

自助ETL-应用示例

- 需求场景
- 实现方案
 - 数据基础
 - 操作步骤
- 数据源
 - 数据预处理
 - 目标源

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

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：

☰ 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

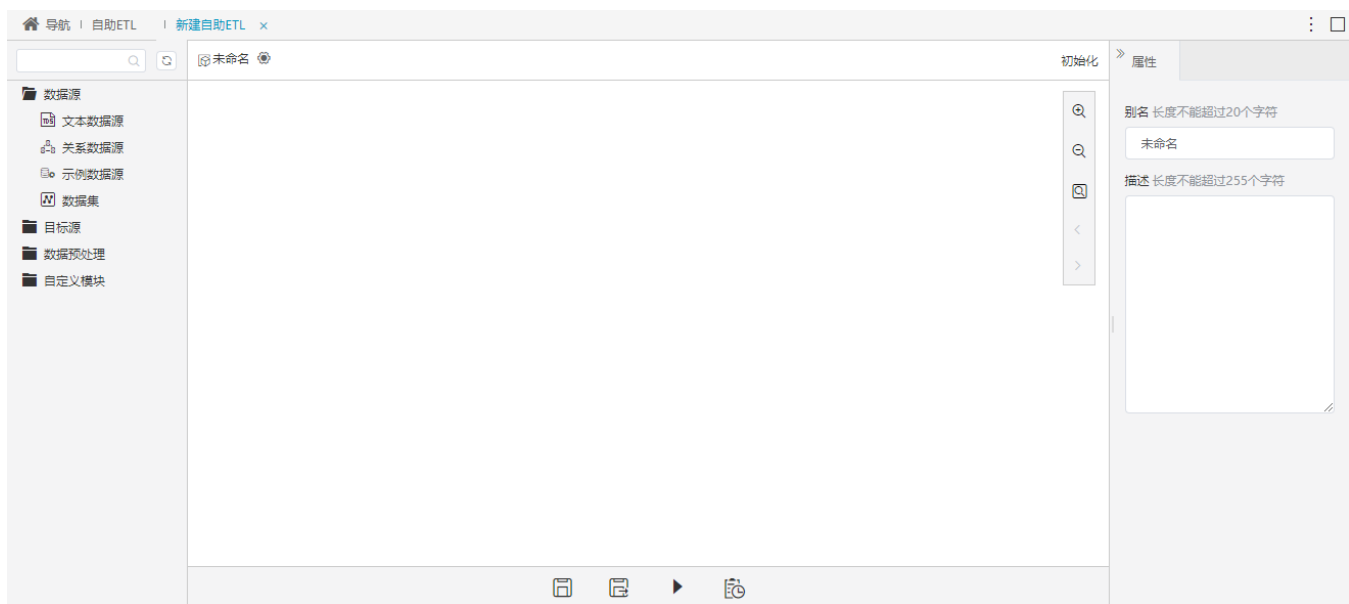
1、在系统主界面的快捷菜单中选择 **数据准备** > **自助ETL**，进入“自助ETL”管理界面：



