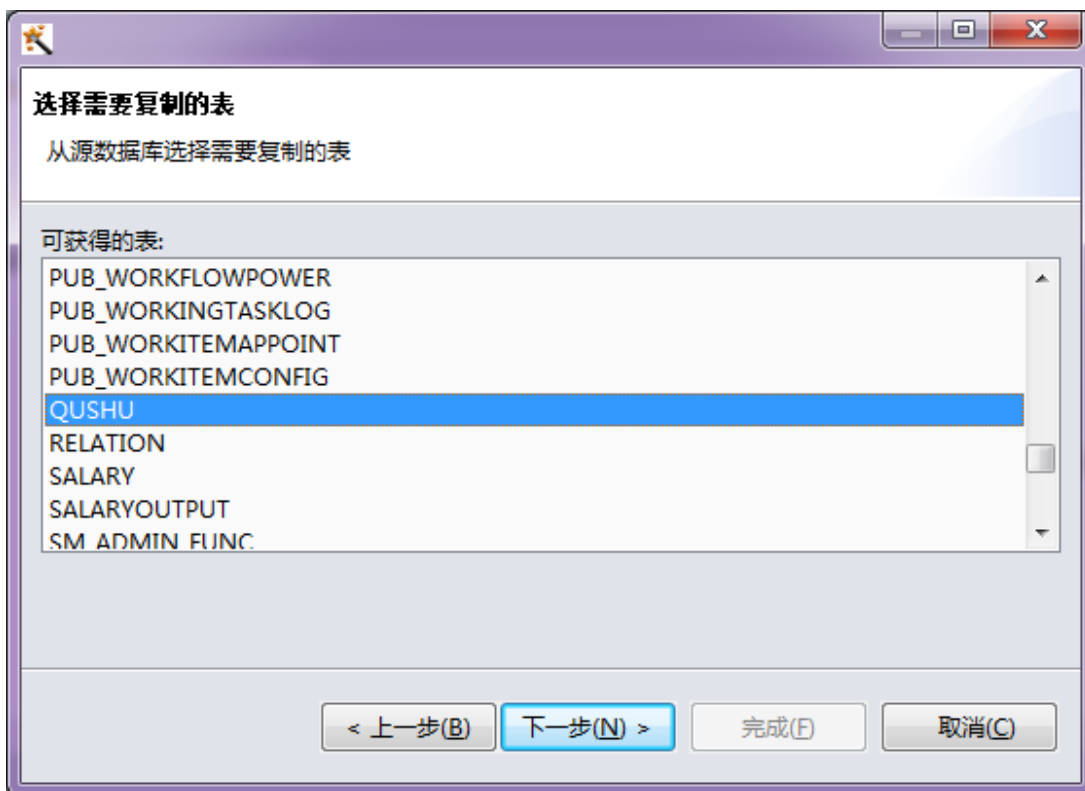
4.7. 复制单表向导

复制单表向导，可以通过向导的方式，生成一个单表复制的转换，此转换完成的功能是，将源数据库中的某个表，输出到目标数据库中。

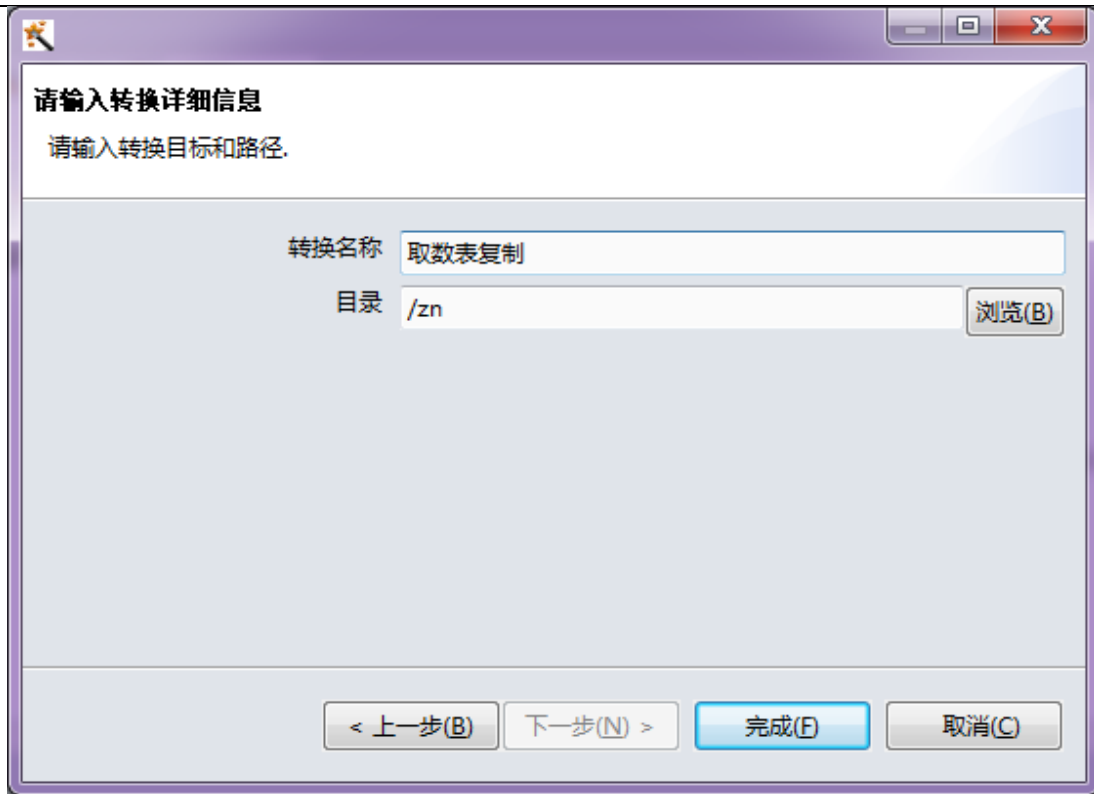
选择<工具>菜单下，<向导>子菜单下的<复制单表向导>菜单项，弹出向导对话框，选择源端数据库和目标端数据库，如下图所示：



点击<下一步>，选择需要复制的表，我们这里选择 QUSHU 表，如下图所示：

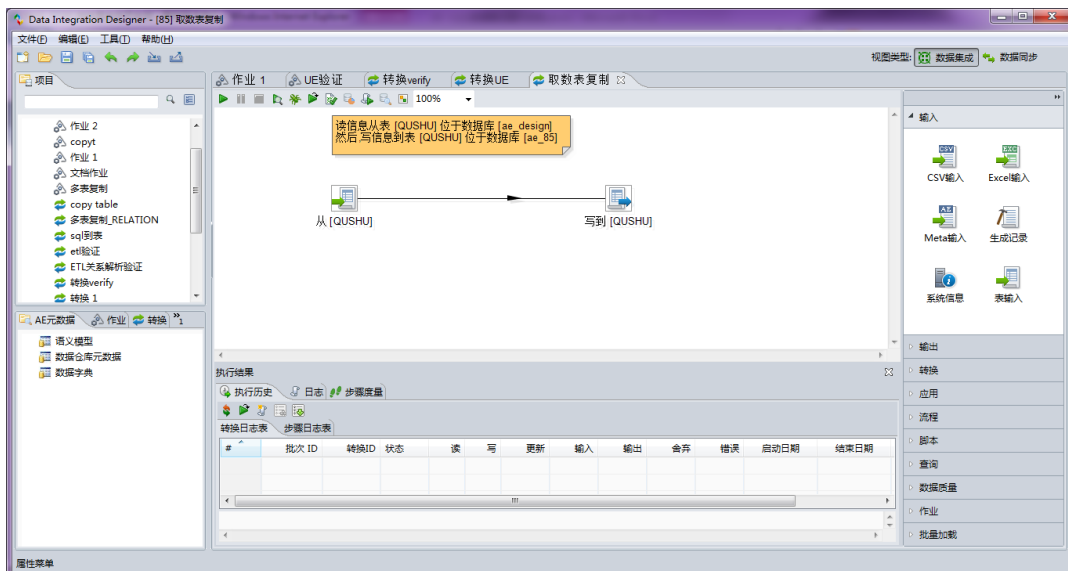


点击<下一步>，输入生成转换的名称，默认保存的目录是当前项目，如下图所示：



点击<完成>, 可以看到“取数表复制”的转换, 已经自动生成, 此转换包括两个步骤, 一个是表输入, 即我们选择的源数据库中的“QUSHU”表, 另一个是表输出, 输出到目标数据库的对应表。

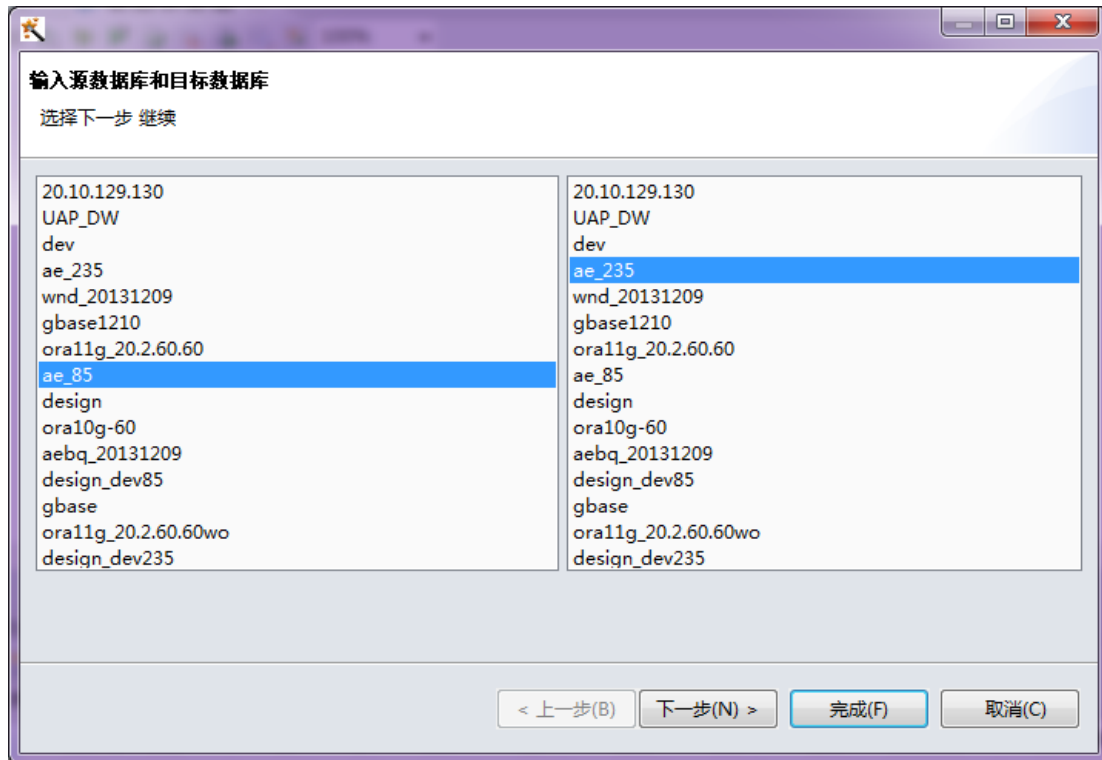
向导还生成了一条备注信息, 对此转换进行说明。



4.8. 复制多表向导

复制多表向导, 可以通过向导的方式, 生成一个作业, 此作业将包含多个单表复制的转换。

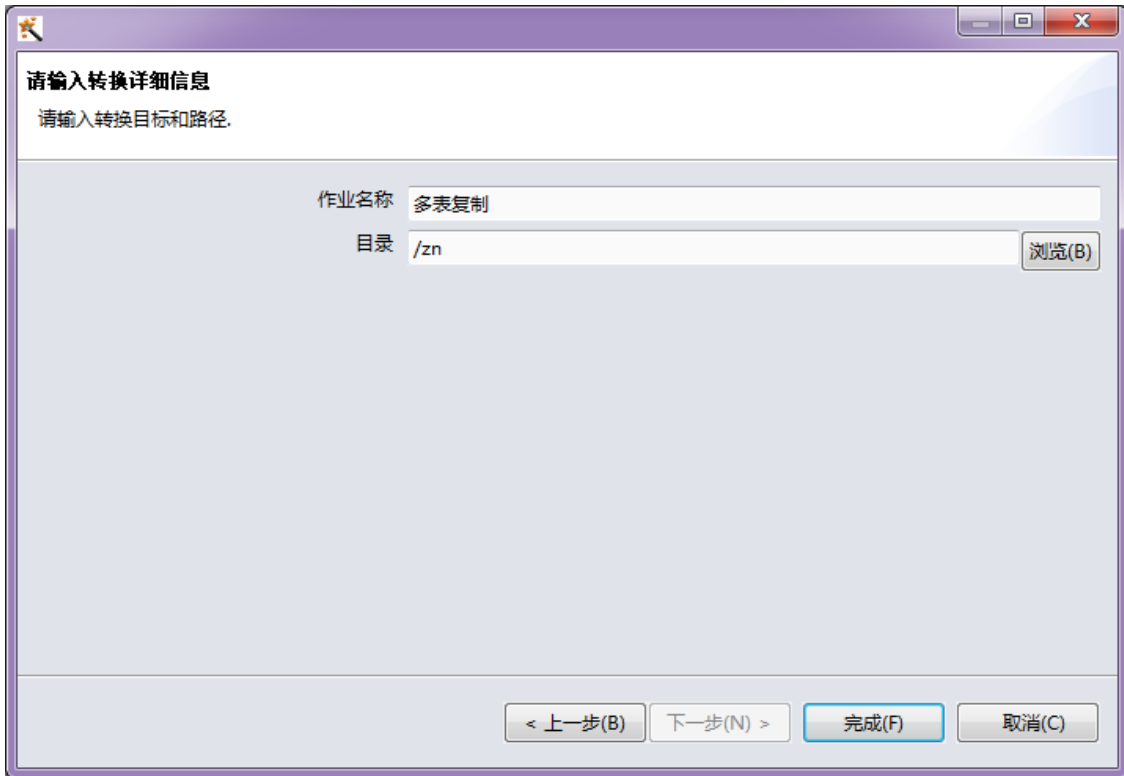
选择<工具>菜单下, <向导>子菜单下的<复制多表向导>菜单项, 弹出向导对话框, 选择源端数据库和目标端数据库, 如下图所示:



点击<下一步>，选择要复制的表，与单表不同的是，此处可以选择多张表，即可以一次性的完成多张源数据库表到目标数据库的复制操作，这里我们选择三张表，如下图所示：



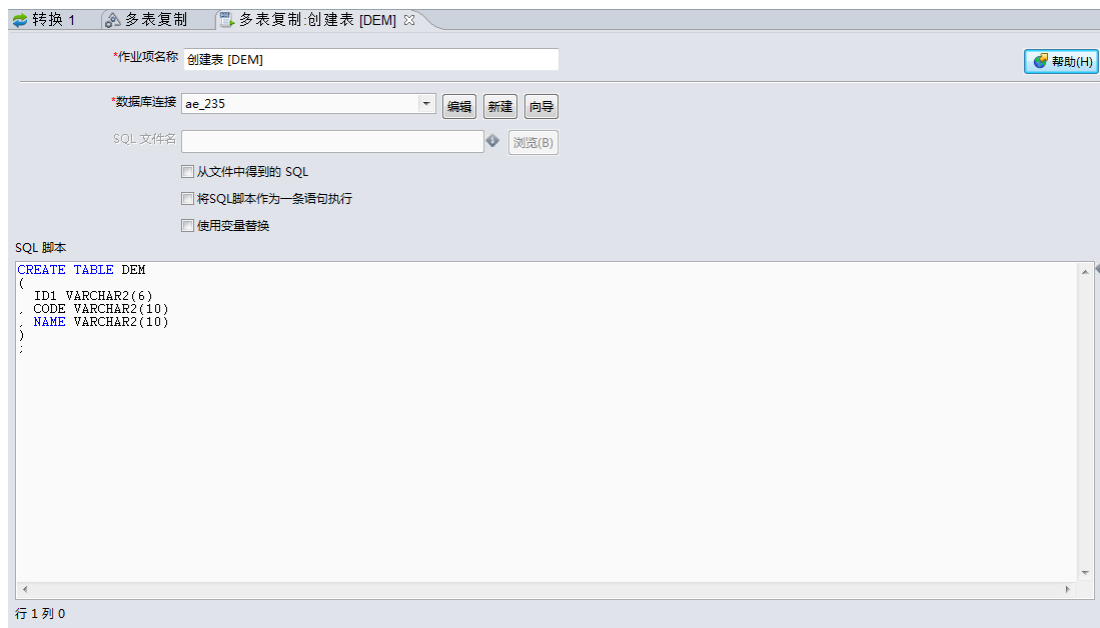
单击<下一步>，输入生成作业的名称和目录，目录默认是当前项目，如下图所示：



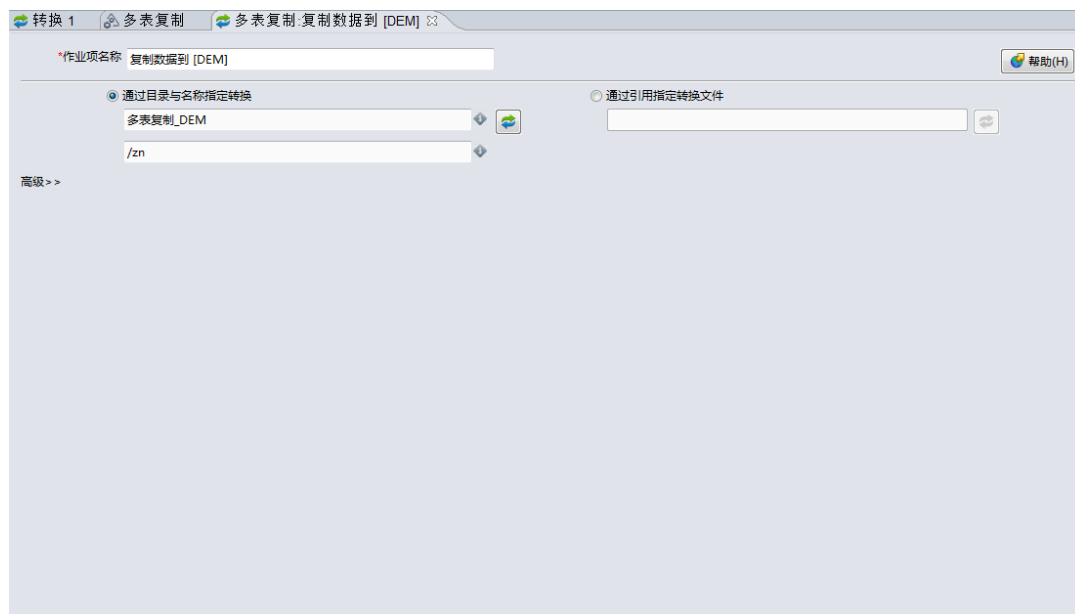
点击<完成>，可以看到“多表复制”作业，已经自动生成，点击保存。如下图所示：



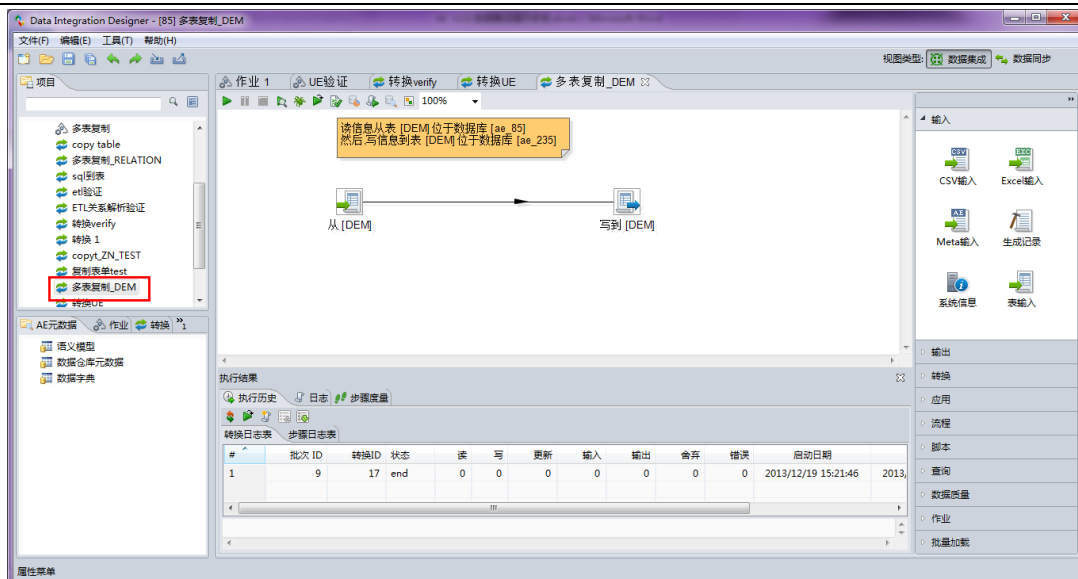
我们具体看下此作业包含的内容，首先包含一个“start”作业项，它代表着一个作业的开始，后面是名为“创建表[DEM]”的 SQL 作业项，此作业项可以书写 DDL 语句，在目标库中创建目标表，有关此作业项的详细信息会在后面介绍，双击“创建表[DEM]”作业项，可以看到目标数据库连接和创建表的 DDL 语句，如下图所示：



“创建表[DEM]”作业项之后，是“复制数据到[DEM]”的作业项，此转换时完成了将源数据库中 DEM 表复制到目标数据库 DEM 表的处理过程，即 DEM 的单表复制功能，双击“复制数据到[DEM]”作业项，可以看到这里的转换只是一个引用，引用了名为“多表复制_DEM”的转换，如下图所示：



在项目区双击“多表复制_DEM”转换，打开此转换，可以看到此转换的内容与上面单表复制的步骤是一样的，这里就不再做说明。同样，此作业中的其他步骤与 DEM 表的类似。



5. 转换步骤介绍

5.1. 示例说明

文本输入的操作步骤最详细。其他组件只介绍组件的参数配置，不再说明如何拖拽组件等具体操作。
全局变量设置：示例中的文件路径应用的是全局变量设置。

5.2. 输入组件：文本输入

5.2.1. 参数说明

“文本输入”步骤用于读取多种文本文件类型的数据。最常用的格式包括电子表格生成的“逗号分隔值”（CSV 文件）和固定宽度平面文件。

“文本输入”步骤可用于指定要读取的文件列表，或以正则表达式形式显示的带有通配符的目录列表。此外，您可以接受上一步骤中的文件名，从而增加了文件名处理的通用性。

以下部分描述了配置“文本输入”步骤可用的相关选项。

转换步骤名称	文本输入
功能描述	
	利用这个步骤可以从 Designer 支持的系统的文本文件里面读取数据。
参数说明	
文件页签	

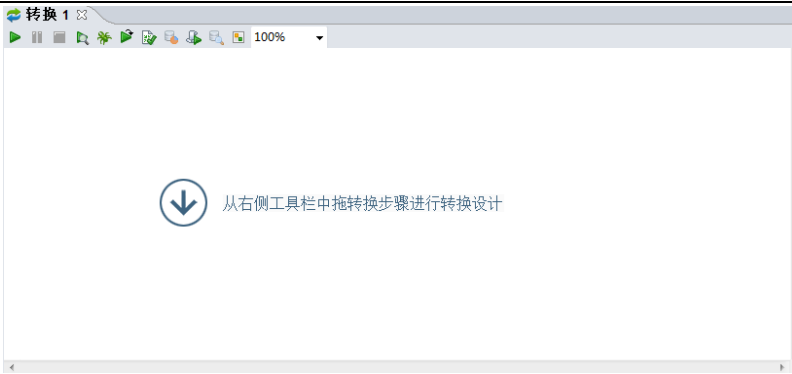
选项	描述
文件或目录	该字段指定输入文本文件的位置和/或名称。注意：按下“添加”按钮，向以下选定的文件（网格）列表中添加文件/目录/通配符组合。
正则表达式	指定要使用的正则表达式，以选择上一个选项中指定的目录中的文件。
选定的文件	该表包含选定文件的列表（或通配符选择）以及一个指定是否需要文件的属性。如果某个文件是必需的却未找到该文件，则将生成错误；否则将跳过该文件名。
显示文件名...	根据当前选定的文件定义显示将要加载的所有文件列表
显示文件内容	显示选定文件的内容。
显示首个数据行的内容	只显示选定文件的首个数据行的内容。
内容页签	
文件类型	可以是 CSV 文件或固定宽度文件。根据此选项，当单击最后的“字段”选项卡中的“获取字段”按钮时，Spoon 将启动另一个帮助 GUI。
分隔符	在单行文本内分隔字段的一个或多个字符。一般使用“;”或 Tab 键。特殊字符（如 CHAR ASCII HEX01）的格式可设置为 \$[value]，如 \$[01] 或 \$[6F, FF, 00, 1F]。
环绕	可使用一对字符串将一些字段环绕起来，以便在这些字段中使用分隔符字符。环绕字符串为可选设置。如果使用重复环绕，就会出现文本行 'Not the nine o'clock news.'。由于使用“'”作为环绕字符串，这句话将被解析为 Not the nine o'clock news。特殊字符（如 CHAR ASCII HEX01）的格式可设置为 \$[value]，如 \$[01] 或 \$[6F, FF, 00, 1F]。
允许在环绕字段中使用换行	还未施行。
转义	如果数据中出现特殊字符，可指定一个或多个转义字符。如果将 \ 指定为转义字符，文本 'Not the nine o'clock news'（带有 ' 环绕符）将被分析为 Not the nine o'clock news。特殊字符（如 CHAR HEX01）的格式可设置为 \$[value]，如 \$[01] 或 \$[6F, FF, 00, 1F]。
页眉和页眉行数	如果文本文件有页眉行（文件的前几行）可启用该功能；可以指定页眉行出现的次数。
页脚和页脚行数	如果文本文件有页脚行（文件的最后一行）可启用该功能；可以指定页脚行出现的次数。
折断行和换行数量	如果您处理的数据行中换行超过了特定的页面限制，则可使用此选项；请注意，页眉和页脚永远不会被视为换行
页面布局、页面大小和文档页眉	只有当需要用行式打印机打印文本时才有必要使用这些选项；使用文档页眉行数以跳过介绍性文本，使用每页行数以定位数据线。
压缩	如果文本文件放置于 Zip 或 GZip 档案中，请启用该选

	项。注意：目前只有档案中的第一个文件会被读取
无空行	不将空行发送到下一步骤。
在输出内容中包含文件名	如需使文件名在输出内容中出现请启用该选项
文件名字段名称	包含文件名的字段名称
输出中带行号	如需在输出中包括行号，则启用该选项
行号字段名	包含行号的字段的名称
按文件显示行号	允许在每个文件中重置行编号
格式	可为 DOS、UNIX 或混合。UNIX 文件中，每行以换行符终止。DOS 文件中，行与行之间以回车符和换行符分隔。如果指定为混合，将不会实施验证。
编码	指定要使用的文本文件编码；将此字段留空以使用系统的默认编码。要使用 Unicode，请指定 UTF-8 或 UTF-16。首次使用时，Spoon 会搜索系统中的可用编码。
限制	设置从文件中读取的行数；0 表示读取所有行。
实施宽松的日期分析	如果要对数据字段执行严格解析，不要启用该选项；如果启用了宽松解析，日期 1 月 32 日会变为 2 月 1 日
日期格式区域设置	此项区域设置用于分析完整写出的日期，如“February 2nd, 2006”；在法语区域设置（fr_FR）运行的系统中分析此日期将无法进行，因为“February”在该地区为 Février。
字段页签	
名称	字段的名称
类型	字段类型可以是字符串、日期或数字
格式	请参阅“数字格式”了解格式符号的完整描述。
长度	对于数字：数字中有效数字的总量；对于字符串：字符串的总长度；对于日期：屏幕上列印的字符串长度（例如，长度为 4 则只返回年份）。
精度	对于数字：浮点数字数量；对于字符串、日期和布尔值：未使用；
货币	用于解译 \$10,000.00 or E5.000,00 此类数字
小数	小数点可以为“.”（10;000.00）或“,”（5.000,00）
分组	分组可以用逗号“,”（10;000.00）或“.”（5.000,00）
空值条件	将此值看作空值
默认	在文本文件中的字段未指定值（为空）时将采用的默认值
修剪	在处理字段之前对其进行修剪，类型包括左、右、两侧
重复	如果本行中的对应值为空，则重复上一次不为空的值（是/否）（确定吗？）
错误处理页签	

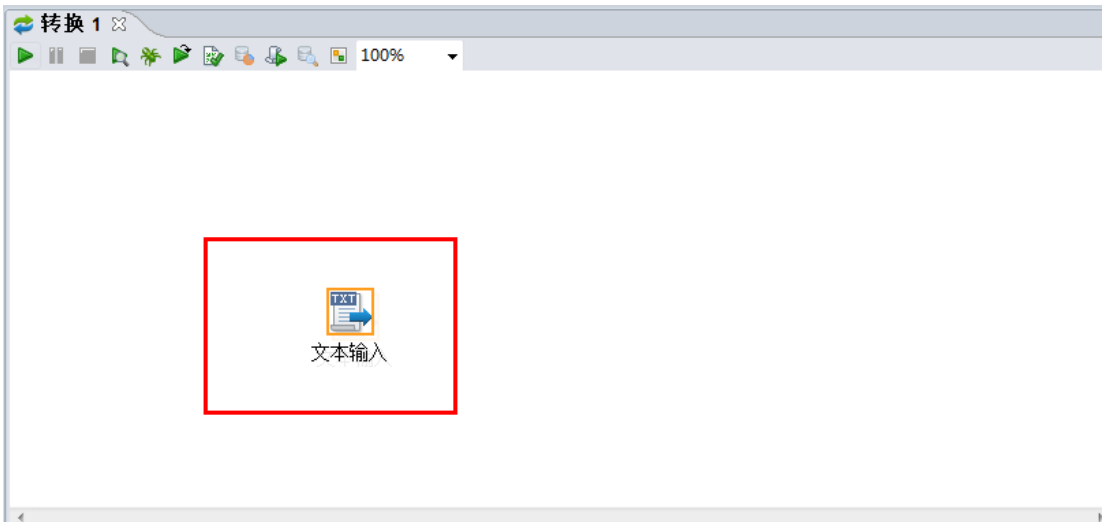
忽略错误	如需在解析中忽略错误，可启用此选项
跳过错误行	如果需要跳过包含错误的行，则启用此选项。可以为出现错误的行编号另外生成一个文件。有错误的行不会跳过，出现解析错误的字段将为空（空值）
错误计数字段名	向输出流的行中添加一个字段；此字段包含错误行的数量
错误字段字段名	向输出流的行中添加一个字段；此字段包含发生错误的字段名称
错误文本字段名	向输出流的行中添加一个字段；此字段包含发生的解析错误的说明
警告文件目录	如果出现警告，则会被放置在此目录中。该文件的名称为 /filename。
错误文件目录	如果出现与文件不存在或不可访问相关的错误，这些错误将保存在此目录中。该文件的名称为 /filename。
失败行编号文件目录	当行出现解析错误时，行编号将保存在此目录中。该文件的名称为 /filename。
过滤页签	
筛选器字符串	要搜索的字符串
筛选器位置	筛选器字符串在行中所处的位置。零 (0) 指行中的第一个位置。如果在此指定的值小于零 (0)，将在整个字符串中搜索筛选器字符串。
遇到筛选器时停止	如果希望当遇到筛选器字符串时停止处理当前文本文件，请指定为“是”。
正匹配	如果要对与筛选器条件匹配的行进行处理，则指定为“是”；如果要忽略这些行则指定为“否”。
文件名页签	
接受上一步骤中的文件名	启用该选项，就可以获取来自前面步骤的文件名。
从中读取文件名的步骤	要从其中读取文件名的步骤
输入中要用作文件名的字段	“文本文件输入” 查看此步骤以确定要使用哪个文件名
使用范例	

5.2.2. 示例

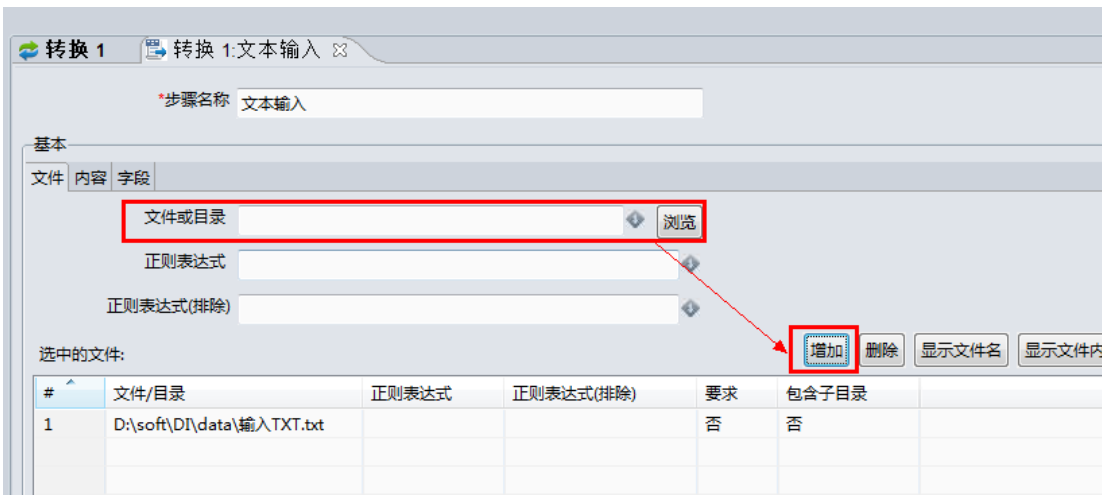
step1. 【文件】—【新建】—【转换】，弹出如下对话框：

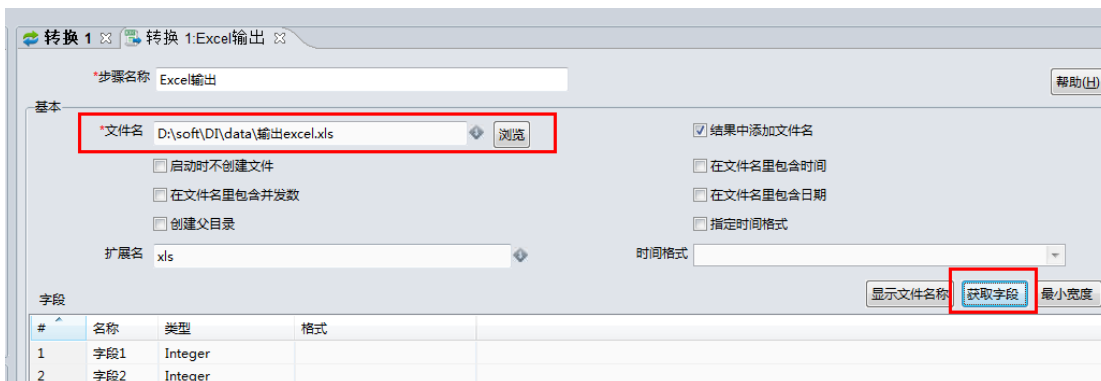
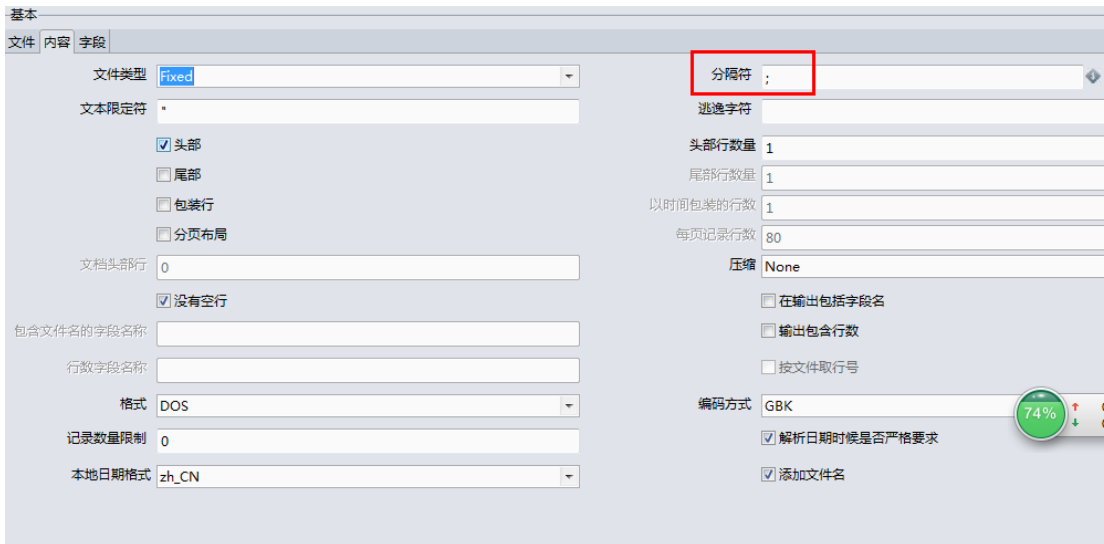


step2. 拖动组件区的输入组件【文本输入】到面板中，如下图所示：



step3. 双击组件“文本输入”，打开组件的属性页签，点击“预览”，打开要输入的文本文件，如下图所示：





5.3. 输入组件：CSV 输入

5.3.1. 参数说明

CSV 逗号分割值文件(Comma Separated Value), 是一种用来存储数据的纯文本格式, 通常用于电子表格或数据库软件。

NIO, 本地系统调用文件意味着更快的性能, 但是仅限于本地文件, 不支持 vfs。

并行运行, 如果设置了此步骤在集群模式或者在多个副本中运行, 并且使用并行执行, 每一个副本都会读取一个来自于文件的不同信息。

简易转换，如果读许多字段，并且这些字段中有很多都是无法操作的，但是仅仅通过转换来加载一些文本文件或其他数据库，懒转换可以阻止 DI Designer 进行一些不必要的操作，例如将他们转换成字符串、日期、或者数字等。

转换步骤名称	CSV 输入
功能描述	
这个步骤可以从 CVS 文件里面读取数据	
参数说明	
选项	描述
步骤名称	步骤的名称，名称必须唯一
文件名	要读取的 csv 文件
列分隔符	用来分割不同列的字符
封闭符	在目标文件中明确的封闭符
NIO>缓存大小	读取数据的缓存大小，允许一次从磁盘中读取的记录数
简易转换	简易转换试图避免不必要的数据类型转换，因此可以在一定程度上提升系统性能，典型的例子就是从文本文件中读取信息，然后再写回文本文件。
包含列头行	如果目标文件的头包含列名的话要选中此参数
将文件增加到结果	在结果中增加 CSV 文件名
行号字段	输出步骤中包含行号（整型）的字段名称
并发运行	并发运行，用以提高系统性能。警告：由于某些原因并发读取文件信息仅仅支持一些字段文件，这些字段文件没有分割线或者附带返回值。
文件编码	确定读取文件的编码方式
字段中有会车行	
字段	表中包含了文件中所有可读取的字段
预览	预览来自文件中的信息
获取字段	获取来自文件中的字段
使用范例	

