

自助ETL-应用示例

- 需求场景1：数据预处理
- 实现方案
 - 数据基础
 - 操作步骤
 - 数据源
 - 数据预处理
 - 目标源
- 需求场景二：增量抽取

自助ETL的主要应用场景是将多来源的异构数据，进行处理后得到具备完整性、一致性的数据模型。使用自助ETL可以提高数据的质量、适应不同的数据分析方法。

需求场景1：数据预处理

想要对产品的销售情况进行分析，数据源中有一张产品销售表，先根据需求对数据进行处理，然后进行数据分析。

实现方案

数据基础

数据源中“产品销售表”的数据预览如下:

三 导航 | 预览数据[产品销售表] x

表名称: 产品销售表

表别名: 产品销售表

产品类别	销售额	采购金额
点心	480.00	411.05
调味品	336.00	295.85
谷类/麦片	2128.00	1522.84
海鲜	372.60	248.92
日用品	1032.00	734.63
肉/家禽	2079.00	1264.54

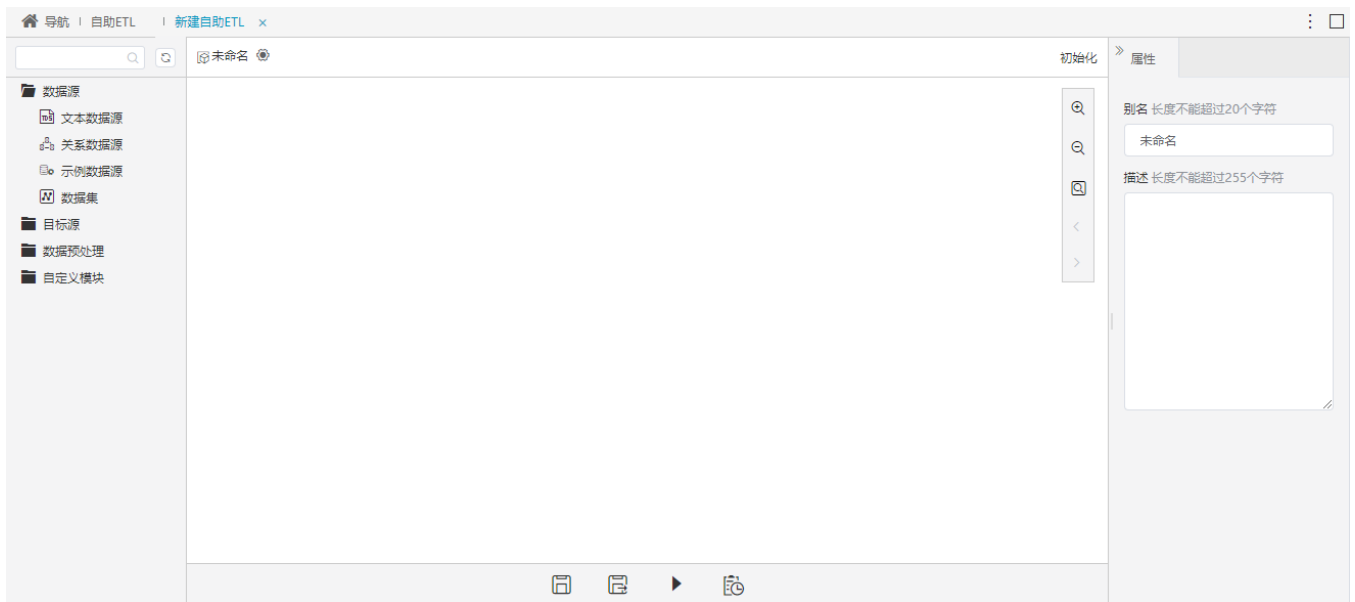
关闭(C)

