

自助数据集-定义流程

- 新建自助数据集
- 选择数据来源
- 定义表集合
- 规范字段
- 保存自助数据集
- 数据抽取


自助数据集是一类基于个性化需求的数据集，它面向各阶层用户提供数据查询和抽取服务。

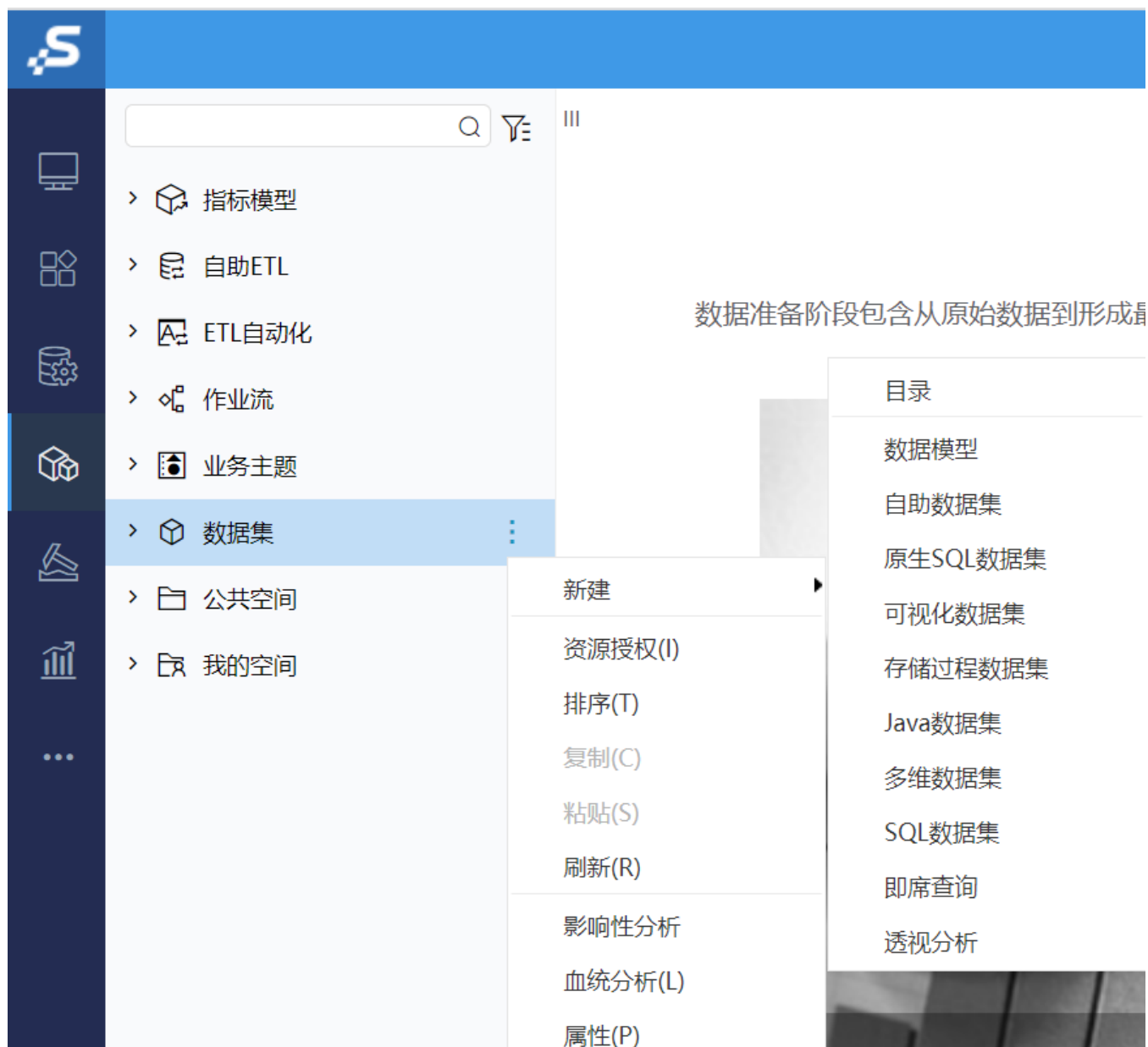
用户可以根据业务要求，通过可视化的方式，使用跨数据库、多表关联、数据转换、复杂逻辑关系计算、数据抽取ETL等功能，将数据提取到自助数据集，这些自助数据集允许被自助仪表盘、透视分析和电子表格进行数据分析应用。