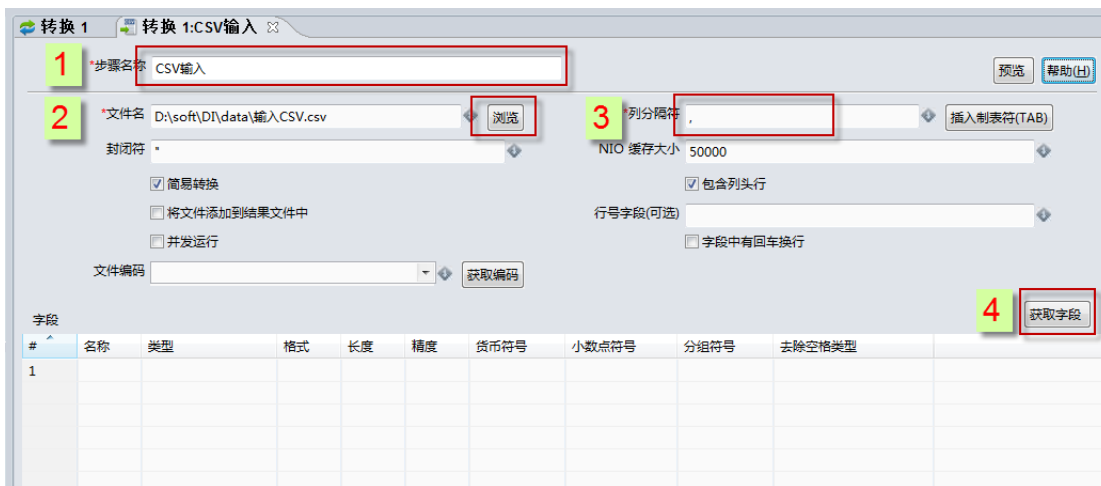
5.3.2. 示例

示例说明：将 CSV 文件作为输入，并输出为 excel 文档。如下图所示：

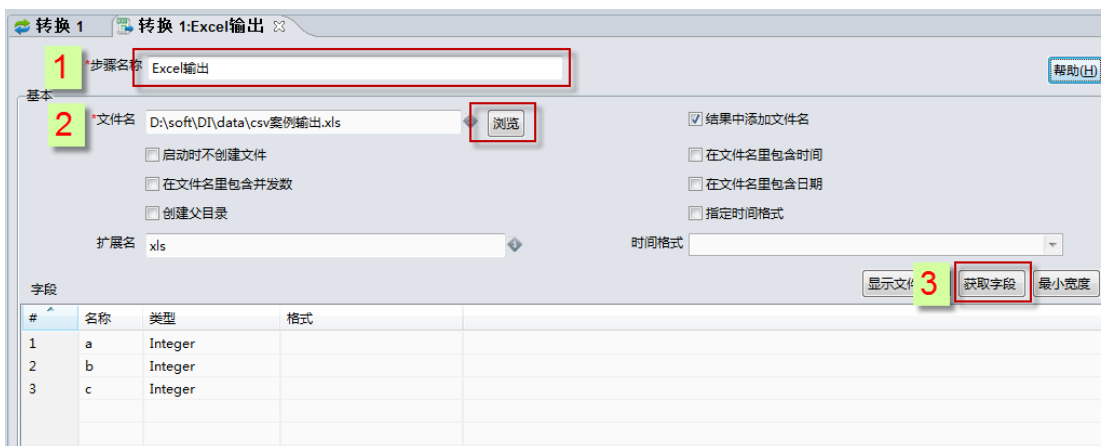


示例步骤：

step1. 编辑转换“CSV 输入”属性面板，如下图所示：



step2. 编辑转换“Excel 输出”属性面板，如下图所示：



step3. 运行转换，即可。

5. 4. 输入组件：Excel 输入

5. 4. 1. 参数说明

转换步骤名称	Excel 输入
--------	----------

功能描述	
<p>利用这个步骤可以从 Designer 支持的系统的 Excel 文件里面读取数据。 备注：默认的工作表格式是 03-07. 当读取 OpenOffice, ODS, Excel 2007 这种文件时, 需要在内容页签中改变工作表类型。</p>	
参数说明	
文件页签	
选项	描述
步骤名称	步骤的名称, 在单个转换中必须唯一。
文件或目录	文件所在的具体位置. 须知: 点击增加按钮, 将选择的 Excel 文件添加到文件列表中。
正则表达式	
正则表达式 (排除)	
预览	点击预览可以查看指定 Excel 文件的内容。
工作表页签	
<p>在这一页签中你可以根据名字确定要操作的工作表, 你可以为每一工作表指定开始的行和列, 行列都是从 0 开始。</p> <p>【备注】: 读取所有工作表示可以的, 如果想读取, 可以通过清除“要读取的工作表”和第一行中开始的行和列, 可以用于所有工作表. 也就是说, 如果你想读取所有的工作表, 一定要注意不要指定任何工作表名称。</p>	
字段页签	
<p>在这里可以指定想从 EXCEL 中读取的字段. 如果工作表有一个头部行, 按钮“获取来自头部的字段”将自动获取并填充可用的字段。</p>	
从哪个步骤读文件名	包含了选中文件名的列表
保存文件名字段	保存的文件名字段列表
使用范例	

5.4.2. 示例

略。

5.5. 输入组件: LDAP 输入

5.5.1. 参数说明

转换步骤名称	LDAP 输入
功能描述	
<p>“LDAP 输入”步骤可用于从 LDAP 服务器中读取有关用户、角色的信息和其他数据。以下部分描述了“LDAP 输入”步骤的可用选项。</p>	

参数说明	
选项	描述
步骤名称	步骤节点的名称，必须唯一
主机	运行 LDAP 服务器的主机
端口	运行 LDAP 服务器的端口编号
协议	
使用授权	授权或不授权
用户名	用户名
密码	密码
测试连接按钮	测试连接
预览按钮	预览 LDAP 信息
搜索库	根据 LDAP 节点搜索 LDAP 内容
筛选器字符串	通过 LDAP 筛选器搜索自定义内容，目前阶段只支持“单一筛选器格式”。例如：邮件 = *
重复	如果 LDAP 未返回某个属性的值，则使用上一行的值
使用范例	

5.5.2. 示例

略。

5.6. 输入组件：Meta 输入

5.6.1. 参数说明

转换步骤名称	Meta 输入
功能描述	
这一步骤主要是从 AE 数据库中读取元数据（语义模型、语义元数据、数据字典）。	
参数说明	
选项	描述
步骤名称	步骤节点的名称，必须唯一
显示名	读取的元数据实体种类
表名	具体的表名
限制	提交记录数量的限制，0 为无限制。

使用范例

5.6.2. 示例

略。

5.7. 输入组件：XML 输入

5.7.1. 参数说明

此步骤可用于使用 StAX 解析器从任何类型的 XML 文件中读取数据。虽然现有的“从 XML 获取数据”步骤更易于使用，但需要使用 DOM 解析器进行内存中处理，当文件过大时即便清除文件的部分内容也无法进行。

XML 输入流 (StAX) 步骤采用完全不同的方法处理数据结构量大且复杂的情况，满足快速加载数据的需要：因为 DI Designer 自身包含很多步骤，可以用不同的方式处理数据，处理逻辑的重点就变成了转化过程，步骤本身提供了原始 XML 数据流以及其他有用的处理信息。

因为有时一些 XML 文件的处理逻辑很难处理，建议在充分掌握 DI Designer 步骤后使用此步骤。请参阅本页末尾的各类示例，了解具体用法。

如果在其他步骤中受到限制或需要在以下情况中解析 XML，请选择此步骤：

速度非常快，不论文件有多大都不受内存影响（由于采用了流式传播方法，可实现超过 GB 字节的传输）

可用不同的方式灵活读取 XML 文件的不同部分（避免了多次解析文件）。

转换步骤名称	XML 输入
功能描述	
利用这个步骤可以从 Designer 支持的系统的 XML 文件里面读取数据。	
参数说明	
选项	描述
步骤名称	步骤的名称，在单个转换中必须唯一。
文件名	指定输入 XML 文件的文件名
在结果中添加文件名	在转换结果中添加处理过的 XML 文件名。内存中会保存一份唯一的列表，可在下次作业输入时使用，如在下次转换时。
跳过（元素/属性）	要跳过的元素/属性数量。此选项可用于从文件的特定位置开始处理过程。解析器仍会加载文件，但不生成行。
限制（元素/属性）	当超过元素/属性限制时，处理过程会结束。因为有跳过和限制属性，可以启用分块加载，此选项要在外层循环定义。
默认字符串长度（元素/属性）	XML 数据名称和值字段的默认字符串长度。

编码	XML 文件的编码
添加命名空间信息	<p>如果选定命名空间信息,XML 数据类型 <i>NAMESPACE</i> 就被添加到了数据流中,包含一个可选的前缀(显示在 XML 数据名称中)和 URI 信息(显示在 XML 数据值中)。XML 数据名称之前还会有一个在“ELEMENT”数据类型中定义的前缀,例如,前缀:产品。</p> <p>性能注意事项:由于要同时进行命名空间处理,此选项会稍微降低处理吞吐量。</p>
修剪字符串	选定后,步骤将修剪所有名称/值元素和属性。还会清除位于字符串首和末尾的空格、制表符和回车符。
输出是否包含文件名/字段名	如果选择此选项,步骤会将处理过的文件名添加到给定的字段名中
输出是否包含行号/字段名	如果选择此选项,步骤会将处理过的行编号(从 1 开始)添加到给定的字段名中。
输出是否包含 XML 数据类型(数字)/字段名	<p>如果选择此选项,步骤将在提供的字段名中添加经过处理的数字格式的数据类型,</p> <p>可设定以下数据类型:</p> <ul style="list-style-type: none"> 0 - “UNKNOWN”(未使用,保留) 1 - “START_ELEMENT” 2 - “END_ELEMENT” 3 - “PROCESSING_INSTRUCTION”(未使用,保留) 4 - “CHARACTERS” 5 - “COMMENT”(未使用,保留) 6 - “SPACE”(未使用,保留) 7 - “START_DOCUMENT” 8 - “END_DOCUMENT” 9 - “ENTITY_REFERENCE”(未使用,保留) 10 - “ATTRIBUTE” 11 - “DTD”(未使用,保留) 12 - “CDATA”(未使用,保留) 13 - “NAMESPACE”(选定命名空间信息后) 14 - “NOTATION_DECLARATION”(未使用,保留) 15 - “ENTITY_DECLARATION”(未使用,保留)
输出是否包含 XML 数据类型(描述)/字段名	<p>如果选择此选项,步骤会将经过处理的文本格式的数据类型添加到给定的字段名中。应使用此类型而非数值数据类型,以使转换获得更好的可读性。请参阅 XML 数据类型(数字)了解值的列表。</p> <p>性能注意事项:由于可能导致字符串处理速度减慢以及占用额外的内存,建议在加载大量数据时使用数字数据类型的格式。</p>
输出是否包含 XML 位置行/字段名	如果选择此选项,步骤会将处理过的源 XML 位置行添加到给定的字段名中。
输出是否包含 XML 位置列/字段名	如果选择此选项,步骤会将处理过的源 XML 位置列添加到给定的字段名中。

输出是否包含 XML 元素 / 字段名	如果选择此选项, 步骤会将处理过的元素编号(从 0 开始) 添加到给定的字段名中。与行号不同, 字段是按新元素(而非按新行) 累加的。这能够保证在不同级别之间实行正确的嵌套。
输出是否包含 XML 父元素 ID/ 字段名	如果选择此选项, 步骤将在给定的字段名中添加父元素编号。 注意: 同时使用 XML 元素 ID 和 XML 父元素 ID, 可生成完整的 XML 元素树以供以后使用
输出是否包含 XML 元素级别/ 字段名	如果选择此选项, 步骤会将处理过的元素级别(从根部的 START_ 和 END_DOCUMENT, 即 0 级开始算起) 添加到给定的字段名中。
输出是否包含 XML 路径 / 字段名	如果选择此选项, 步骤会将处理过的 XML 路径添加到给定的字段名中。
输出是否包含父路径/ 字段名	如果选择此选项, 步骤会将处理过的 XML 父路径添加到给定的字段名中。
输出是否包含 XML 数据名称/ 字段名	如果选择此选项, 步骤将在给定的字段名中添加处理过的元素的数据名称、属性和可选的命名空间前缀。
输出是否包含 XML 数据值/ 字段名	如果选择此选项, 步骤将在给定的字段名中添加处理过的元素的数据值、属性和可选的命名空间 URI。
示例页签	
分发包中提供了演示此步骤的功能的示例转换(在示例文件夹中)。	
使用范例	

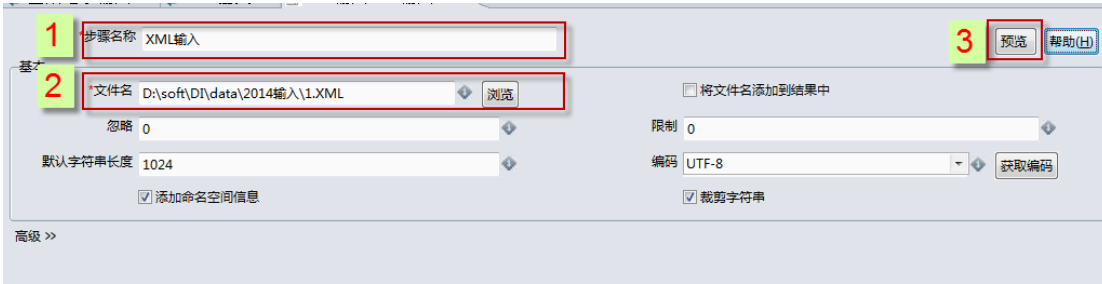
5.7.2. 示例

示例说明: 将 XML 文件作为输入, 并输出为 excel 文档。转换如下图所示:

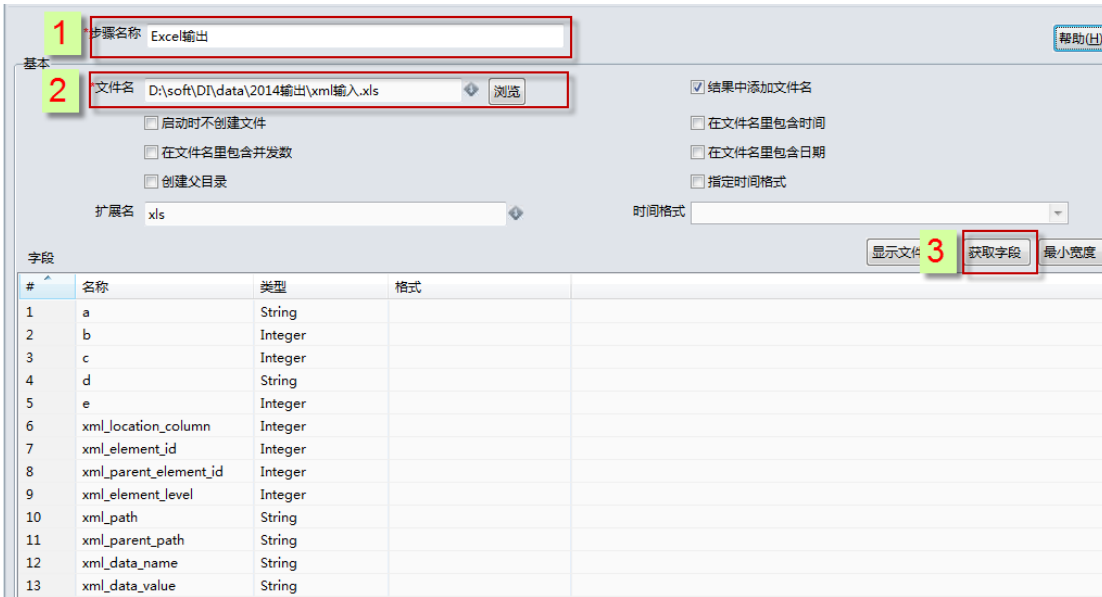


示例步骤:

step1. 编辑转换“XML 输入”属性面板, 如下图所示:



step2. 编辑转换“Excel 输出”属性面板，如下图所示：



step3. 运行转换，即可。

5.8. 输入组件：生成记录

5.8.1. 参数说明

转换步骤名称	生成记录
功能描述	
这个步骤输出一定数量的行，缺省为空。可选包括一定数量的静态字段。	
参数说明	
选项	描述
步骤名称	步骤的名称，在单个转换中必须唯一
限制	设置你要生成行的最大数量
字段	在输出行中你想要包含的静态字段
从不停止生成记录	这选项是用于实时控制
间隔(ms)	间隔的时间（毫秒）
目前时间字段值	目前的时间字段
先前的时间字段值	先前的时间字段

使用范例

5.8.2. 示例

示例说明：自动生成记录文件，并输出为 excel 文档。转换如下图所示：



示例步骤：

step1. 编辑输入转换“生成记录”属性面板，如下图所示：

#	名称	类型	格式	长度	精度	货币类型	小数	分组	值	设为空串
1	a	Integer							10	否

step2. 编辑输出转换“Excel 输出”属性面板，如下图所示：

#	名称	类型	格式
1	a	Integer	0

step3. 运行转换，即可。

5.9. 输入组件：系统信息

5.9.1. 参数说明

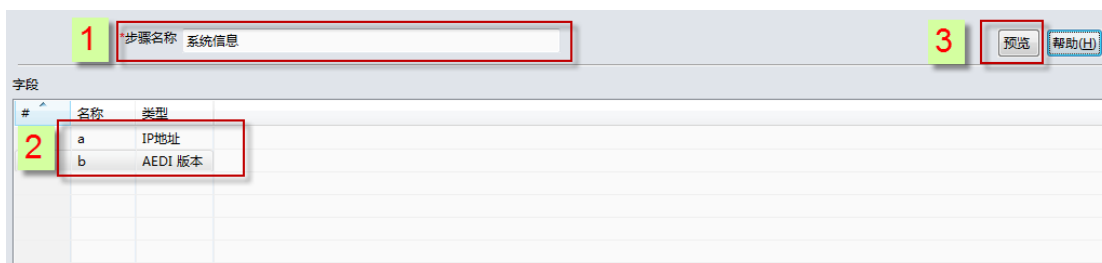
转换步骤名称	系统信息
功能描述	
这一步骤可以获取 DI Designer 环境的系统信息，下表中为系统信息的具体含义	
参数说明	
选项	描述
步骤名称	转换的步骤名称，必须唯一
字段	要输出的字段
使用范例	

5.9.2. 示例

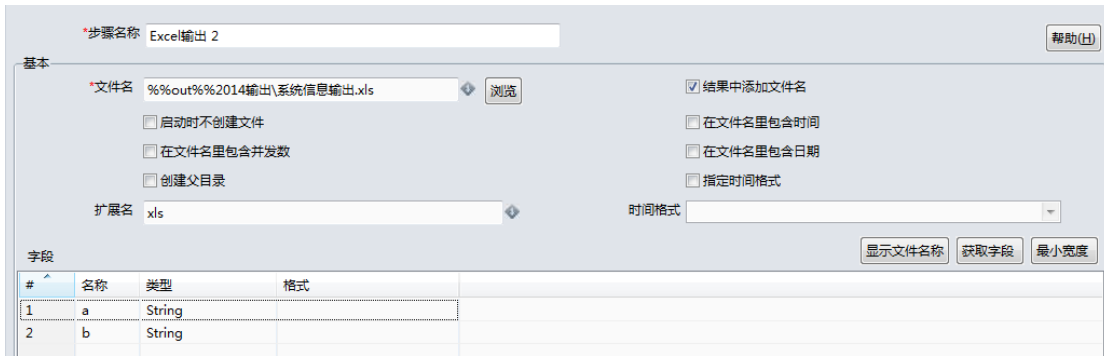
示例说明：将系统信息输出为 excel 文档。转换如下图所示：



step1. 编辑转换“系统信息”属性面板，如下图所示：



step2. 编辑转换“Excel 输出”属性面板，如下图所示：



step3. 运行转换，即可。

5. 10. 输入组件：表输入

5. 10. 1. 参数说明

转换步骤名称	表输入
功能描述	
这一步骤通过数据库连接和 SQL 语句读取数据库中的数据，一些基本的 SQL 语句可以自动生成	
参数说明	
选项	描述
步骤名称	步骤的名称，在单一的步骤中，名称必需
数据库连接	读取数据的数据库连接
SQL	SQL 语句用来从数据库连接中读取数据
允许懒加载	如果允许，懒加载避免了一些不必要的数据类型，提升执行性能
是否变量取代	可以在 SQL 脚本中使用变量
从步骤插入数据	指定我们期待读取数据的步骤名称。这些信息能被插入到 SQL 语句
每一行执行	允许为每一行分别插入数据
限制	设置从数据库中读取的行数。0 所有行
使用范例	

5. 10. 2. 示例

略。

5. 11. 输入组件：配置输入

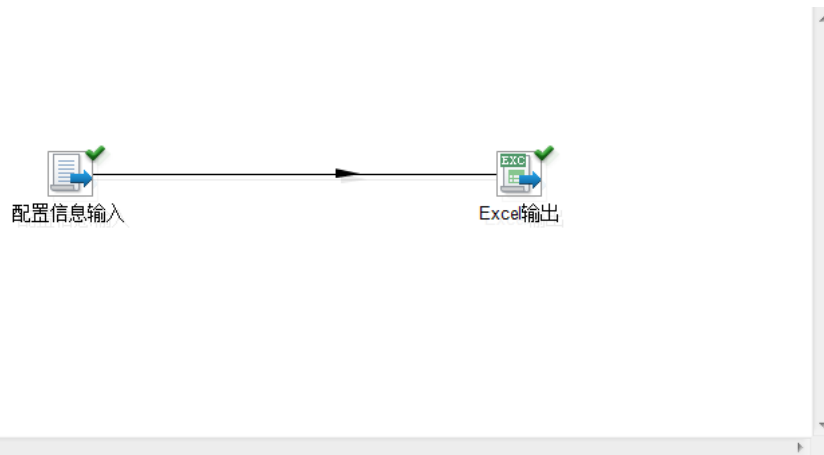
5. 11. 1. 参数说明

转换步骤名称	配置输入
功能描述	
此步骤从 Java 属性文件中读取。有关文件格式的更多信息	
文件页签	
选项	描述
步骤名称	步骤的名称，在单一的步骤中，名称必需
字段中的文件名	<p>是否在字段中定义文件名：上一步使用输入流的特定字段作为文件名。这些将被读取。</p> <p>从字段中获取文件名：指定文件名来源的字段。</p>
文件或目录	<p>指定输入文本文件的位置和/或名称。</p> <p>重要提示：单击“添加”，在下方的选定文件（网格）列表中添加文件/目录/通配符组合。</p>
正则表达式	指定要使用的正则表达式，以选择上一个选项中指定的目录中的文件。
选定的文件	包含选定文件的列表（或通配符选项）以及一个指定是否需要文件的属性。如果需要某个文件但却未找到该文件，则会产生错误；否则将跳过该文件名。
显示文件名	根据当前选定的文件定义显示将要加载的所有文件列表
内容页签	
输出是否包含文件名	用于指定一个字段名以便在此步骤的输出中包括文件名(字符串)。
输出中带行号	用于指定一个字段名以便在此步骤的输出中包括行号(整数)。
是否每个文件重置一次行号	如果希望在每个独立文件的开头将生成的行数(可选)重置(为1)，请启用此选项。
限制	将行数量限制为此数字（零（0）表示所有行）。
结果文件名	在结果文件名中添加文件：将读取的文件名添加到转换的结果中。内存中会保存一份唯一的列表，可在下次作业输入时使用，如在下次转换时。
字段页签	
名称	输入字段的名称
列	要读取的属性
类型	要转换成的数据类型
格式	数据类型转换所使用的格式掩码或转换掩码
长度	输出数据类型的长度
精度	输出数据类型的精度

货币	数据类型转换中要使用的货币符号
小数	数据类型转换中要使用的小数符号
分组	数据类型转换中要使用的数字分组符号
修剪类型	数据类型转换中要使用的修剪类型
重复	如果列中的值为空（空值），则重复上一行的列值
使用范例	

5.11.2. 示例

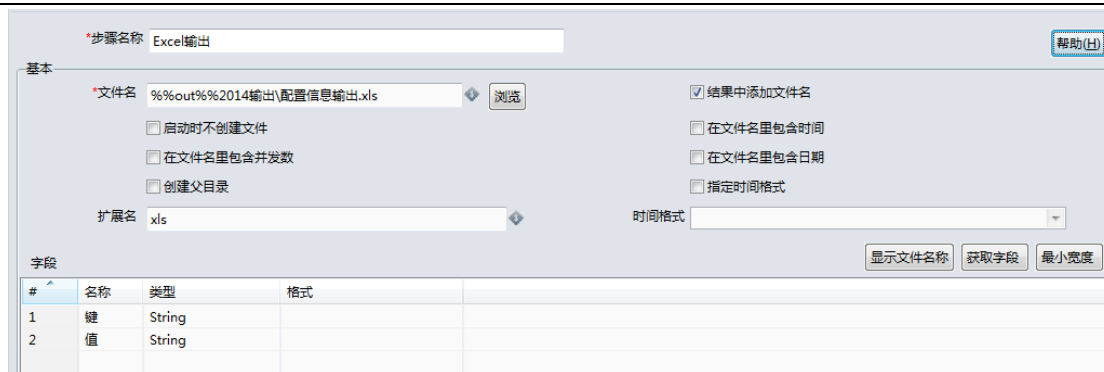
示例说明：将配置信息输出为 excel 文档。转换如下图所示：



示例步骤：

1. 编辑“配置信息”转换的属性面板，如下图所示：

2. 编辑转换“Excel 输出”属性面板，如下图所示：



3. 运行转换，即可。

5.12. 输出组件：Excel 输出

5.12.1. 参数说明

转换步骤名称	Excel 输出
功能描述	
利用这个步骤,在 DI Designer 支持的系统中,你可以写入数据到一个或者多个 Excel 文件中。	
参数说明	
基本	
选项	描述
步骤名称	步骤的名称,在单一的步骤中,名称必需
文件名	文件名,可以用变量
扩展名	扩展名,可以用变量,用于输出文件
包含步骤数	这参数主要是在多个文件执行中时需选中
包含日期	允许文件名中包含系统日期
包含时间	允许文件名中包含系统时间
显示文件	展现步骤运行后,创建的文件名集合,这是一个虚拟的集合,有时候它取决于每一个文件中的记录数目等等
字段	
你可以指定你想输出到 EXCEL 文件中的字段。如果电子表单的头部,按钮“获取字段”将自动获取可以使用的字段,可以指定 EXCEL 使用的格式。	
高级	
高级的组包含了一些用户不经常设置的信息,主要的参数如下:	
选项	描述
头部	检查表单是否有头部
尾部	检查表单是否有尾部
编码	指定表单使用的编码
分割记录	分割数据到几个输出文件中
工作表名称	指定要写入数据的工作表名称
保护工作表	输出的工作表是否使用密码保护
密码	指定保护的工作表的密码

使用模板	输出的 EXCEL 数据是否使用模板
Excel	用来格式模板名称的 excel
模板	Excel 输出文件
附加 Excel 模板	是否将 EXCEL 输出添加到指定的 EXCEL 模板中
使用范例	

5.12.2. 示例

略。

5.13. 输出组件：Meta 输出

5.13.1. 参数说明

转换步骤名称	Meta 输出
功能描述	
这个步骤可以存储信息到 AE 元数据数据库表中	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
显示名	读取的元数据实体种类名称
目标表	要写数据的 AE 元数据表名
提交记录数量	在 AE 元数据表中用事务插入行。如果 N 比 0 大，每 N 行提交一次连接。否则，不使用事务，速度会慢一些。
裁剪表	在第一行数据插入之前裁剪元数据表
忽略插入错误	使 AE Designer 忽略比如违反主键约束之类的插入错误，最多 20 个警告将被日志记录。在批量插入的时候这个功能不可用。
表分区数据	使用这个选项可以在多个表之间拆分数据。例如用把数据放到表 SALES_200510, SALES_200511, SALES_200512, 中来代替直接插入数据到表 SALES 中。以下的视图 SALES 展示完整的订单： CREATE OR REPLACE VIEW SALES AS SELECT * FROM SALES_200501 UNION ALL SELECT * FROM SALES_200502 UNION ALL SELECT * FROM SALES_200503 UNION ALL SELECT * FROM SALES_200504 ...
表名定义在一个字段里	使用这些选项可以拆分数据到一个或者多个 AE 元数据表里，目标表名可以用你指定的字段来定义。例如如果你想存储顾客性别数据，这些数据可能会存储到表 M 和表 F 里面 (female 女性和 male 男性表)。这个选项可以阻止这些字段插入到对应的表里。
返回一个自动	往表中插入行时，是否产生一个关键字

产生的关键字	
产生关键字的 字段名称	指定包含关键字的输出字段的字段名称
SQL	生成自动创建输出表的 SQL 脚本
使用范例	

5.13.2. 示例

略。

5.14. 输出组件：删除

5.14.1. 参数说明

转换步骤名称	删除
功能描述	
这个步骤利用查询关键字在表中搜索行。如果行找到，就删除记录。	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	目标表所在的数据库连接名称
目标模式	要写入数据的表的 Schema 名称。允许表名中包含“.”是很重要的
目标表	想插入或者更新的表的名称
提交数量	提交之前要改变（插入/更新）的行数。
用来查询的关键字	可以指定字段值或者比较符。可以用以下比较符：=, <>, <, <=, >, LIKE, BETWEEN, IS NULL, IS NOT NULL。
使用范例	

5.14.2. 示例

示例说明：将 excel 作为输入，并搜索文档中字段为 a 的进行删除。示例结果图如下所示：



示例步骤:

step1. 编辑转换“excel 输入”属性面板，如下图所示:

step2. 编辑转换“删除”属性面板，如下图所示:

运行转换，即可。

5. 15. 输出组件：插入更新

5. 15. 1. 参数说明

转换步骤名称	插入更新
功能描述	这个步骤利用查询关键字在表中搜索行。如果行没有找到，就插入行。如果能被找到，并且要被更新的字段没有任何改变，就什么也不做。如果有不同，行就会被更新。
参数说明	

选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	目标表所在的数据库连接名称
目标模式	要写入数据的表的 Schema 名称。允许表名中包含“.”是很重要的
目标表	想插入或者更新的表的名称
提交数量	提交之前要改变（插入/更新）的行数。
不执行任何更新	如果被选择，数据库的值永远不会被更新。仅仅可以插入
用来查询的关键字	可以指定字段值或者比较符。可以用以下比较符：=, <>, <, <=, >, LIKE, BETWEEN, IS NULL, IS NOT NULL。
更新字段	指定你想要插入/更新的字段
SQL 按钮	点击 SQL 按钮可以生成需要的 SQL 脚本
使用范例	

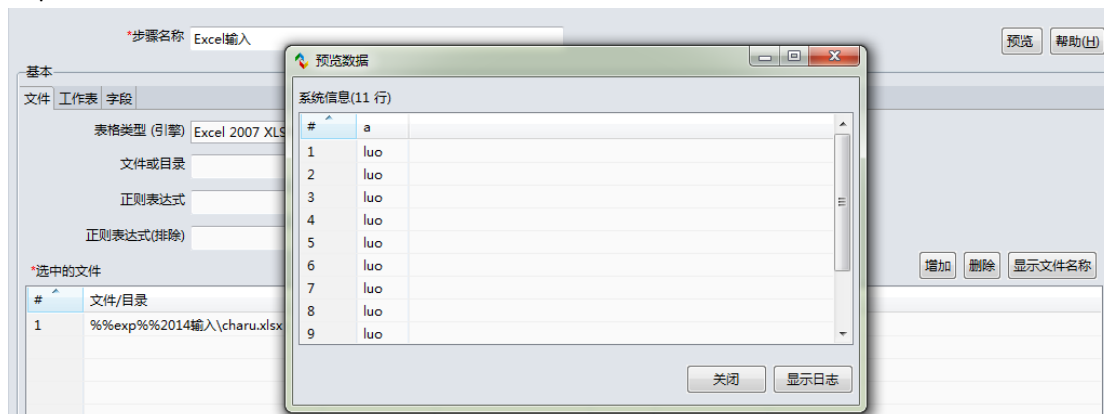
5.15.2. 示例

示例说明：将 excel 作为输入并插入更新记录。转换如下图所示：

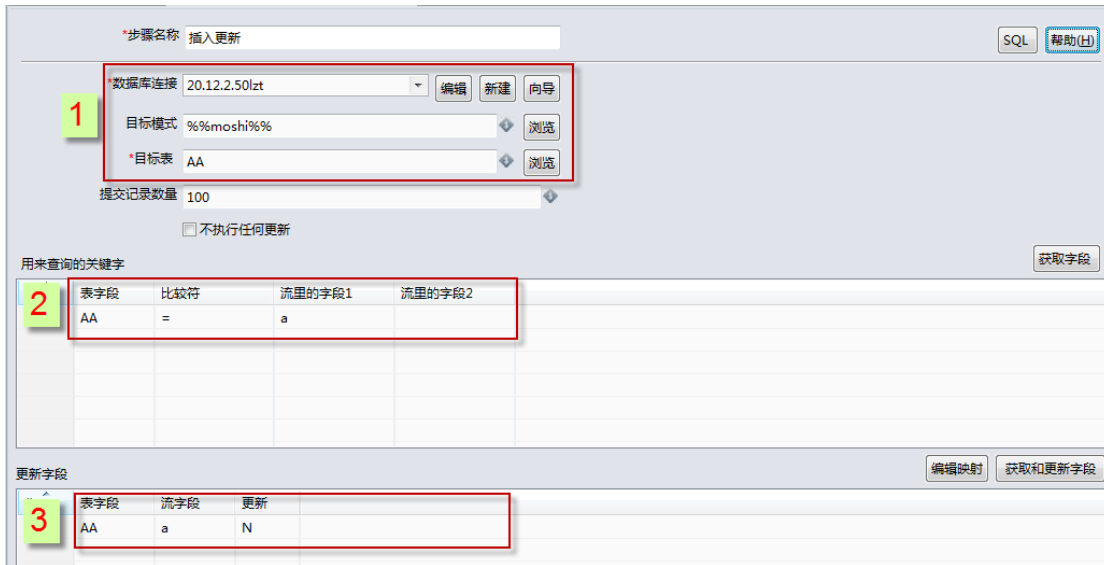


示例步骤：

step1. 编辑转换“excel 输入”属性面板，如下图所示：



step2. 编辑转换“插入更新”属性面板，如下图所示：



step1. 运行转换，即可。

5. 16. 输出组件：表输出

5. 16. 1. 参数说明

转换步骤名称	表输出
功能描述	
这个步骤可以存储信息到数据库表中。	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	用来写数据的数据库连接
目标模式	要写入数据的表的 Schema 名称。允许表名中包含“.”是很重要的
目标表	想插入或者更新的表的名称
提交记录数量	在数据表中用事务插入行。如果 N 比 0 大，每 N 行提交一次连接。否则，不使用事务，速度会慢一些。
裁剪表	在第一行数据插入之前对表进行 truncate 操作，清空表。
忽略插入错误	使 AE Designer 忽略比喻违反主键约束之类的插入错误，最多 20 个警告将被日志记录。在批量插入的时候这个功能不可用。
使用批量插入	如果你想批量插入的话，就使用这个选项。这个选项的速度最快，默认被选上。
表分区数据	使用这个选项可以在多个表之间拆分数据。例如用把数据放到表 SALES_200510, SALES_200511, SALES_200512, ... 中来代替直接插入数据到表 SALES 中。以下的视图 SALES 展示完整的订单：

	CREATE OR REPLACE VIEW SALES AS SELECT * FROM SALES_200501 UNION ALL SELECT * FROM SALES_200502 UNION ALL SELECT * FROM SALES_200503 UNION ALL SELECT * FROM SALES_200504 ...
表名定义在一个字段里	使用这些选项可以拆分数据到一个或者多个表里，目标表名可以用你指定的字段来定义。例如如果你想存储顾客性别数据，这些数据可能会存储到表 M 和表 F 里面（female 女性和 male 男性表）。这个选项可以阻止这些字段插入到对应的表里。
返回一个自动产生的关键字	往表中插入行时，是否产生一个关键字
自动产生关键字的字段名称	指定包含关键字的输出字段的字段名称
SQL	生成自动创建输出表的 SQL 脚本
使用范例	

5.16.2. 示例

略。

5.17. 转换组件：值映射

5.17.1. 参数说明

转换步骤名称	值映射
功能描述	
这个步骤简单的映射字符串，从一个值映射到另一个值。通常你想解决存储一个数据库中转换表的问题，不管怎么说，这是一种可选的方案：简单的将转换表作为值映射对话框的一部分。 例如：如果你想替换 Language codes，你可以：使用的字段名：LanguageCode，目标字段名：LanguageDesc。源值/目标值： EN/English, FR/French, NL/Dutch, ES/Spanish, DE/German	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
使用的字段名	使用的字段，源字段
目标字段名	使用的字段，目标字段
不匹配的默认值	定义一个默认值用于源字段不为空，但是没有字段匹配
字段列表	包含了源字段到目标字段的具体映射
使用范例	

5.17.2. 示例

示例说明：将 excel 输入文件中字段 a 的值为 2 转换为 4，并输出 excel 文件。转换如下图所示：



示例步骤：

step1. 编辑转换“Excel 输入”属性面板，如下图所示：

1. 步骤名称: Excel输入

2. 文件类型 (引擎): Excel 2007 XLSX (Apache POI)

3. 增加 删除 显示文件名称

#	文件/目录	正则表达式	正则表达式(排除)	要求	包含子目录
1	%%exp%%excelshuru*.xls			否	否

step2. 编辑转换“值映射”属性面板，如下图所示：

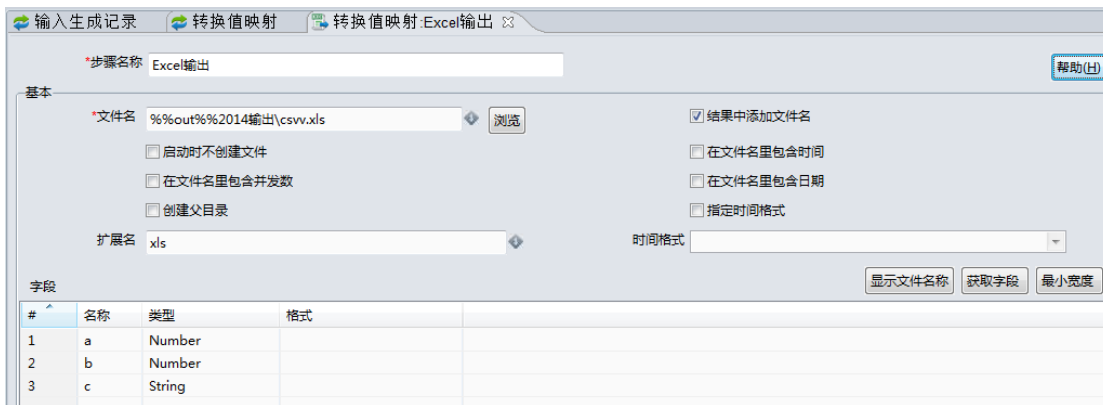
1. 步骤名称: 值映射

2. 使用的字段名: a

3. 源值 目标值

源值	目标值
2	4

step3. 编辑转换“Excel 输出”属性面板，如下图所示：



step4. 运行转换，即可。

5. 18. 转换组件：行转列

5. 18. 1. 参数说明

转换步骤名称	列转行										
功能描述	这个步骤允许你通过查询键值对来反向规格化数据。也可以立即转换数据类型。										
参数说明	<table border="1"> <thead> <tr> <th>选项</th> <th>描述</th> </tr> </thead> <tbody> <tr> <td>步骤名称</td> <td>步骤名称，在单个转换中必须唯一</td> </tr> <tr> <td>关键字段</td> <td>指定关键字段</td> </tr> <tr> <td>分组字段</td> <td>指定构成分组的字段</td> </tr> <tr> <td>目标字段</td> <td>指定反向规格化的字段</td> </tr> </tbody> </table>	选项	描述	步骤名称	步骤名称，在单个转换中必须唯一	关键字段	指定关键字段	分组字段	指定构成分组的字段	目标字段	指定反向规格化的字段
选项	描述										
步骤名称	步骤名称，在单个转换中必须唯一										
关键字段	指定关键字段										
分组字段	指定构成分组的字段										
目标字段	指定反向规格化的字段										
使用范例											

5. 18. 2. 示例

示例说明：行转列。转换如下图所示：

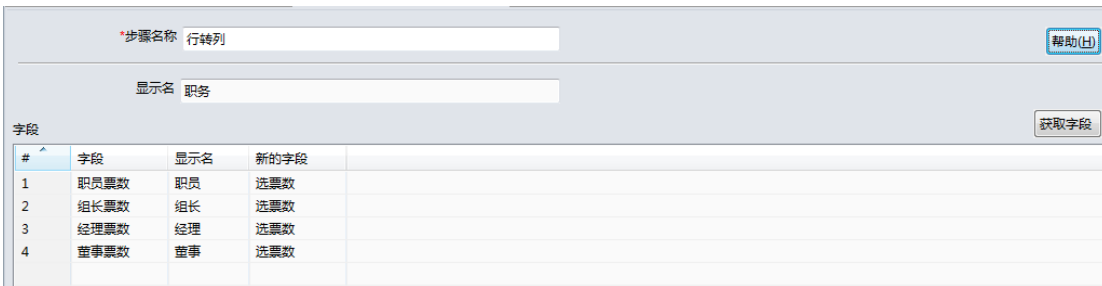


示例步骤:

