

# 快速创建自助数据集

## 概述说明

自助数据集通过简单拖拉操作对原始数据进行处理，支持跨数据库查询、数据集嵌套，支持关联、筛选、分组、建立层次、格式处理、逻辑计算、数据抽取等操作。

- 支持表拖拽，自动或手动创建表关系进行表关联查询。
- 支持添加筛选器，过滤数据结果。
- 不需要熟悉技术语言，用户可以很轻松地通过拖拽字段的方式创建具有复杂计算公式的业务逻辑计算字段。
- 系统支持将有时间信息和地理信息的字段创建成时间维或者地理维，便于对数据在时间或地理上的不同维度分析。
- 可以随时调整修改表或字段的分组、别名，设置表或字段的可见性。
- 通过数据行权限控制，保证所有用户抽取到其权限内的数据。

自助数据集可以被自助仪表盘、透视分析和电子表格使用。

## 流程定义

业务示例中的销售情况仪表盘，我们将基于“DEMO”数据源来实现。在定制该示例之前，我们先要对要求具备的数据和已有数据进行对比，得出如下结论：

- 1、“DEMO”数据源中没有年、月维度，但“订单表”中有“订单日期”字段，我们可以将该字段分解成“年、季、月、日”四层时间维。
- 2、“DEMO”数据源中没有“销售额”数据，但我们可以通过“单价\*数据\*(1-折扣)”的公式定义“销售额”计算字段。
- 3、地区的销售额分布和下钻，需要将表示区域的字段定义地理维实现。

通过以上分析，我们将按照如下流程定义符合要求的自助数据集：



## 详细步骤

### 1、新建自助数据集

1) 在定制管理界面选择快捷菜单 **自助数据集**（入口1），或在“资源目录区”的“资源定制”节点或其子孙目录节点的右键菜单中选择 **新建查询** > **自助数据集**（入口2）。

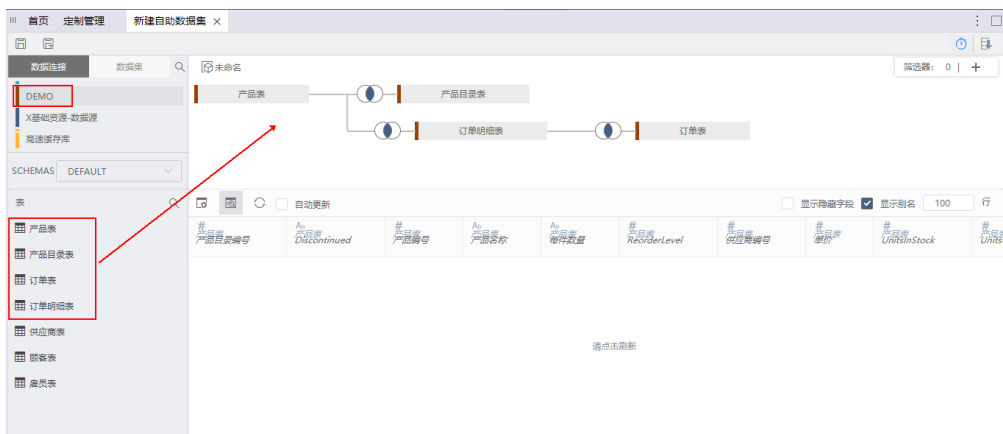


2) 进入“新建自助数据集”界面，如下图所示：

- 概述说明
- 流程定义
- 详细步骤
  - 1、新建自助数据集
  - 2、创建时间维
  - 3、创建地区维
  - 4、字段分类及调整
  - 5、创建计算字段
  - 6、保存自助数据集
  - 7、数据抽取
- 示例视频