自助数据集的定义流程如下：



新建自助数据集

在系统导航栏中选择 **数据准备**，在数据集的更多  中选择 **新建>自助数据集**。



关于自助数据集的创建入口、资源显示和使用，统一通过V10.5 license控制。如希望在V10.5版本中继续使用自助数据集，则需更新授权包含自助数据集的license文件。

选择数据来源

自助数据集的数据来源可以来自数据源，也可以是数据集。

来自数据源	来自数据集
-------	-------



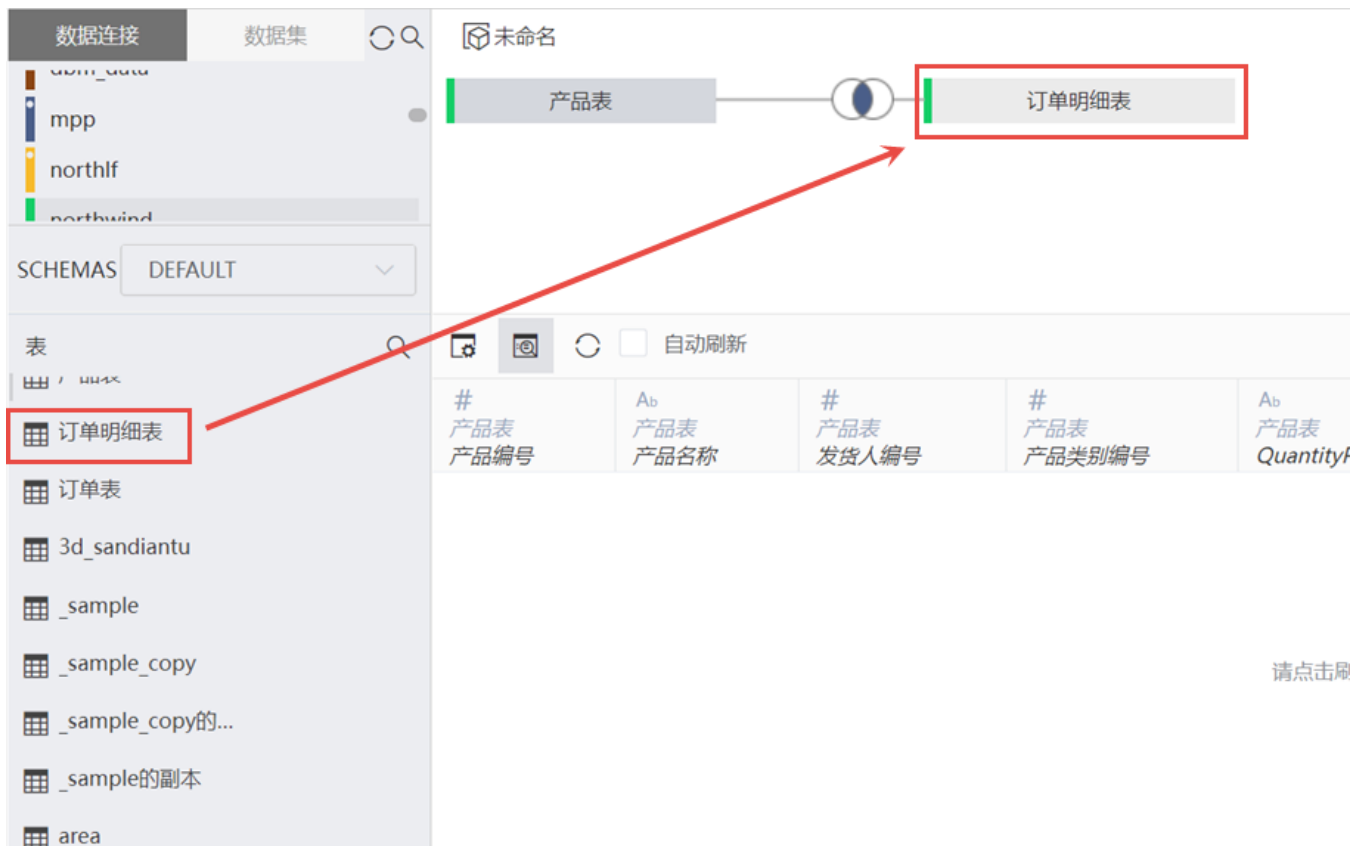
详情请参考 [自助数据集-数据来源](#)