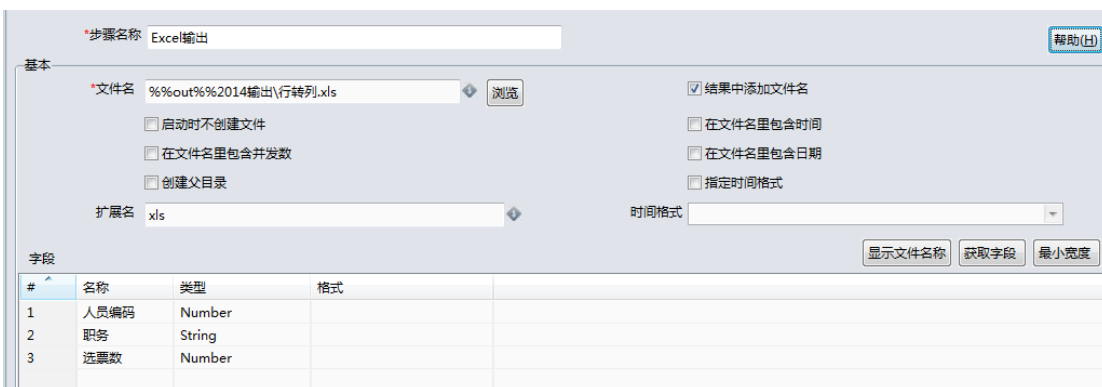
step1. 编辑转换“excel 输入”属性面板，如下图所示:



step2. 编辑转换“行转列”属性面板，如下图所示:



step3. 编辑转换“excel 输出”属性面板，如下图所示:



step4. 运行转换，即可。

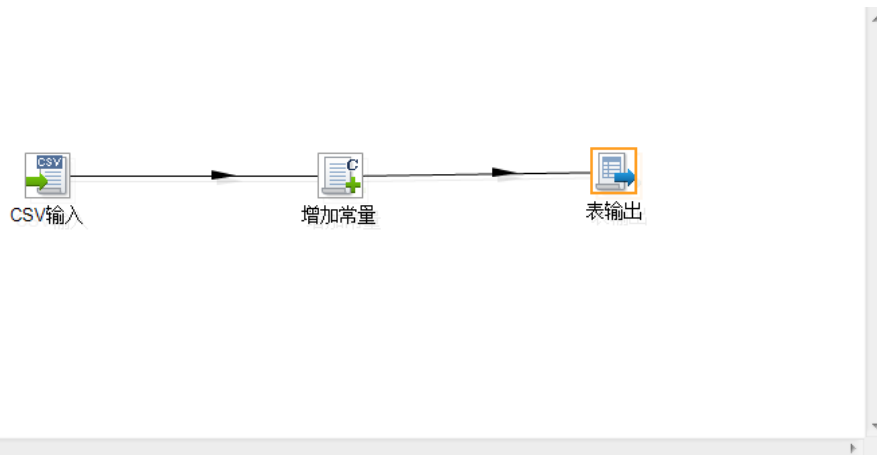
5. 19. 转换组件：增加常量

5. 19. 1. 参数说明

转换步骤名称	增加常量
功能描述	
这个步骤很简单，主要是添加常量到流中。它的使用也很容易：用字符串形式指定名称，类型和值。利用选择的数据类型指定转换格式。 .	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
字段	确定字段的名称，类型和值，然后可以指定格式用于不同类型之间的数据转换
使用范例	

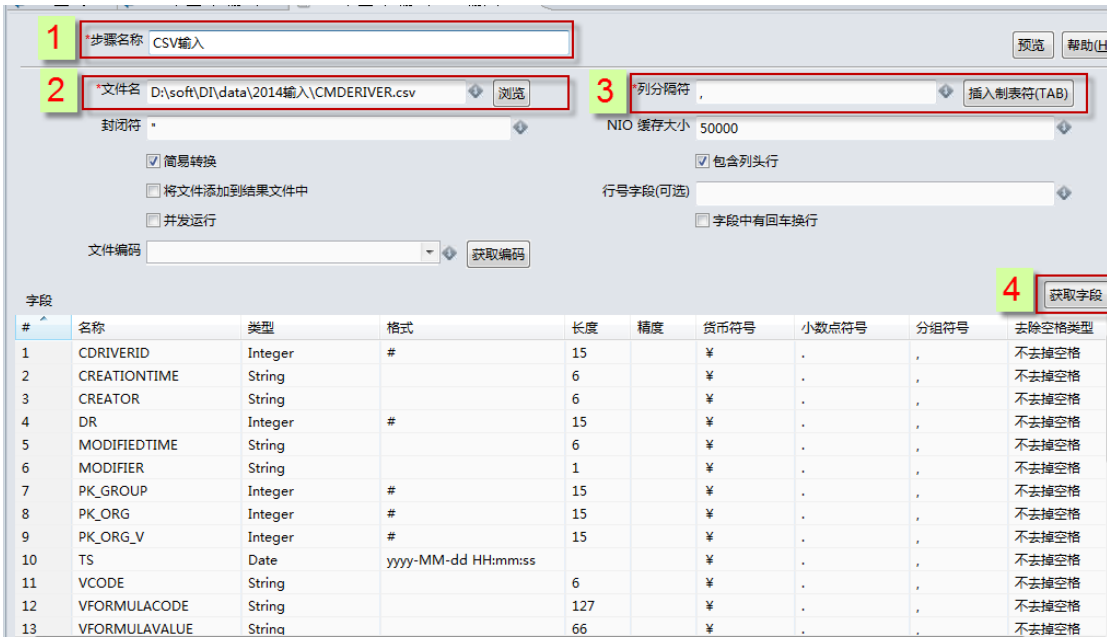
5. 19. 2. 示例

示例说明：增加常量。转换如下图所示：

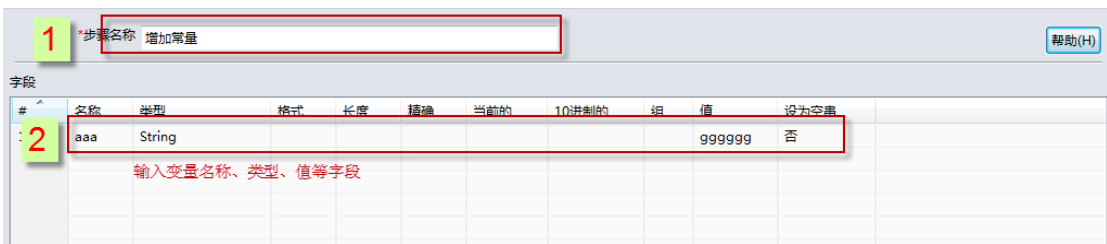


示例步骤：

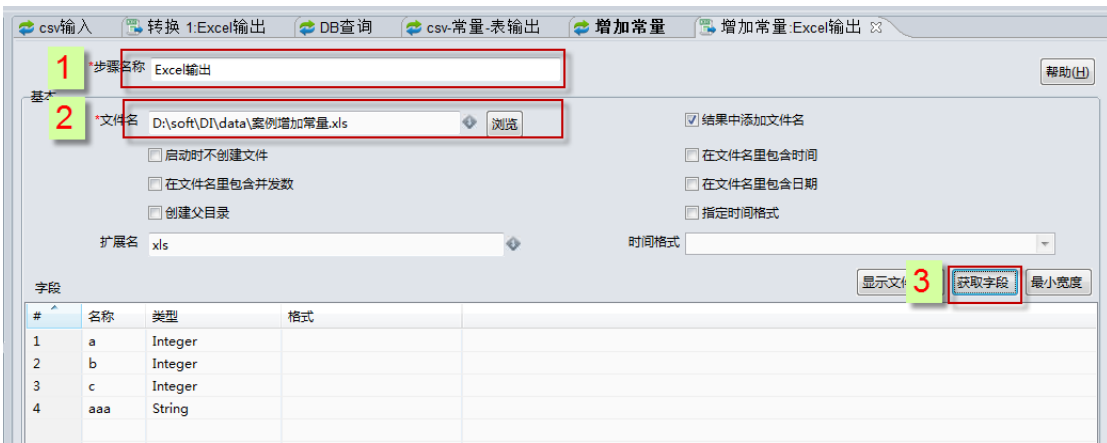
step1. 编辑转换“CSV 输入”属性面板，如下图所示：



step2. 编辑转换“增加常量”属性面板，如下图所示：



step1. 编辑转换“excel 输出”属性面板，如下图所示：



运行转换，即可。

5. 20. 转换组件：增加序列

5. 20. 1. 参数说明

转换步骤名称	增加序列
功能描述	

<p>这个步骤在流中增加一个序列。一个序列是在某个起始值和增量的基础上，经常改变的整数值。你可以使用数据库的序列，也可以使用 DI Designer 决定的序列。 备注：DI Designer 序列在同一个转换中是唯一使用的。每一次转换运行的时候，序列的值又会重新循环一次（从开始值开始）</p>	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
值的名称	新序列值的名称，将被添加到流中
使用数据库获取序列	连接名称：选择数据库序列存在的连接名称。 序列名称：数据库序列的名称。
使用计数器来计算序列	计数器名称（可选）：如果一个转换中多个步骤生成同样的值名称，这个选项允许你指定计数器的名称，避免按照先后顺序通过多个步骤 起始值：序列的起始值。 增量：序列的增量。 最大值：序列的最大值，如果超过，将重新开始
使用范例	
<p>Start at = 1, increment by = 1, max value = 3 将会产生序列：1, 2, 3, 1, 2, 3, 1, 2...</p> <p>Start at = 0, increment by = -1, max value = -2 将会产生序列：0, -1, -2, 0, -1, -2, 0... </p>	

5.20.2. 示例

略。

5.21. 转换组件：字段选择

5.21.1. 参数说明

选择字段步骤可用于在数据流中选择、删除、重命名、更改数据类型以及配置字段的长度和精度。这些操作可分为不同类别：

选择并更改——指定在输出行显示的字段的确切顺序和名字

删除——指定需要从输出行中删除的字段

元数据——更改一个或多个字段的名称、类型、长度和精度（元数据）

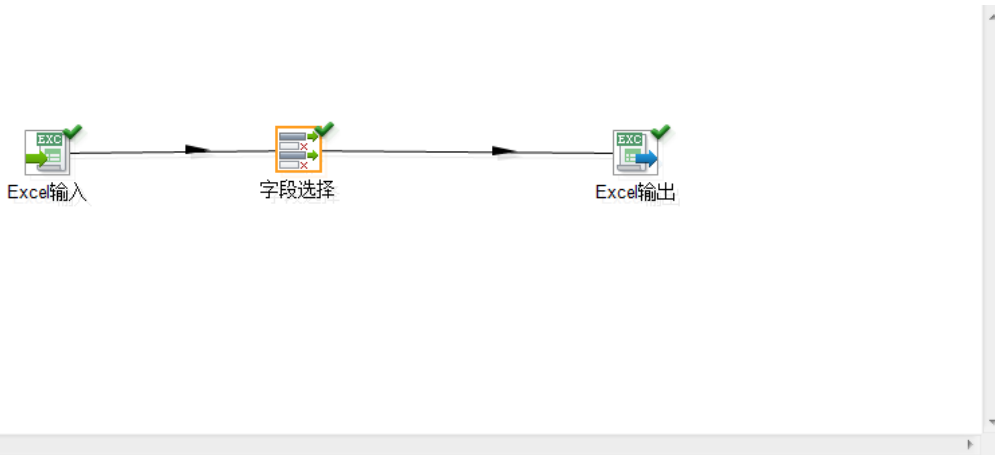
转换步骤名称	字段选择
功能描述	
选择字段步骤可用于在数据流中选择、删除、重命名、更改数据类型以及配置字段的长度和精度。	
参数说明	
选择和修改页签	

选项	描述
步骤名称	步骤名称，在单个转换中必须唯一，您也可以根据需要更改此步骤的名称。
字段名	单击以将所有输入流中的字段插入步骤中
重命名为	单击以将所有输入流中的字段插入步骤中
长度	如果要从输入流中隐式选择所有未在“字段”部分明确选择的其他字段，请启用此选项。
精度	精度选项取决于字段类型，只支持数字；它返回的是浮点数字的数量。
包括未指定字段，根据名称排序	如果要从输入流中隐式选择所有未在“字段”部分明确选择的其他字段，请启用此选项
移除页签	
填写需要从输入流中删除的字段名。 注意：字段移除将使内部的字段顺序不断变化，从而减慢流程运行的速度。	
元数据页签	
使用此选项卡中的选项，可以进行重命名和更改数据类型，以及更改进入“选择值”步骤的字的长度和精度。单击“获取字段”更改为从上一步中导入字段。此选项卡也可用于进行许多数据类型转换。	
字段名	导入字段的名称
重命名为	如果要为字段重命名，可在此填入新名称
类型	此字段的数据类型
长度	字段长度
精度	精度选项取决于字段类型，只支持数字；它返回的是浮点数字的数量
二进制到常规	在合适的时候将字符串转换为数字数据类型
格式	格式掩码（数字类型或日期格式）
日期格式宽松	用于指定日期解析算法是严格还是宽松。宽松意味着会对无效的日期值进行处理。如果设置为“否”，将只接受严格有效的日期值；如果指定为“是”，解析算法会试图找出不正确日期的意图，还可能对其进行更正。
日期区域设置	指定日期转换和计算使用的日期区域设置。将其留空以使用系统的默认编码；或相应地从已填充的列表中选择。
日期时区	指定日期转换和计算使用的日期时区。将其留空以使用系统的默认编码；或相应地从已填充的列表中选择。
宽松数字转换	如果设置为“是”，将对数字进行解析直到找到非数值（如破折号或斜线），此时停止解析而不报告错误。如果设定为“否”，则对数字进行严格解析；如果输入中有无效数字，则报告错误。默认设置为“否”，可以通过将变量 KETTLE_LENIENT_STRING_TO_NUMBER_CONVERSION 设置为“是”来进行更改。
编码	指定要使用的文本文件编码。将此字段留空以使用系统的默认编码。要使用 Unicode，请指定 UTF-8 或 UTF-16。首次使用时，Spoon 搜索系统中的可用编码，并相应地填充该列表。

小数	小数点；可以是点号或逗号
分组	一种分隔四位数或更多位数的三位一组的分隔方法。可用小数点或逗号。
货币	用于代表货币的符号
使用范例	

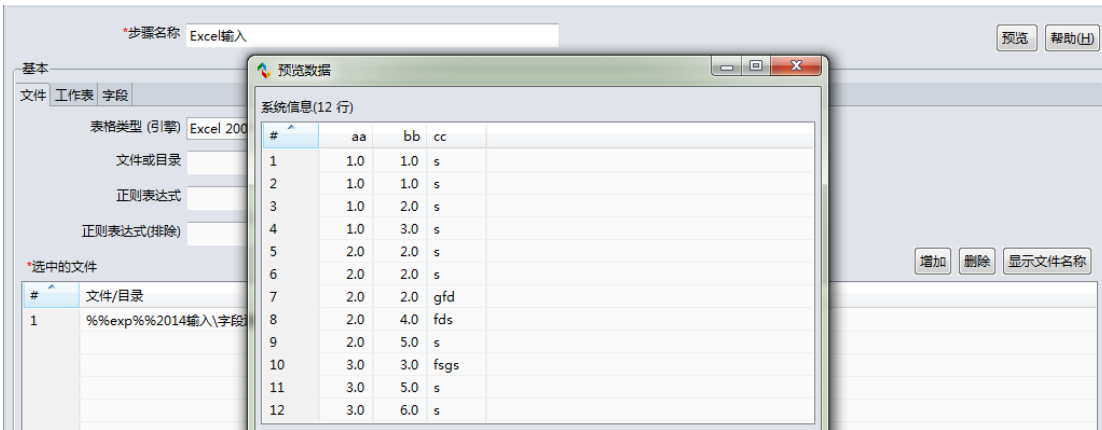
5.21.2. 示例

示例说明：有选择的输出字段。示例结果图如下所示：

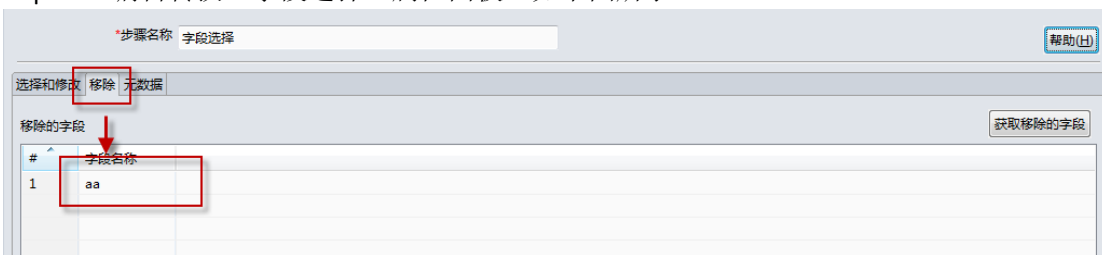


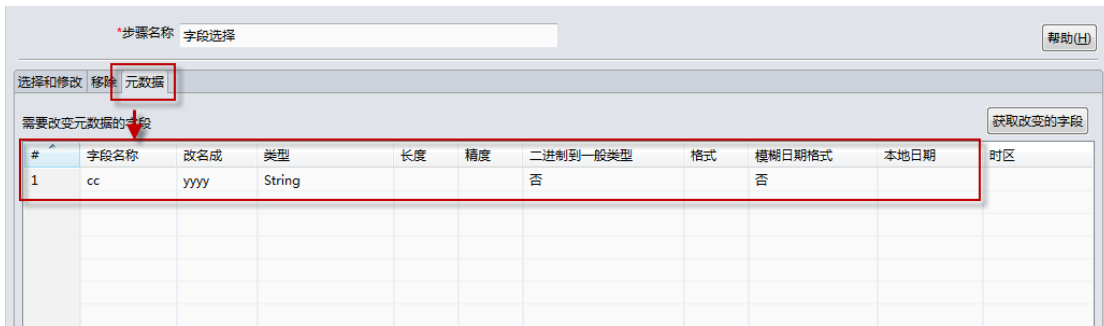
示例步骤：

step1. 编辑转换“excel 输入”属性面板，如下图所示：

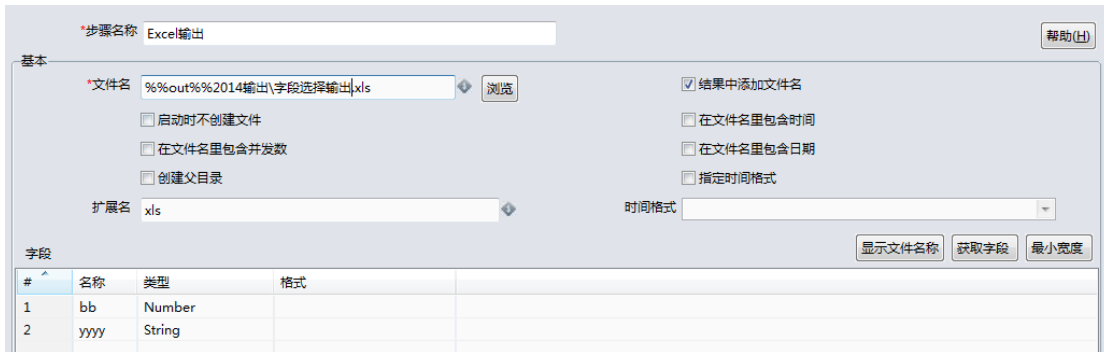


step2. 编辑转换“字段选择”属性面板，如下图所示：





step1. 编辑转换“excel 输出”属性面板，如下图所示：



step2. 运行转换，结果如下所示：

	A	B	C	D	E	F	G	H	I	J	K
1	bb	yyyy									
2	1.00	s									
3	1.00	s									
4	2.00	s									
5	3.00	s									
6	2.00	s									
7	2.00	s									
8	2.00	gfd									
9	4.00	fds									
10	5.00	s									
11	3.00	fsgs									
12	5.00	s									
13	6.00	s									
14											

5.22. 转换组件：拆分字段

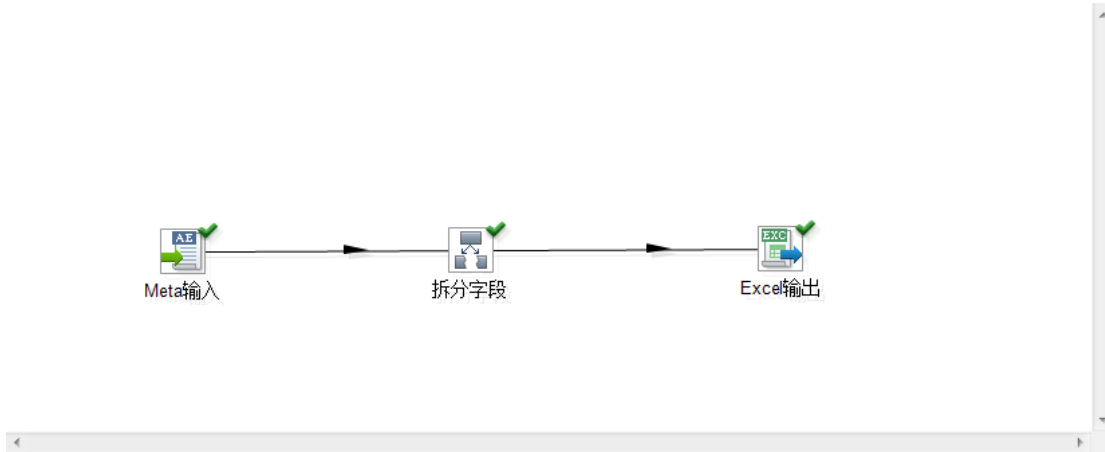
5.22.1. 参数说明

转换步骤名称	拆分字段
功能描述	
这个步骤允许你根据分隔符来拆分字段	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
需要拆分的字段	想要拆分的字段的名
分隔符	决定字段结束的分隔符
字段表格	拆分形成的字段列表

使用范例

5.22.2. 示例

示例说明：将字段拆分并输出为 Excel 文档。示例结果图如下所示：



示例步骤：

step1. 编辑转换“meta 输入”属性面板，如下图所示：

1 步骤名称 Meta输入

2 *显示名 DB2模型数据3级次1(DB2)

3 *表名 BI_EXCEL_IMPORT_OERB3

记录数量限制 0

#	显示名	列名	数据类型
1	name1	NAME1	String
2	values1	VALUES1	BigNumber
3	riqi	RIQI	String
4	pk_group	PK_GROUP	String
5	地区	DIQU	String
6	pk_org	PK_ORG	String
7	wnd_gai_pk_org	WND_GAI_PK_ORG	String
8	主键	GUID	String
9	code	CODE	String
10	ass	ASS	String
11	city	CITY	String

4 重新获取字段

step2. 编辑转换“拆分字段”属性面板，如下图所示：

1 步骤名称 拆分字段

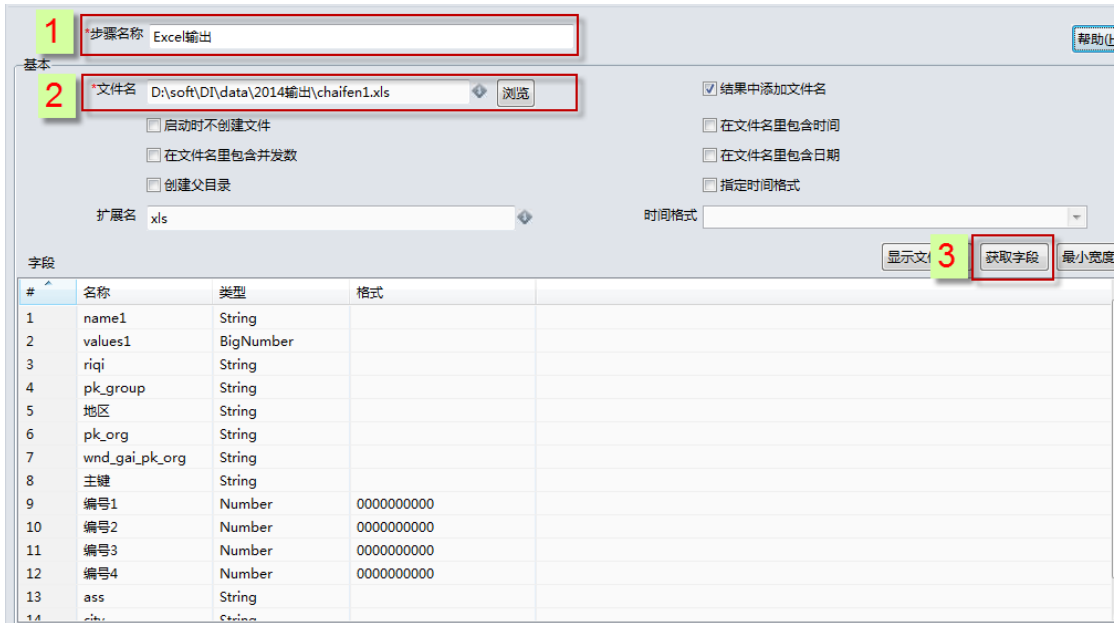
2 需要拆分的字段 code

3 分隔符 1

字段

#	新的字段	ID	移除ID?	类型	长度	精度	格式	分组符号	小数点符号	货币符号	空	缺省	去除
1	编号1		N	Number	10								不
2	编号2		N	Number	10								不
3	编号3		N	Number	10								不
4	编号4		N	Number	10								不

step1. 编辑转换“excel 输出”属性面板，如下图所示：



step2. 运行转换，即可。

5. 23. 转换组件：排序记录

5. 23. 1. 参数说明

“排序记录”步骤可根据指定的字段以及是按升序还是降序对行进行排序。

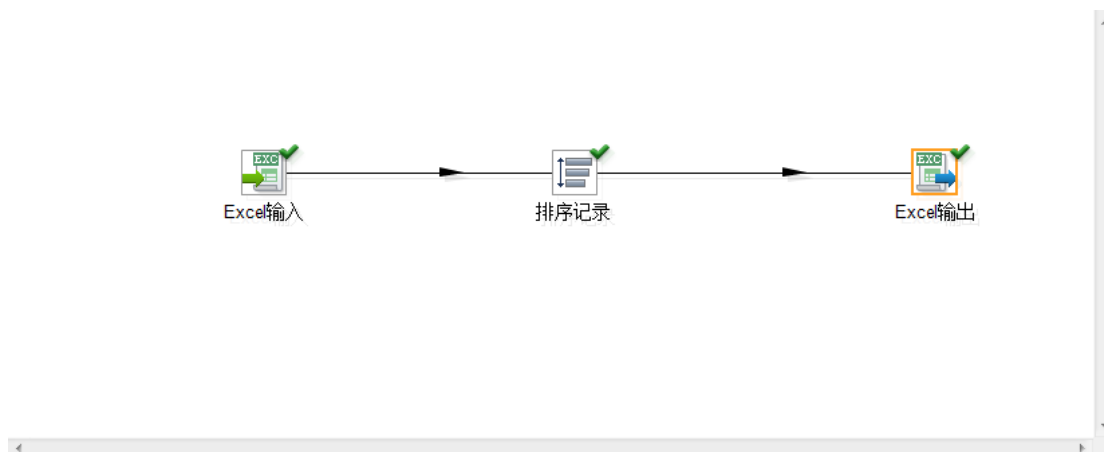
注意：当行数超过了指定的排序大小时（默认设置为一百万行），DI Designer 将使用临时文件对行进行排序。当遇到内存不足异常 (OOM)，则需要减小此大小或更换可用的内存。当使用并行步骤的多个副本时（含有“改变副本数量以开始”的本地 JVM 或在使用 Carte 的集群环境），需要将每个经过排序的数据块合并起来，以保证适当的排序顺序。完成此项操作需要之后添加“排序合并”步骤（无需开启多个副本的本地 JVM 或在主机上的集群）。

转换步骤名称	排序记录
功能描述	
“排序记录”步骤可根据指定的字段以及是按升序还是降序对行进行排序。	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
排序目录	保存临时文件以备所需的目录；默认设置为系统的标准临时目录
TMP 文件前缀	选择一个易于辨识的前缀，以便在临时文件目录中找到文件。
排序大小	存储在内存中的行数越多，排序流程进行得越快，因为只需使用较少的临时文件且产生的 I/O 较少
空闲内存阈值（按百分比计）	如果排序算法发现可用的内存比少于指示的数值小，则

	会将数据分页传递到磁盘中。 注意：但这无法精确计算，因为：每一千行执行一次检查。考虑到行的大小和复杂转换中的其他步骤，仍旧有可能造成内存不足问题。在 Java 虚拟机中不可能知道确切的空闲内存量。因此我们建议您不要将其用包含使用大量内存的步骤和流程的非常复杂的转换。
压缩 TMP 文件	压缩临时文件以完成排序过程。
只传递唯一行	如果希望只将唯一行传入输出流，则将其启用。
“字段”表	指定要排序的字段和方向（升序/降序）。您可以指定是否执行区分大小写的排序（可选设置）
获得字段	单击以检索数据流带来的所有字段的列表。
使用范例	

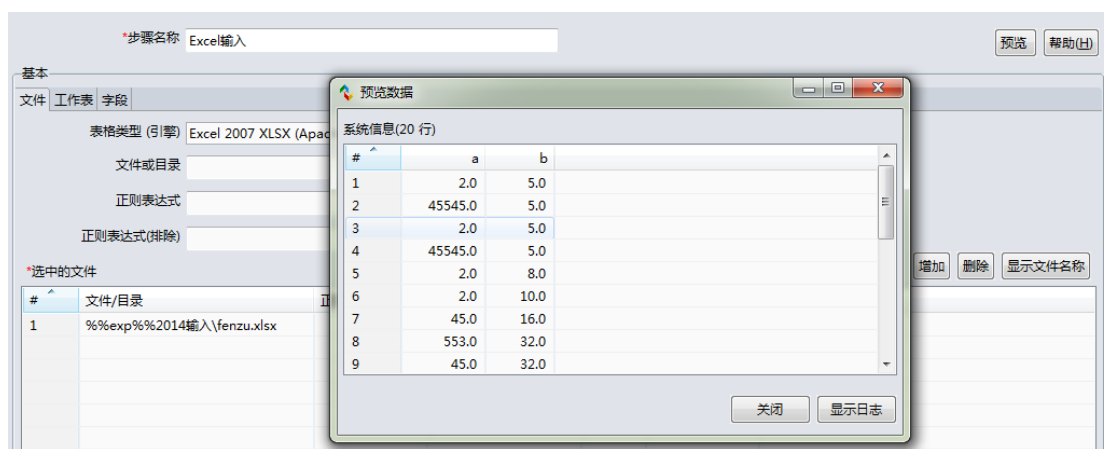
5.23.2. 示例

示例说明：对输入的字段进行排序处理并输出。示例结果图如下所示：

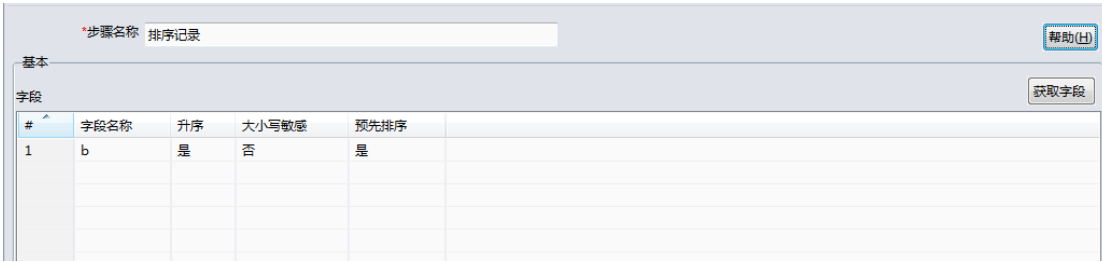


示例步骤：

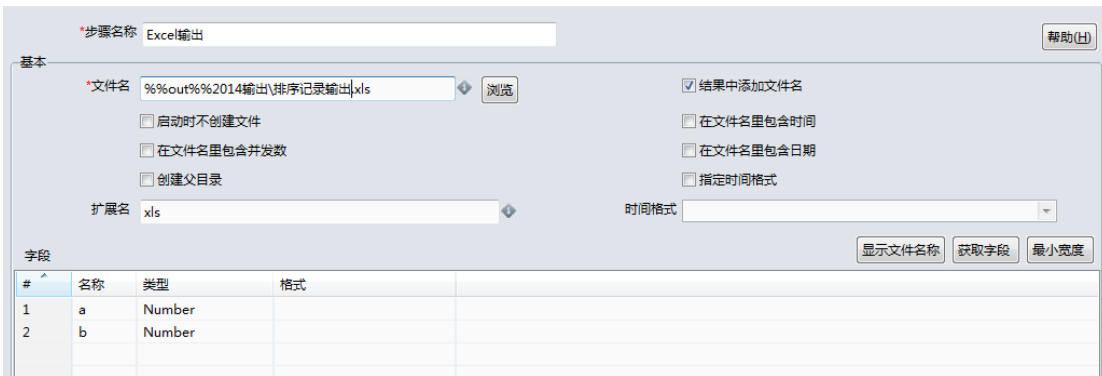
step1. 编辑转换“excel 输入”属性面板，如下图所示：



step2. 编辑转换“排序记录”属性面板，如下图所示：



step3. 编辑转换“Excel 输出”属性面板，如下图所示：



step4. 运行转换，结果如下图所示：

	A	B	C	D	E	F	G	H	I	J	K
1	a	b									
2	2.00	5.00									
3	2.00	5.00									
4	2.00	8.00									
5	2.00	10.00									
6	24.00	65.00									
7	24.00	65.00									
8	45.00	16.00									
9	45.00	32.00									
10	45.00	32.00									
11	45.00	45.00									
12	45.00	323.00									
13	45.00	323.00									
14	232.00	323.00									
15	232.00	323.00									
16	553.00	32.00									
17	553.00	32.00									
18	7,845.00	65.00									
19	7,845.00	65.00									
20	45,545.00	5.00									
21	45,545.00	5.00									

5. 24. 转换组件：替换 NULL

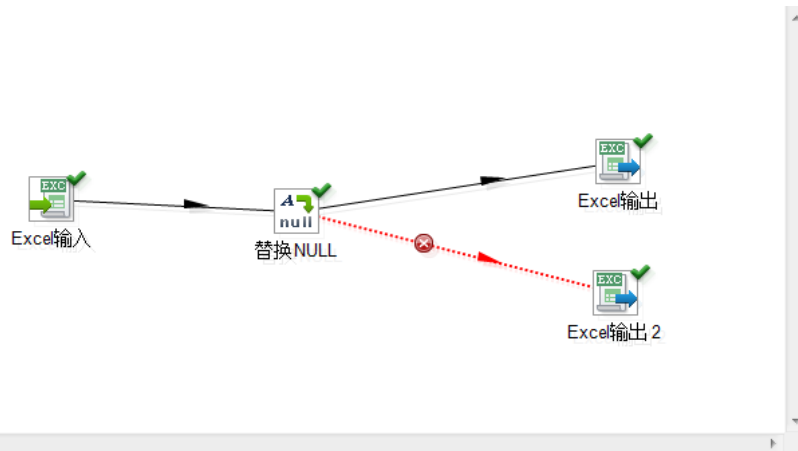
5. 24. 1. 参数说明

转换步骤名称	替换 NULL
功能描述	
替换 null 值步骤允许根据一定的条件来替换 null 值	
参数说明	

选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
值替换为	替换所有的 null 为输入的值
选择值类型	根据值得类型来替换 null 值 (Number, String, Date etc.)
选择字段	根据指定字段的名称替换 null 值
使用范例	

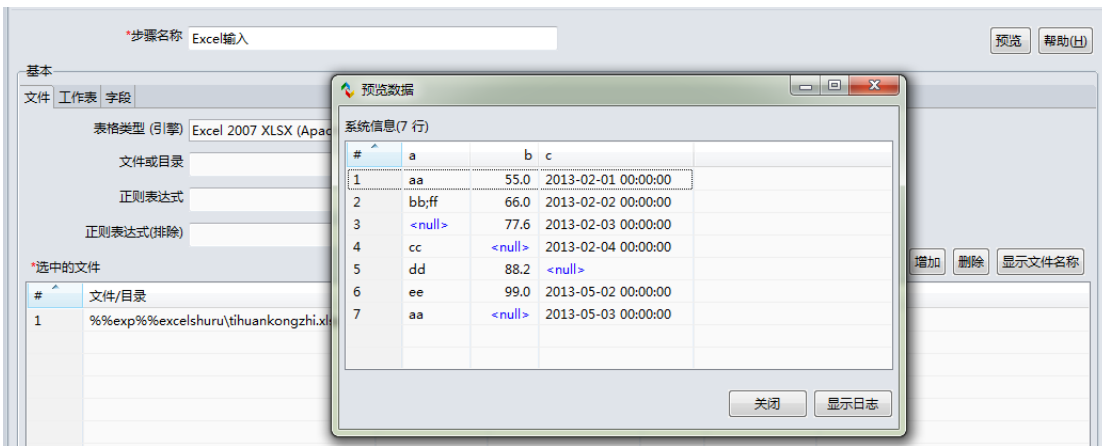
5.24.2. 示例

示例说明：对字段为空的进行替换处理，示例结果图如下所示：

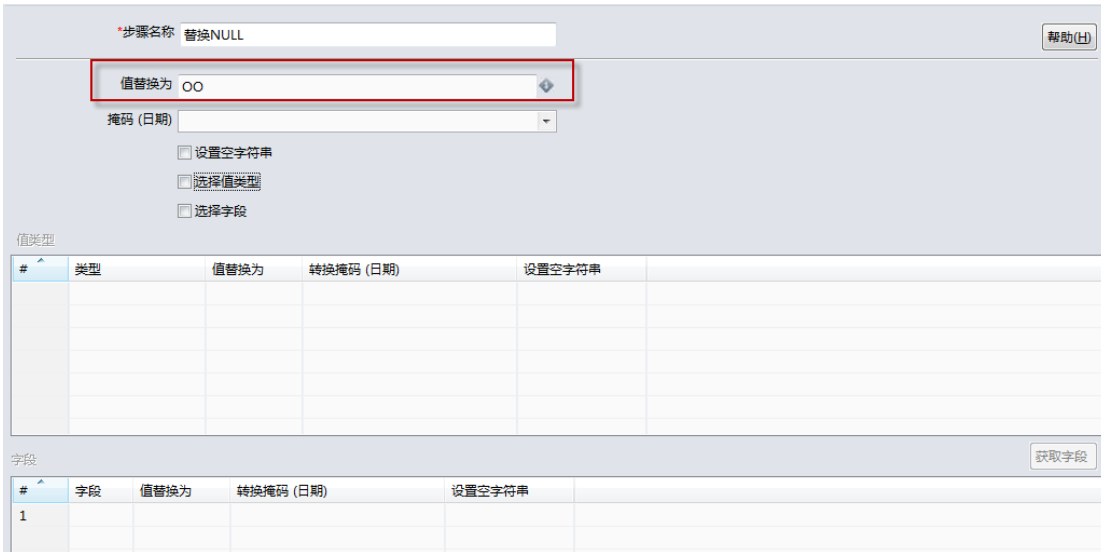


示例步骤：

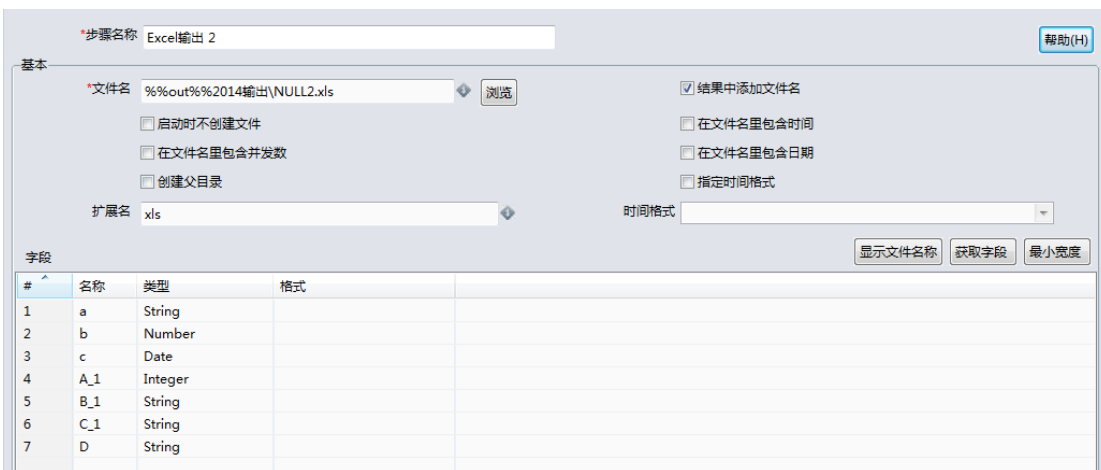
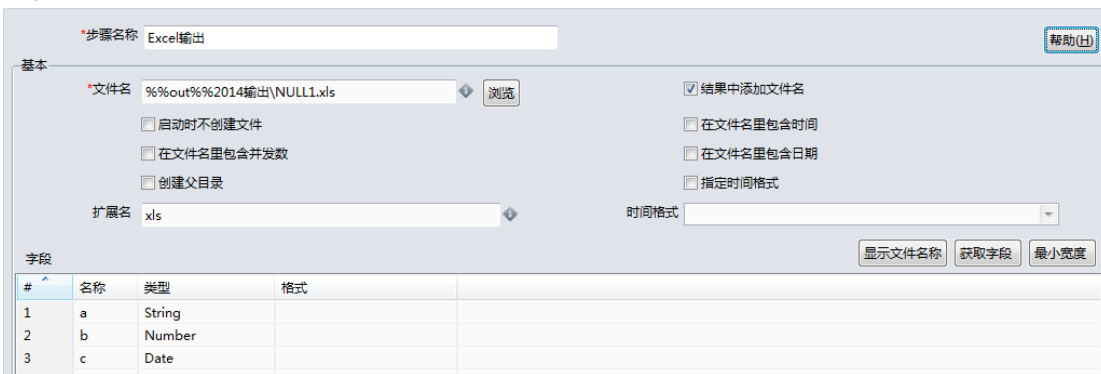
step1. 编辑转换“excel 输入”属性面板，如下图所示：



step2. 编辑转换“替换 NULL”属性面板，如下图所示：



step3. 编辑转换“Excel 输出”属性面板，如下图所示：



step4. 运行转换，结果如下图所示：

	A	B	C	D	E	F	G	H	I	J	K
1	a	b	c								
2	aa	55.00	#####								
3	bb,ff	66.00	#####								
4	OO	77.60	#####								
5	ee	99.00	#####								
6											
7											
8											
9											