3) 选择之前创建的数据源“DEMO”作为数据来源，从该数据源中依次拖拽“产品表”、“产品目录表”、“订单明细表”、“订单表”到表集合区。如下图所示：



4) 单击 **表属性** 按钮，显示表属性列表：



## 2、创建时间维

由于数据源“DEMO”中没有“年、季、月、日”的时间维层次，因此我们需要分解“订单表”中的“订单日期”字段进行时间维层次定义。

1) 在表属性列表的维度中找到“订单表”中的“订单日期”字段，在其右键菜单中选择 **创建时间维度**：

名称	别名	数据类型	数据格式	可见性	元数据
As CustomerID	顾客编号	STRING	默认值	👁	DEMO.default.default.orders.CustomerID
# EmployeeID	雇员编号	INTEGER	默认值	👁	DEMO.default.default.orders.EmployeeID
📅 OrderDate	订单日期	DATE	默认值	👁	DEMO.default.default.orders.OrderDate
📅 RequiredDate	交货日期	DATE	默认值	👁	DEMO.default.default.orders.RequiredDate
📅 ShippedDate	发货日期	DATE	默认值	👁	DEMO.default.default.orders.ShippedDate
As ShipName	ShipName	STRING	默认值	👁	DEMO.default.default.orders.ShipName
As ShipAddress	ShipAddress	STRING	默认值	👁	DEMO.default.default.orders.ShipAddress
As ShipCity	ShipCity	STRING	默认值	👁	DEMO.default.default.orders.ShipCity
As ShipProvince	ShipProvince	STRING	默认值	👁	DEMO.default.default.orders.ShipProvince

2) 弹出“创建时间维度”对话框，选择“年季月日”层次：

创建时间维度

☐ 年月日
 ☐ 年月
 ☐ 年季月
 ☒ 年季月日

取消

确定

3) 单击 **确定** 按钮，列表中自动生成时间维层次，单击时间维层次目录呈编辑状态，修改目录别名为“时间”：

名称	别名	数据类型	数据格式	可见性	元数据
📅 时间	时间	DATE	默认值	👁	DEMO.default.default.orders.OrderDate
# OrderDate_Year	年	INTEGER	默认值	👁	DEMO.default.default.orders.OrderDate_Year
# OrderDate_Quarter	季	INTEGER	默认值	👁	DEMO.default.default.orders.OrderDate_Quarter
# OrderDate_Month	月	INTEGER	默认值	👁	DEMO.default.default.orders.OrderDate_Month
# OrderDate_Day	日	INTEGER	默认值	👁	DEMO.default.default.orders.OrderDate_Day
📦 products	产品表			👁	DEMO.default.default.products
📦 categories	产品目录表			👁	DEMO.default.default.categories

### 3、创建地区维

1) 在表属性列表空白区的右键菜单中选择 **新建层次**：

名称	别名	数据类型	数据格式	可见性	元数据
As ShipAddress	ShipAddress	STRING	默认值	👁	DEMO.default.default.orders.ShipAddress
As ShipCity	ShipCity	STRING	默认值	👁	DEMO.default.default.orders.ShipCity
As ShipProvince	ShipProvince	STRING	默认值	👁	DEMO.default.default.orders.ShipProvince
As ShipRegion	ShipRegion	STRING	默认值	👁	DEMO.default.default.orders.ShipRegion
As ShipPostalCode	ShipPostalCode	STRING	默认值	👁	DEMO.default.default.orders.ShipPostalCode
As ShipCountry	ShipCountry	STRING	默认值	👁	DEMO.default.default.orders.ShipCountry
📦 products	产品表			👁	DEMO.default.default.products

2) 弹出“新建层次”对话框，输入层次名称为“地区”：

新建层次

\* 名称:

地区

取消

确定

3) 单击 **确定** 按钮，列表区增加“地区”层次目录，将“订单表”中的字段“ShipRegion”、“ShipProvince”、“ShipCity”三个字拖拽到该维层次目录下，并且分别修改别名为“发货区域”、“发货省份”、“发货城市”：