操作步骤

1、在系统主界面的快捷菜单中选择 **数据准备** > **自助ETL**，进入“新建ETL”定制界面。

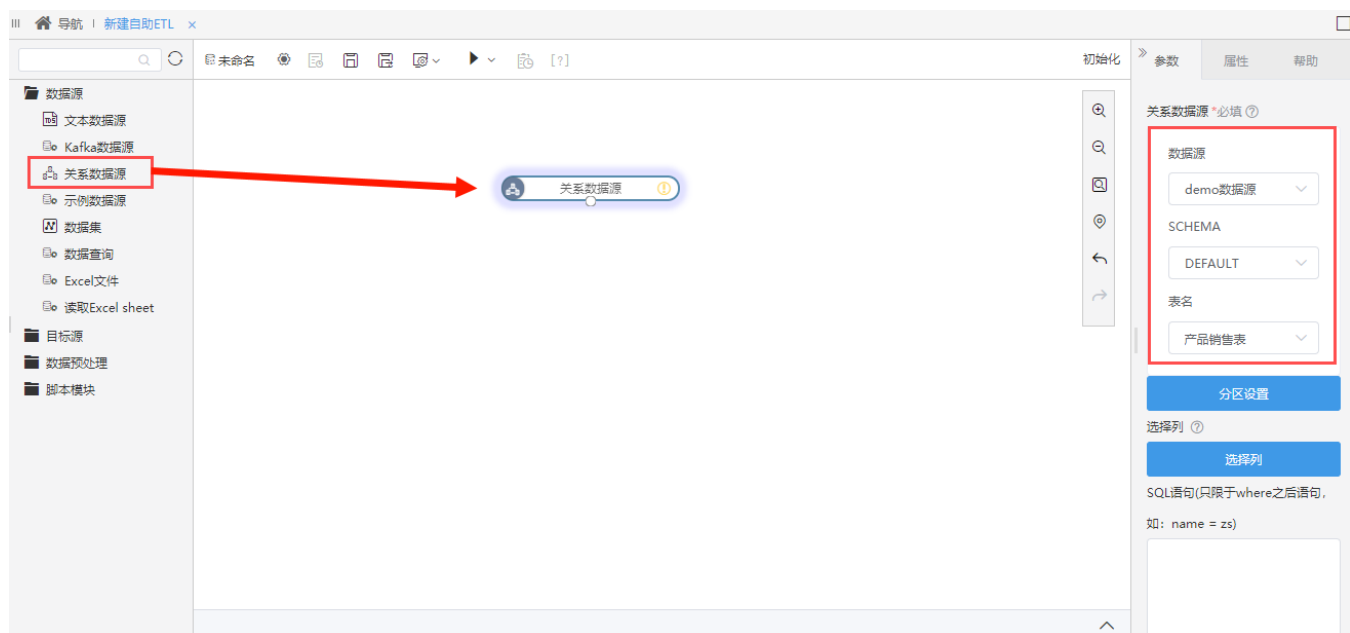


2、在此界面可以设计定制ETL workflow。

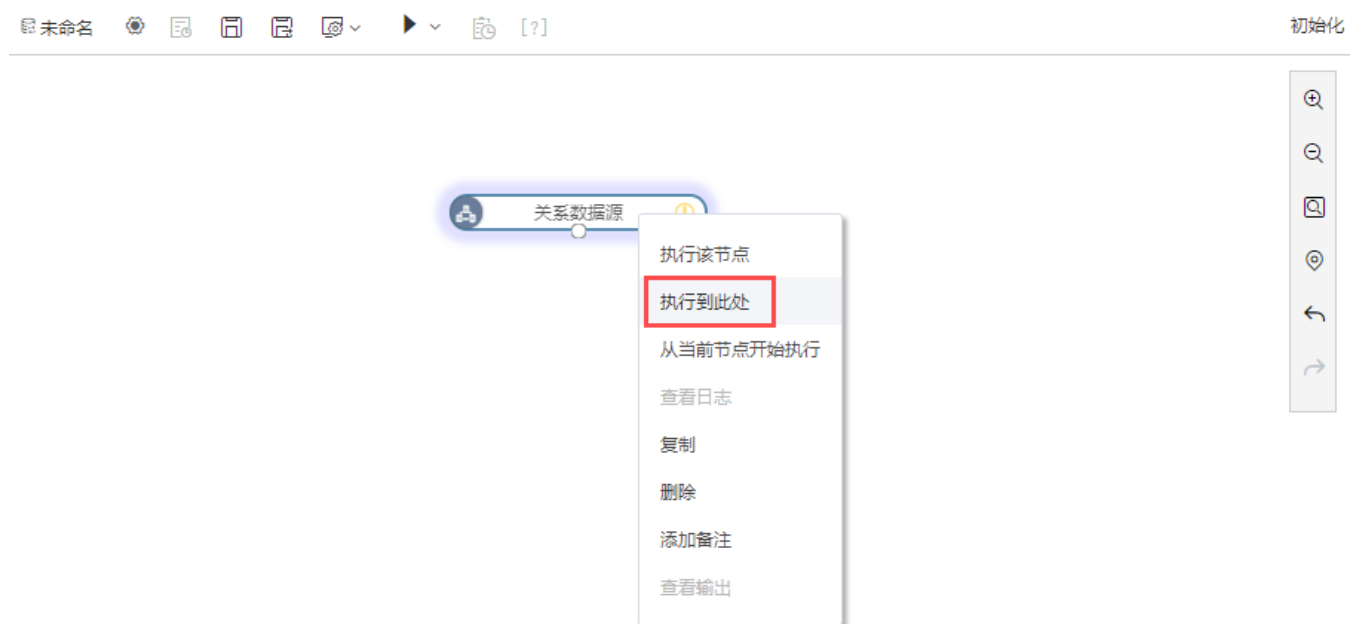


数据源

1、将关系数据源节点拖入画布区，在节点配置区设置关系数据源的参数。



2、在关系数据源上右键选择 **执行到此处**，运行当前节点。



3、点击查看输出，可查看关系数据源节点输出的数据，如下：

⌚ 当前显示 6 条 / 总共有 6 条数据 [提示:点击单元格可查看超出的内容](#)

产品类别	销售额	采购金额
点心	480.00	411.05
调味品	336.00	295.85
谷类/麦片	2128.00	1522.84
海鲜	372.60	248.92
日用品	1032.00	734.63
肉/家禽	2079.00	1264.54