A1 fx a

	A	B	C	D	E	F	G	H	I	J	K	L
1	a	b	c	A_1	B_1	C_1	D					
2	cc		#####	1.00	uap.impl.d		IFNULL001					
3	dd	88.20		1.00	uap.impl.d		IFNULL001					
4	aa		#####	1.00	uap.impl.d		IFNULL001					
5												
6												
7												
8												
9												

5. 25. 转换组件：生成 GUID

5. 25. 1. 参数说明

转换步骤名称	生成 GUID
功能描述	GUID 步骤用来生成指定长度 GUID，可用来生成数据库表主键
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
值的名称	新序列值的名称，将被添加到流中。
GUID 长度	生成 GUID 的固定长度，GUID 长度必须在 8-128 之间。
使用范例	

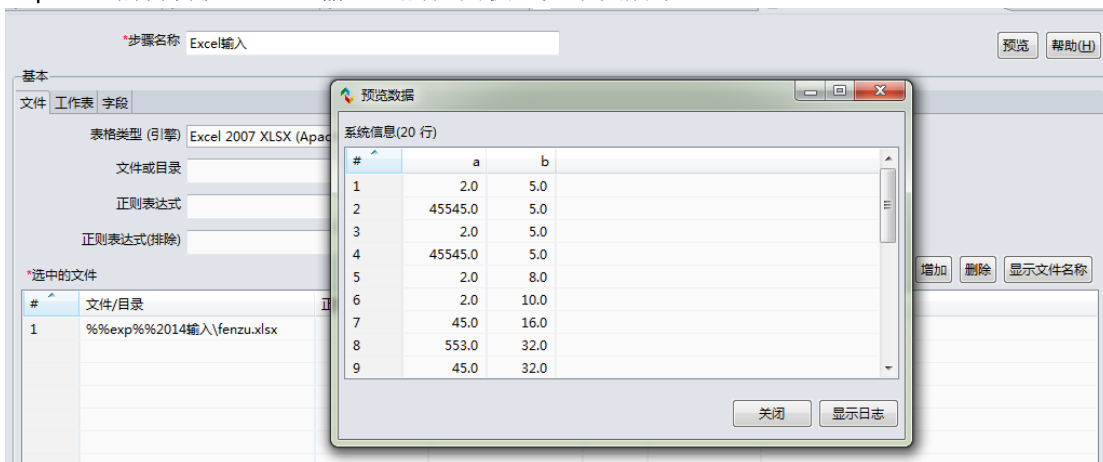
5. 25. 2. 示例

示例说明：新增字段 GUID，示例结果图如下所示：

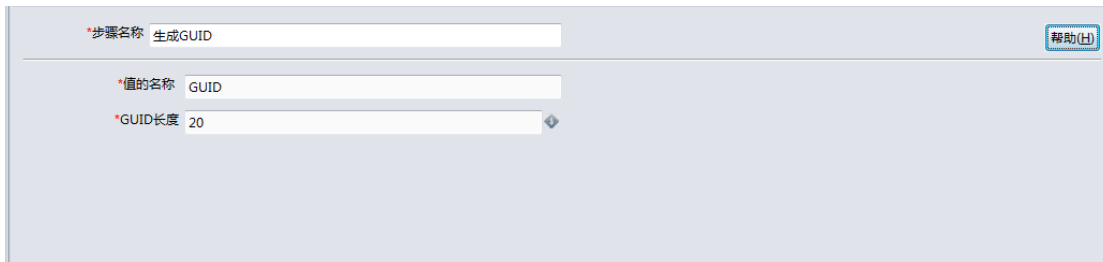


示例步骤:

step1. 编辑转换“excel 输入”属性面板，如下图所示:



step2. 编辑转换“生成 GUID”属性面板，如下图所示:



step3. 编辑转换“Excel 输出”属性面板，图略。

step4. 运行转换，结果如下图所示:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	a	b	GUID										
2	2.00	5.00	5fz0w0a8510ym2w0ick										
3	45,545.00	5.00	88ygxl9dx11m25v9u8gb										
4	2.00	5.00	e8ygxl9dx12wqng3xz4c										
5	45,545.00	5.00	u8ygxl9dx13nvrzgmtiq										
6	2.00	8.00	d8ygxl9dx145qt88gw0i										
7	2.00	10.00	t8ygxl9dx15lp59mn02t										
8	45.00	16.00	z8ygxl9dx165zaz8cjgr										
9	553.00	32.00	t8ygxl9dx17efxm62sln										
10	45.00	32.00	y8ygxl9dx18a8wpp7imi										
11	553.00	32.00	c8ygxl9dx196o2dypyfj										
12	45.00	32.00	q8ygxl9dx1abgm0wbq1w										
13	45.00	45.00	y8ygxl9dx1bewareuzk79										
14	24.00	65.00	y8ygxl9dx1ct8m906uzl										
15	7,845.00	65.00	m8ygxl9dx1dhgkk2o7lp										
16	24.00	65.00	h8ygxl9dx1eeeybknye7										
17	7,845.00	65.00	o8ygxl9dx1f384t0d10q										
18	232.00	323.00	o8ygxl9dx1gmchw2jhcv										
19	45.00	323.00	p8ygxl9dx1hmoh9pogpx										
20	232.00	323.00	k8ygxl9dx1ihidel1pg8										
21	45.00	323.00	j8ygxl9dx1j79ds07s6w										
22													

5. 26. 转换组件：行转列

5. 26. 1. 参数说明

转换步骤名称	行转列																																														
功能描述	将行数据转化为不同列																																														
参数说明	<table border="1"> <thead> <tr> <th>选项</th> <th>描述</th> </tr> </thead> <tbody> <tr> <td>步骤名称</td> <td>步骤名称，在单个转换中必须唯一</td> </tr> <tr> <td>字段类型</td> <td>字段类型的名称。</td> </tr> <tr> <td>字段列表</td> <td> 字段：标准化的字段列表： 字段名称：标准化的字段名称 类型：用字符串分类字段 新字段：你可以指定一个或者多个新值将要转入的字段。备注：可以一次转换多个字段 </td> </tr> <tr> <td>获取字段</td> <td>填充输入流中所有可用的字段</td> </tr> </tbody> </table>						选项	描述	步骤名称	步骤名称，在单个转换中必须唯一	字段类型	字段类型的名称。	字段列表	字段：标准化的字段列表： 字段名称：标准化的字段名称 类型：用字符串分类字段 新字段：你可以指定一个或者多个新值将要转入的字段。备注：可以一次转换多个字段	获取字段	填充输入流中所有可用的字段																															
选项	描述																																														
步骤名称	步骤名称，在单个转换中必须唯一																																														
字段类型	字段类型的名称。																																														
字段列表	字段：标准化的字段列表： 字段名称：标准化的字段名称 类型：用字符串分类字段 新字段：你可以指定一个或者多个新值将要转入的字段。备注：可以一次转换多个字段																																														
获取字段	填充输入流中所有可用的字段																																														
使用范例	<table border="1"> <thead> <tr> <th>DATE</th> <th>PR1_NR</th> <th>PR_SL</th> <th>PR2_NR</th> <th>PR2_SL</th> <th>PR3_NR</th> <th>PR3_SL</th> </tr> </thead> <tbody> <tr> <td>20030101</td> <td>5</td> <td>100</td> <td>10</td> <td>250</td> <td>4</td> <td>150</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table> <p>最终行转列的结果如下：</p> <table border="1"> <thead> <tr> <th>DATE</th> <th>TYPE</th> <th>Product Sales</th> <th>Product Number</th> </tr> </thead> <tbody> <tr> <td>20030101</td> <td>product1</td> <td>100</td> <td>5</td> </tr> <tr> <td>20030101</td> <td>product2</td> <td>250</td> <td>10</td> </tr> <tr> <td>20030101</td> <td>product3</td> <td>150</td> <td>4</td> </tr> <tr> <td>...</td> <td>...</td> <td>...</td> <td>...</td> </tr> </tbody> </table>						DATE	PR1_NR	PR_SL	PR2_NR	PR2_SL	PR3_NR	PR3_SL	20030101	5	100	10	250	4	150	DATE	TYPE	Product Sales	Product Number	20030101	product1	100	5	20030101	product2	250	10	20030101	product3	150	4
DATE	PR1_NR	PR_SL	PR2_NR	PR2_SL	PR3_NR	PR3_SL																																									
20030101	5	100	10	250	4	150																																									
...																																									
DATE	TYPE	Product Sales	Product Number																																												
20030101	product1	100	5																																												
20030101	product2	250	10																																												
20030101	product3	150	4																																												
...																																												

5.26.2. 示例

略。

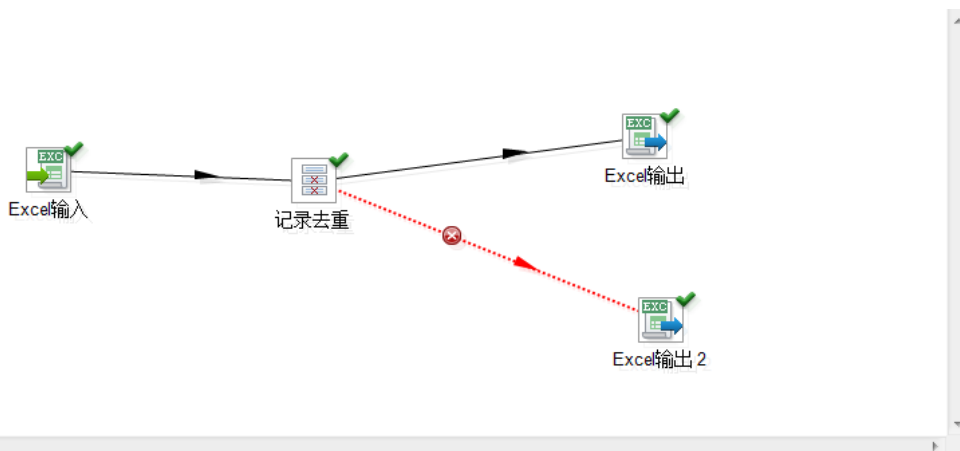
5.27. 转换组件：记录去重

5.27.1. 参数说明

转换步骤名称	记录去重
功能描述	此步骤用来去掉输入流中的相同记录
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
增加计数器到输出	增加一计数器列。
计数器字段	计数器字段的名称。
重定向重复记录	将处理重复记录作为错误，在错误的转换流中重定向记录。需要你为这一步设置错误处理。
错误描述	在错误处理描述中显示重复的行。只有当重定向重复记录选中时才有效。
比较字段	需要去重的字段
使用范例	

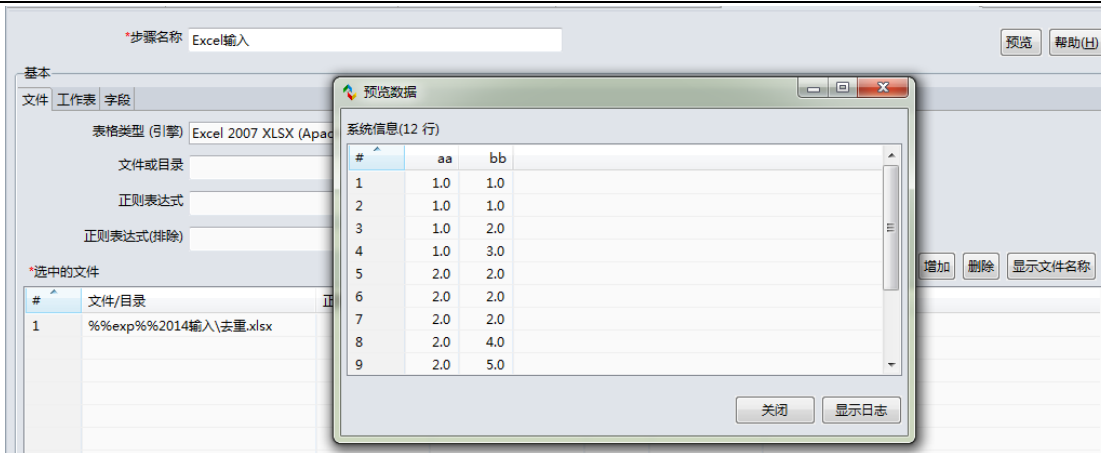
5.27.2. 示例

示例说明：示例结果图如下所示：

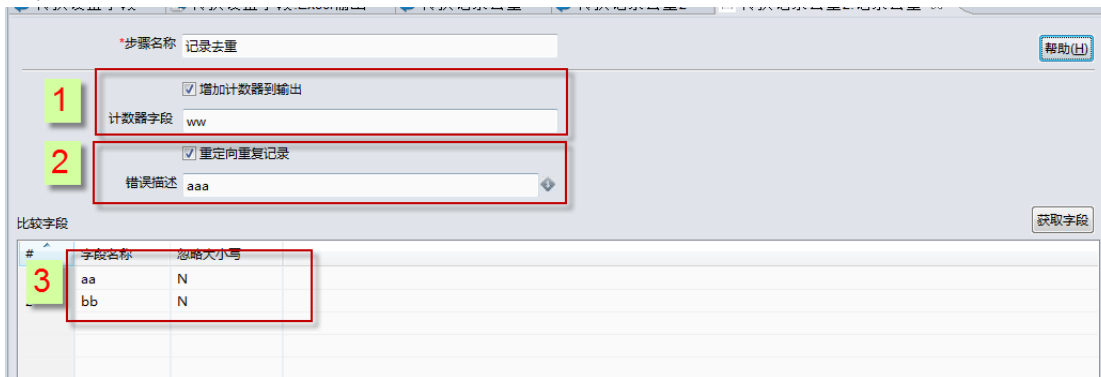


示例步骤：

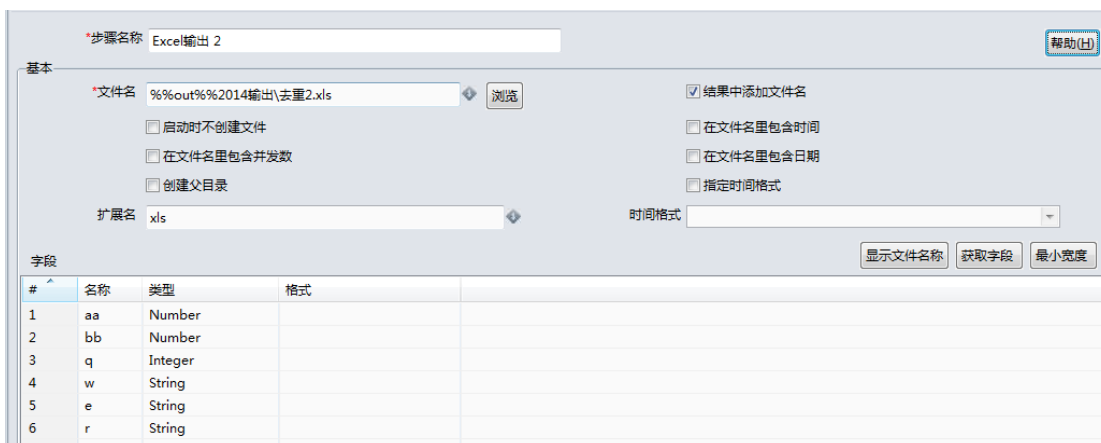
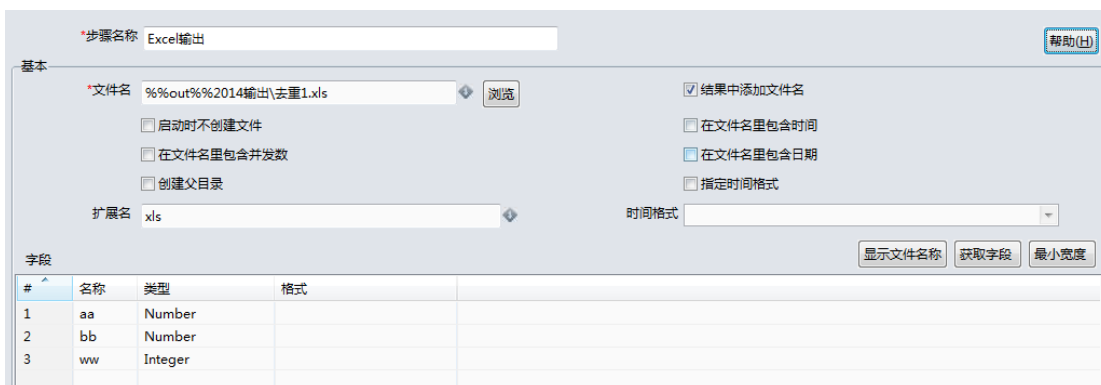
step1. 编辑转换“excel 输入”属性面板，如下图所示：



step2. 编辑转换“记录去重”属性面板，如下图所示：



step3. 编辑转换“Excel 输出”和“Excel 输出 2”属性面板，如下图所示：



step4. 运行转换，结果如下图所示：

	A	B	C	D	E	F	G	H	I	J	K
1	aa	bb	ww								
2	1.00	1.00	2.00								
3	1.00	2.00	1.00								
4	1.00	3.00	1.00								
5	2.00	2.00	3.00								
6	2.00	4.00	1.00								
7	2.00	5.00	1.00								
8	3.00	3.00	1.00								
9	3.00	5.00	1.00								
10	3.00	6.00	1.00								
11											
12											

	A	B	C	D	E	F	G	H	I	J	K
1	aa	bb	q	w	e	r					
2	1.00	1.00	1.00	aaa	aa,bb	UNR001					
3	2.00	2.00	1.00	aaa	aa,bb	UNR001					
4	2.00	2.00	1.00	aaa	aa,bb	UNR001					
5											
6											
7											
8											
9											
10											

5. 28. 转换组件：设置字段

5. 28. 1. 参数说明

转换步骤名称	设置字段
功能描述	
这个步骤允许你可以根据需要设置字段值	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
字段列表	包含了源字段到目标字段的具体映射
获取字段	获取流中的字段
使用范例	

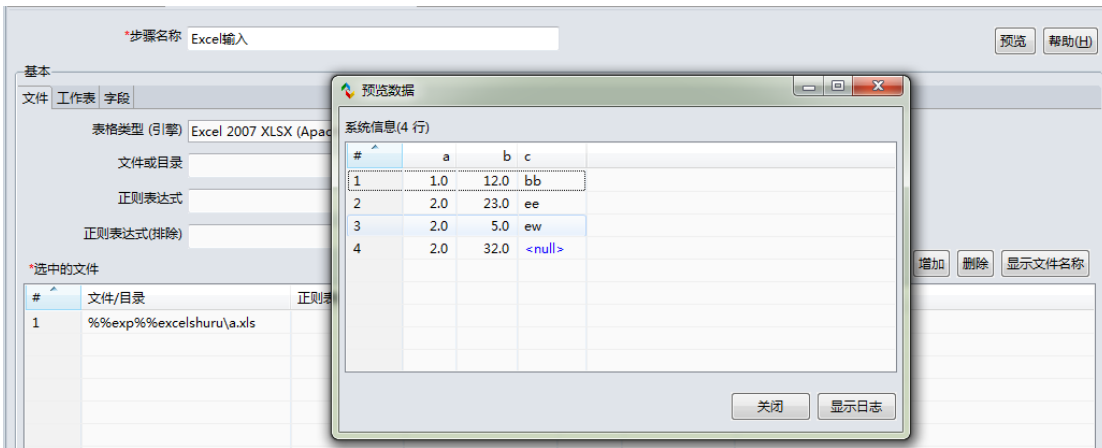
5. 28. 2. 示例

示例说明：将 Excel 输入的 a 列字段值设置为 b 列字段值，示例结果图如下所示：

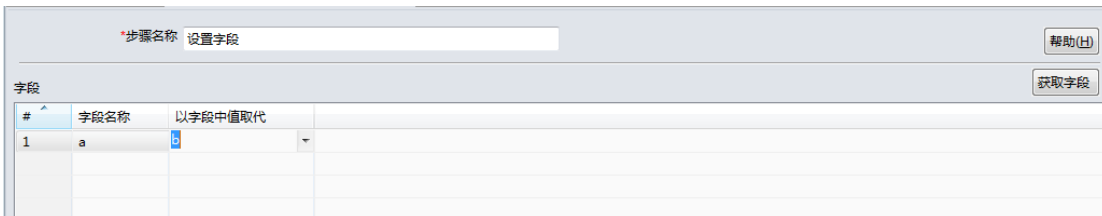


示例步骤:

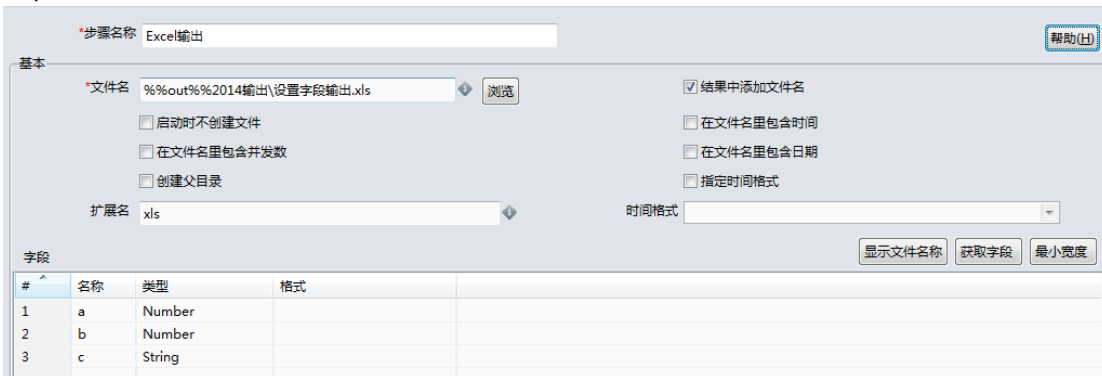
step1. 编辑转换“excel 输入”属性面板，如下图所示:



step2. 编辑转换“设置字段”属性面板，如下图所示:



step3. 编辑转换“Excel 输出”属性面板，如下图所示:



运行转换，结果如下图所示:

	A	B	C	D	E	F	G	H	I	J	K	L
1	a		c									
2	12.00	12.00	bb									
3	23.00	23.00	ee									
4	5.00	5.00	ew									
5	32.00	32.00										
6												
7												
8												
9												
10												
11												
12												

5. 29. 转换组件：连接字段

5. 29. 1. 参数说明

“连接字段”步骤用于将多个字段连接为一个目标字段。字段可用分隔符分开，且环绕逻辑与“文本文件输出”步骤完全兼容。

注意：

因为与“文本文件输出”步骤兼容，也可创建可保存在其他位置的相同字段和布局（包括页眉、页脚、结束行）。如果选中“连接由…分隔的字符串”并选择换行符（如 CR/LF），则使用“分组依据”步骤可以创建一个包含与输出字段相同内容的字段

性能注意事项：勾选“快速数据转储”选项，以实现最大吞吐量。您也可以考虑勾选“禁用环绕修复”选项。可在下面的“高级”选项中了解更多详细信息。

转换步骤名称	连接字段
功能描述	
这个步骤允许你可以根据需要设置字段值	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
目标字段名称	目标字段的名称（字符串类型）
目标字段长度	字符串类型的长度（设置字符串类型的元数据，用于性能优化的“快速数据转储”选项也有此项设置）
分隔符	指定单行文本内分隔字段的字符。一般使用“;”或 Tab 键。
环绕	一对字符串可以环绕一些字段。通过这种方式，可在字段中使用分隔符字符。环绕字符串为可选设置。
字段	
名称	字段名称
类型	字段类型可以是字符串、日期或数字。
格式	转换使用的格式掩码。请参阅“数字格式”了解格式符号的完整描述。
长度	长度选项取决于下述字段类型：

	<ul style="list-style-type: none"> • 数字 - 数字中有效数字的总量 • 字符串 - 字符串总长度 • 日期 - 屏幕上列印的字符串长度（例如，长度为 4 则只返回年份）
精度	<p>精度选项取决于下述字段类型：</p> <ul style="list-style-type: none"> • 数字 - 浮点数字的小数位 • 字符串 - 未使用 • 日期 - 未使用
货币	用于代表货币的符号，如 \$10,000.00 或 E5.000,00
小数	小数点可以为“.” (10,000.00) 或“,” (5.000,00)
分组	分组可以用“,” (10,000.00) 或“.” (5.000,00)
修剪类型	要应用于字符串的修剪方法。仅当未指定字段长度时才可以使用修剪。
空值	如果字段值为空，则在文本文件中插入此字符串
获取	单击以从输入字段流中检索字段列表
最小宽度	调整字段选项卡中的选项，使文本文件的行宽为最小值。这样就不会保存 0000001，而是写入 1，以此类推。字符串字段不再需要填充到指定长度。
高级	
删除选定的字段	勾选此选项以便从输出流中删除所有选定的字段。
是否强制在字段周围使用环绕	此选项强制使用在上方“环绕”属性中指定的字符环绕所有字段名称。
是否禁用环绕修复	用于设置与环绕符和分隔符相关的向后兼容属性（1.0 以后版本）。1.0 版本以后的逻辑为：如果字符串字段包含环绕符，则会被环绕；而环绕本身将转义。如果字符串字段含有分隔符，则分隔符将被环绕。如果不需要此逻辑，请勾选此选项。此外，由于要对字符串进行环绕符和分隔符扫描，这将加重性能负担。因此如果已确定字符串不包含这些字符，无需此项逻辑，而且您想要提升性能，请取消勾选此选项。
页眉	如果需要页眉行（数据流的首行）请启用该选项。注意：生成此行时，其他所有输出流字段会设为空值。
页脚	如果需要页脚行（数据流的最后一行）请启用此选项。注意：生成此行时，其他所有输出流字段会设为空值
编码	指定要使用的字符串编码。将此字段留空以使用系统的默认编码。要使用 Unicode，请指定 UTF-8 或 UTF-16。首次使用时，Spoon 会搜索系统中的可用编码。注意：当把不同编码字段连接为拥有唯一编码的目标字段时，这项设置必不可少。这也适用于因为延迟转换而以二进制存储的字符串字段。
右侧填充字段	在字段尾部添加空格（或移走尾部的字符），直到达到指定长度。
快速数据转储（无格式化）	当连接大量数据时，通过不包括任何格式化的信息来提升性能。请考虑将“目标字段长度”选项设置为接近目标字段长度的最大值。

	<p>由于会对内部缓冲器进行分配，且当数量不足时无需再次分配，性能将会有所提升。</p> <p>注意：注意：当“目标字段长度”选项为“0”时，内部缓冲区大小是以连接字段量的 50 倍来计算，例如 250 的内部缓冲区大小默认供 5 个字段使用。</p>
每...行分隔一次	<p>如果数字 N 大于 0，则将生成的数据流分隔为多个部分，每个部分 N 行。注意：只有当使用与“文本文件输出”步骤所得结果兼容的页眉/页脚时，才需要使用此项功能。</p>
添加文件结束行	<p>用于向输出流指定替代的结束行。注意：生成此行时，其他所有输出流字段会设为空值。</p>
使用范例	

5.29.2. 示例

略。

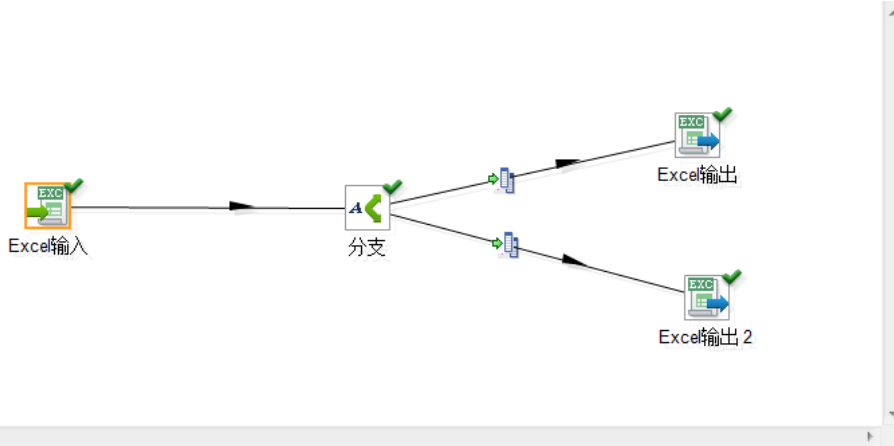
5.30. 流程组件：分支

5.30.1.1. 参数说明

转换步骤名称	分支
功能描述	
这个步骤节点所做的事情类似编程语言中 Switch/Case 所做的事情，Switch/Case 步骤节点根据不同的值会进行不同的操作	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
Switch 字段	根据 Switch 字段可以选择目标步骤。
使用字符串包含比较	
Case 值数据类型	指定 switch 字段的数据类型
Case 值转换掩码	switch 字段的转换掩码 (numeric / date values)
Case 值小数点符号	switch 字段的小数点符号 (numeric values)
Case 值分组标志	switch 字段的分组标志 (numeric values)
默认目标步骤	记录不匹配时默认输出的目标步骤。
Case 值	在这里，你可以指定不同的 case 值对应不同的目标步骤
使用范例	

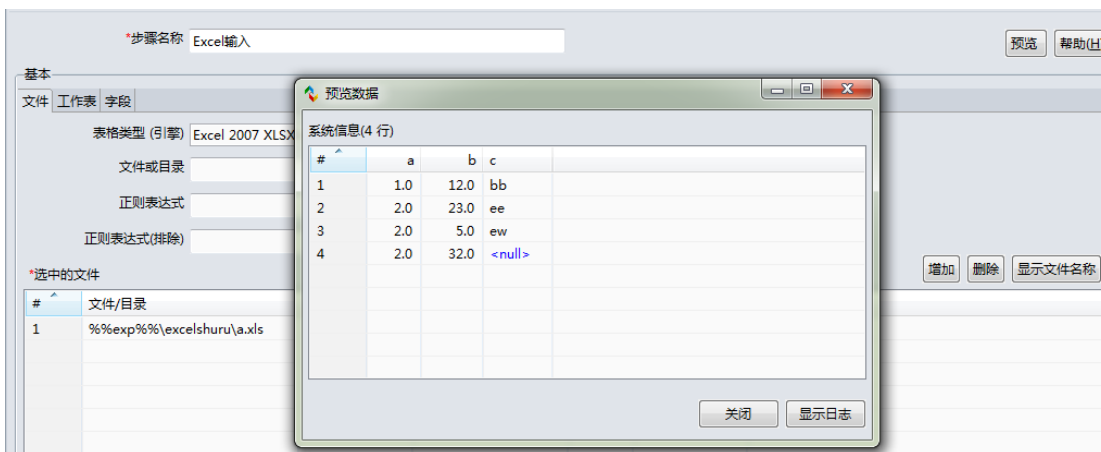
5.30.1.2. 示例

示例说明：应用分支组件，将符合不同条件的文件输出到不同的 excel 文档中。示例中将字段 a 值为 2 的文件输出到 excel2 中，最终结果图如下所示：

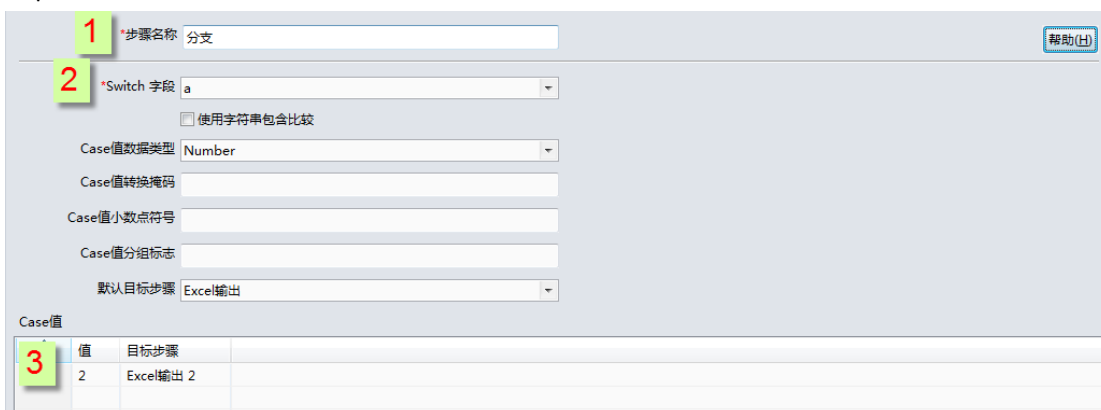


示例步骤：

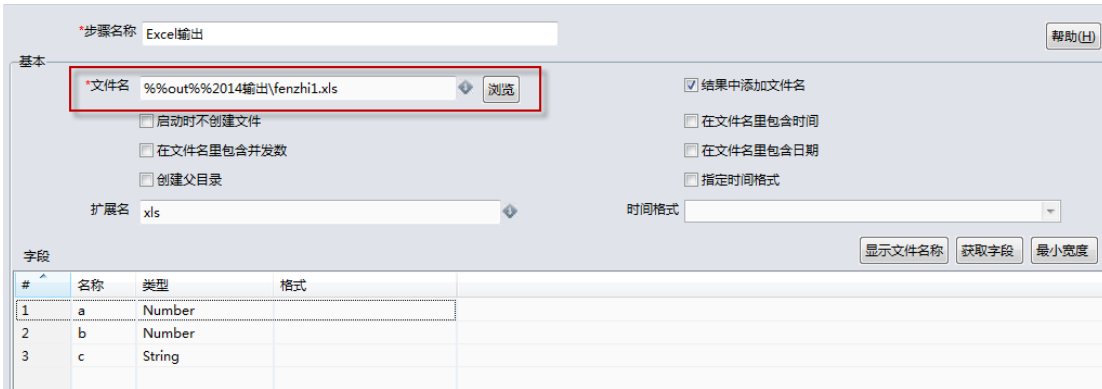
step1. 编辑转换“excel 输入”属性面板，如下图所示：



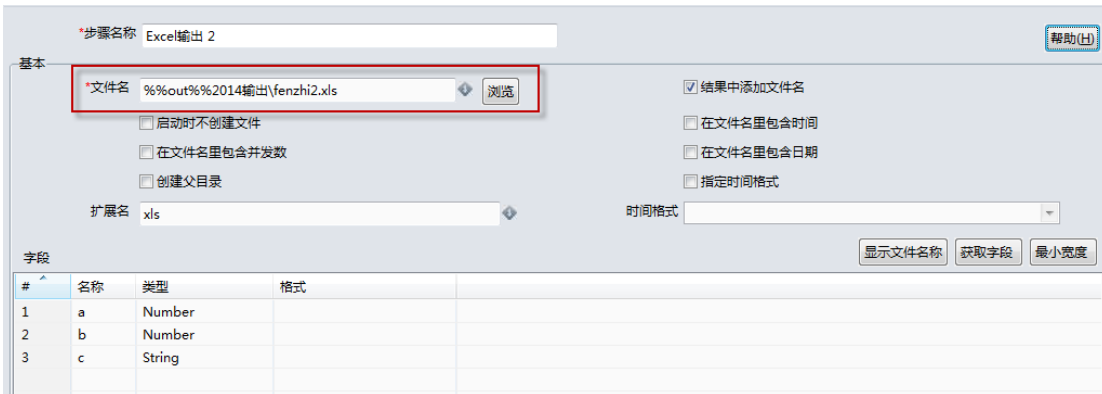
step2. 编辑转换“分支”属性面板，如下图所示：



step3. 编辑转换“excel 输出”属性面板，如下图所示：

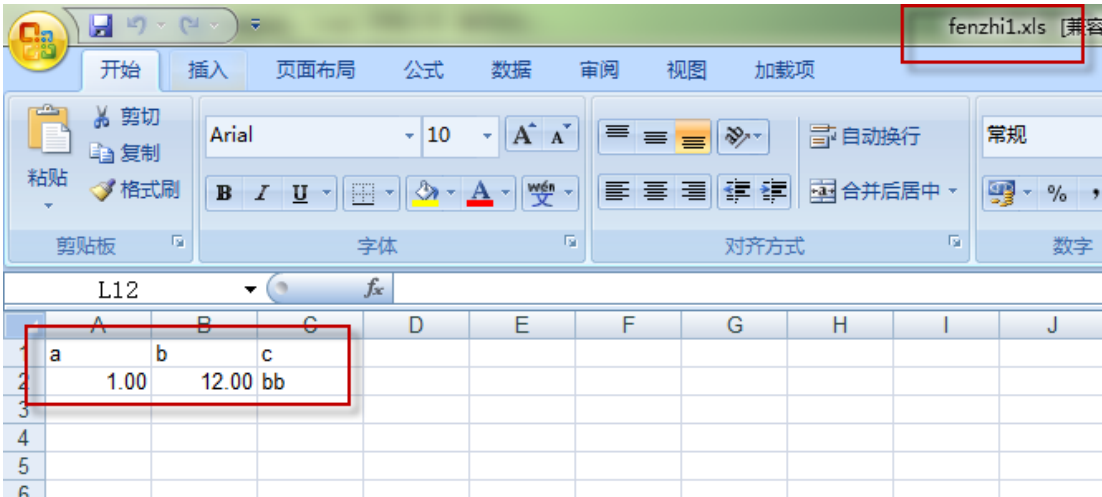


step4. 编辑转换“excel 输出 2”属性面板，如下图所示：

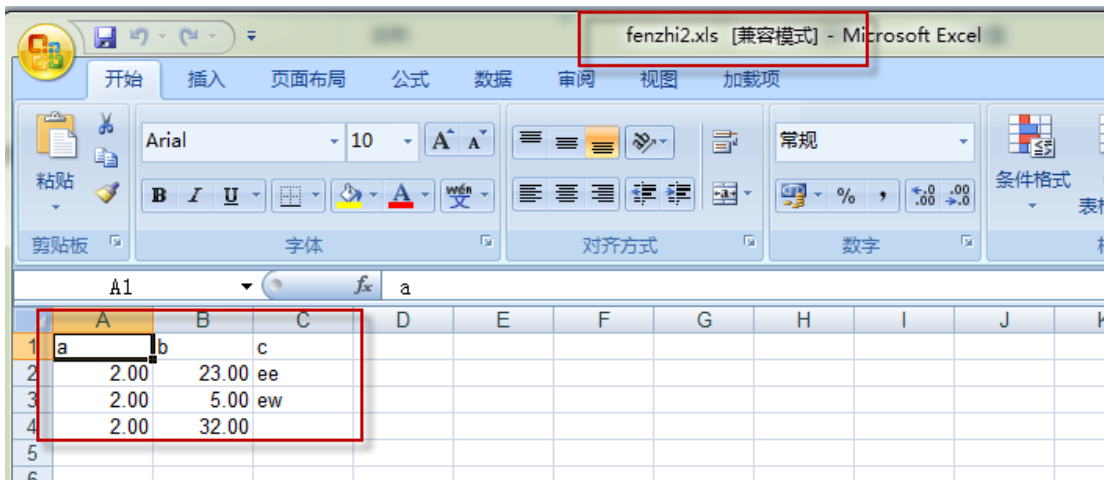


step5. 运行转换，结果如下图所示：

文件“fenzhi1.xls”内容：



文件“fenzhi2.xls”内容如下：



5.31. 流程组件：空操作

5.31.1. 参数说明

转换步骤名称	空操作
功能描述	空操作不做任何事。它的主要作用是，在你想测试什么的时候，充当一个占位符。例如有一个转换，你至少需要两个彼此连接的步骤。如果你想测试文本文件输入步骤，你可以将它连接到一个空操作步骤。
参数说明	空操作，什么都不用做。
使用范例	