名称	别名	数据类型	数据格式	可见性	元数据
维度					
custom1535197585557	地区				
A ShipRegion	发货区域	STRING	默认值		DEMO.default.default.orders.ShipRegion
A ShipProvince	发货省份	STRING	默认值		DEMO.default.default.orders.ShipProvince
A ShipCity	发货城市	STRING	默认值		DEMO.default.default.orders.ShipCity
custom-OrderDate	时间				
# OrderDate_Year	年	INTEGER	默认值		
# OrderDate_Quarter	季	INTEGER	默认值		

4) 分别在字段“发货城市”、“发货省份”、“发货区域”的右键菜单中选择 **标记地理维度**。

顺序如下：“发货区域” > “发货省份” > “发货城市”。

名称	别名	数据类型	数据格式	可见性	元数据
维度					
custom1535197585557	地区				
ShipRegion	发货区域	STRING	默认值		DEMO.default.default.orders.ShipRegion
ShipProvince		STRING	默认值		DEMO.default.default.orders.ShipProvince
A ShipCity		STRING	默认值		DEMO.default.default.orders.ShipCity
custom-OrderDate					
# OrderDate_Year		INTEGER	默认值		
# OrderDate_Quarter		INTEGER	默认值		

4、字段分类及调整

- 1) 把“产品目录表”中的字段“产品目录名称”拖拽到“产品表”下。
- 2) 把“订单表”别名改成“发货”。
- 3) 在维度中隐藏部分不需要的表和字段：

隐藏的表：“产品目录表”、“订单明细表”；

隐藏的字段：“产品表”中的“产品编号”、“产品目录编号”、“供应商编号”、“QuantityPerUnit”、“Discontinued”；

“订单表”（即“发货”）中的“订单编号”、“顾客编号”、“雇员编号”；

名称	别名	数据类型	数据格式	可见性	元数据
products	产品表				
A CategoryName	产品目录名称	STRING	默认值		DEMO.default.default.categories.Categor...
# ProductID	产品编号	INTEGER	默认值		DEMO.default.default.products.ProductID
A ProductName	产品名称	STRING	默认值		DEMO.default.default.products.ProductN...
# SupplierID	供应商编号	INTEGER	默认值		DEMO.default.default.products.SupplierID
# CategoryID	产品目录编号	INTEGER	默认值		DEMO.default.default.products.CategoryID
A QuantityPerUnit	每件数量	STRING	默认值		DEMO.default.default.products.QuantityP...
A Discontinued	Discontinued	UNKNOWN	默认值		DEMO.default.default.products.Discontin...

一旦设置“表”不可见则该表下的所有字段都默认隐藏。

4) 在度量中把“发货”目录下的“运费”字段调整拖拽到“订单明细表”下。

名称	别名	数据类型	数据格式	可见性	元数据
度量					
products	产品表				
orderdetails	订单明细表				
# UnitPrice2	单价	DOUBLE	默认值		DEMO.default.default.orderdetails.UnitPri...
# Quantity	数量	INTEGER	默认值		DEMO.default.default.orderdetails.Quantity
# Discount	折扣	DOUBLE	默认值		DEMO.default.default.orderdetails.Discount
# Freight	运费	DOUBLE	默认值		DEMO.default.default.orders.Freight
orders	发货				

5) 把度量下的“发货”和“产品表”目录隐藏。

5、创建计算字段

1) 在度量列表中找到“订单表”中的“数量”字段，在其右键菜单中选择 新建计算字段：

名称	别名	数据类型	数据格式	可见性	元数据
度量					
products	产品表				
orderdetails	订单明细表				
# UnitPrice2	单价	DOUBLE	默认值		DEMO.default.default.orderdetails.UnitPri...
# Quantity	数量		默认值		DEMO.default.default.orderdetails.Quantity
# Discount	折扣		默认值		DEMO.default.default.orderdetails.Discount
# Freight	运费		默认值		DEMO.default.default.orders.Freight
orders	发货				

2) 弹出“新建计算字段”对话框，进行如下设置：

名称 销售额

数据类型 DOUBLE

字段

1 [数量]\*[单价]\*(1-[折扣])