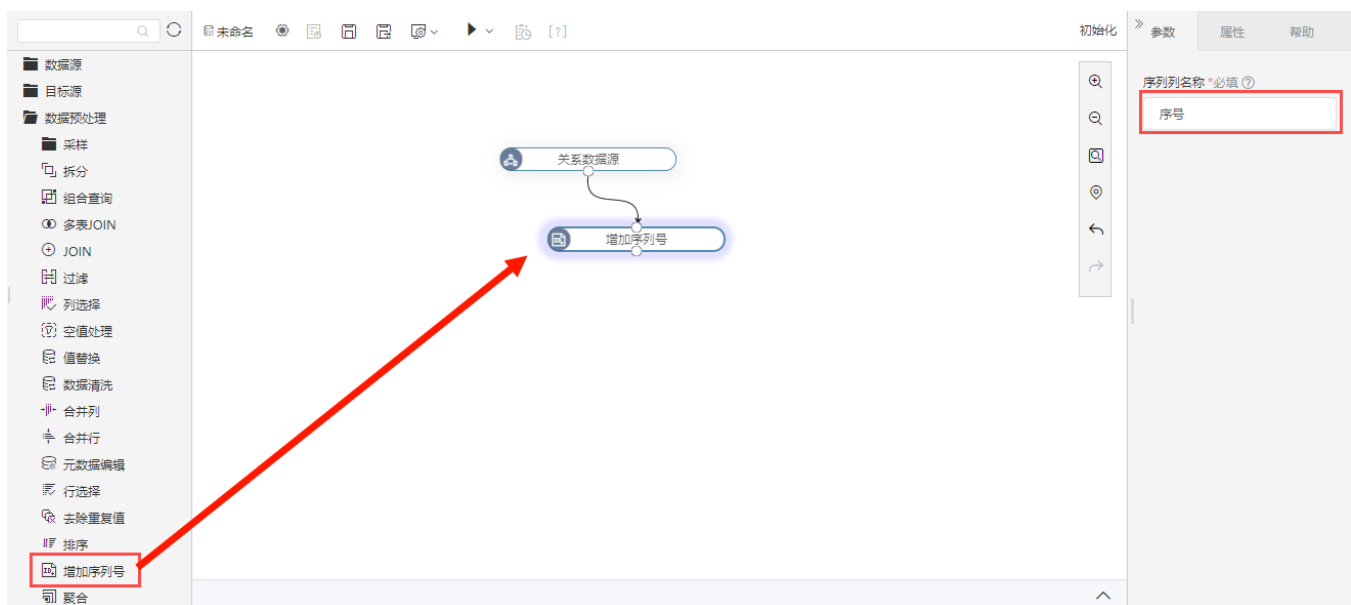
表头真名 ☒ 表头别名

将表的数据导入当前工作流完成，下面对数据进行 **数据预处理**。

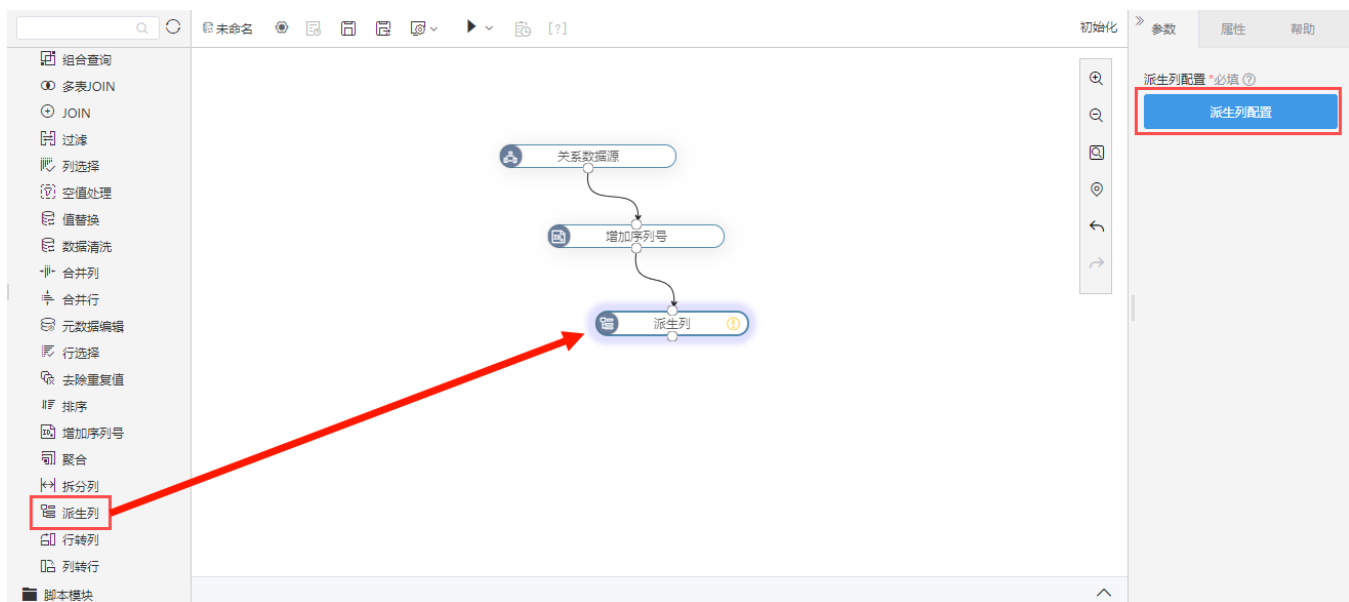
数据预处理

需要在表中添加序号，并去除采购金额字段的小数位。

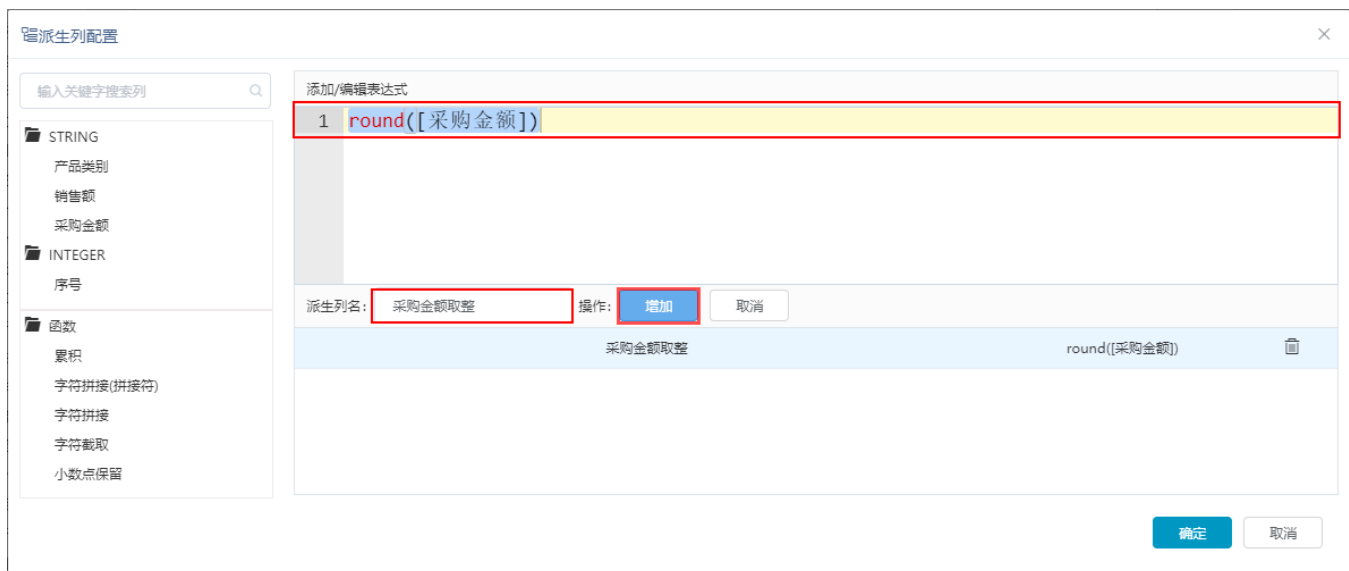
- 1、将增加序列号节点拖入画布区，连接两个节点，设置序列列名称。



- 2、将派生列节点拖入画布区，连接两个节点。在节点配置区选择 **派生列配置** 按钮。



3、将对应函数和字段分别拖入添加/编辑表达式区，点击 **确定** 按钮，添加一条“采购金额取整”的派生列。



4、在派生列节点上右键，选择 **执行到此处**，运行当前节点。查看派生列节点的输出数据如下：

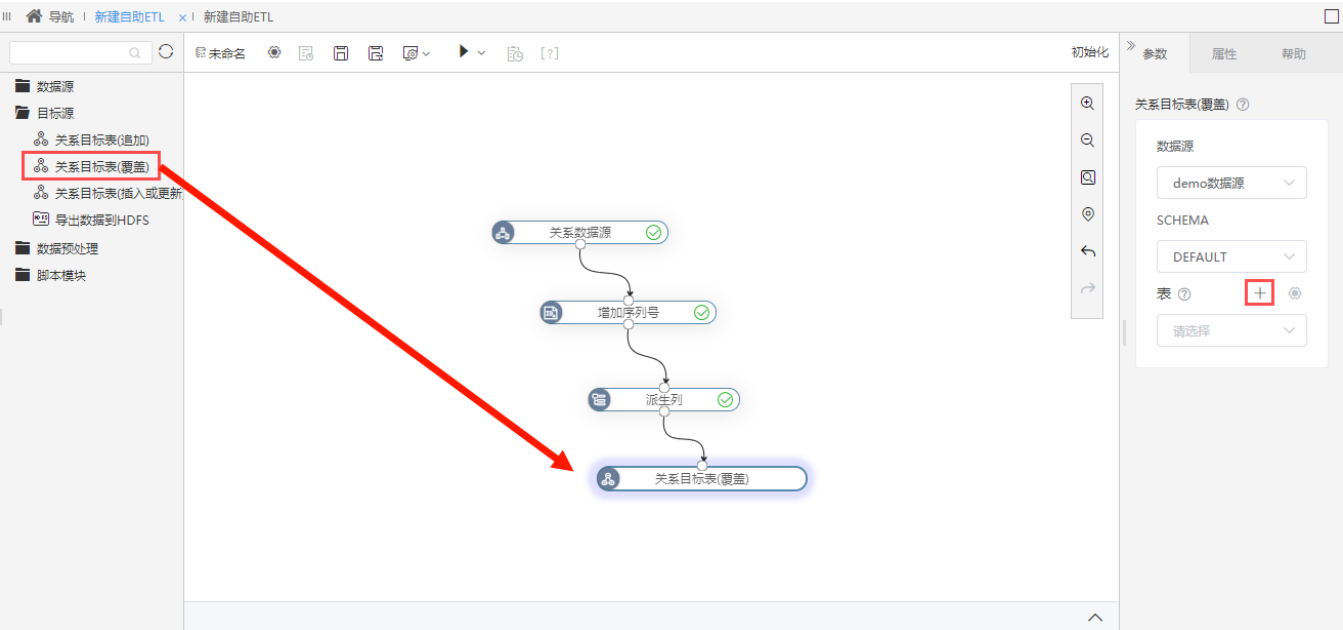
派生列				
当前显示 5 列 / 总共 5 列, 12 条 / 总共有 12 条数据				