5.31.2. 示例

略。

5.32. 流程组件：过滤记录

5.32.1. 参数说明

转换步骤名称	过滤记录
功能描述	

这个步骤允许你根据条件和比较符来过滤记录。一旦这个步骤连接到先前的步骤中，你可以简单的单击 “>”，“=” 和 “<” 区域来构建条件

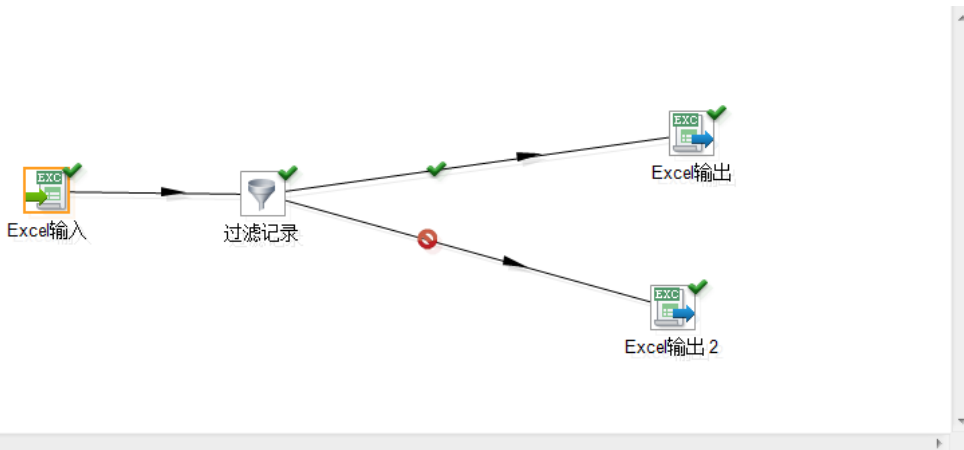
参数说明

选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
发送 true 数据给步骤	指定条件返回 true 的数据将发送到此步骤
发送 false 数据给步骤	指定条件返回 false 的数据将发送到此步骤
条件	点击 Not 按钮用以取反条件 点击<Field>来选择条件字段. 点击<value>在条件判断中输入一确定的值. 右键删除条件
添加	点击以增加条件

使用范例

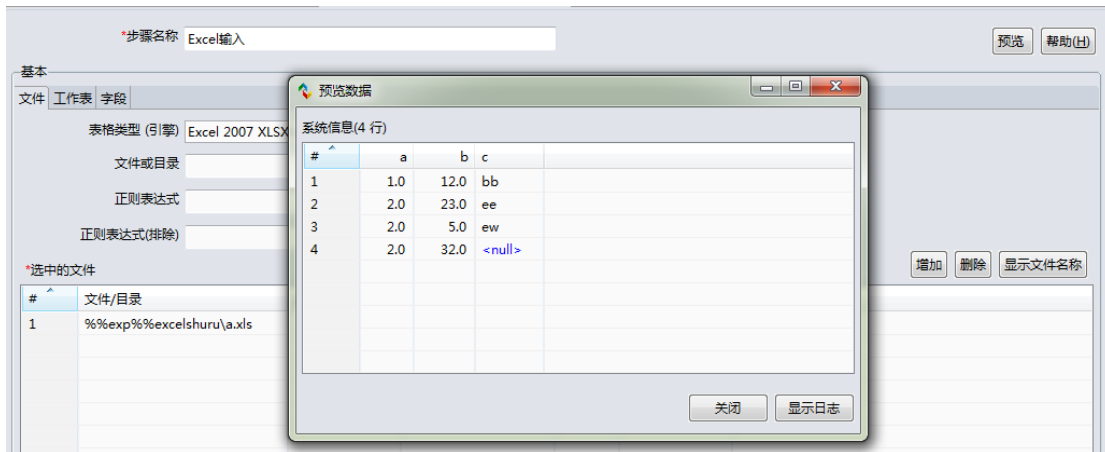
5.32.2. 示例

示例说明：示例结果图如下所示：

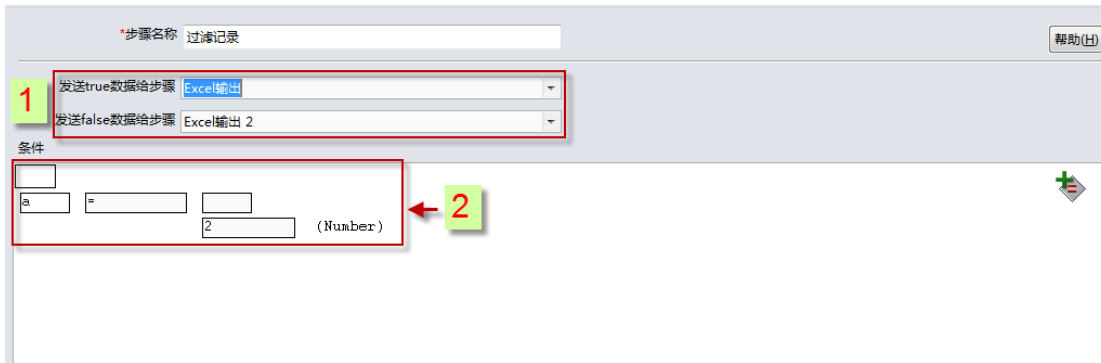


示例步骤：

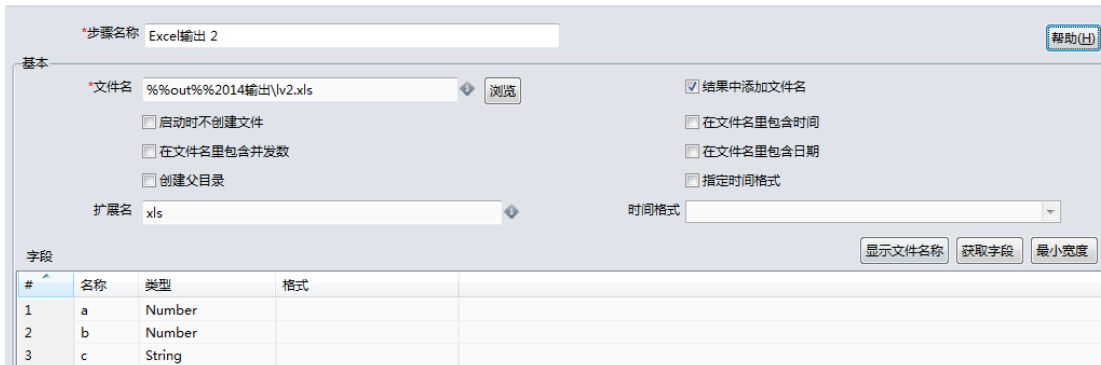
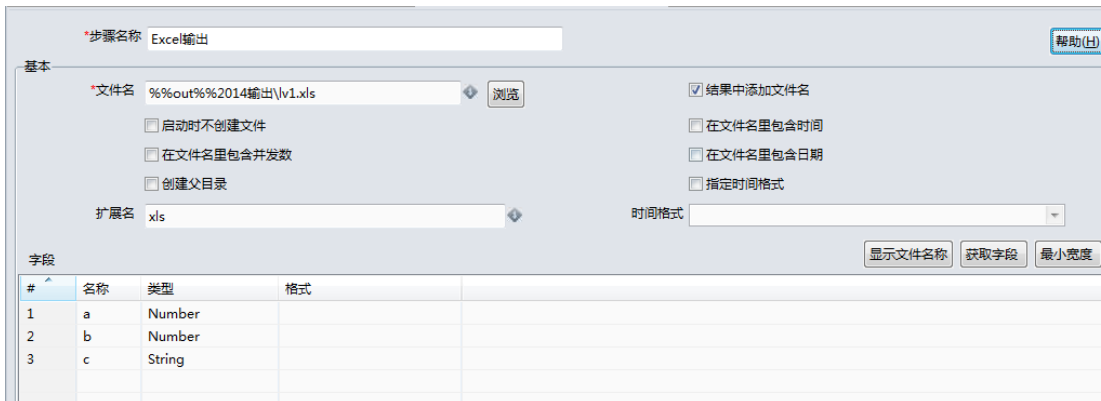
step1. 编辑转换“excel 输入”属性面板，如下图所示：



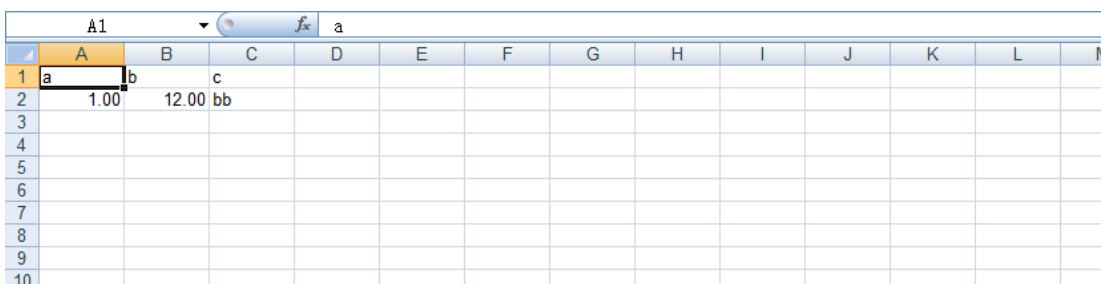
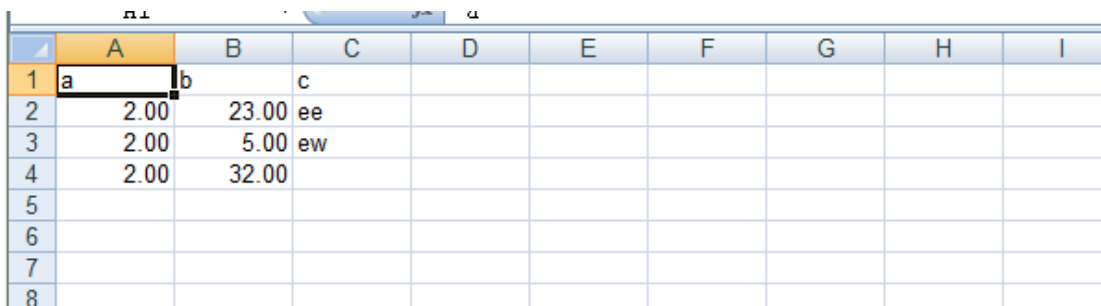
step2. 编辑转换“过滤记录”属性面板，如下图所示：



step3. 编辑转换“Excel 输出”和“Excel 输出 2”属性面板，如下图所示：



step4. 运行转换，结果如下图所示：



5.33. 脚本组件：SQL 脚本

5.33.1. 参数说明

转换步骤名称	SQL 脚本
功能描述	
在这个步骤中你可以执行 SQL 脚本，或者在转换初始化的时候执行，或者在步骤的每一个输入行执行。	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	选择一个执行脚本的数据库连接
SQL 脚本	执行的 SQL 脚本，用“;”分割不同的语句
执行每一行	选中则执行每一行. 在这种情形下，参数可以被用，不选中时，SQL 语句会在初始阶段执行.
作为一条语句执行	在此选项不会根据“;”区分不同的语句，将会发送整个 SQL 脚本到数据库中
变量替换	如果你想在 SQL 脚本中使用变量需要选中此选项
绑定参数	选择在使用 SQL 语句时是否需绑定参数，否则这一步骤会产生一原意的字符串。 Note: Since PDI version 5.0
引用字符串	此选项会根据数据库增加引用字符串。 Note: Since PDI version 5.0
参数	定义一些步骤所需参数
包含插入状态的字段	在插入数据时，如果你想选择一额外的字段，在此定义字段名称
包含更新状态的字段	与插入类似，只不过是更新
包含删除状态的字段	与插入类似，只不过是删除
包含读状态的字段	与插入类似，只不过是读
使用范例	

5.33.2. 示例

略。

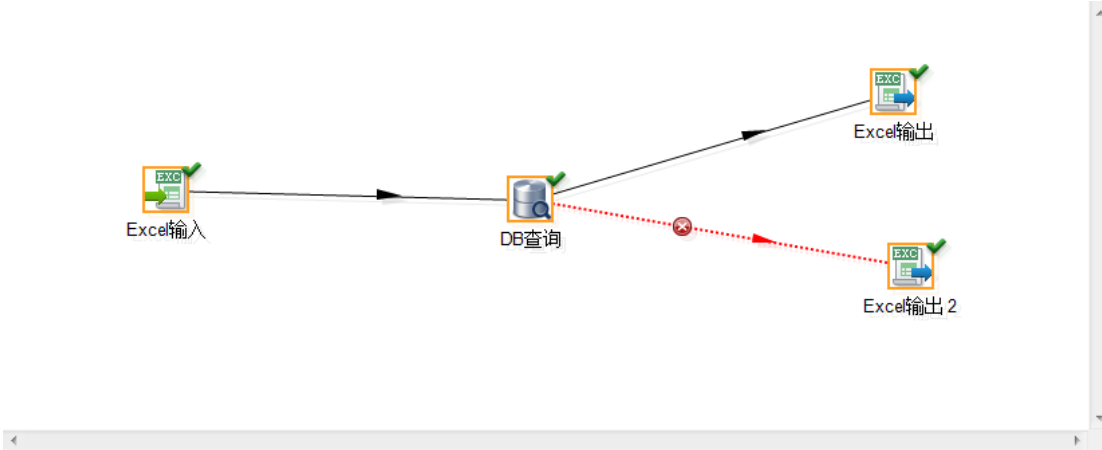
5.34. 查询组件：DB 查询

5.34.1. 参数说明

转换步骤名称	DB 查询
功能描述	
这个步骤类型允许你在数据库表中查找值	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	数据库连接
模式	查询表所在的模式
表	想要查询的表名。
使用缓存号	数据库查询时缓存的大小
从表中加载所有数据	是不是加载所有的表中数据
查询关键字表	关键字和条件用以数据库查询
查询表返回值	查询表结果需要额外添加的字段 “新名称”允许你使用一个不同的新名称 “默认”查询失败是返回的值。 “类型”输出字段的类型
如果查询失败不要忽略这一行	当查询失败时，是否选择忽略跳过行
返回多行失败	如果查询返回多行记录，强制步骤失败
排序	如果查询返回多行记录时，排序会帮助你选择根据某一字段进行排序
获取字段	点击以从输入流中获取以可用的字段列表
获取查询自选	点击可从目标查询表中获取一可用的的字段列表
使用范例	

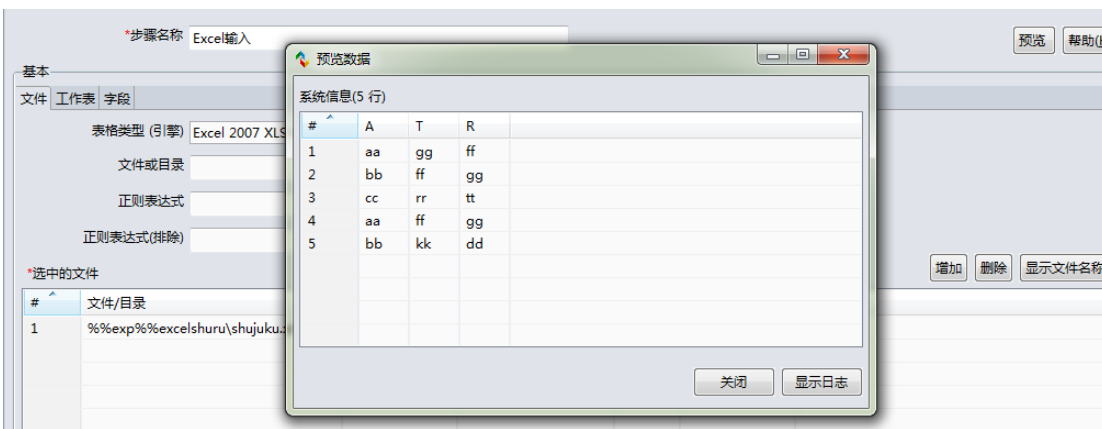
5.34.2. 示例

示例说明：示例结果图如下所示：

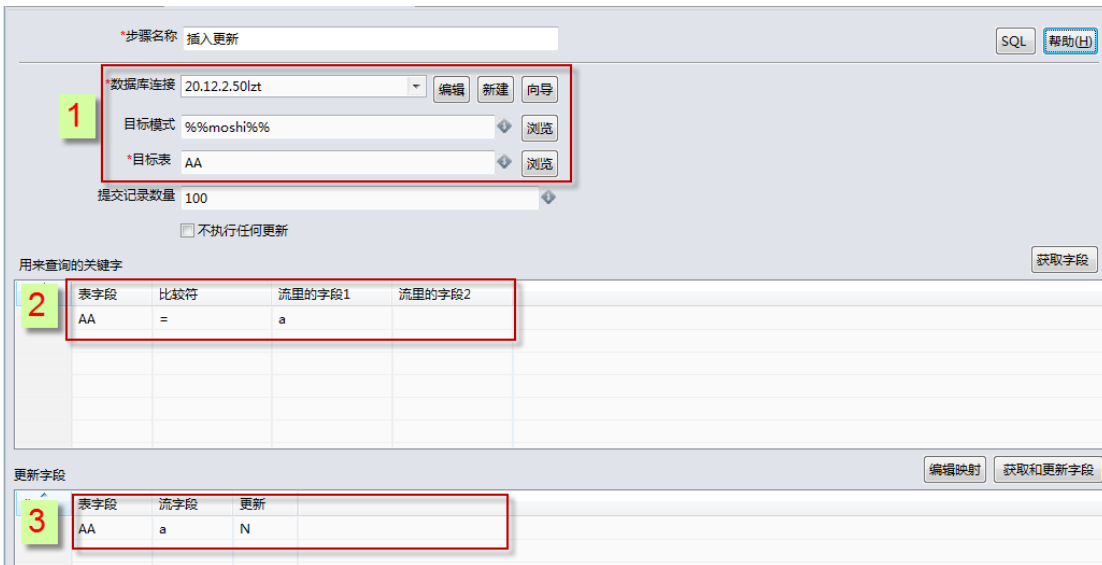


示例步骤:

step1. 编辑转换“excel 输入”属性面板，如下图所示：



step2. 编辑转换“DB 查询”属性面板，如下图所示：



step3. 编辑转换“Excel 输出”属性面板，略。

运行转换，结果如下图所示：

	A	B	C	D	E	F	G	H	I	J	K
1	A	T	R	A_1	B						
2	aa	gg	ff	aa	aa						
3	bb	ff	gg	bb	ee						
4	aa	ff	gg	aa	aa						
5	bb	kk	dd	bb	ee						
6											
7											
8											
9											

	A	B	C	D	E	F	G	H	I	J	K
1	A	T	R								
2	cc	rr	tt								
3											
4											
5											
6											

5. 35. 查询组件：Meta 查询

5. 35. 1. 参数说明

转换步骤名称	Meta 查询
功能描述	
这个步骤类型允许你在 AE 元数据数据库表中查找值	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	数据库连接
模式	查询表所在的模式
表	想要查询的表名。
使用缓存号	数据库查询时缓存的大小
从表中加载所有数据	是不是加载所有的表中数据
查询关键字表	关键字和条件用以数据库查询
查询表返回值	查询表结果需要额外添加的字段 “新名称”允许你使用一个不同的新名称 “默认”查询失败是返回的值。 “类型”输出字段的类型
如果查询失败不要忽略这一行	当查询失败时，是否选择忽略跳过行
返回多行失败	如果查询返回多行记录，强制步骤失败
排序	如果查询返回多行记录时，排序会帮助你选择根据某一字段进行排序
获取字段	点击以从输入流中获取以可用的字段列表
获取查询自选	点击可从目标查询表中获取一可用的字段列表
使用范例	

5.35.2. 示例

略。

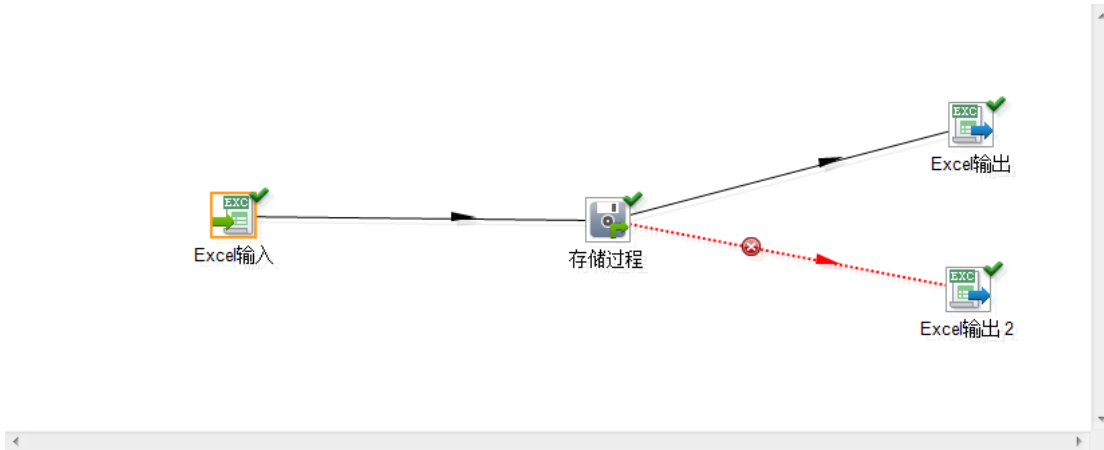
5.36. 查询组件：存储过程

5.36.1. 参数说明

转换步骤名称	存储过程
功能描述	
这个步骤允许你运行一个数据库存储过程，获取返回结果	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	存储过程所有的数据库的连接名称
存储过程名称	调用的存储过程或者函数名称
查找	点击来根据数据库链接来查找指定的存储过程名称
自动	在某些情况下，您希望更新
提交	在运行存储过程的时候自动提交
结果名称	调用存储过程或者函数返回结果的名称
结果类型	调用存储过程或者函数返回结果的类型
参数	存储过程或者函数需要的参数列表 名称：字段的名称 方向：可以是 IN, OUT 或者 INOUT 类型：使用输出参数，以便 Designer 知道返回了什么
获取字段	获取字段并将字段填入表中
使用范例	

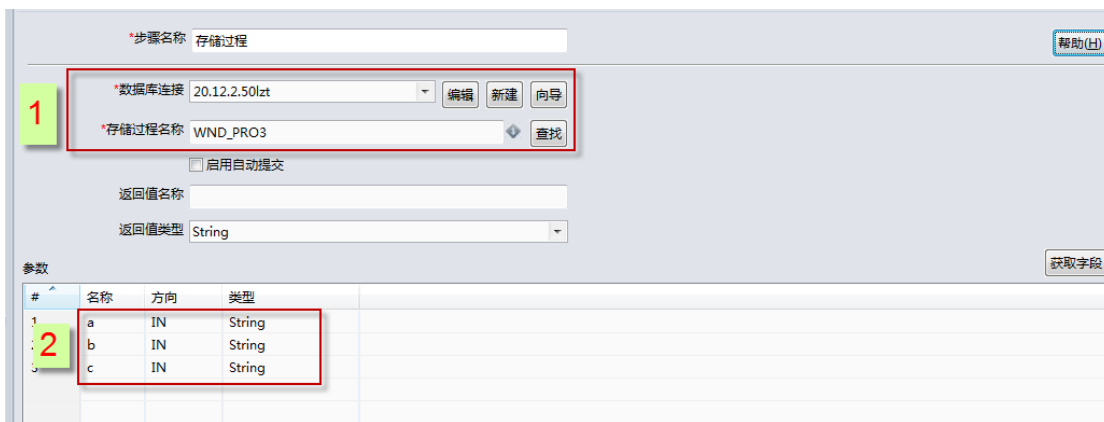
5.36.2. 示例

示例说明：示例结果图如下所示：



示例步骤:

step1. 编辑转换“存储过程”属性面板，如下图所示:



step2. 编辑转换“excel 输入”属性面板，图略。

step3. 运行转换，结果如下图所示:

	A	B	C	D	E	F	G	H	I
1	a	b	c						
2	abcd	abcd	abcd						
3									
4									
5									
6									
7									

5. 37. 查询组件：检查表

5. 37. 1. 参数说明

转换步骤名称	检查表
功能描述	
你可以用此步骤来检查某一表或者某些表在指定数据库、指定模式中是否存在	
参数说明	

选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
连接	需要使用的数据库连接
模式	需要查询的模式名称
表名	表的名称，用来查询的表名字段
结果字段名	查询表是否存在的结果字段名称
使用范例	

5.37.2. 示例

略。

5.38. 查询组件：维度查询

5.38.1. 参数说明

转换步骤名称	维度查询
功能描述	
这个步骤类型允许查询维度	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
数据库连接	数据库连接
目标模式	查询表所在的模式
目标表	想要查询的表名。
使用缓存	数据库查询时缓存的大小
预加载缓存	是否要启用加载缓存
缓存行数	启用加载缓存的行数
业务键字段	描述
维字段	维字段
流里的字段	查询表所在的模式
查询返回字段	描述
维字段	维字段
输出字段的新名称	输出字段的新名称
返回字段类型	数据库连接
代理键字段	想要查询的表名。
代理键新名称	数据库查询时缓存的大小
流日期字段	流日期字段
开始日期字段	想要查询的表名。
截止日期字段	数据库查询时缓存的大小

使用范例

5.38.2. 示例

略。

5.39. 连接组件：合并记录

5.39.1. 参数说明

合并行可用于比较两行数据流。这对于比较两个不同时间的数据很有用。当数据仓库的源系统不包含最近更新的日期时，也常用到此项功能。

它可将两行数据流，即参考数据（旧数据）和比较数据流（新数据）合并。每次只有行的最新版本才会传递到下面的步骤。行标记如下：

- “相同” - 两个数据流中都能找到此键，且比较的值相同；
- “改变” - 两个数据流中都能找到此键，但有一个或多个值不同；
- “新” - 参考数据流中未找到此键；
- “删除” - 比较数据流中未找到此键。

比较数据流中的行将传递到后续步骤，除非被标记为“删除”或“相同”。

重要提示：两种数据流都必须根据指定的键排序。当使用“排序”步骤时此项操作可正常执行。如果对 AEDI 以外的数据进行排序，可能会遇见与内部区分大小写/不区分大小写标记相关的问题。

转换步骤名称	合并记录	
功能描述		
合并行可用于比较两行数据流。这对于比较两个不同时间的数据很有用。当数据仓库的源系统不包含最近更新的日期时，也常用到此项功能		
参数说明		
选项	描述	
步骤名称	步骤名称，在单个转换中必须唯一	
参考行来源	指定参考行的步骤源 <- 含有原有行的数据流，或要与新行相比较的行。	
比较行来源	指定比较行的步骤源。<- 含有新行的数据流	
标记字段名	指定输出流标记字段的名称。	
要匹配的键	指定要用于匹配的包含键的字段；单击“获得键字段”插入所有来源于参考行步骤的字段	
要比较的值	指定包含要比较的值的字段；单击“获得值字段”，插入所有来源于值行步骤的字段。这里无需指定键字段。	
使用范例		

5.39.2. 示例

示例说明：将输入合并后再输出为 excel 文档。示例结果图如下所示：



示例步骤：

step1. 编辑转换“excel 输入”和“excel 输入 2”属性面板，如下图所示：

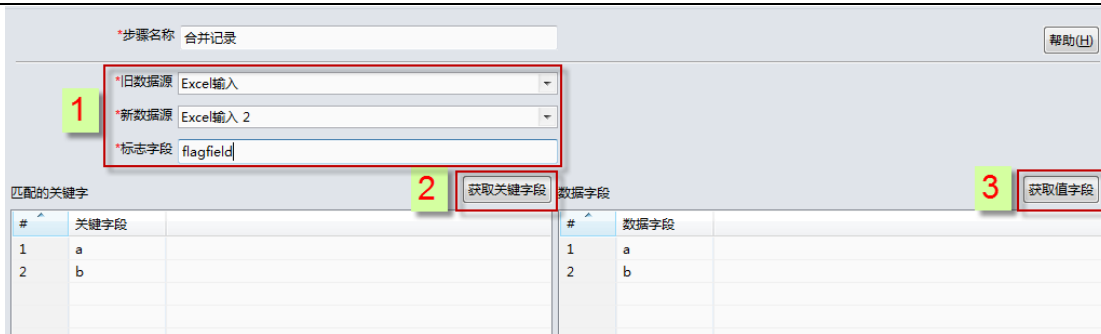
The first screenshot shows the configuration for 'Excel输入'. The '预览数据' window displays the following data:

#	a	b
1	2.0	5.0
2	45545.0	5.0
3	2.0	5.0
4	45545.0	5.0
5	2.0	8.0
6	2.0	10.0
7	45.0	16.0
8	553.0	32.0
9	45.0	32.0

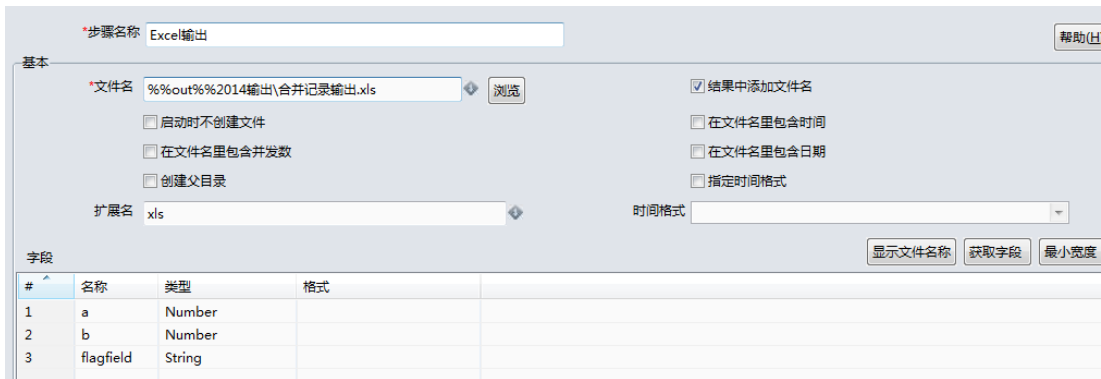
The second screenshot shows the configuration for 'Excel输入 2'. The '预览数据' window displays the following data:

#	c	d
1	45545.0	5.0
2	201.0	8.0
3	100.0	11.0
4	101.0	12.0
5	12.0	13.0
6	553.0	32.0
7	15.0	158.0

step2. 编辑转换“合并记录”属性面板，如下图所示：



step3. 编辑转换“excel 输出”属性面板，如下图所示：



step4. 运行转换，结果如下图所示：

	A	B	C	D	E	F	G	H	I	J	K	L
1	a	b	flagfield									
2	2.00	5.00	deleted									
3	45,545.00	5.00	identical									
4	2.00	5.00	deleted									
5	201.00	8.00	new									
6	100.00	11.00	new									
7	101.00	12.00	new									
8	12.00	13.00	new									
9	553.00	32.00	new									
10	15.00	158.00	new									
11	45,545.00	5.00	deleted									
12	2.00	8.00	deleted									
13	2.00	10.00	deleted									
14	45.00	16.00	deleted									
15	553.00	32.00	deleted									
16	45.00	32.00	deleted									
17	553.00	32.00	deleted									
18	45.00	32.00	deleted									
19	45.00	45.00	deleted									
20	24.00	65.00	deleted									
21	7,845.00	65.00	deleted									
22	24.00	65.00	deleted									
23	7,845.00	65.00	deleted									
24	232.00	323.00	deleted									
25	45.00	323.00	deleted									
26	232.00	323.00	deleted									
27	45.00	323.00	deleted									

5. 40. 连接组件：记录连接

5. 40. 1. 参数说明

“合并连接”步骤对具有来自两个不同输入步骤所得的数据的数据集执行标准合并连接操作。合并选项包括 INNER、LEFT OUTER、RIGHT OUTER 和 FULL OUTER。

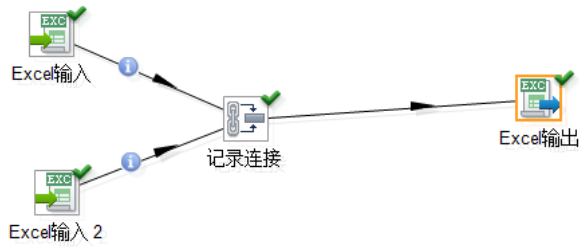
注意：在此步骤中，最好根据指定键字段对行进行排序。当使用“排序”步骤时此项操作可正常执行。如果对 AEDI 以外的数据进行排序，可能会遇见与内部区分大小写/不区分大小写标记相关的问题。

注意：如果键字段出现重名情况（如 ID），则结果集中将创建名为 \$key_1（如 ID_1）的第二个键字段。

转换步骤名称		记录连接
功能描述		
合并连接”步骤对具有来自两个不同输入步骤所得的数据的数据集执行标准合并连接操作。合并选项包括 INNER、LEFT OUTER、RIGHT OUTER 和 FULL OUTER。		
参数说明		
选项	描述	
步骤名称	步骤名称，在单个转换中必须唯一	
第一步	指定合并连接的第一个输入步骤（左）	
第二步	指定合并连接的第二个输入步骤（右）	
连接类型	<p>可从以下连接类型中选择：</p> <ul style="list-style-type: none"> • FULL OUTER: 结果集中包含两个来源中的所有行，两个数据流中键不匹配的行为空值 • LEFT OUTER: 结果集中包括第一个源中的所有行，第二个数据流中键不匹配的行为空值 • RIGHT OUTER: 结果集中包括第二个源中的所有行，第一个数据流中键不匹配的行为空值 • INNER JOIN: 结果集中只包括两个源中具备相同键的行 	
第一步中的键	指定用于对传入数据进行排序的键字段；单击“获得键字段”，在指定的步骤检索一列字段。	
第二步中的键	指定用于对传入数据进行排序的键字段；单击“获得键字段”，在指定的步骤检索一列字段。	
使用范例		

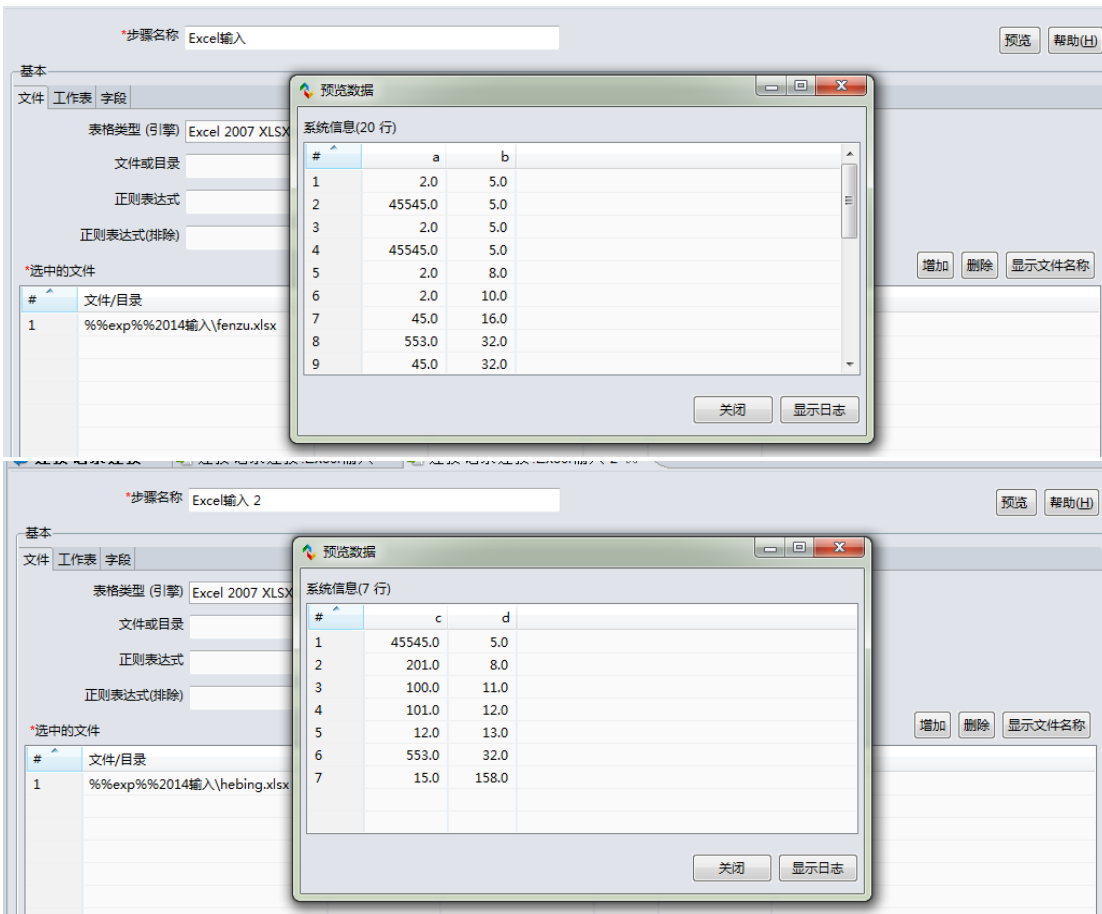
5.40.2. 示例

示例说明：将两个 excel 输入字段进行左连接，并输出为 excel 文档。示例结果图如下所示：

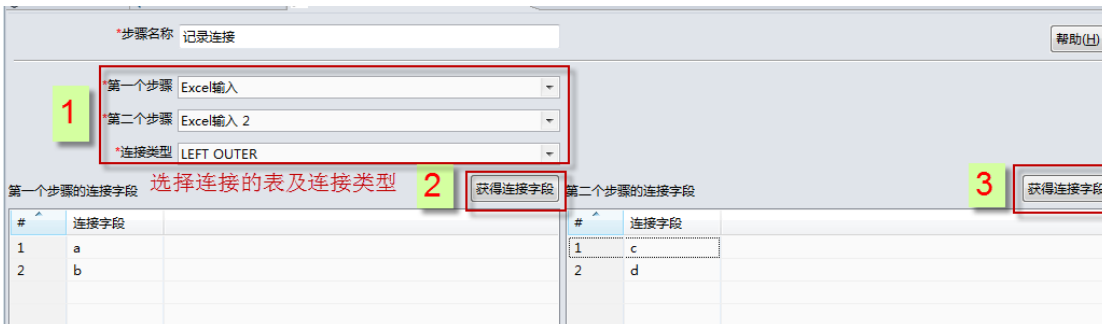


示例步骤:

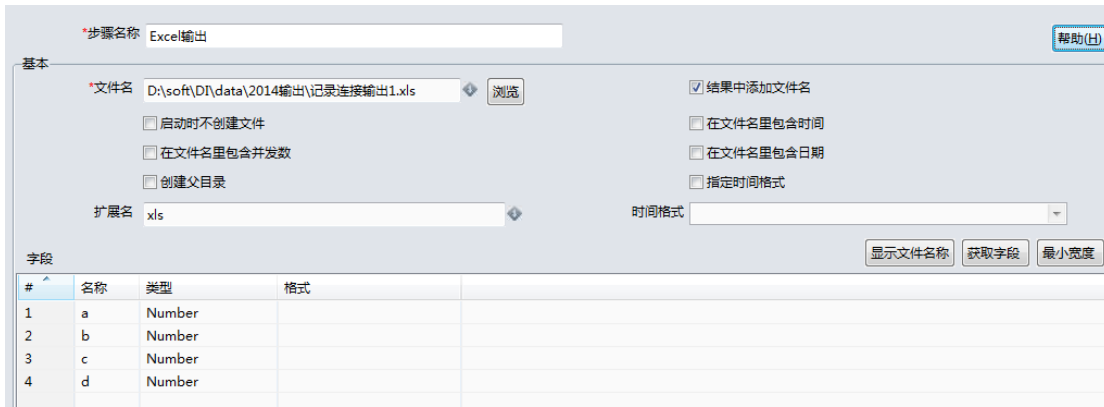
step1. 编辑转换“excel 输入”和“excel 输入 2”属性面板，如下图所示：



step2. 编辑转换“记录连接”属性面板，如下图所示：



step3. 编辑转换“excel 输出”属性面板，如下图所示：



step4. 运行转换，结果如下图所示：

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	a	b	c	d									
2	2.00	5.00											
3	45,545.00	5.00	45,545.00	5.00									
4	2.00	5.00											
5			201.00	8.00									
6			100.00	11.00									
7			101.00	12.00									
8			12.00	13.00									
9			553.00	32.00									
10			15.00	158.00									
11	45,545.00	5.00											
12	2.00	8.00											
13	2.00	10.00											
14	45.00	16.00											
15	553.00	32.00											
16	45.00	32.00											
17	553.00	32.00											
18	45.00	32.00											
19	45.00	45.00											
20	24.00	65.00											
21	7,845.00	65.00											
22	24.00	65.00											
23	7,845.00	65.00											
24	232.00	323.00											
25	45.00	323.00											
26	232.00	323.00											
27	45.00	323.00											