表达式中的字段可以拖拽也可以双击选择

字符串

lower

max

min

upper

fx 的数描述

说明：

示例：

取消

确定

注：表达式必须在英文输入法状态下定义。

3) 在度量中将计算字段“销售额”拖拽到“订单明细表”下，隐藏“计算字段”目录。

名称	别名	数据类型	数据格式	可见性	元数据
orders	发货				
度量					
custom-calcFiled_measure	计算字段(度量)				
products	产品表				
orderdetails	订单明细表				
# custom1535192025739	销售额	DOUBLE	默认值		
# Quantity	数量	INTEGER	默认值		DEMO.default.default.orderdetails.Quantity
# UnitPrice2	单价	DOUBLE	默认值		DEMO.default.default.orderdetails.UnitPri...

6、保存自助数据集

保存自助数据集到指定目录下，命名为“自助数据集-快速入门”。

7、数据抽取

数据抽取是根据当前自助数据集的数据结果从原始数据库将数据抽取到系统内置的高速缓存数据库中，以保证秒级获取大级别量的数据结果，提高查询速度。

1) 在自助数据集编辑界面，单击工具栏中的 **抽取** 按钮：



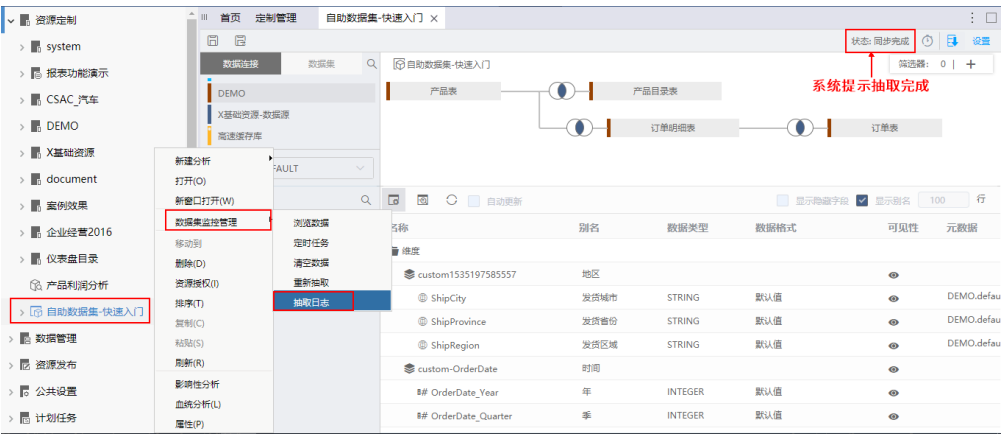
2) 显示数据抽取设置按钮，如下图所示：



3) 单击该 **设置** 按钮，弹出“抽取设置”窗口，保持默认设置，单击 **立即抽取** 按钮，再单击 **确定** 按钮，系统执行数据抽取操作。



4) 系统提示数据抽取完成后，我们在“资源目录区”找到该自助数据集，在其右键菜单中选择 **数据集监控管理** > **抽取日志**：



5) 在抽取日志界面查看数据是否抽取成功。

开始时间	结束时间	日志	抽取时间	类型	状态	抽取条数	抽取失败信息
2018-08-26 17:31:24	2018-08-26 17:31:28	数据抽取成功	4秒	全量	同步完成	2154	---
2018-08-26 17:31:24	2018-08-26 17:31:30	数据抽取成功	6秒	全量	同步完成	2154	---

抽取日志界面可以查看抽取的时间、抽取数量等详情，若抽取失败还可以查看错误日志。

## 示例视频

视频链接地址：[自助数据集](#)

视频资源与文档资源对应关系说明：

视频资源	文档资源
数据源“培训演示”	数据源“DEMO”
产品分类	产品目录表
产品	产品表
订单明细	订单明细表
订单	订单表