2、在“自助ETL”管理界面中，选择 **新建自助ETL** 按钮，进入“新建ETL”定制界面。

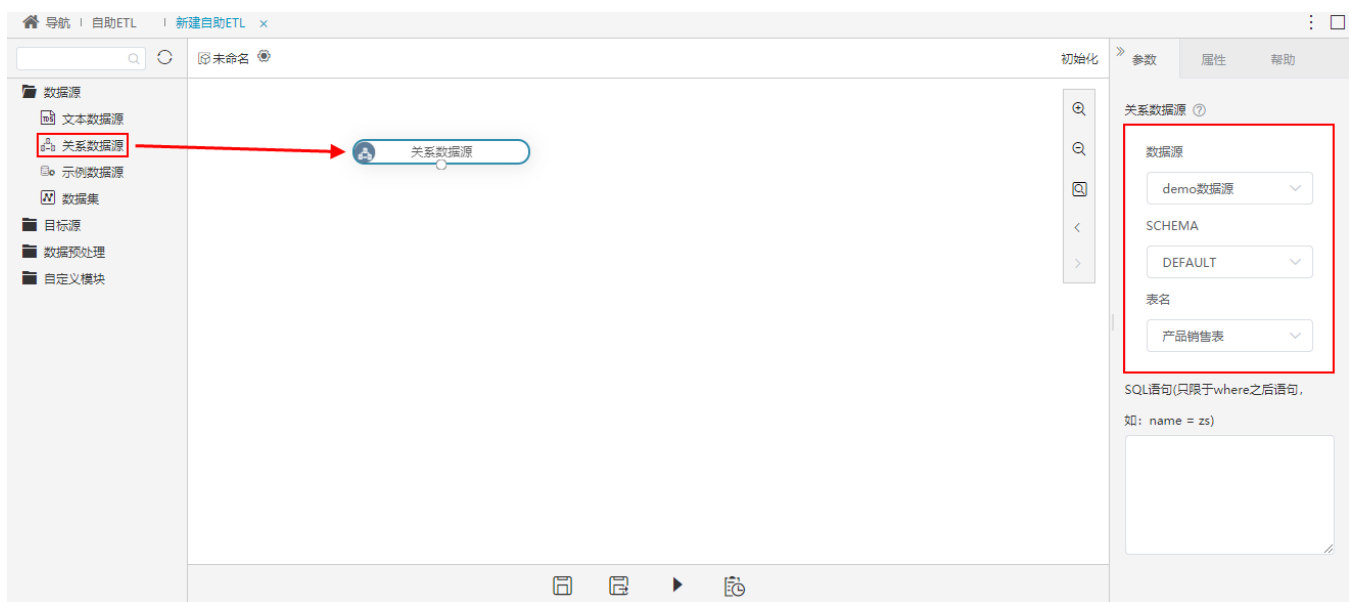


3、在此界面可以设计定制ETL workflow。

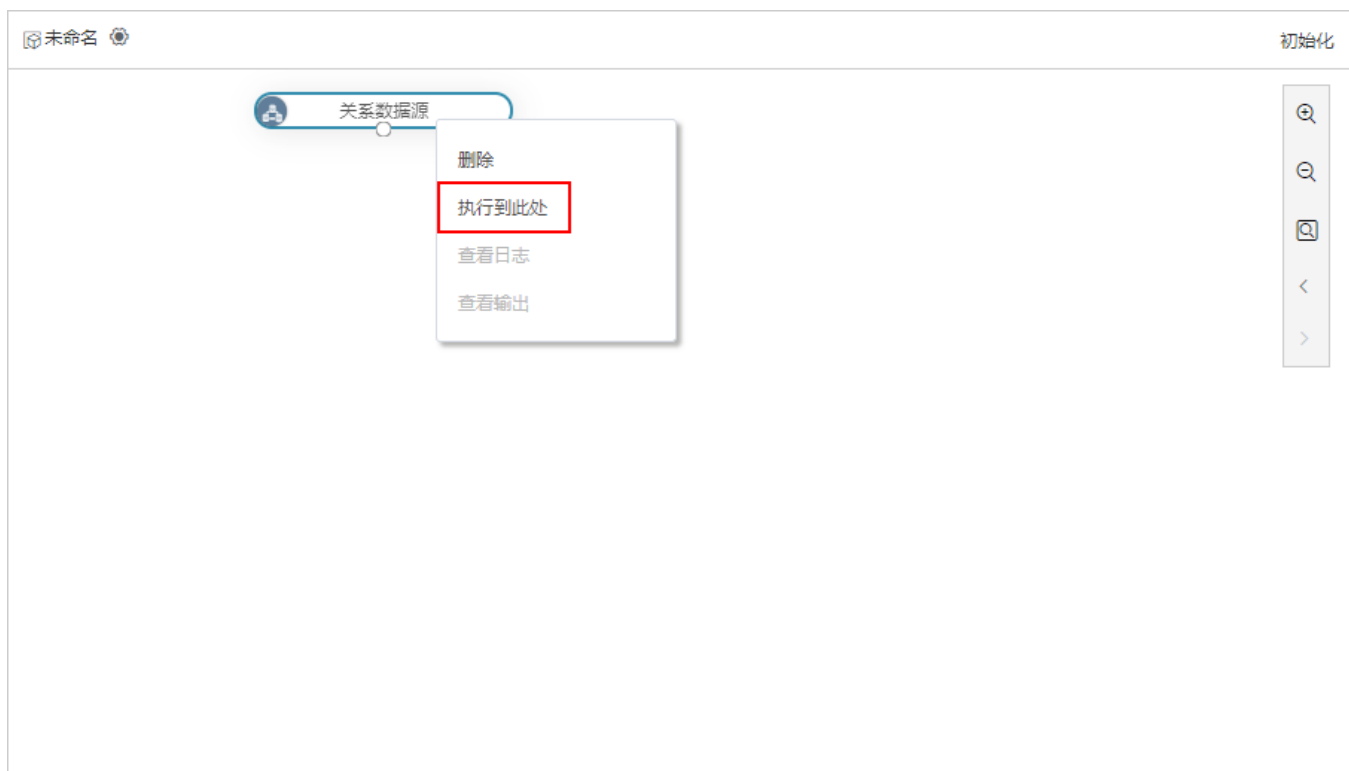


数据源

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 |

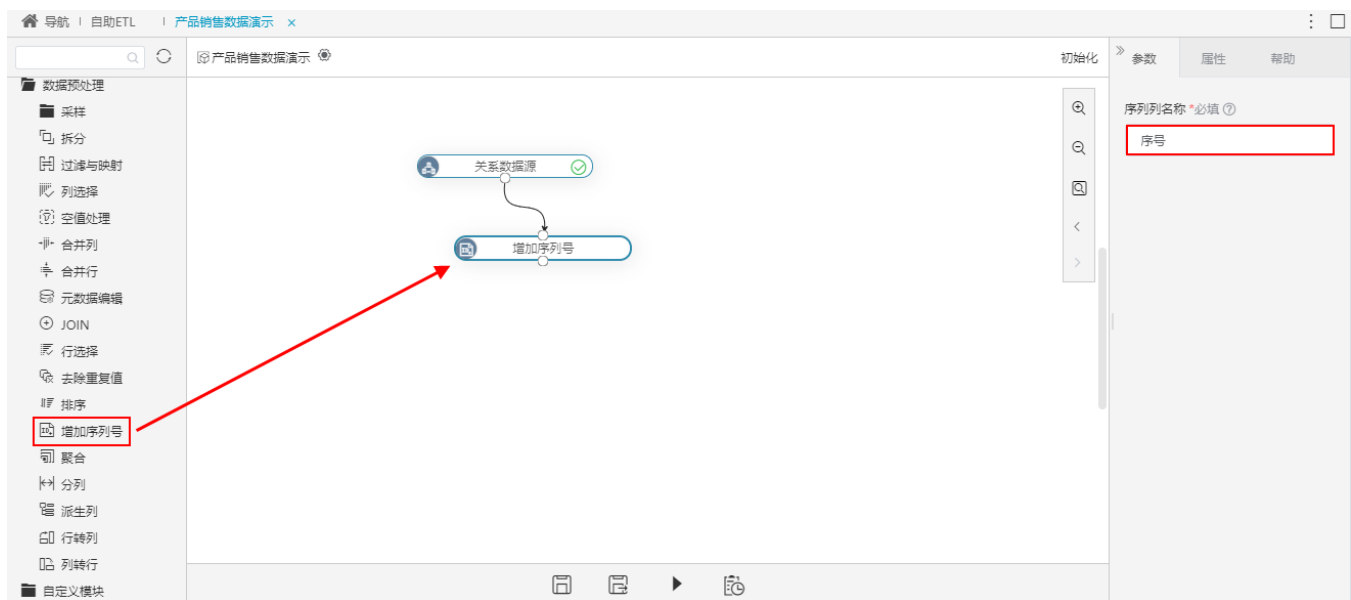
表头真名 ☒ 表头别名

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**。

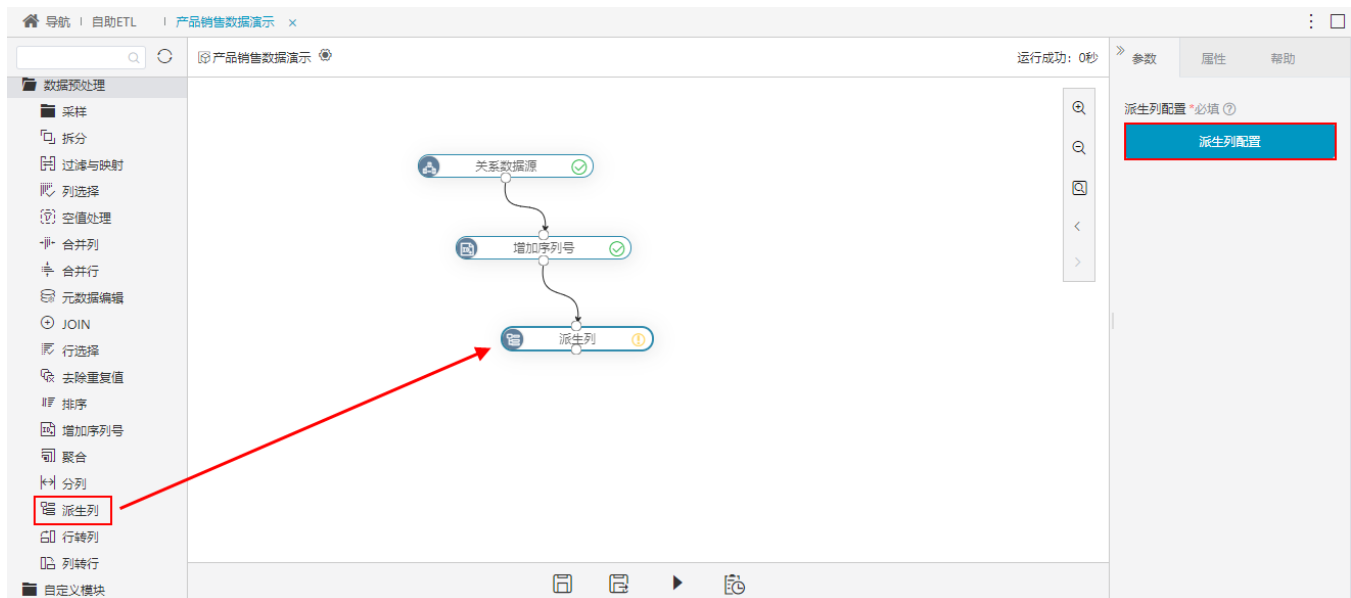