5. 41. 质量与统计组件：分组

5. 41. 1. 参数说明

此步骤可用于计算限定字段组的值。常见使用情形如下：计算每件产品的平均销售量或得到库存中的黄衬衫的数量。

注意：只有经过排序的输入才可执行此项操作。如果输入未经排序，则只有双连续行才能够得到正确处理。当使用“排序”步骤时此项操作可正常执行。如果对 DI Designer 以外的数据进行排序，可能会遇见与内部区分大小写/不区分大小写标记相关的问题。

欲了解如何处理未经排序的输入，另请参阅“内存分组”步骤。

转换步骤名称	分组
功能描述	此步骤可用于计算限定字段组的值。常见使用情形如下：计算每件产品的平均销售量或得到库存中的黄衬衫的数量。

参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
包括所有行	如果需要输出所有行而不只是聚合的行，请启用该选项；如需在输出中区分两种类型的行，则需要在输出中加入标记。这种情况下，必须指定标记字段的名称（类型为布尔型）。
临时文件目录	存储临时文件的目录（当选项“包括所有行”启用、且分组的行数超过 5000 时，需要此目录）；默认设置是系统的标准临时目录
TMP 文件前缀	指定为临时文件命名时使用的文件前缀
添加行号，每一组重新从头开始	启用以添加行编号功能，使每一组从 1 开始编号
行编号字段名	启用以添加行编号功能，使每一组从 1 开始编号
总是返回一行	如果启用该选项，即便没有输入行，“分组”步骤也会返回一个结果行。此项功能可用于计算总行数。如果不勾选此选项，永远不会得到 (0) 行的结果。
分组字段表	指定要依据分组的字段。单击“获得字段”，添加输入流中的所有字段。
聚合表	<p>指定必须聚合的字段、方式以及生成的新字段的名称。以下是可用的聚合方法：</p> <ul style="list-style-type: none"> • 总和 • 平均值（均值） • 中位数 • 百分数 • 最小值 • 最大值 • 值的总量 (N) • 连接用 “,”（逗号）分隔的字符串 • 第一个非空值 • 最后一个非空值 • 第一个值（包括空值） • 最后一个值（包括空值） • 累积和（所有行都选中此选项时才可使用！） • 累加平均（所有行都选中此选项时才可使用！） • 标准差

	<ul style="list-style-type: none"> • 连接用“<Value>:”分隔的字符串在“值”过去中指定分隔符 • 不等值的数量 • 行数（不含字段变量）
使用范例	

5.41.2. 示例

略。

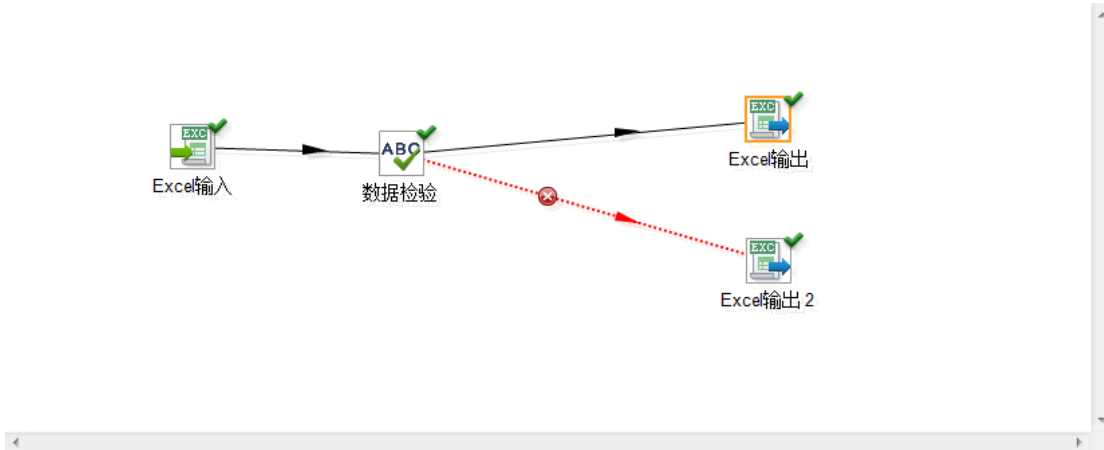
5.42. 质量与统计组件：数据检验

5.42.1. 参数说明

转换步骤名称	数据检验
功能描述	
数据检验典型的应用就是用来检查输入的数据质量，数据检验经常会因不同需要发生，例如你怀疑现在来的数据没有很好的质量，或者仅仅是因为你有一个 SLA 位置，数据检验允许你自己定义一定的规则去描述输入的数据应该是什么样子。	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一
检验的名称	检验的名称
增加检验	增加一个检验
删除检验	删除一个检验
报告所有错误	检验中有可能出现许多错误，选中会报告所有的错误，不止第一个错误
使用范例	

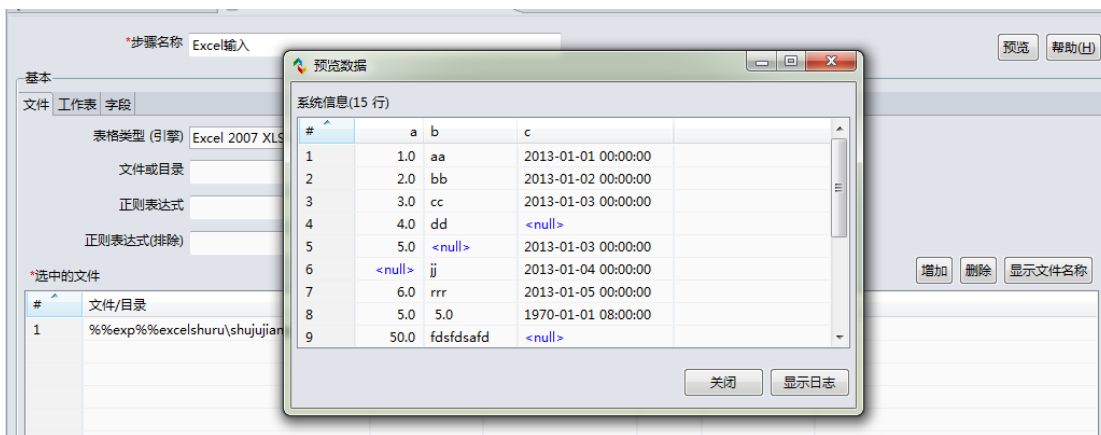
5.42.2. 示例

示例说明：对 excel 输入的字段进行数据类型校验，并输出 excel 文档，示例结果图如下所示：

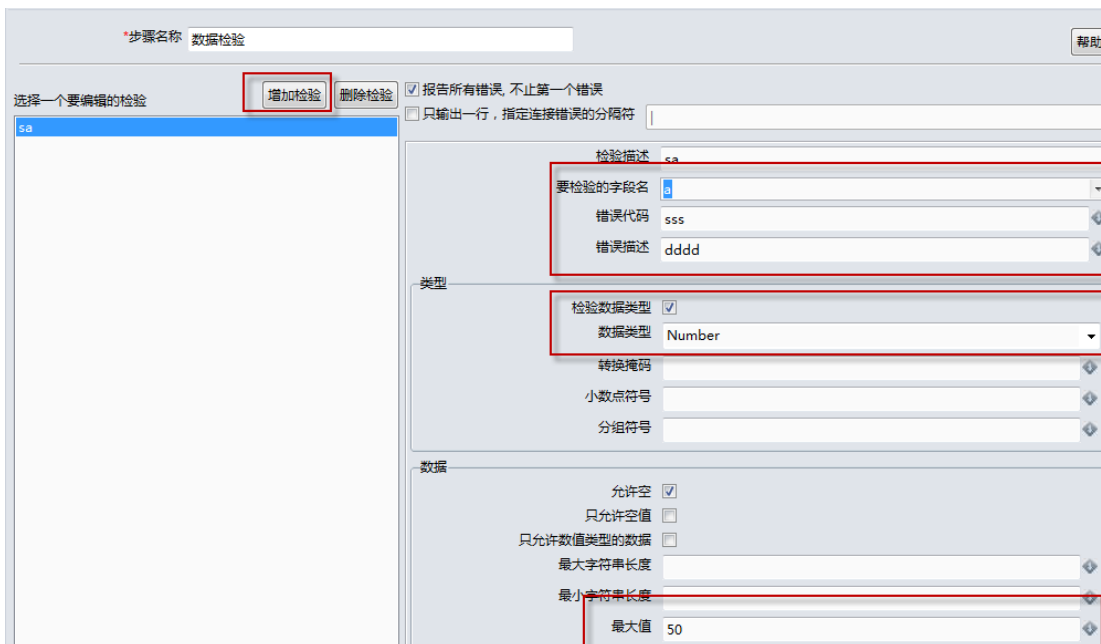


示例步骤:

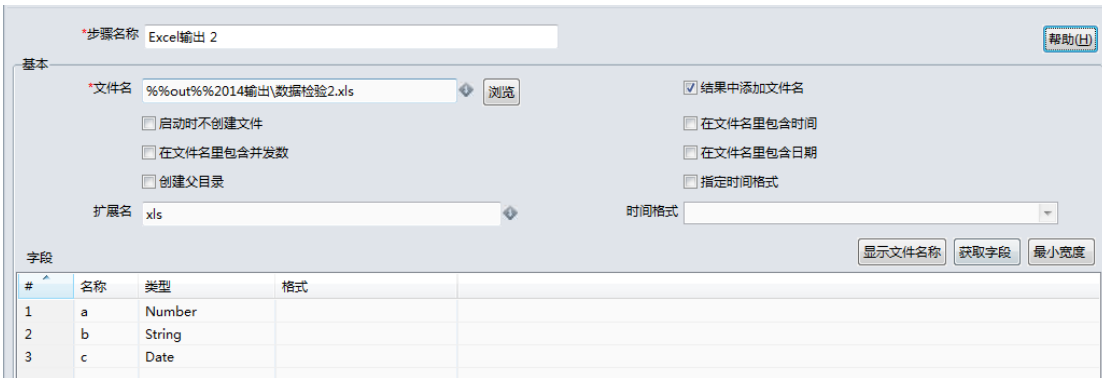
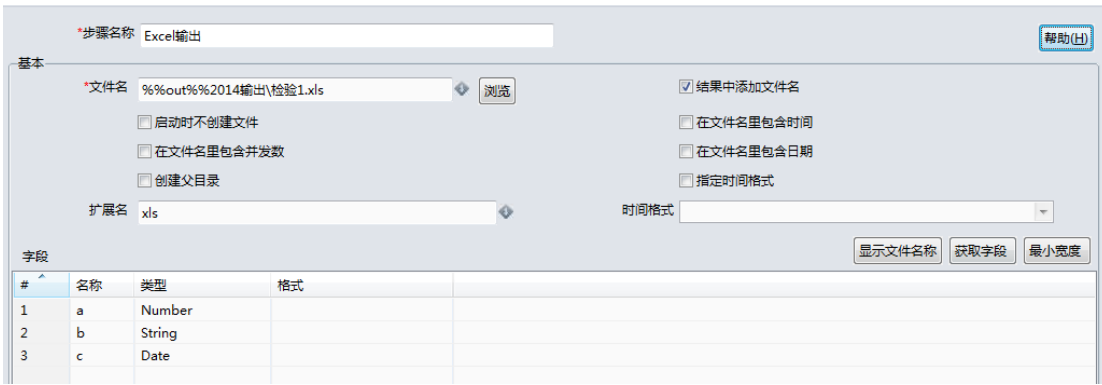
step1. 编辑转换“excel 输入”属性面板，如下图所示:



step2. 编辑转换“数据检验”属性面板，如下图所示:



step3. 编辑转换“Excel 输出”和“Excel 输出 2”属性面板，如下图所示:



step4. 运行转换，结果如下图所示：

	A	B	C	D	E	F	G	H	I	J	K
1	a	b	c								
2		5.00	#####								
3			jj								
4		6.00	rrr								
5		5.00	5.0								
6		50.00	fdsfsafd								
7											
8											
9											

	A	B	C	D	E	F	G	H	I	J	K
1	a	b	c								
2		1.00	aa								
3		2.00	bb								
4		3.00	cc								
5		4.00	dd								
6		56.00	fdsa								
7		57.00	fdsgdsg								
8		58.00	fds								
9		59.00	fgdsgfda								
10		60.00	bfsdagds								
11		61.00	62.0								
12											

5. 43. 作业组件：获取变量

5. 43. 1. 参数说明

转换步骤名称	获取变量
功能描述	

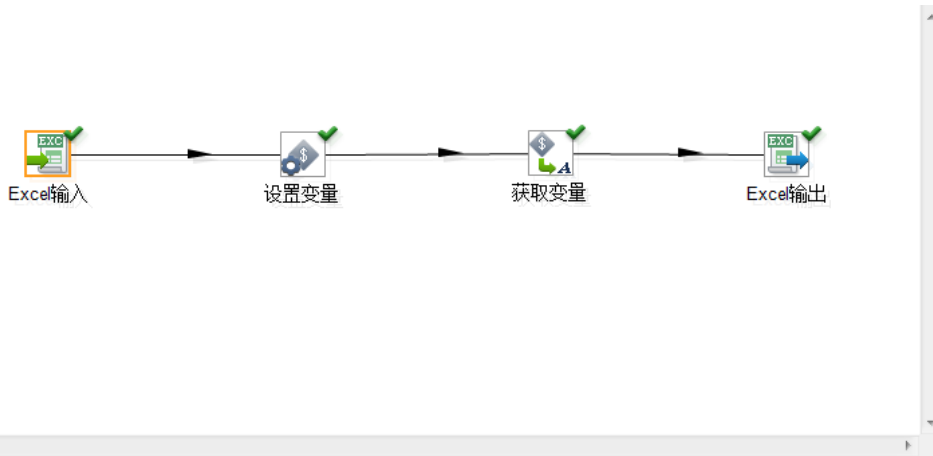
这一步骤允许你从任务中或者虚拟机获取变量

参数说明

使用范例

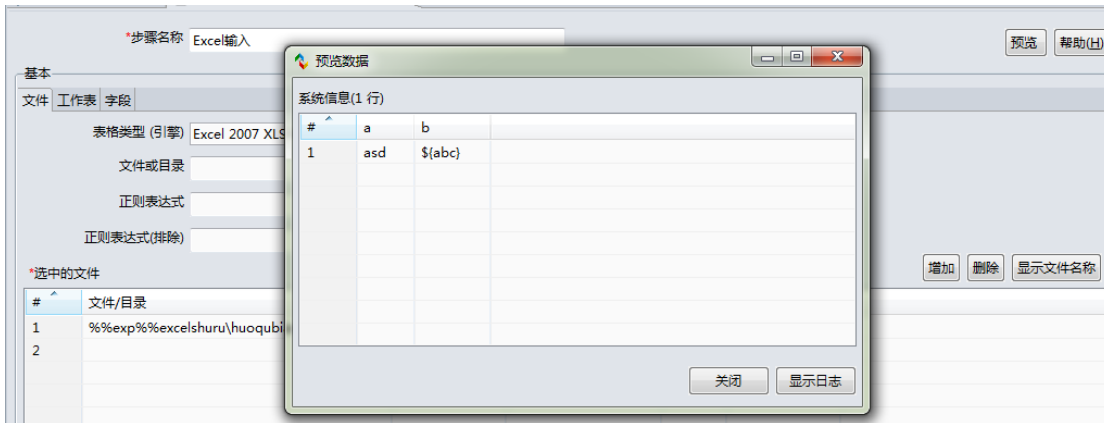
5.43.2. 示例

示例说明：示例结果图如下所示：

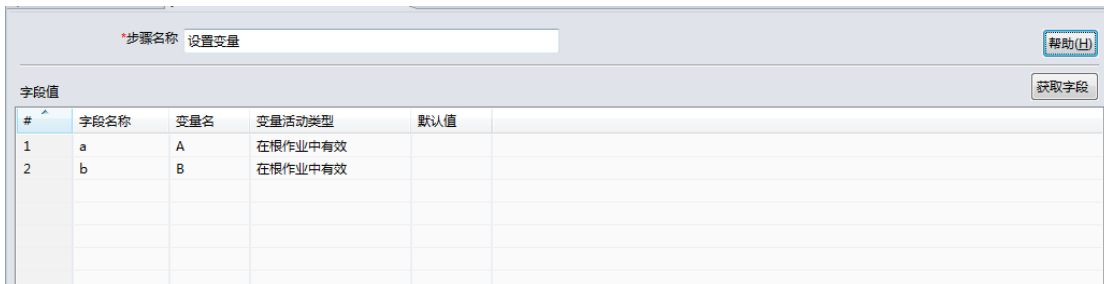


示例步骤：

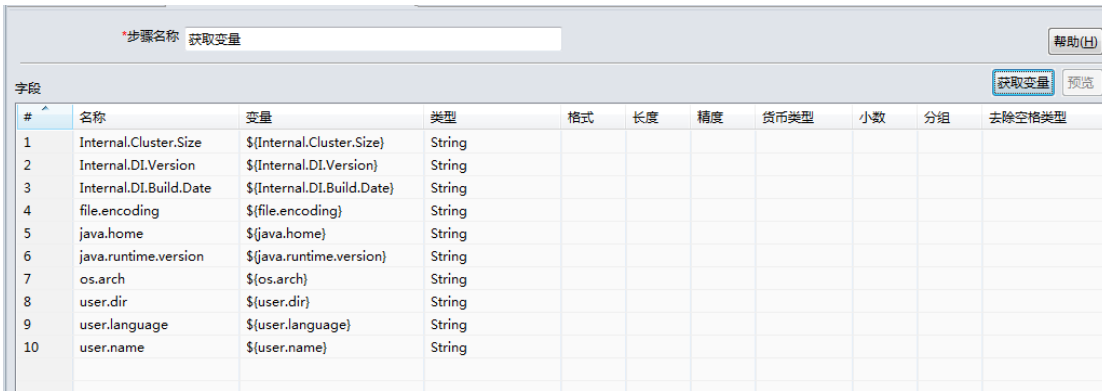
step1. 编辑转换“excel 输入”属性面板，如下图所示：



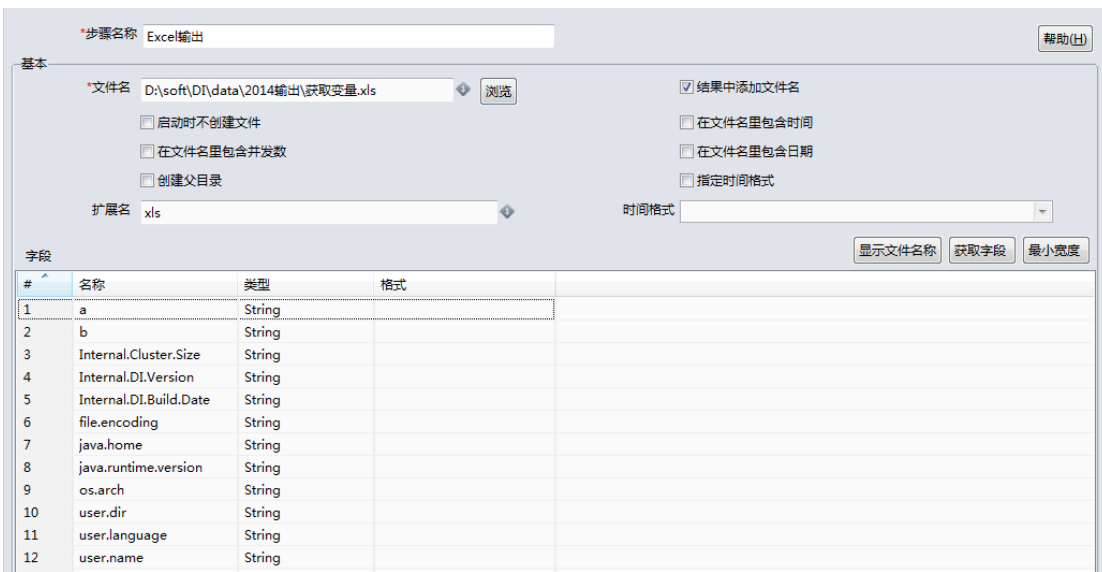
step2. 编辑转换“设置变量”属性面板，如下图所示：



step3. 编辑转换“获取变量”属性面板，如下图所示：



step4. 编辑转换“Excel 输出”属性面板，如下图所示：



step5. 运行转换，即可。

5. 44. 作业组件：设置变量

5. 44. 1. 参数说明

转换步骤名称	设置变量
功能描述	
这个步骤允许你在一个任务中或者虚拟机中设置变量。它仅仅可以用一行数据来设置变量值	
参数说明	
选项	描述
步骤名称	步骤名称，在单个转换中必须唯一。
变量有效范围	设置属性文件名得到的变量的有效范围。
变量	逐个设置所需的变量，包括变量名，变量值 以及每个变量的有效范围。
使用范例	

5.44.2. 示例

略。查看转换《获取变量》示例。

5.45. 批量加载组件：Oracle 加载

5.45.1. 参数说明

转换步骤名称	Oracle 加载
功能描述	
这个步骤允许你大批量加载数据到 Oracle 数据库,它将用一个正确的装载格式,然后调用 Oracle 的 SQL*Loader 数据加载工具加载到指定的表中	
参数说明	
选项	描述
步骤名称	步骤名称, 在单个转换中必须唯一
连接	维表所在数据库的连接。
模式	要写入数据的表的 schema 名称, 表中可以包含 “.”
表	目标表的名称
数据库加载工具路径	SQL*Loader 数据加载工具的全路径
加载方式	“自动加载” 或者 “手动加载”。如果是 “自动加载”, 步骤将在用指定的参数收到所有的输入后运行 Sqlldr 工具。如果是 “手动加载”, 将生成一个可以作为后面使用的控制和数据文件
加载动作	Append、Insert、Replace、Truncate。这些会映射到 sqlldr 的动作来执行。
错误的最大行数	发生错误的行数, 这时 sqlldr 将被中断。与 sqlldr 的 “ERROR” 属性来通信
提交	提交的行数, 与 sqlldr 的 “ROWS” 属性通信。
绑定大小	与 sqlldr 的 “BINDSIZE” 属性通信
读取大小	与 sqlldr 的 “READSIZE” 属性通信。
控制文件	sqlldr 控制文件的名称。
数据文件	数据将被写入的数据文件的名称
日志文件	日志文件的名称, 可选。
坏文件	坏文件的名称, 可选。
废文件	废文件的名称, 可选
编码	数据指定的编码格式, 可以从下拉列表框中选择
直接路径	直接路径加载的开关, 与 sqlldr 的 “DIRECT=TRUE” 通信。
使用后删除 cfg/dat 文件	加载后是否删除控制文件和数据文件的开关。 .
加载字段	这个表包含了需要从配置文件读取的一个字段值列表 表字段:ORACLE 表中加载的表字段。

流字段:输入行中带来的字段; 日期掩码:决定了 oracle 数据库中日期和时间戳加载的方式
使用范例

5.45.2. 示例

略。

6. 作业项介绍

作业项是构成作业的基本单位，通过作业项间的连接，我们可以完成不同的业务处理流程，下面将对每个作业项做详细说明。

6.1. 通用组件：开始

6.1.1. 参数说明

作业项名称	开始
功能描述	
该作业项是作业执行的开始步骤，并且只能有一个。该作业项还可以设置调度策略	
参数说明	
选项	说明
作业项名称	该作业项的名称
是否加入调度序列	设置的调度策略是否生效
发生频率	设置调度执行的频率是天、周或月，以及间隔的次数。周或月还可以设置周或月中是哪几天执行
一天内(服务器时间)	设置一天内调度一次的时间或者周期性调度的频率，按照分钟或小时，并设置开始及结束时间。
有效期(服务器时间)	设置整个调度的有效时间范围。
清除	清除设置好的调度策略
使用范例	

6.1.2. 示例

示例说明：示例结果图如下所示：



示例步骤:

step1. 编辑作业项“START”属性面板，如下图所示：

*作业项名称 START 帮助(H)

加入调度序列

发生频率

按天 按周 按月

间隔 1 在有效期的时间段内，两次调度执行所间隔的天数

一天内(服务器时间)

发生一次时间 00:00:00 周期发生频率 1 分钟

起始时间 00:00:00 终止时间 00:00:00

有效期(服务器时间)

开始日期 2014-07-30 15:30:25 结束日期 清除(C)

step2. 编辑作业项“设置变量”属性面板，如下图所示：

*作业项名称 设置变量 帮助(H)

属性文件名 浏览

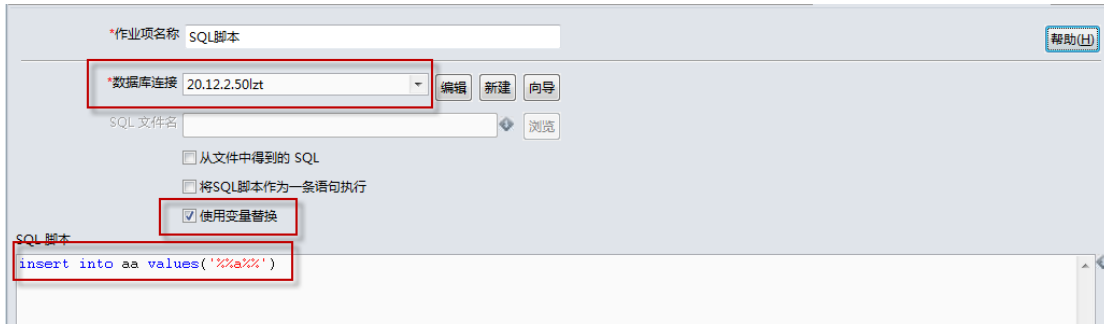
变量有效范围 在JVM中有效

变量替换

变量

#	变量名	值	类型	变量有效范围
1	a	xue	值	在JVM中有效

step3. 编辑作业项“SQL脚本”属性面板，如下图所示：



step4. 运行。

6.2. 通用组件：空操作

6.2.1. 参数说明

此作业项什么都不做，可以用作流程分支。

6.2.2. 示例

略。

6.3. 通用组件：作业

6.3.1. 参数说明

作业项名称	作业
功能描述	
该作业项可以引用已有的作业并执行	
参数说明	
选项	说明
作业项名称	该作业项的名称
资源库:通过名称指定	通过名称指定引用的作业，可用变量
资源库:通过引用指定	通过 ID 指定引用的作业
复制上一步结果到命名参数	执行过程中复制上一步结果到命名参数
对每个输入行执行一次	对输入的内容逐行执行
将所有参数下发到子作业	在子作业中可以使用命名参数
获取命名参数	获取指定作业中的命名参数
命名参数	设置命名参数的名称，列名和值
使用范例	

6.3.2. 示例

略。

6.4. 通用组件：成功

6.4.1. 参数说明

作业项名称	成功
功能描述	
该步骤清除前面步骤的失败状态，并强制转换为成功	
参数说明	
无	
使用范例	

6.4.2. 示例

略。

6.5. 通用组件：设置变量

6.5.1. 参数说明

作业项名称	设置变量
功能描述	
该作业项是根据设计需要用来定义作业和转换中用到的变量	
参数说明	
选项	说明
作业项名称	该作业项的名称
属性文件名	可以文件中批量读取自定义变量。
变量有效范围	设置属性文件名得到的变量的有效范围。
变量替换	是否覆盖前面步骤中设置过的同名的变量。
变量	逐个设置所需的变量，包括变量名，变量值以及每个变量的有效范围
使用范例	

6.5.2. 示例

略。参看作业项《开始》示例。

6.6. 通用组件：转换

6.6.1. 参数说明

作业项名称	转换
功能描述	
该作业项可以引用已有的转换并执行	
参数说明	
选项	说明
作业项名称	该作业项的名称
通过目录与名称指定转换	通过名称指定引用的转换，可用变量
通过引用指定转换文件	通过 ID 指定引用的转换
复制上一步结果到命名参数	执行过程中复制上一步结果到命名参数
对每个输入行执行一次	对输入的内容逐行执行
在执行前清除结果行列表	在执行前清除结果中的行列表信息
在执行前清除结果文件列表	在执行前清除结果中的文件列表信息
将所有参数下发到子转换	在子转换中可以使用命名参数
获取参数	获取指定转换中的命名参数
命名参数	设置命名参数的名称，流列名和值
使用范例	

6.6.2. 示例

略。

6.7. 邮件组件：邮件

6.7.1. 参数说明

作业项名称	邮件
功能描述	
该作业项可以用于发送邮件	
参数说明	

选项	说明
作业项名称	该作业项的名称
收件人	设置收件人的地址，抄送人的地址，暗送人的地址
发件人	设置邮件的发件人地址，回复名称，回复地址，联系人及联系电话
邮件服务器	设置 SMTP 服务器地址，端口，是否需要用户验证的名称和密码，是否需要安全验证以及安全连接的类型
消息设置	设置消息中是否有日期，是否只发送邮件正文，是否使用 HTML 邮件格式，设置 HTML 邮件格式的编码，设置是否有管理优先级以及设置管理的优先级，重要性和敏感度
消息	设置邮件的标题和正文
预览文件名称	设置是否带附件，以及附件的文件类型，附件是否要压缩成统一格式以及文件名
内嵌图片	设置内嵌图片的文件名和 ID，可设置多个
使用范例	

6.7.2. 示例

略。

6.8. 邮件组件：UAP 消息

6.8.1. 参数说明

作业项名称	UAP 消息
功能描述	
该作业项可以发送执行结果到 UAP 消息中心。	
参数说明	
选项	说明
作业项名称	该作业项的名称
接收用户	选择需要接收消息的用户。
标题	消息的正文。
附件	选择需要上传的附件。
使用范例	

6.8.2. 示例

略。

6.9. 条件组件：检查表

6.9.1. 参数说明

作业项名称	检查表
功能描述	
该作业项可以用于检查表是否存在。	
参数说明	
选项	说明
作业项名称	该作业项的名称
数据库连接	选择一个可用的数据库连接
模式名称	选择一个可用的模式名称
表名	输入一个需要检查的表
使用范例	

6.9.2. 示例

略。

6.10. 条件组件：检查连接

6.10.1. 参数说明

该作业项可以用于检查某一数据库连接是否有效可用

6.10.2. 示例

略。

6.11. 条件组件：检验值

6.11.1. 参数说明

作业项名称	检查表
功能描述	
该步骤可以用于检验结果流里字段或变量的值。	

参数说明	
选项	说明
作业项名称	该作业项的名称
检验	选择要检验的是上一步结果的字段还是变量
字段名/变量名	输入字段或变量的名称
类型	输入字段或变量的名称
变量设置成功	选择后，当然变量设置时走成功的分支
成功条件	选择一种可以成功的情况
值	输入一个和字段或变量比较的值
使用范例	

6.11.2. 示例

略。

6.12. 脚本组件：SQL 脚本

6.12.1. 参数说明

作业项名称	SQL 脚本
功能描述	
该作业项可以用执行 SQL 语句	
参数说明	
选项	说明
作业项名称	该作业项的名称
数据库连接	设置一个数据库连接
从文件得到 SQL	是否从文件得到 SQL 语句
SQL 文件名	从文件得到的 SQL 文件名
将 SQL 脚本作为一条语句执行	将 SQL 语句作为一个语句块执行
使用变量替换	执行 SQL 语句中使用变量替换
SQL 脚本	编写要执行的 SQL 脚本
使用范例	

6.12.2. 示例

略。参看作业项《开始》示例。

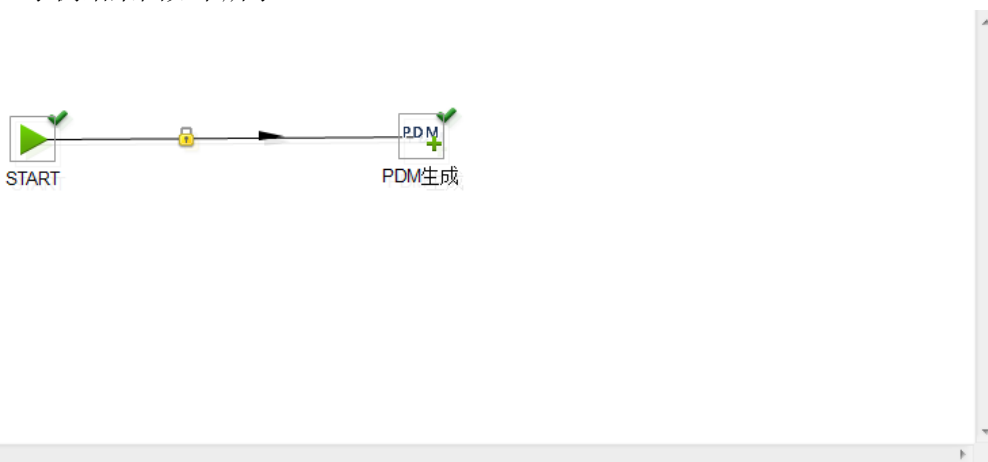
6.13. 脚本组件：PDM 生成

6.13.1. 参数说明

作业项名称	PDM 生成
功能描述	
该步骤可以通过 PDM 文件生成数据库对象。当物理模型与目标库的数据类型一致时支持 Oracle, gbase, sqlserver, db2. 当物理模型与目标库的数据类型不一致时只支持物理模型为 Sqlserver2005	
参数说明	
选项	说明
作业项名称	该作业项的名称
数据库连接	选择一个需要生成数据库对象的目标连接
模式	选择一个需要生成数据库对象的目标模式
选择文件	选择一个 PDM 文件用来生成数据库对象
删除已存在的表或视图	选择是否要删除库中已存在的表或视图
生成类型	选择需要生成的数据库对象类型，包括表、视图、索引和外键
使用范例	

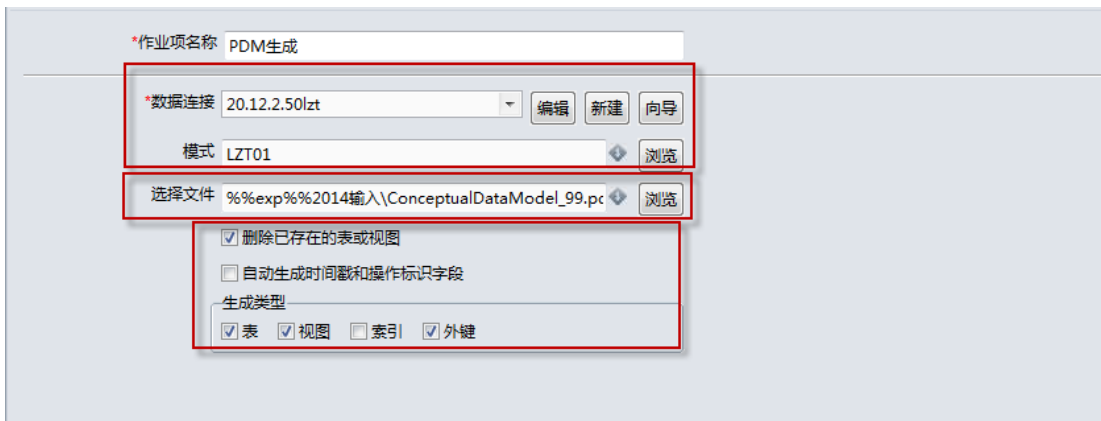
6.13.2. 示例

示例说明：示例结果图如下所示：



示例步骤：

step1. 编辑作业项“PDM 生成”属性面板，如下图所示：



step2. 运行即可。

6. 14. 批量加载组件：Oracle 到 GBase

6. 14. 1. 参数说明

作业项名称	Oracle 到 GBase
功能描述	
该作业项可以用于从 Oracle 批量载入数据到 GBase	
参数说明	
选项	说明
作业项名称	该作业项的名称
源端数据库连接	选择一个源端的数据连接
模式	选择一个源端的模式
数据表	选择一个源端的数据表
目标端数据库连接	选择一个目标端的数据连接
GBloader	设置一个 GBloader 进行加载的端口
数据库	获取 GBase 的数据库
重新构造目标表	删除目标端与源端同名的表，并重建该表
追加方式同步数据	不删除原表中的数据直接载入新数据
表映射	以列表的形式展现需要进行批量载入的源端和目标端的表
使用范例	

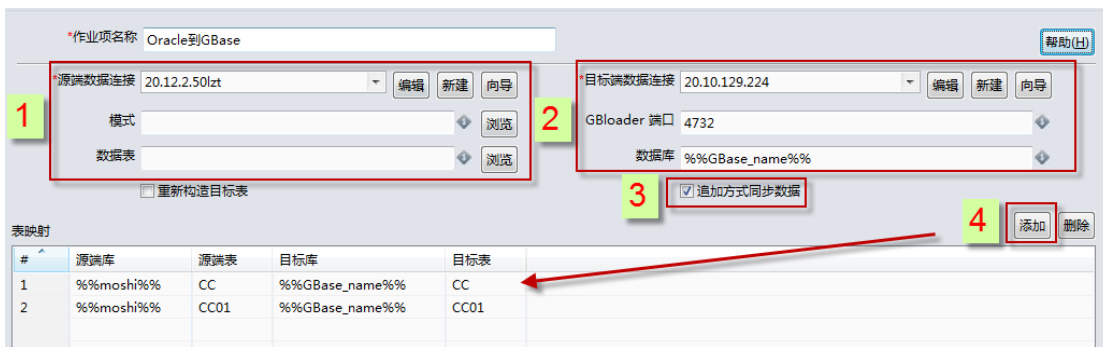
6. 14. 2. 示例

示例说明：示例结果图如下所示：



示例步骤:

step1. 编辑作业项“Oracle 到 GBase”属性面板，如下图所示:



step2. 运行。

6. 15. XML 组件: XSD 校验

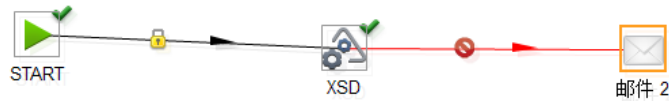
6. 15. 1. 参数说明

作业项名称	XSD 校验								
功能描述	验证 Xml 的结构								
参数说明	<table border="1"> <thead> <tr> <th>选项</th> <th>说明</th> </tr> </thead> <tbody> <tr> <td>作业项名称</td> <td>该作业项的名称，名称必须唯一</td> </tr> <tr> <td>XML 文件</td> <td>XML 文件名称</td> </tr> <tr> <td>XSD 文件</td> <td>XSD 文件名称</td> </tr> </tbody> </table>	选项	说明	作业项名称	该作业项的名称，名称必须唯一	XML 文件	XML 文件名称	XSD 文件	XSD 文件名称
选项	说明								
作业项名称	该作业项的名称，名称必须唯一								
XML 文件	XML 文件名称								
XSD 文件	XSD 文件名称								
使用范例									

6. 15. 2. 示例

示例说明: 示例结果图如下所示:

作业 XSD 后台：组成部分，嵌套了作业“XSD”