定义表集合

定义表集合包含两部分内容：

- 从数据源中拖拽表或者拖拽数据集到“表集合区”。
- 在“表集合区”定义表关系，系统会根据表间相同别名的字段创建表关系，若无法自动创建需要用户手动创建表关系。

“表集合区”中的表从“数据区”中拖拽而来：



“表集合区”中手动创建表关系的方法：

- 1、单击表间连接点，如下图所示：



- 2、弹出“设置表连接”窗口，如下图所示：

设置表连接

Inner join

Left join

Right join

Full join

产品 . ProductID

=

订单明细 . ProductID

增加表达式

请选择

=

请选择

删除表达式

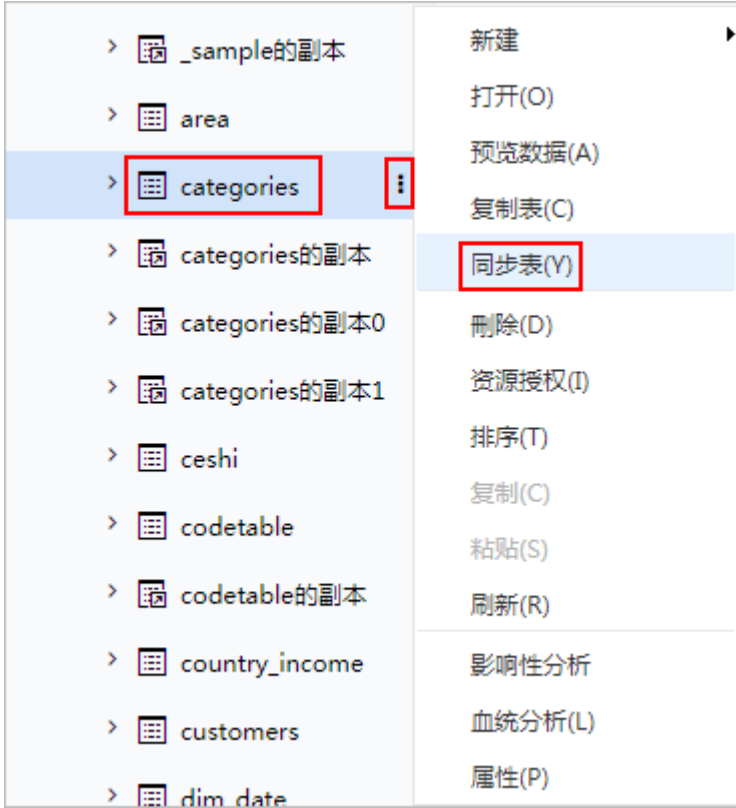

取消

确定

- 选择连接方式：内连接、左连接、右连接、全连接。
- 确定表间连接字段。

同步：原表或原数据集字段增加或删除时，自助数据集利用同步功能更新。

数据来源	操作
------	----

关系数据源/高速缓存库	<p>自助数据集引用的表字段增加、删除时，实现同步更新，需要进行如下两个步骤：</p> <p>1、数据库某表字段增加、删除时，在数据源节点下，选中该表，更多操作 > 同步表，如图：</p>  <p>2、在自助数据集中选中该表，右键 > 同步，如图：</p> 
数据集	<p>选中带有字段变更的数据集，右键 > 同步，如图：</p> 