数据预处理

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。

派生列配置

输入关键字搜索列

STRING

产品类别

销售额

采购金额

INTEGER

序号

函数

累积

字符拼接(拼接符)

字符拼接

字符截取

小数点保留

添加/编辑表达式

1 round([采购金额])

派生列名: 采购金额取整 操作: 确定 取消

采购金额取整 round([采购金额])

确定 取消

4、在派生列节点上右键，选择 **执行到此处** ，运行当前节点。派生列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

| 序号 | 产品类别 | 销售额 | 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 |

增加序列号

派生列

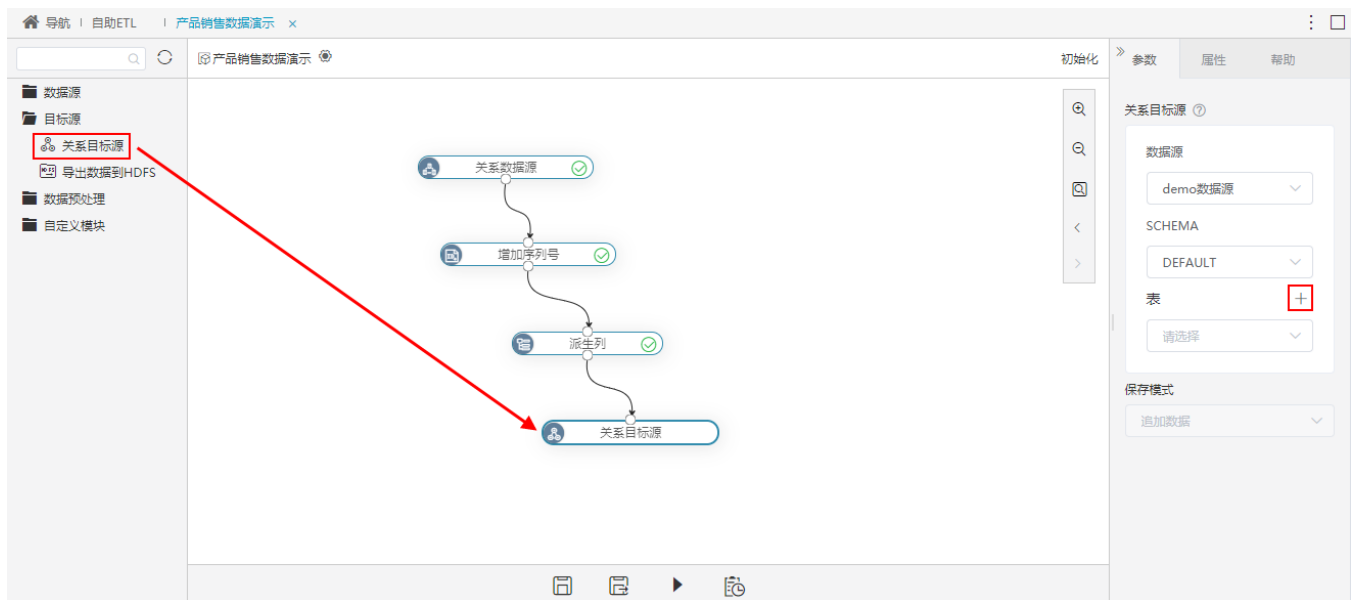
表头真名 表头别名

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、将关系目标源家待拖入画布区，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 |
|-------------------------------------|--------|--------|---------|---------|-----|----|
| <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 | 10 | 4 |