作业 XSD：组成部分如下图所示：



示例步骤：

step1. 创建作业“XSD”，并编辑作业项“XSD 验证器”属性面板，如下图所示：

step2. 创建作业“后台 XSD”，并编辑嵌套作业项“XSD”属性面板，如下图所示：

step3. 编辑作业“后台 XSD”的作业项“邮件”属性面板，如下图所示：

运行。

6. 16. 应用组件：终止作业

6. 16. 1. 参数说明

作业项名称	终止作业
功能描述	
该作业项可以用于中止正在执行的作业	
参数说明	
选项	说明
作业项名称	该步骤的名称
消息	对于中止的描述
使用范例	

6. 16. 2. 示例

略。

6. 17. 应用组件：QDI 调用

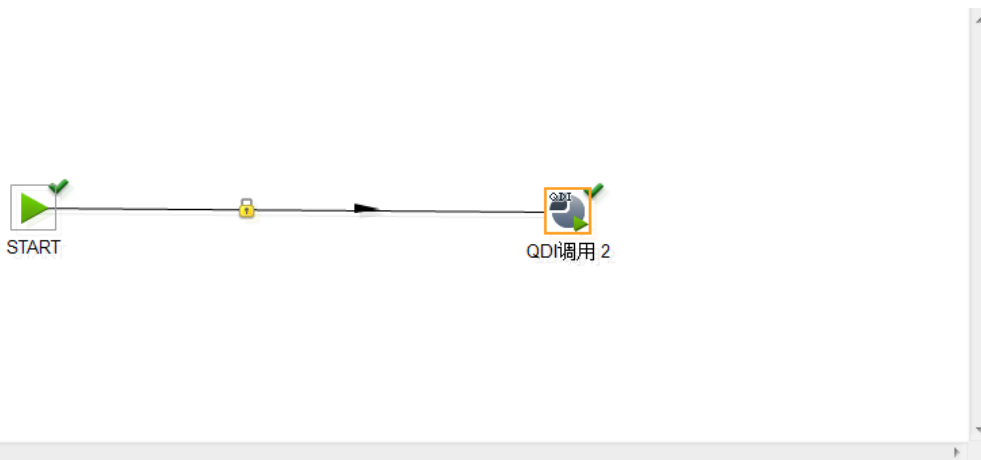
6. 17. 1. 参数说明

作业项名称	QDI 调用
功能描述	
该作业项可以用于执行 QDI 作业。	
参数说明	

选项	说明
作业项名称	该步骤的名称
选择 QDI 作业	选择需要执行的 QDI 作业。
执行单个任务	选择是否执行单个 QDI 任务。
选择 QDI 任务	选择需要执行的 QDI 任务。
使用范例	

6.17.2. 示例

示例说明：示例结果图如下所示：



示例步骤：

step1. 编辑作业项“QDI 调用 2”属性面板，如下图所示：

step2. 运行。

6.18. 应用组件：表对比

6.18.1. 参数说明

作业项名称	表对比
功能描述	
该作业项可以用于执行全表对比	
参数说明	

选项	说明
作业项名称	该步骤的名称
源数据连接	选择一个源端的数据连接
源表模式	选择一个源端的模式
源表名	选择一张需要对比的源表
目标数据连接	选择一个目标端的数据连接
目标端模式	选择一个目标端的模式，GBASE 数据库则不需要选择模式
目标表名	选择一个目标端的表
主键类型	选择一种表的主键类型，数据库定义主键或是没有主键
对比列表	列出添加好的需要对比的表信息
是否输出 SQL	表对比不同时是否输出差异的 SQL 语句
输出路径	设置一个输出 SQL 语句的保存路径
不相同设置结果为失败	对比结果不相同则设置执行结果为失败
使用范例	

6.18.2. 示例

示例说明：示例结果图如下所示：



示例步骤：

step1. 编辑作业项“表对比”属性面板，如下图所示：

#	源表模式	源表名	目标表模式	目标表名	主键类型
1	%%moshi...	CC	%%moshi%%	CC01	使用数据库...

step2. 运行。

6. 19. 文件传输组件：文件上传

6. 19. 1. 参数说明

作业项名称	文件上传
功能描述	
该步骤可以用于执行文件上传到文件服务器	
参数说明	
选项	说明
作业项名称	该步骤的名称
IP 地址	设置一个需要上传的文件服务器地址
端口	设置一个需要上传的文件服务器端口
模块名	设置一个需要上传的文件服务器的模块名
测试连接	测试文件服务器是否能正常连接
上传文件后删除本地文件	上传文件后把本地文件删除掉
浏览	浏览本地要上传的文件
上传列表	列出要上传的文件列表
使用范例	

6. 19. 2. 示例

略。

6. 20. 文件传输组件：文件下载

6. 20. 1. 参数说明

作业项名称	文件下载
功能描述	
该步骤可以用于执行从文件服务器下载文件到本地	
参数说明	
选项	说明
作业项名称	该步骤的名称
IP 地址	设置一个需要上传的文件服务器地址
端口	设置一个需要上传的文件服务器端口
模块名	设置一个需要上传的文件服务器的模块名

测试连接	测试文件服务器是否能正常连接
浏览	浏览文件服务器上可以下载的文件
文件名称	列出要下载的文件列表
目标目录	选择一个文件下载的地址
在文件名中包含日期	是否在下载的文件名中包含下载时的系统日期
在文件名中包含时间	是否在下载的文件名中包含下载时的系统时间
指定日期时间格式	是否指定需要的日期时间格式
日期时间格式	选择一种需要的日期时间格式
在扩展名前添加日期	名称中的日期添加到扩展名之前
不能覆盖文件	是否需要覆盖目录下已存在的同名文件
如果文件已存在	选择一种不覆盖文件情况下处理同名文件的处理方法
在结果中添加文件名	在执行结果中添加下载的文件名
使用范例	

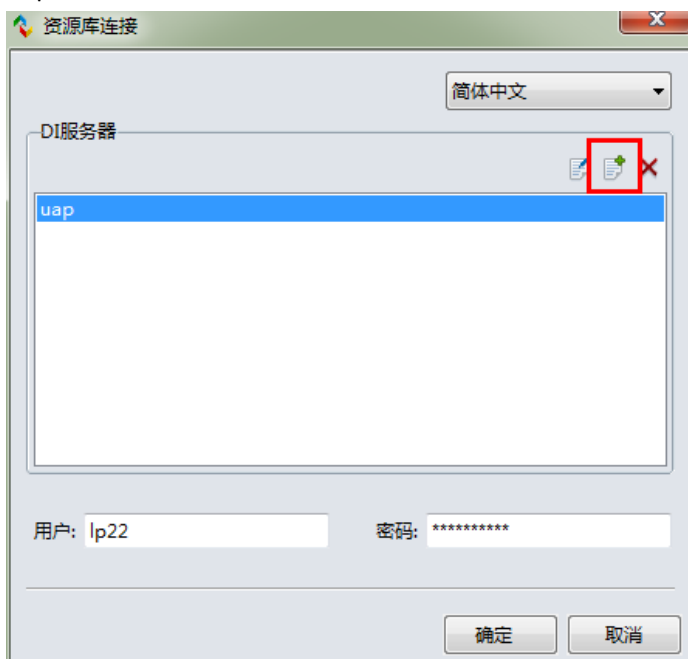
6. 20. 2. 示例

略。

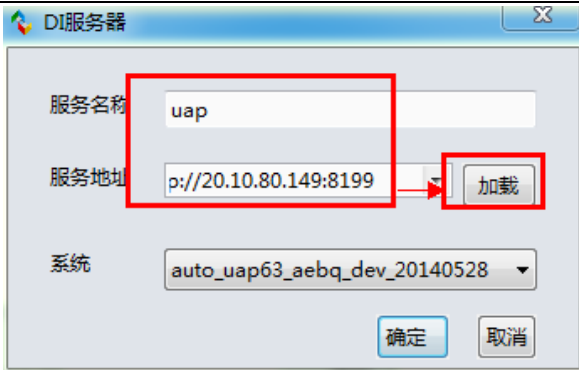
7. 开始使用 DI

7. 1. 登录服务器

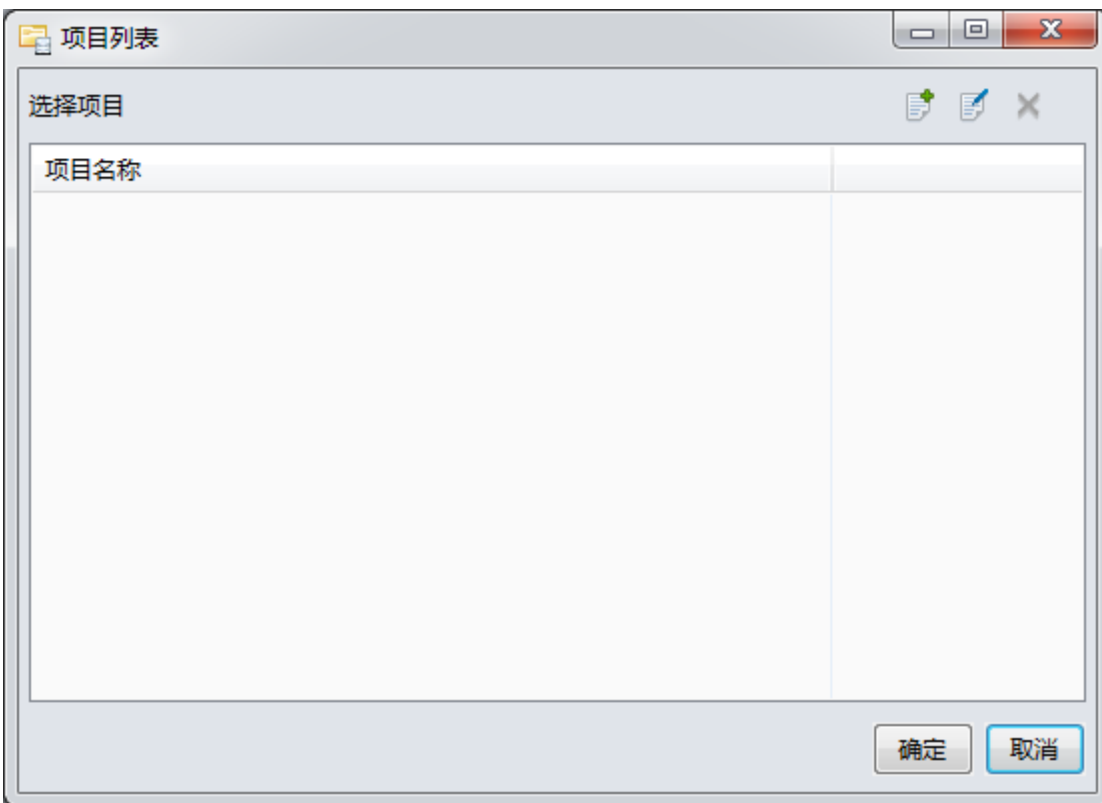
step1. 选择【文件】—【登录】，弹出资源库连接对话框，选择【新增】，如下图所示：



step2. 输入“服务器名称”和“服务器地址”，选择【加载】—【确定】。

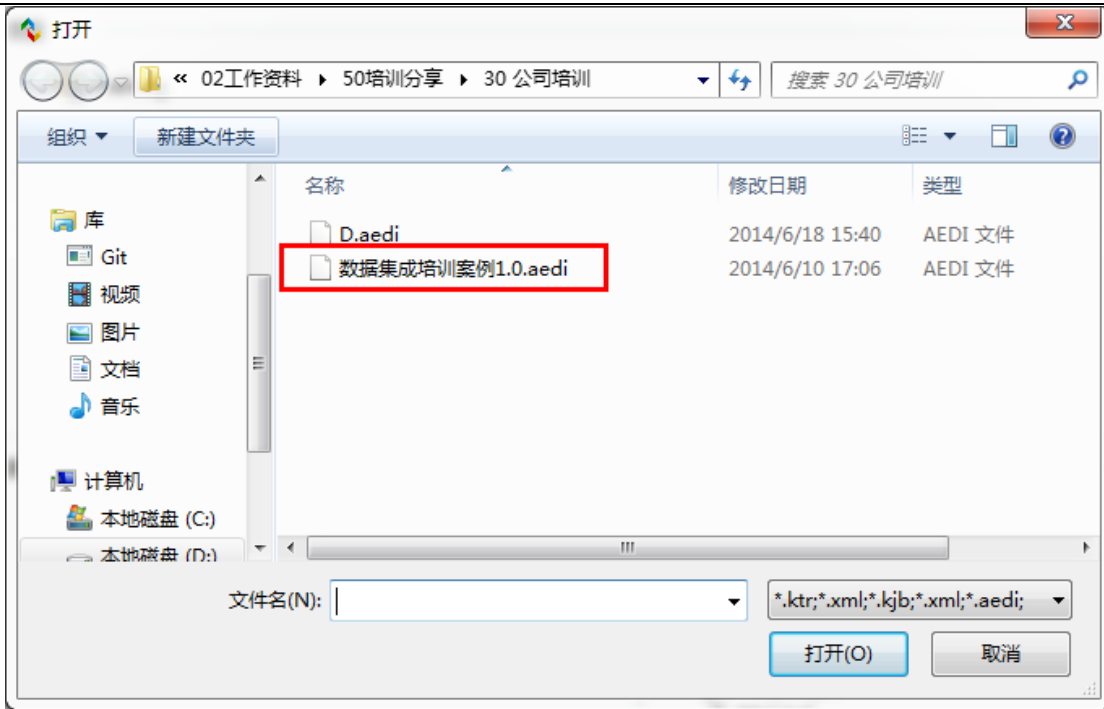


7.2. 如何新建一个项目

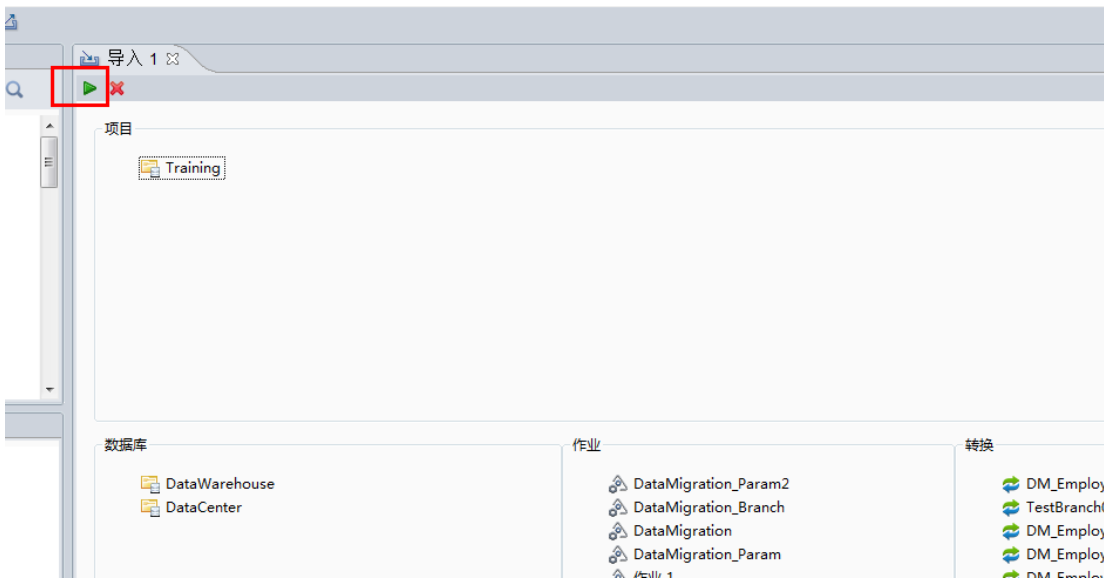


7.3. 如何导入一个项目

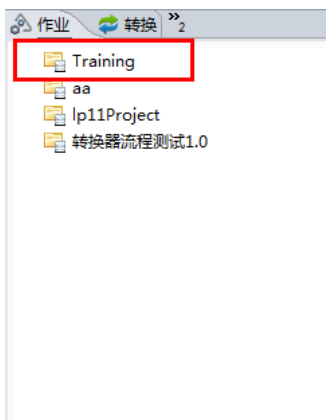
step1. 选择【文件】—【导入】，弹出对话框，选择要导入的文件，如下图所示：



step2. 选中要导入的项目，选择 .

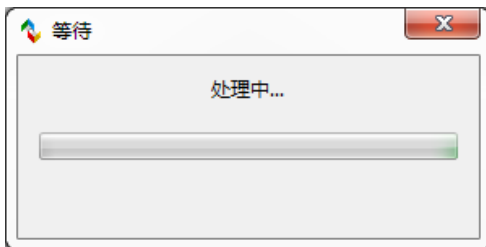
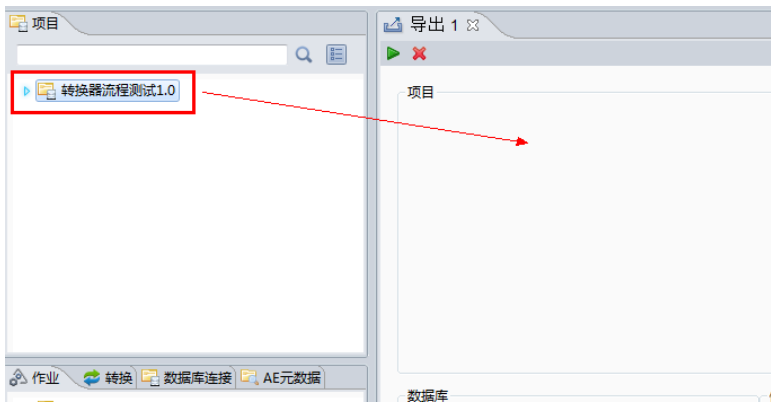



step3. 在左侧资源管理区可以浏览到导入的数据，如下图所示。

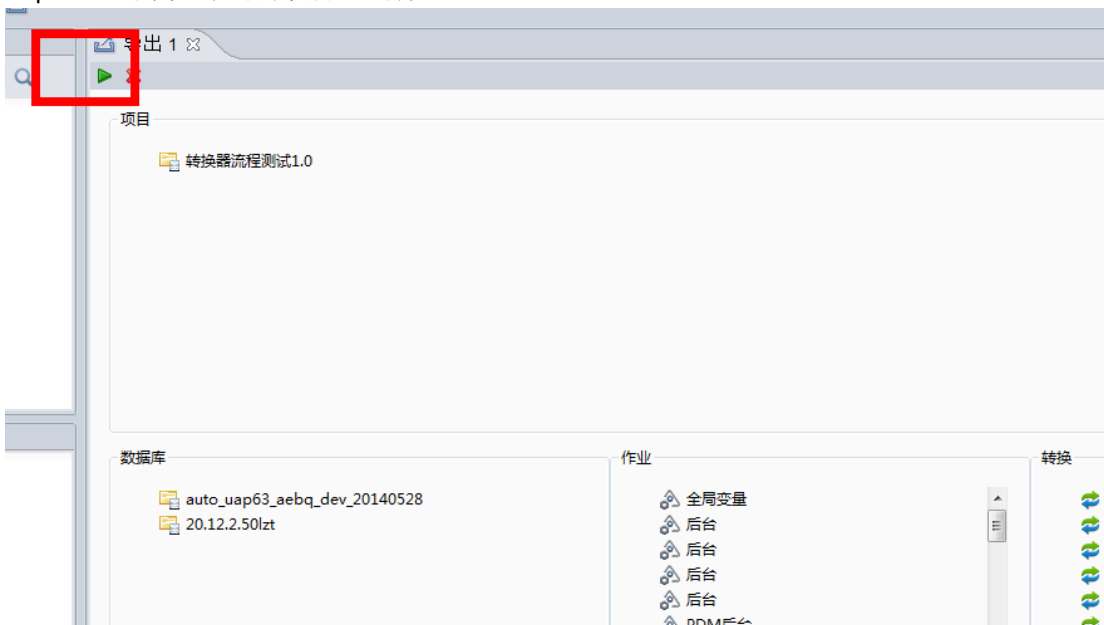


7.4. 如何导出一个项目

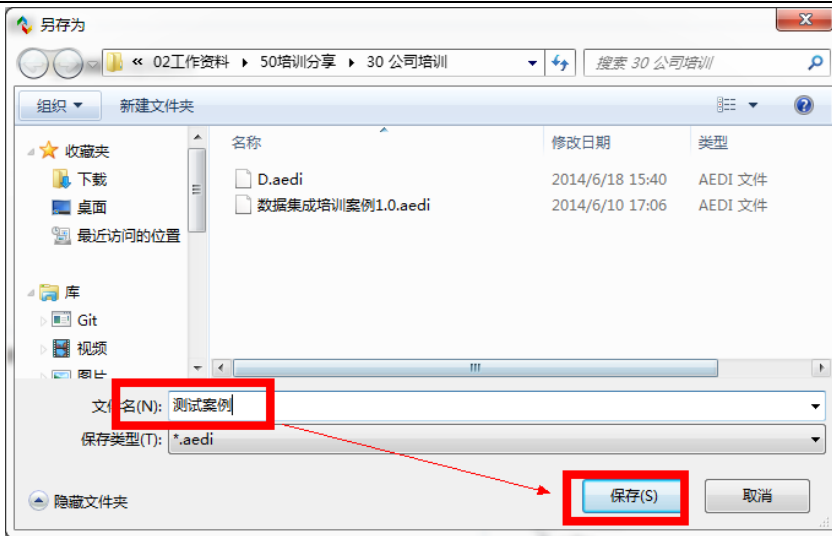
step1. 选择【文件】—【导出】，拖动左侧需要导出的项目到右侧面板中。



step2. 选中要导出的项目，选择 。

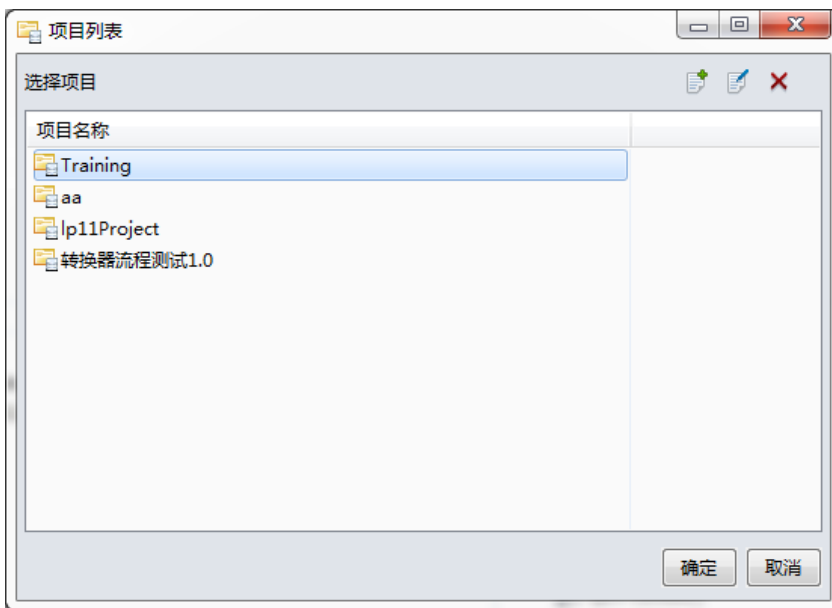


step3. 选择保存路径，默认文件后缀为.aedi，保存即可。

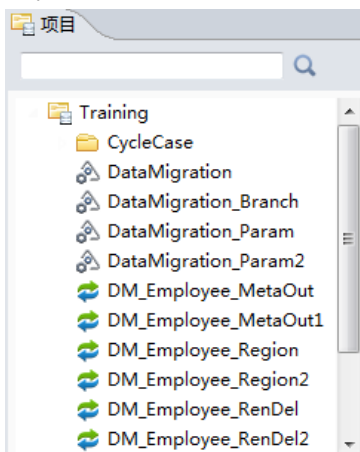


7.5. 如何打开一个项目

step1. 选择【文件】—【打开项目】，选中要打开的项目。



step2. 在项目区可以看到打开的项目。



7.6. 如何创建一个转换

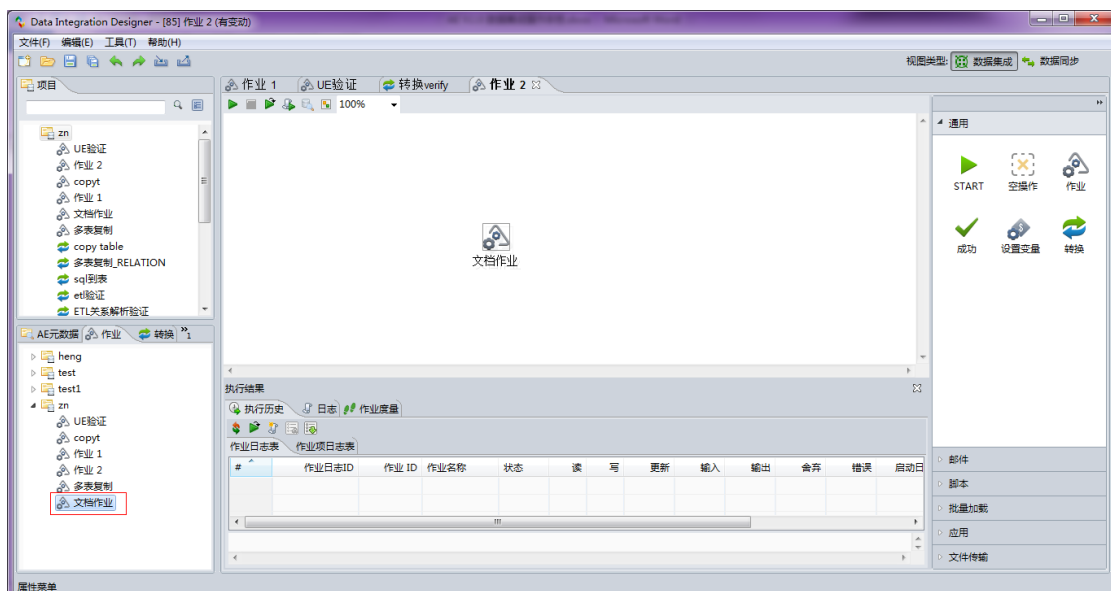
参看章节《5.2.2 示例》。

7.7. 如何创建一个作业

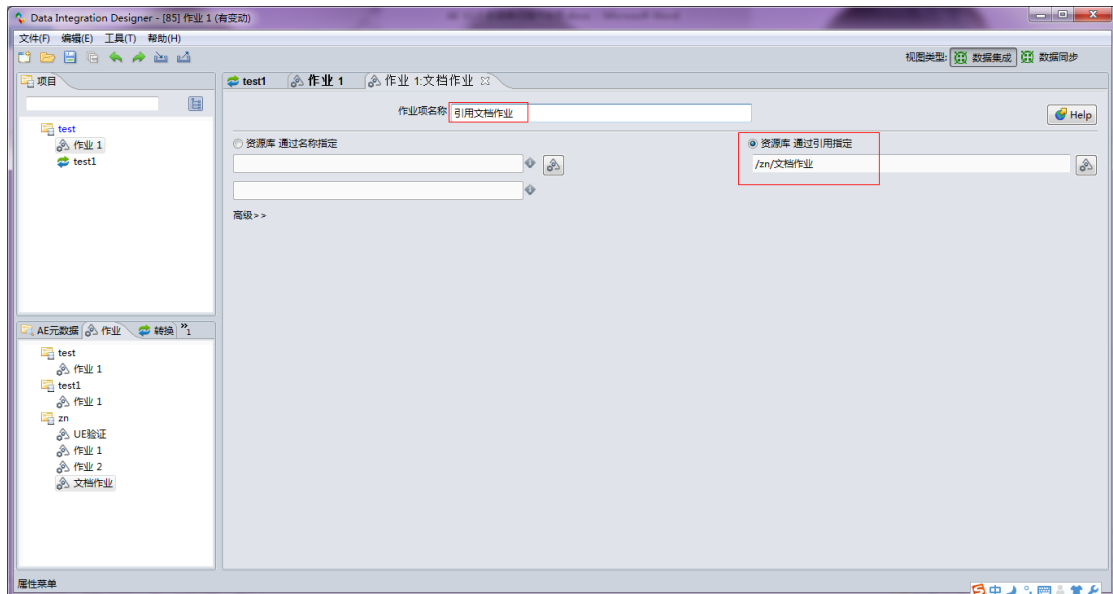
同新建转换。略。

7.8. 如何嵌套引用作业？

每个作业都可以通过拖拽的方式，完成作业的引用，类似项目区中作业的引用，例如，在当前项目新建作业 2，选中“zn”项目中的“文档作业”，拖拽到作业 2 的设计区，如下图所示：



双击“文档作业”，可以看到这个作业是通过资源库引用指定的，将此作业重命名为“引用文档作业”，如下图所示：

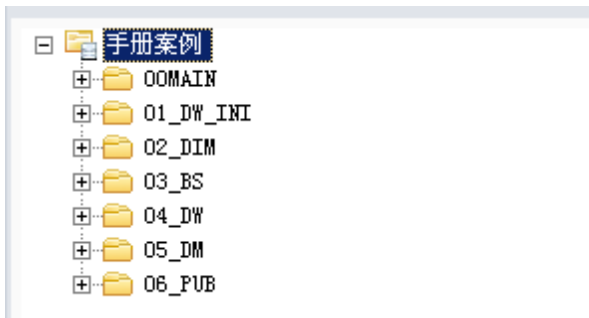


点击保存，关闭，则完成了作业的嵌套引用。

7.9. 典型案例

7.9.1. 案例说明

如下图所示，按照数据仓库分层建立文件夹。



其中，

- 00MAIN 为主调文件夹。
- 01_DW_INT 为数据仓库初始化文件夹。
- 02_DIM 为
- 03_BS 为
- 04_DW 为
- 05_DM 为
- 06_PUB 为

7.9.2. 作业调度

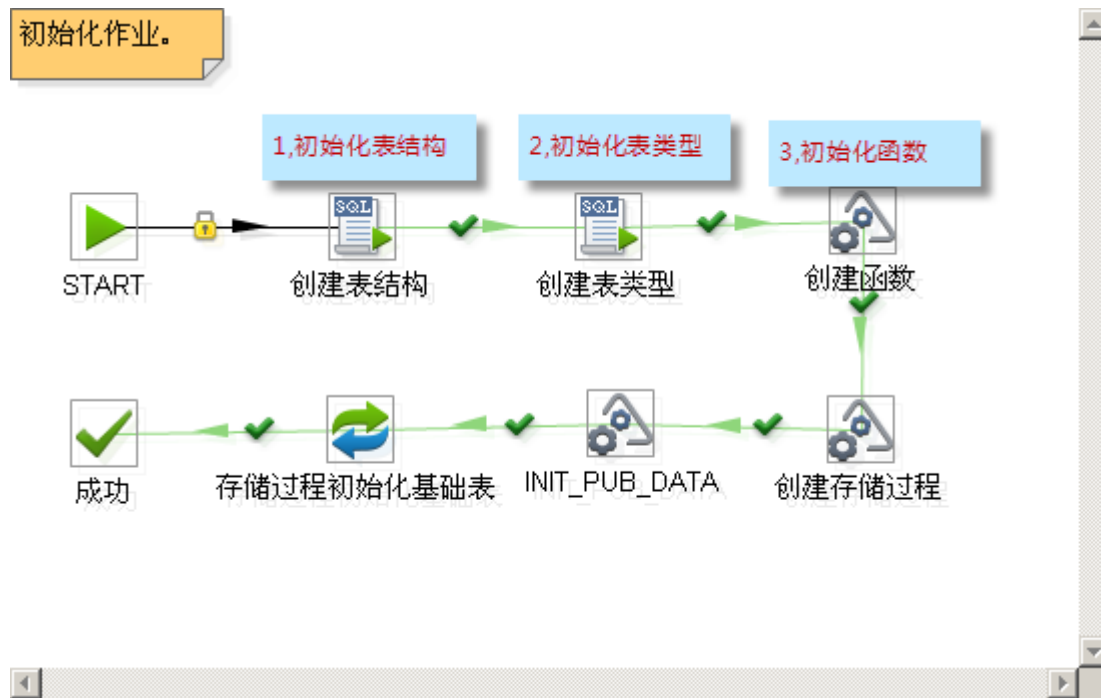
7.9.2.1. 主调度作业



作业“初始化”参看章节 7.10.3.2。

作业“运行业务数据”参看章节 7.10.3.3。

7.9.2.2. 初始化作业



一，作业项“创建表结构”

*作业项名称 帮助(H)

*数据库连接 编辑 新建 向导

SQL 文件名 浏览

从文件中得到的 SQL

将SQL脚本作为一条语句执行

使用变量替换

SQL 脚本

```

/*-----*/
/* DBMS name:      ORACLE Version 11g          */
/* Created on:     2014/3/7 13:25:43          */
/*-----*/

/*-----*/
/* Table:          */
/*-----*/
create table BA_SCM_BS_ARAPTALLY
(
  ID          CHAR(7)

```

创建表类型

*作业项名称 帮助(H)

*数据库连接 编辑 新建 向导

SQL 文件名 浏览

从文件中得到的 SQL

将SQL脚本作为一条语句执行

使用变量替换

SQL 脚本

```

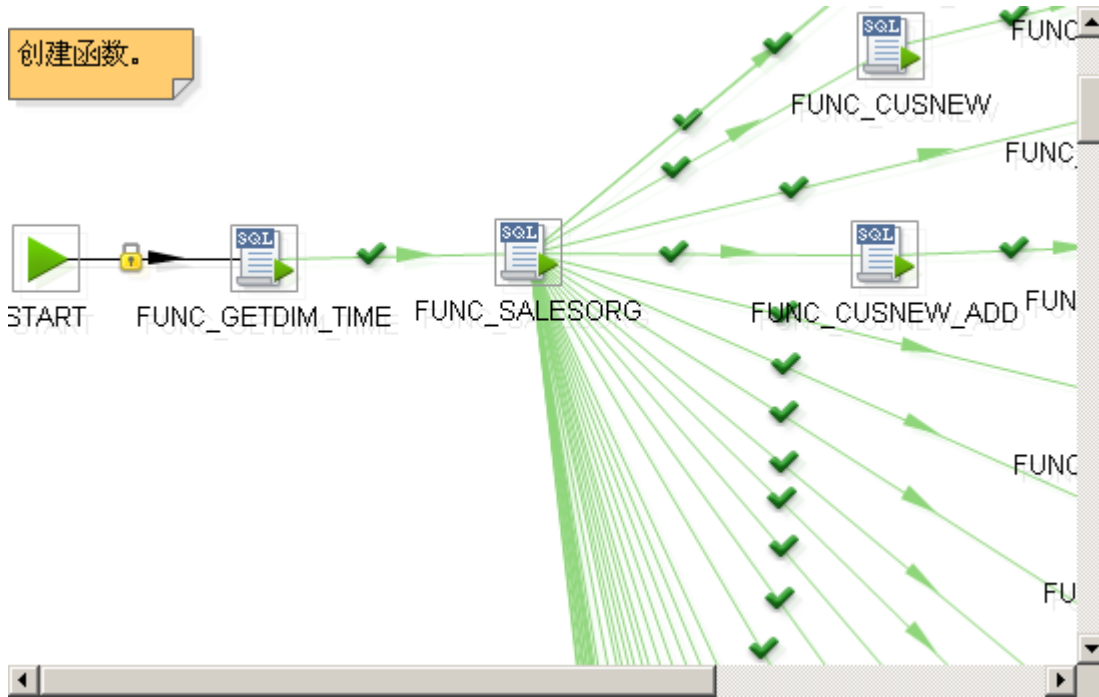
/*-----*/
/* DBMS name:      ORACLE Version 11g          */
/* Created on:     2013912 16:06:18          */
/*-----*/

create or replace type TABLE1 as object (
  t1 VARCHAR2 (50),
  t2 VARCHAR2(20),
  t3 VARCHAR (50),
  t4 NUMBER,
  t5 NUMBER,
  t6 NUMBER
);

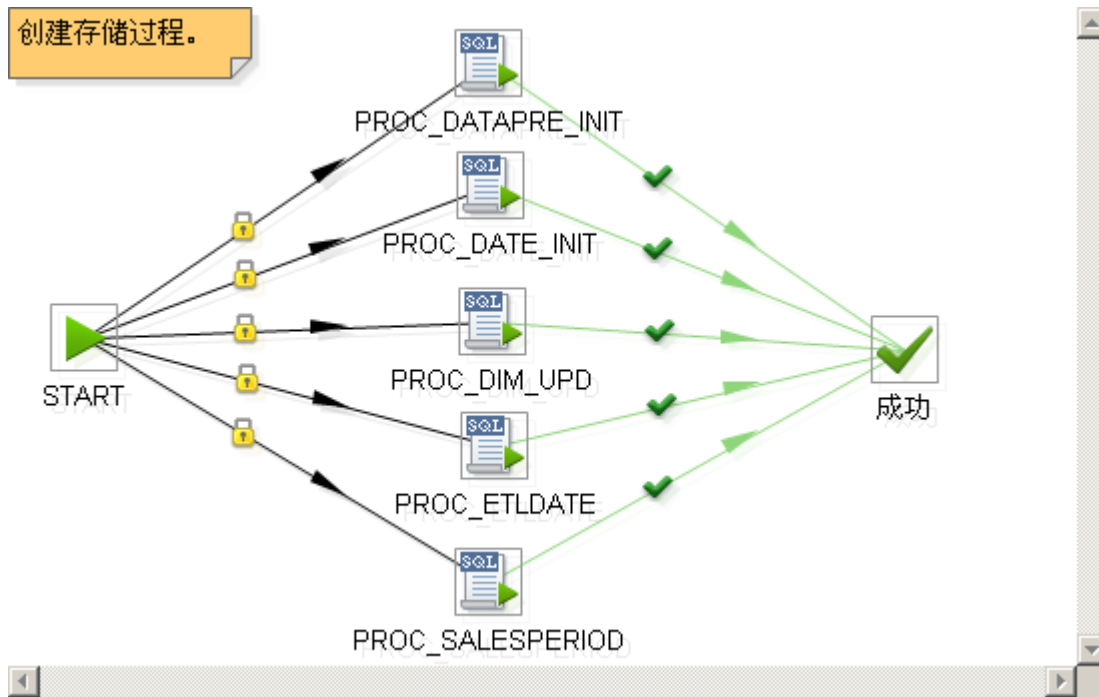
create or replace type t1 as table of TABLE1;