注：目前暂不支持字段类型更改的同步。




使用“同步”功能需保存，再次打开自助数据集才生效。

筛选器：当数据集数据量比较大时，可添加筛选器，对数据集的数据进行筛选过滤，以便得到用户真正所需的数据。

详情请参考 [自助数据集-筛选器](#)

规范字段

自助数据集的“表属性区”对应有维度和度量，系统支持对表和字段进行重新组织、分类、设计等规范化设置操作：[修改别名](#)、[修改数据类型](#)、[设置可见性](#)、[拖拽表和字段](#)、[新建目录](#)、[新建层次](#)、[新建计算字段](#)、[创建时间维](#)、[标记地理维](#)、[删除](#)。详细说明如下：

功能	说明	示例效果
修改别名	修改表或字段别名。	<p>手动输入别名，修改“customers”的别名为“顾客表”，如图：</p> 
修改数据类型	修改字段的数据类型。	<p>选择字段的数据类型，如图：</p> 
设置可见性	<p>设置表或字段为“可见”或“不可见”。</p> <ul style="list-style-type: none">为了快速查找表或字段，可将自助数据集在特定场景中不需要用到的表或字段设置为“不可见”。将表设置为“不可见”时，表的所有字段都为“不可见”。	<p>点击可见性的图标切换为“可见”或“不可见”：</p> 