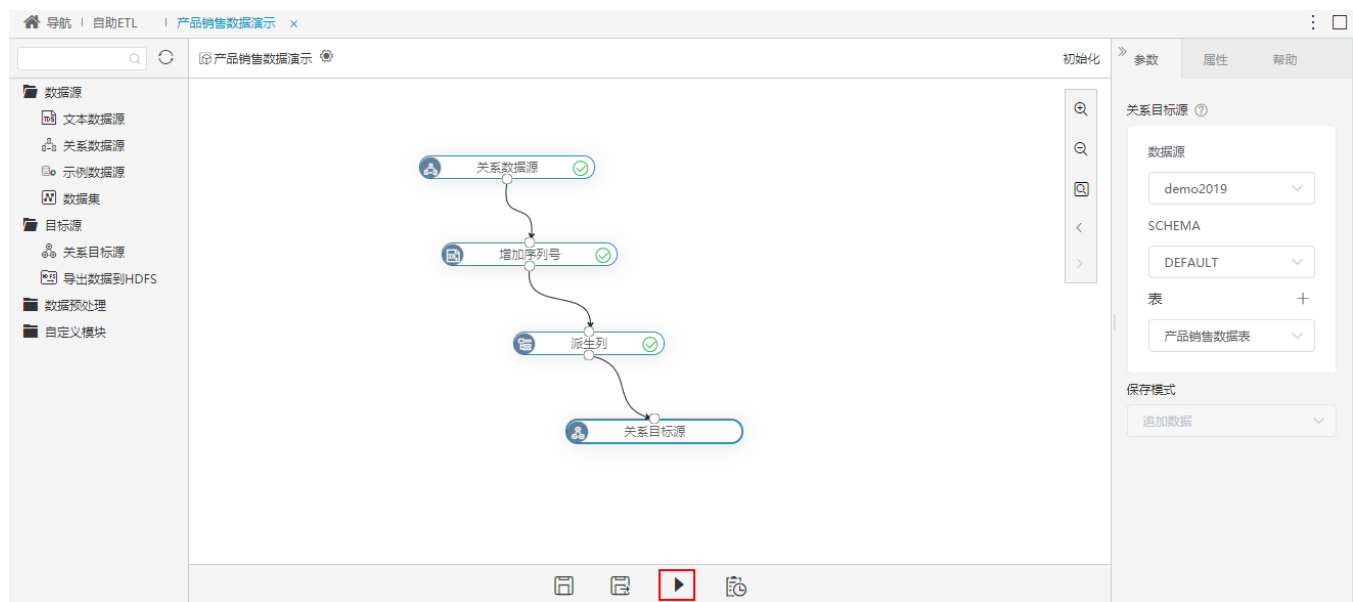
* 产品销售数据表

1~20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

创建表

取消

3、点击 **运行** 按钮，创建一张新的表。



流程运行完成后，关系目标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

| 序号 | 产品类别 | 销售额 | 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 |

表头真名 ☒ 表头别名

在数据源中查看刚才创建的“产品销售数据表”表，预览表中的数据如图：

导航 | 预览数据[产品销售数据表] x

表名称: 产品销售数据表

表别名: 产品销售数据表

| 序号 | 产品类别 | 销售额 | 采购金额 | 采购金额取整 |
|----|-------|---------|---------|----------|
| 1 | 点心 | 480.00 | 411.05 | 411.00 |
| 2 | 调味品 | 336.00 | 295.85 | 296.00 |
| 3 | 谷类/麦片 | 2128.00 | 1522.84 | 1,523.00 |
| 4 | 海鲜 | 372.60 | 248.92 | 249.00 |
| 5 | 日用品 | 1032.00 | 734.63 | 735.00 |
| 6 | 肉/家禽 | 2079.00 | 1264.54 | 1,265.00 |

关闭(C)

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 [分析展现](#) 。