```

创建函数



创建存储过程



*步骤名称

*数据库连接

*存储过程名称

启用自动提交

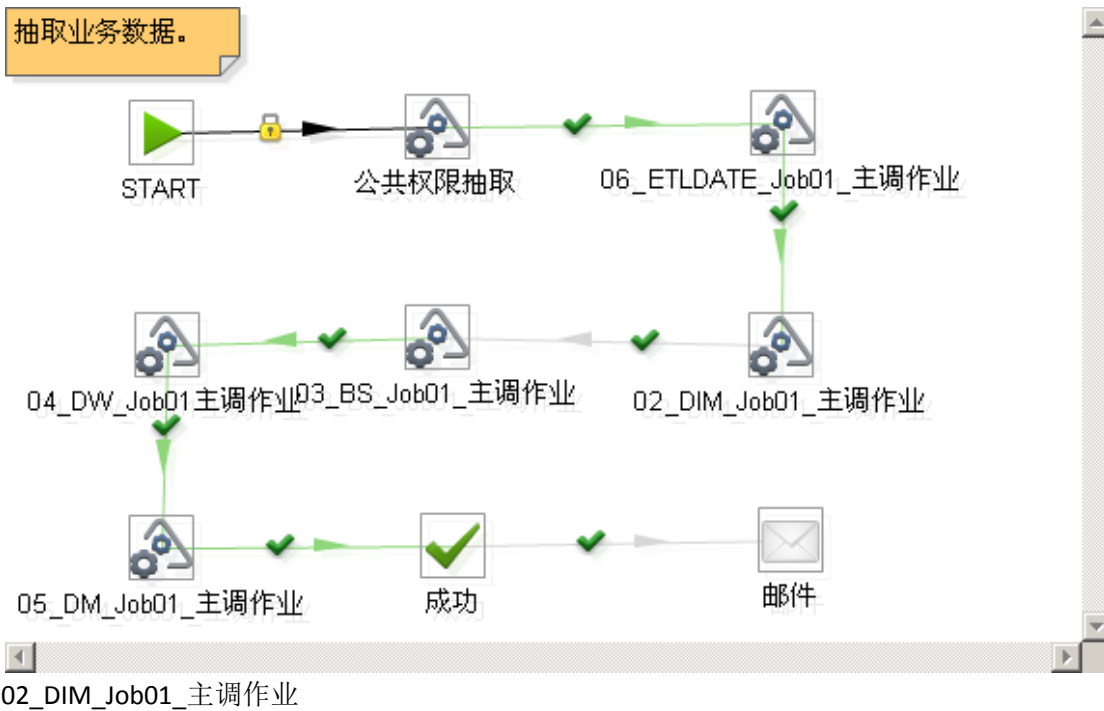
返回值名称

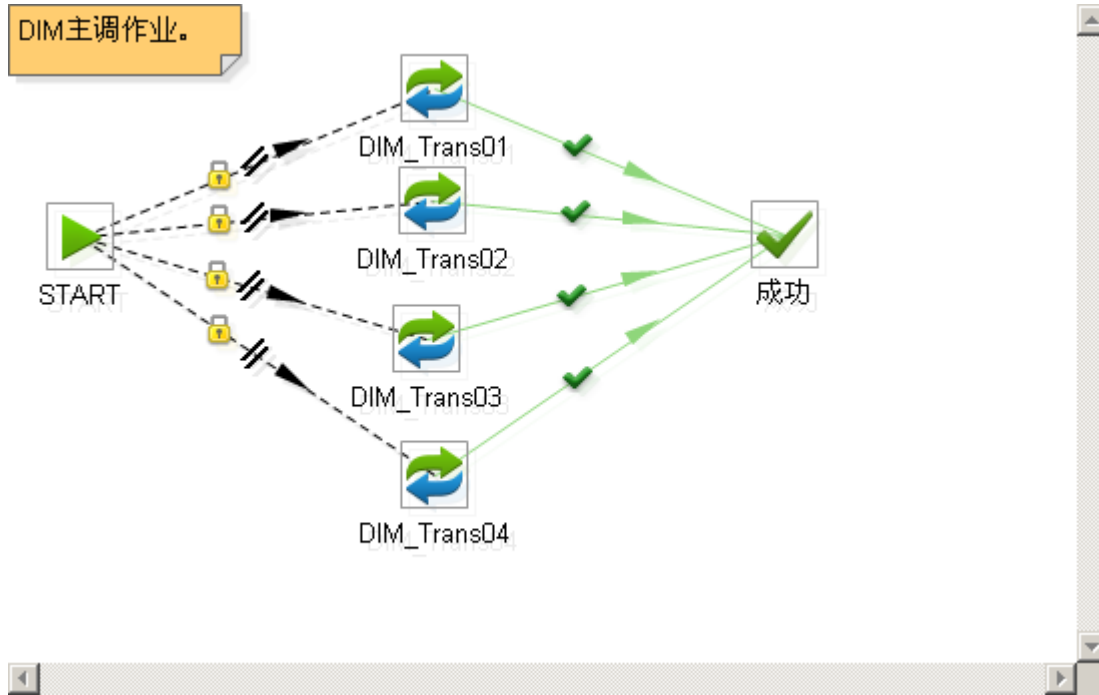
返回值类型

参数

#	名称	方向	类型
1	P_O_SUCCEED	OUT	String

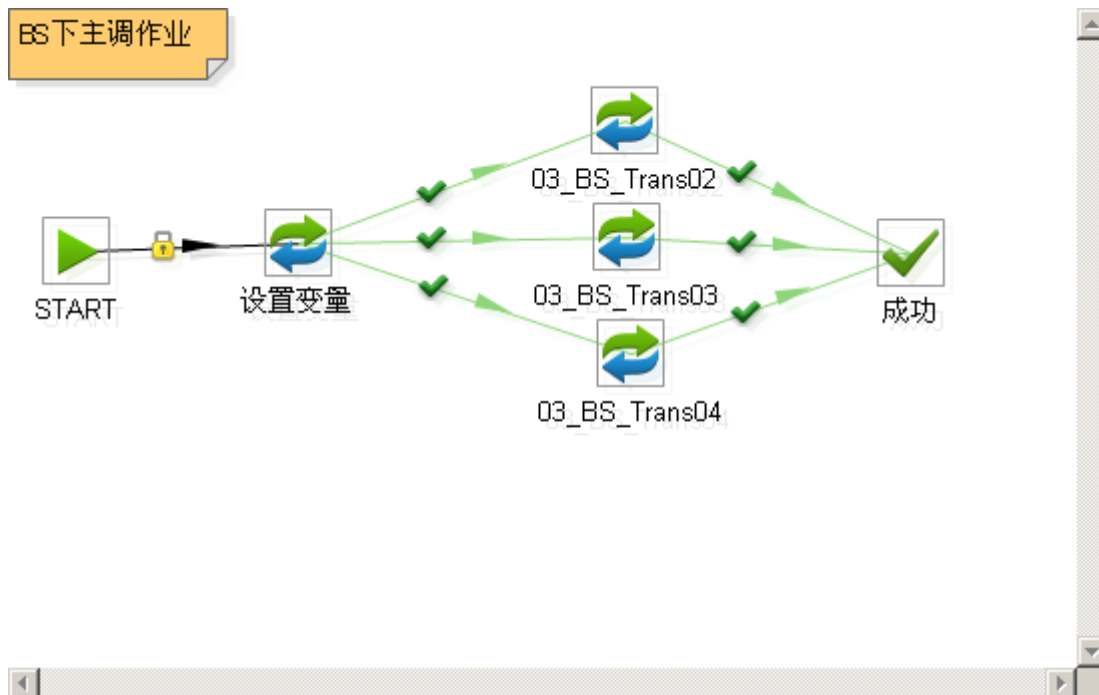
7.9.2.3. 增量数据作业





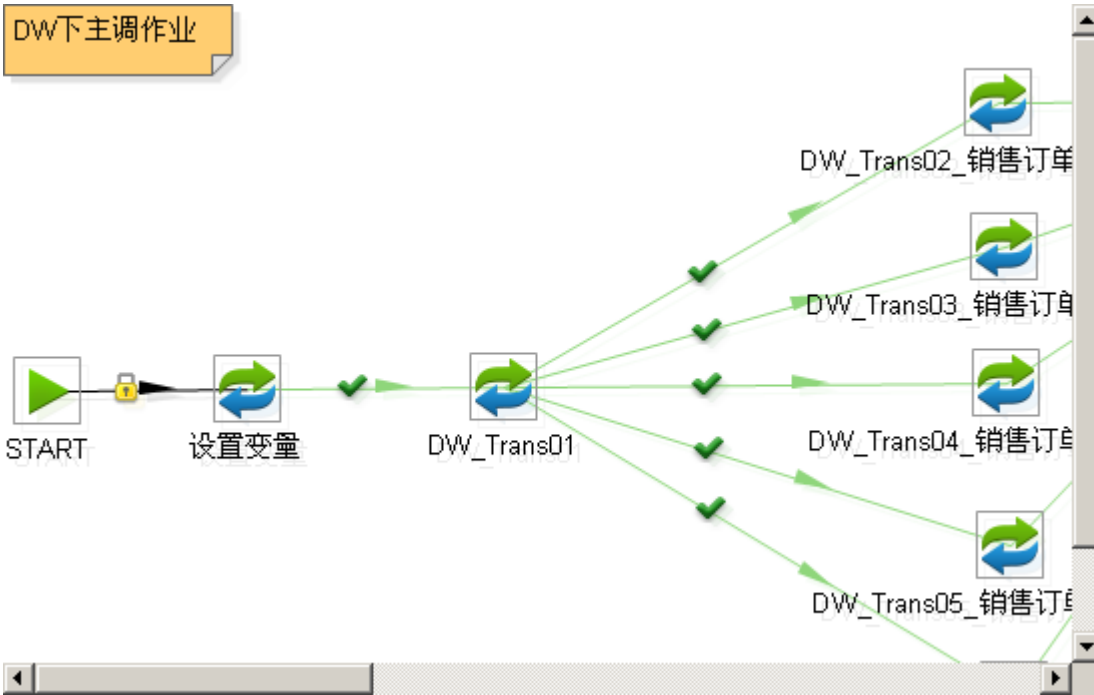
其中，DIM_Trans01 是一个简单的由“表输入”和“Meta 输出”组成的转换。

03_BS_Job01_主调作业



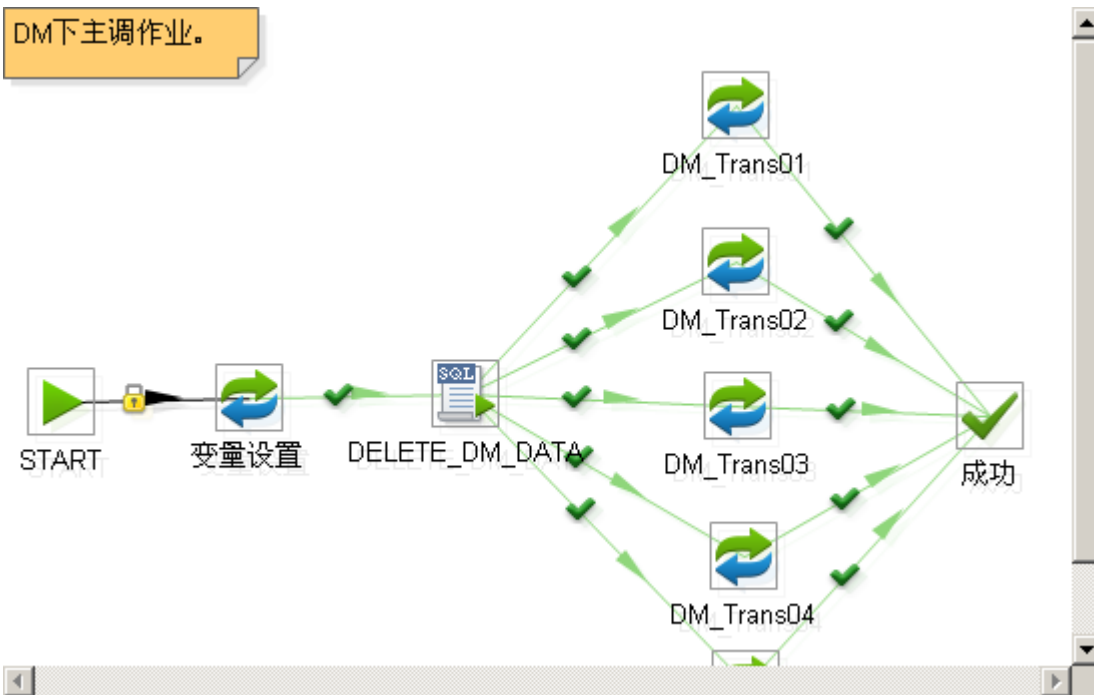
04_DW_Job01 主调作业

DW下主调作业



05_DM_Job01_主调作业

DM下主调作业。



7.9.2.4. 其他作业：数据仓库重新初始化

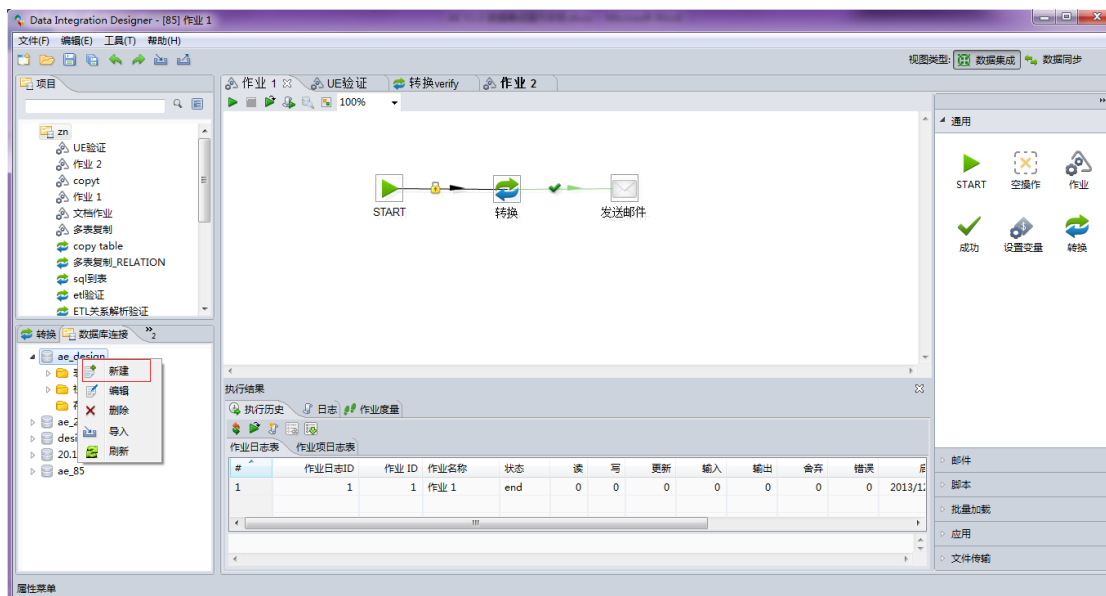
删除数据库对象，需要清除数据库对象时候启动。

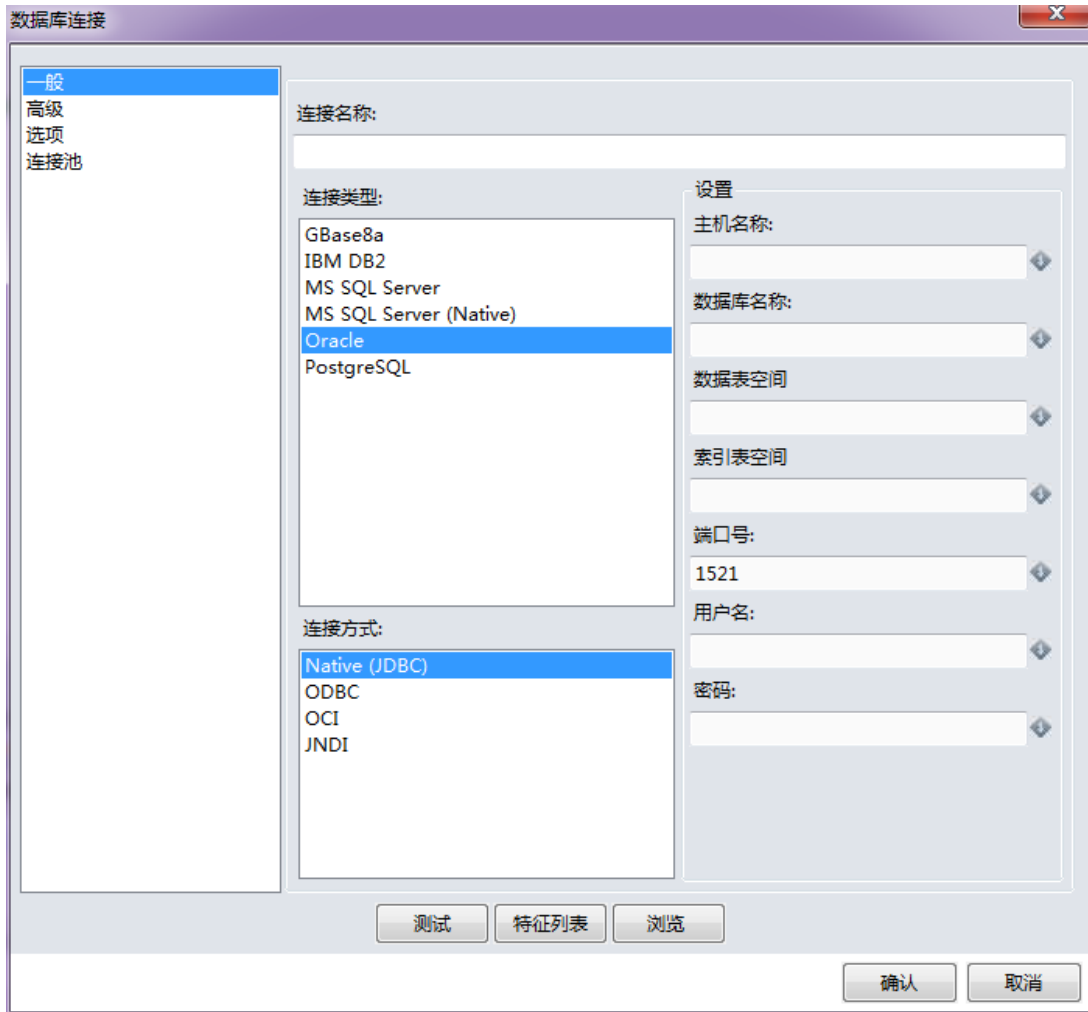


8. 数据库连接管理

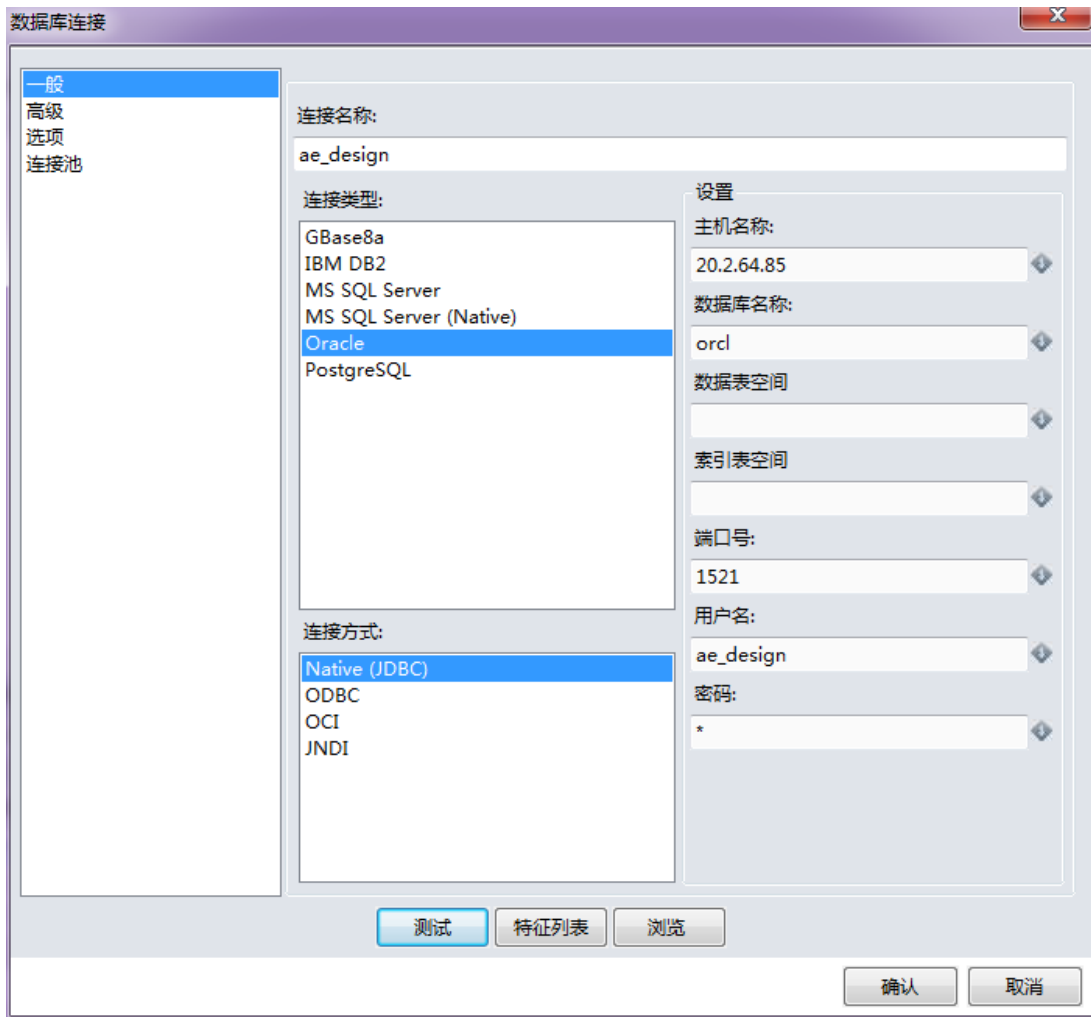
8.1. 如何新建数据库连接？

在数据连接上单击鼠标右键，在弹出的快捷菜单上，选择<新建>，弹出新建数据库连接的对话框，如下图所示：

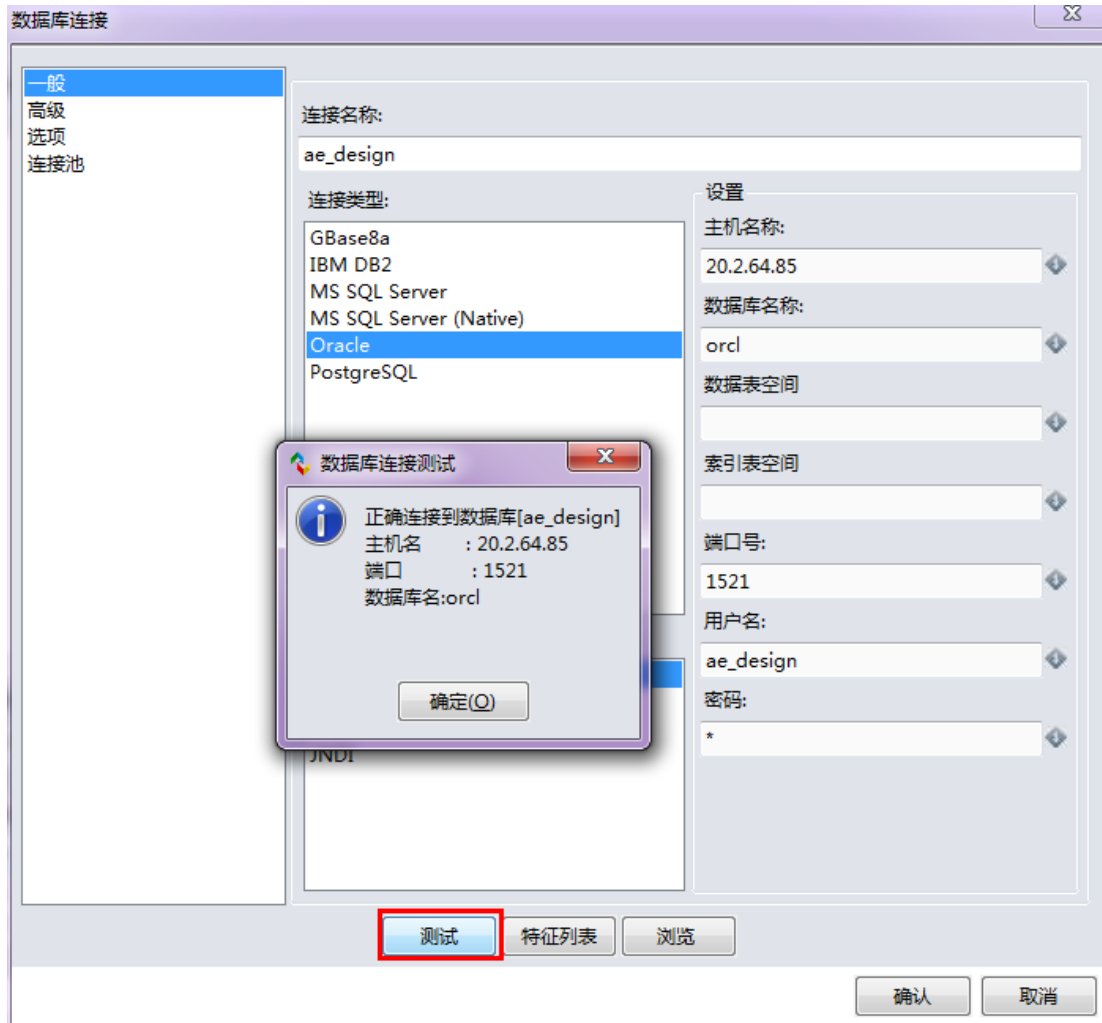




DI 目前支持的数据库，包括 GBase, DB2, SQL Server, Oracle, PostgreSQL, 连接方式由 JDBC, ODBC, OCI 和 JNDI 四种方式。在这里我们以 Oracle 为例，输入主机名称，数据库名称，端口号，用户名和密码，如下图所示：



配置好之后，可以测试刚才配置的数据库连接是否正确，点击<测试>按钮，弹出“连接测试”的对话框，证明我们刚才的数据库配置是正确的，如下图所示：



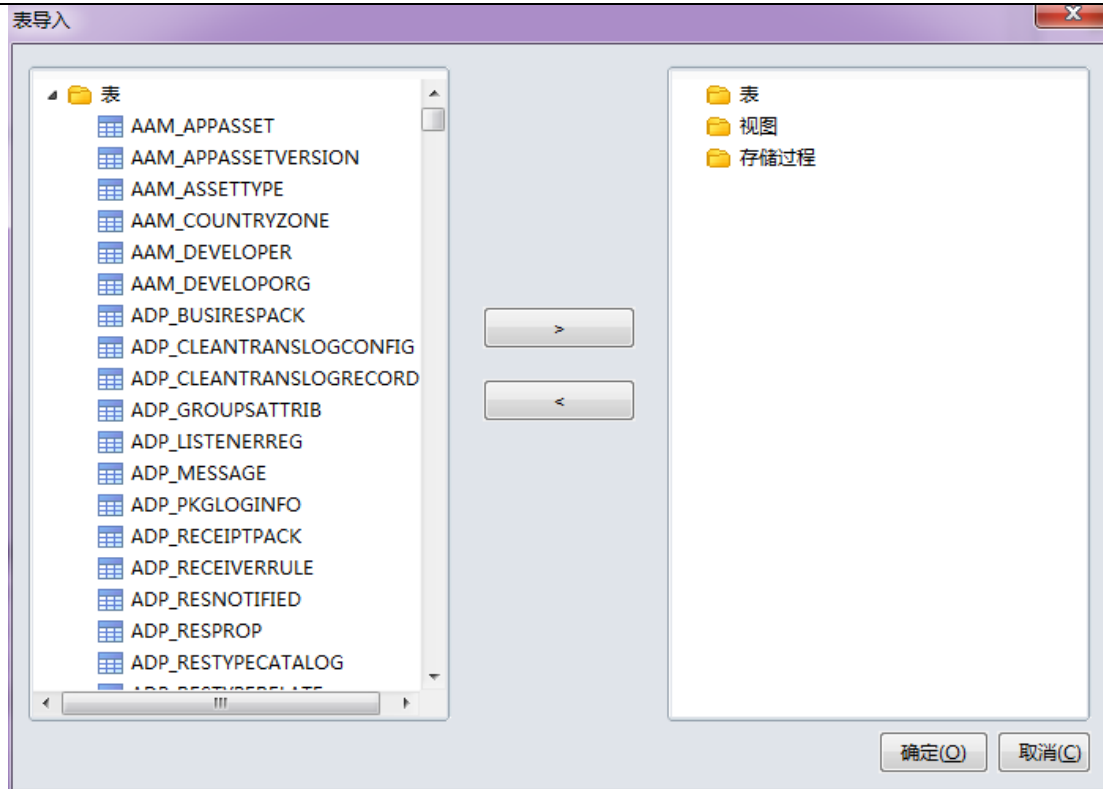
8.2. 如何编辑/删除数据库连接？

同样，通过在数据库连接上单击鼠标右键，在弹出的快捷菜单中，选择<编辑>和<删除>菜单项，可以对已经建立好的连接进行编辑或删除。


8.3. 如何导入数据库对象？

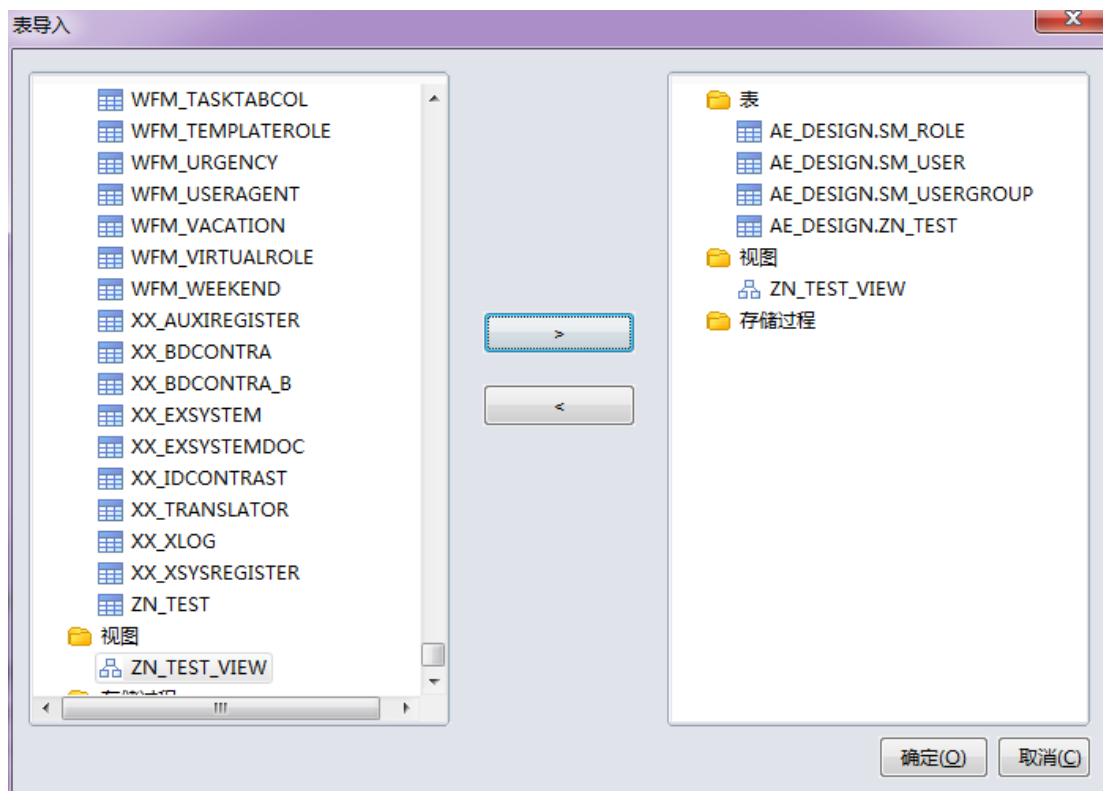
由于数据库连接下默认是不导入任何对象的，所以当在设计 ETL 流程时，需要使用到某个数据库连接下的某些表，视图或存储过程的时候，需要将这些对象手动的导入进来。

在数据库连接“ae_design”上单击鼠标右键，选择<导入>菜单项，弹出导入对话框，如下图所示：

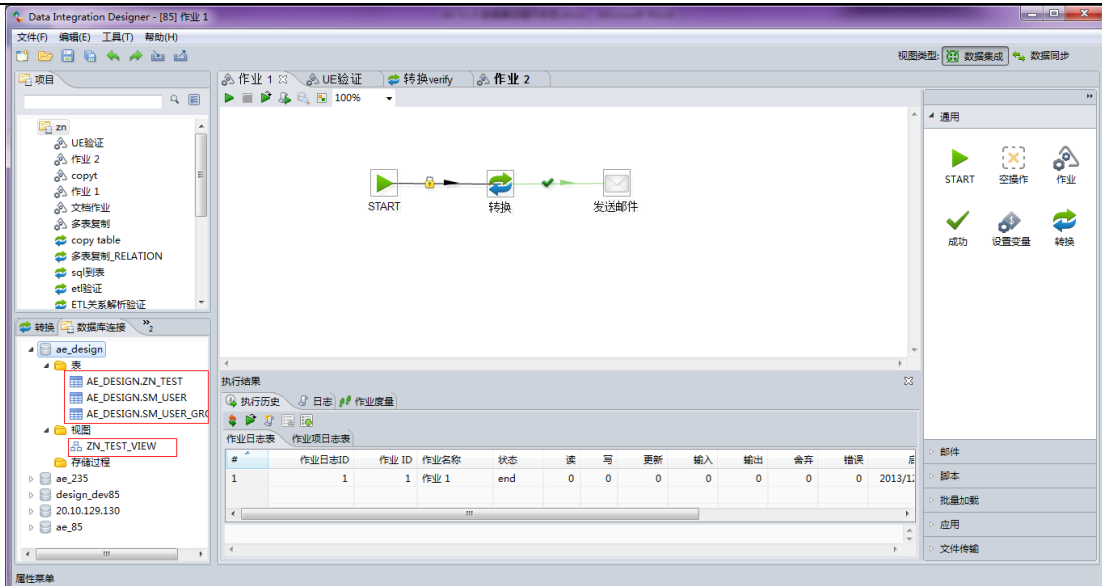


点击  按钮，可以将需要的表，视图或存储过程添加到我们的数据库连接中。

单击  按钮，可以讲已经添加的表，视图或存储过程，删除。



添加完毕之后，点击<确定>按钮将表加入到资源管理区，如下图所示，这些表，视图或存储过程就可以作为资源被作业或转换进行引用。引用方式和 AE 元数据相同，这里就不再举例说明。



9. 调度管理

通过调度管理我们可以启动和终止作业调度，可以配置调度策略，查看调度的执行历史和执行日志，对 DI 进行统一的后台管理。



数据处理平台-> 数据集成-> 调度管理，进入调度管理界面，此界面列出的是作业最新的执行状态，选中某个作业，单击右键，可以立即执行或是终止执行该作业，如下图所示：

序号	作业名称	项目	作业类别	最新执行时间	执行任务状态	用户	创建时间	修改时间
1	文档作业	zn	D作业				2013-12-03 15:51:50.0	2013-12-10 14:33:50.0
2	多	zn	D作业				2013-12-13 15:20:36.0	2013-12-17 10:09:28.0
3		zn	D作业	2013-12-19 00:00:02 578	✓	#UAP#	2013-11-27 09:40:24.0	2013-12-19 16:02:30.0
4		test	D作业				2013-12-03 16:14:42.0	2013-12-03 16:14:46.0
5		zn	D作业	2013-12-19 02:00:02 218	✗	#UAP#	2013-11-27 15:40:37.0	2013-12-19 16:18:55.0
6		test1	D作业				2013-12-03 16:17:57.0	2013-12-03 16:30:11.0
7		zn	D作业				2013-12-18 11:27:01.0	2013-12-18 11:30:23.0
8		zn	D作业				2013-12-02 15:30:49.0	2013-12-19 11:07:29.0
9	AAAA	heng	D作业				2013-12-13 12:36:48.0	2013-12-13 12:37:00.0
10	Job		OD作业	2013-11-06 16:52:34 953	✓	test1	2013-11-06 16:49:07	2013-11-06 16:49:07

9.1. 设置调度策略

选中某个作业，单击鼠标右键，选择<设置调度策略>菜单项，弹出调度策略配置对话框，在此可以完成该作业的调度策略配置，此配置和在 DI 客户端中，start 作业里配置调度策略的方法一致，并且两者的设置是保持一致的。用户可以根据自己的需要设置调度策略，按天、周、月执行，从 AE2.0.1 开始，在按月设置调度策略中增加了以第几周为单位的调度策略。如下图所示：

调度策略配置

发生频率

天
 周
 月

间隔：

星期一 星期二 星期三 星期四
 星期五 星期六 星期日

一天内(服务器时间)

发生一次,时间: : :

周期发生,频率: 小时 开始于: : :

终止于: : :

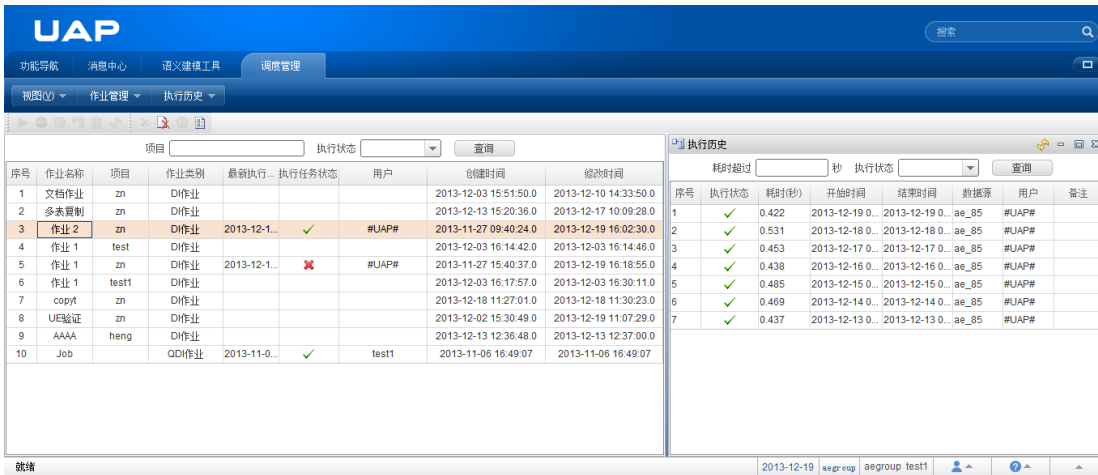
有效期(服务器时间)

开始日期: 终止日期:

清除 确定 取消

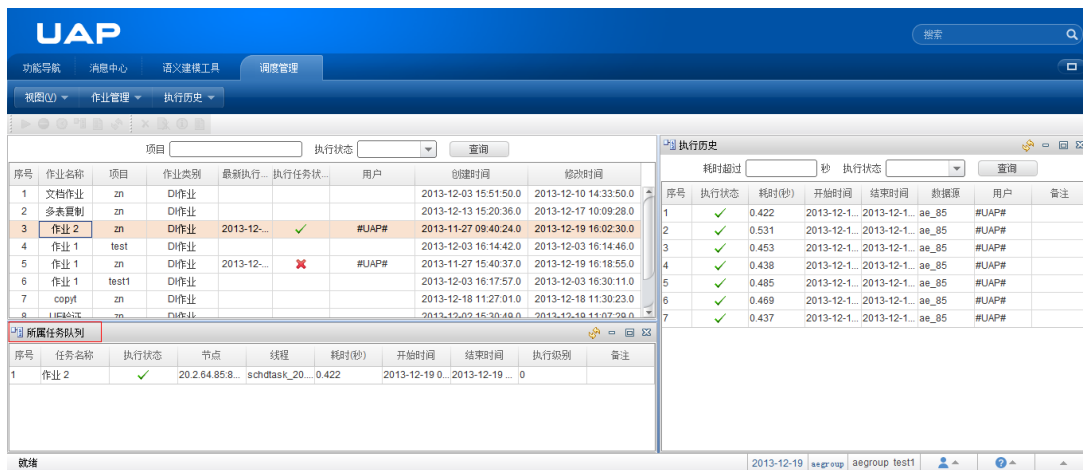
9.2. 查看执行历史

选中作业 2，单击鼠标右键，选择<查看执行历史>菜单项，在调度管理的右侧会出现执行历史的窗口，在此窗口可以查看到作业 2 所有的执行情况，包括每次执行的开始和结束时间，每次执行的状态和耗时等，如下图所示：



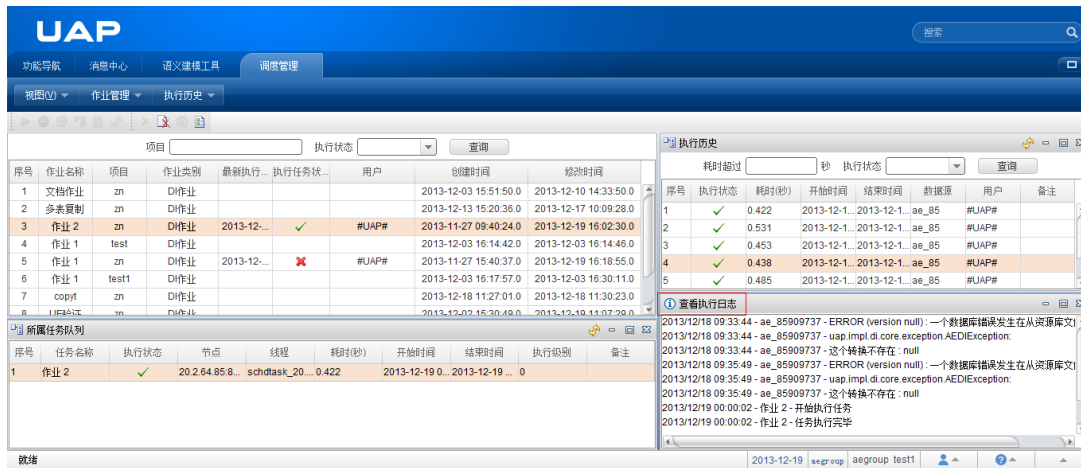
9.3. 查看所属任务

选中作业 2，单击鼠标右键，选择<查看所属任务>菜单项，或在执行历史窗口，单击右键，选择<查看所属任务>菜单项，在窗口的左下方会出现“所属任务”窗口，如果是 QDI 的作业，在此窗口会列出此作业下的所有任务，如下图所示：



9.4. 查看执行日志

在“所属任务队列”窗口，选中作业 2，单击鼠标右键，选择<查看执行日志>菜单项，在窗口的右下方会出现“查看执行日志”窗口，在此窗口会列出该作业最新执行的详细日志。如果在执行历史窗口，选中作业某次执行的历史，单击右键，选择<查看执行日志>菜单项，则在日志窗口显示的是，该次作业执行的日志，如下图所示：



附录

作业节点状态说明

节点连接状态	说明
	无条件执行下一个节点。即，不管上一个节点结果如何，都将执行下一个节点。
	置灰。不可用。
	当上一个节点成功时，执行下一个节点。
	置灰。不可用。
	当上一个节点失败时，执行下一个节点。
	置灰。不可用。