拖拽表和字段	拖拽表或字段移动其位置。	<p>拖拽字段改变字段所属的表，将“产品目录表”的“产品类别”拖拽到“产品表”，如图：</p> <table><tr><th>名称</th><th>别名</th><th>数据类型</th><th>数据格式</th><th>可见性</th><th>脱敏规则</th><th>元数据</th></tr><tr><td>A ShipPostalCode</td><td>ShipPostalCode</td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td>A ShipCountry</td><td>ShipCountry</td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td>📁 orderdetails</td><td>订单明细表</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td>📁 customers</td><td>顾客表</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td>📁 employees</td><td>雇员表</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td>📁 products</td><td>产品表</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td>📁 categories</td><td>产品类别表</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td>A CategoryName</td><td>产品类别</td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td># CategoryID2</td><td>产品类别编号2</td><td>INTEGER</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>A Description</td><td>Description</td><td>ASCII</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr></table>	名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据	A ShipPostalCode	ShipPostalCode	STRING	默认值	👁	请选择	demo数据源.default.default	A ShipCountry	ShipCountry	STRING	默认值	👁	请选择	demo数据源.default.default	📁 orderdetails	订单明细表			👁			📁 customers	顾客表			👁			📁 employees	雇员表			👁			📁 products	产品表			👁			📁 categories	产品类别表			👁			A CategoryName	产品类别	STRING	默认值	👁	请选择	demo数据源.default.default	# CategoryID2	产品类别编号2	INTEGER	默认值	👁		demo数据源.default.default	A Description	Description	ASCII	默认值	👁		demo数据源.default.default
名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据																																																																									
A ShipPostalCode	ShipPostalCode	STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
A ShipCountry	ShipCountry	STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
📁 orderdetails	订单明细表			👁																																																																											
📁 customers	顾客表			👁																																																																											
📁 employees	雇员表			👁																																																																											
📁 products	产品表			👁																																																																											
📁 categories	产品类别表			👁																																																																											
A CategoryName	产品类别	STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
# CategoryID2	产品类别编号2	INTEGER	默认值	👁		demo数据源.default.default																																																																									
A Description	Description	ASCII	默认值	👁		demo数据源.default.default																																																																									
新建目录	用于在“维度”或“度量”中新建目录。	<p>任意区域右键菜单选择 新建目录，新建一个“订单表”目录，如图：</p> <table><tr><th>名称</th><th>别名</th><th>数据类型</th><th>数据格式</th><th>可见性</th><th>脱敏规则</th><th>元数据</th></tr><tr><td>📁 维度</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>📁 orders</td><td>订单表</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td># OrderID</td><td>订单编号</td><td>INTEGER</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>A CustomerID</td><td>顾客编号</td><td></td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td># EmployeeID</td><td>雇员编号</td><td></td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📅 OrderDate</td><td>订单日期</td><td></td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📅 RequiredDate</td><td>到货日期</td><td>DATETIME</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📅 ShippedDate</td><td>发货日期</td><td>DATETIME</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr></table>	名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据	📁 维度							📁 orders	订单表			👁			# OrderID	订单编号	INTEGER	默认值	👁		demo数据源.default.default	A CustomerID	顾客编号		默认值	👁	请选择	demo数据源.default.default	# EmployeeID	雇员编号		默认值	👁		demo数据源.default.default	📅 OrderDate	订单日期		默认值	👁		demo数据源.default.default	📅 RequiredDate	到货日期	DATETIME	默认值	👁		demo数据源.default.default	📅 ShippedDate	发货日期	DATETIME	默认值	👁		demo数据源.default.default														
名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据																																																																									
📁 维度																																																																															
📁 orders	订单表			👁																																																																											
# OrderID	订单编号	INTEGER	默认值	👁		demo数据源.default.default																																																																									
A CustomerID	顾客编号		默认值	👁	请选择	demo数据源.default.default																																																																									
# EmployeeID	雇员编号		默认值	👁		demo数据源.default.default																																																																									
📅 OrderDate	订单日期		默认值	👁		demo数据源.default.default																																																																									
📅 RequiredDate	到货日期	DATETIME	默认值	👁		demo数据源.default.default																																																																									
📅 ShippedDate	发货日期	DATETIME	默认值	👁		demo数据源.default.default																																																																									
新建层次	用于新建层次目录，存储层次字段，这些层次字段主要实现自动下钻应用，例如： 功能示例-时间层次维下钻 、 功能示例-地理层次维下钻 、 功能示例-自定义层次维下钻 、 交互式仪表盘-展开/收缩 等。	<p>任意区域右键菜单选择 新建层次，新建“新建地理层次”层次，按层次顺序将字段拖拽到“新建地理层次”下。如图：</p> <table><tr><th>名称</th><th>别名</th><th>数据类型</th><th>数据格式</th><th>可见性</th><th>脱敏规则</th><th>元数据</th></tr><tr><td>📁 维度</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>📁 custom1563432691986</td><td>新建地理层次</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td>A ShipCity</td><td>发货城市</td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td>A ShipProvince</td><td>发货省份</td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td>A ShipRegion</td><td></td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td>📁 orders</td><td></td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td># OrderID</td><td></td><td>INTEGER</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>A CustomerID</td><td></td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr></table>	名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据	📁 维度							📁 custom1563432691986	新建地理层次			👁			A ShipCity	发货城市	STRING	默认值	👁	请选择	demo数据源.default.default	A ShipProvince	发货省份	STRING	默认值	👁	请选择	demo数据源.default.default	A ShipRegion		STRING	默认值	👁	请选择	demo数据源.default.default	📁 orders				👁			# OrderID		INTEGER	默认值	👁		demo数据源.default.default	A CustomerID		STRING	默认值	👁	请选择	demo数据源.default.default														