列筛选 请选择				
表头真名 表头别名				
# 序号	A ₀ 产品类别	A ₀ 销售额	A ₀ 采购金额	# 采购金额取整
1	点心	480.00	411.05	411.0
2	调味品	336.00	295.85	296.0
3	谷类/麦片	2128.00	1522.84	1523.0
4	海鲜	372.60	248.92	249.0
5	日用品	1032.00	734.63	735.0
6	肉/家禽	2079.00	1264.54	1265.0

数据预处理完成，下面将处理好的数据输出。

目标源

将处理好的数据输出成一张新的表。

1、将关系目标源节点拖入画布区，选择输出表的数据源、SCHEMA，并选择 **新建表** 按钮。



2、新的表命名为“产品销售数据表”，点击 **创建表** 按钮。

新建表

<input checked="" type="checkbox"/>	名称	别名	源数据类型	目标数据类型	长度	精度	是否设置为主键
<input checked="" type="checkbox"/>	序号	序号	integer	INTEGER			
<input checked="" type="checkbox"/>	产品类别	产品类别	string	VARCHAR	255		
<input checked="" type="checkbox"/>	销售额	销售额	string	VARCHAR	255		
<input checked="" type="checkbox"/>	采购金额	采购金额	string	VARCHAR	255		
<input checked="" type="checkbox"/>	采购金额取整	采购金额取整	double	DOUBLE	22	6	

产品销售数据表

1~30个字符，可使用字母、数字、下划线，需以字母开头，下划线不能结尾。支持中文表名的数据库也可以使用中文作为表名

创建表

取消

3、点击 **运行** 按钮，创建一张新的表。

使用自助ETL处理数据完成，下面可对产品销售情况进行数据分析，详情请参考 [可视化与分析数据](#) 。

需求场景二：增量抽取

详情请参考：[数据挖掘-参数设置](#)