名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据																																																																									
📁 维度																																																																															
📁 custom1563432691986	新建地理层次			👁																																																																											
A ShipCity	发货城市	STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
A ShipProvince	发货省份	STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
A ShipRegion		STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
📁 orders				👁																																																																											
# OrderID		INTEGER	默认值	👁		demo数据源.default.default																																																																									
A CustomerID		STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
新建计算字段	<p>新建一个计算字段。</p> <p>计算字段的数据类型决定了新建的计算字段位于维度或度量。</p> <ul style="list-style-type: none">维度：字符串、日期数据、以ID结尾的字段。度量：数值型数据。 <p>应用详情请参考 自助数据集-计算字段 章节。</p>	<p>任意区域或字段右键菜单选择 新建计算字段，新建“销售额”计算字段，如图：</p> <table><tr><th>名称</th><th>别名</th><th>数据类型</th><th>数据格式</th><th>可见性</th><th>脱敏规则</th><th>元数据</th></tr><tr><td>📁 10 Picture</td><td>Picture</td><td>BINARY</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📁 度量</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>📁 custom-calcFiled_measure</td><td>计算字段(度量)</td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td># custom1563431500249</td><td>销售额</td><td>DOUBLE</td><td>默认值</td><td>👁</td><td></td><td></td></tr><tr><td>📁 orders</td><td></td><td></td><td></td><td>👁</td><td></td><td></td></tr><tr><td># ShipVia</td><td></td><td>INTEGER</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td># Freight</td><td></td><td>DOUBLE</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📁 orderdetails</td><td></td><td></td><td></td><td>👁</td><td></td><td></td></tr></table>	名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据	📁 10 Picture	Picture	BINARY	默认值	👁		demo数据源.default.default	📁 度量							📁 custom-calcFiled_measure	计算字段(度量)			👁			# custom1563431500249	销售额	DOUBLE	默认值	👁			📁 orders				👁			# ShipVia		INTEGER	默认值	👁		demo数据源.default.default	# Freight		DOUBLE	默认值	👁		demo数据源.default.default	📁 orderdetails				👁																
名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据																																																																									
📁 10 Picture	Picture	BINARY	默认值	👁		demo数据源.default.default																																																																									
📁 度量																																																																															
📁 custom-calcFiled_measure	计算字段(度量)			👁																																																																											
# custom1563431500249	销售额	DOUBLE	默认值	👁																																																																											
📁 orders				👁																																																																											
# ShipVia		INTEGER	默认值	👁		demo数据源.default.default																																																																									
# Freight		DOUBLE	默认值	👁		demo数据源.default.default																																																																									
📁 orderdetails				👁																																																																											
创建时间维	<p>用于以维层次的形式标识日期信息字段。</p> <p>入口：日期信息字段的右键菜单选择 创建时间维度。</p> <p>创建时间维度后会自动生成一个“时间维度”的层次目录。</p> <p>应用详情请参考 功能示例-时间层次维下钻 章节。</p>	<table><tr><th>名称</th><th>别名</th><th>数据类型</th><th>数据格式</th><th>可见性</th><th>脱敏规则</th><th>元数据</th></tr><tr><td>A CustomerID</td><td>顾客编号</td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td># EmployeeID</td><td>雇员编号</td><td>INTEGER</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📅 OrderDate</td><td>订单日期</td><td>DATETIME</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📅 RequiredDate</td><td></td><td>DATETIME</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>📅 ShippedDate</td><td></td><td>DATETIME</td><td>默认值</td><td>👁</td><td></td><td>demo数据源.default.default</td></tr><tr><td>A ShipName</td><td></td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td>A ShipAddress</td><td></td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr><tr><td>A ShipCity</td><td>发货城市</td><td>STRING</td><td>默认值</td><td>👁</td><td>请选择</td><td>demo数据源.default.default</td></tr></table>	名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据	A CustomerID	顾客编号	STRING	默认值	👁	请选择	demo数据源.default.default	# EmployeeID	雇员编号	INTEGER	默认值	👁		demo数据源.default.default	📅 OrderDate	订单日期	DATETIME	默认值	👁		demo数据源.default.default	📅 RequiredDate		DATETIME	默认值	👁		demo数据源.default.default	📅 ShippedDate		DATETIME	默认值	👁		demo数据源.default.default	A ShipName		STRING	默认值	👁	请选择	demo数据源.default.default	A ShipAddress		STRING	默认值	👁	请选择	demo数据源.default.default	A ShipCity	发货城市	STRING	默认值	👁	请选择	demo数据源.default.default														
名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据																																																																									
A CustomerID	顾客编号	STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
# EmployeeID	雇员编号	INTEGER	默认值	👁		demo数据源.default.default																																																																									
📅 OrderDate	订单日期	DATETIME	默认值	👁		demo数据源.default.default																																																																									
📅 RequiredDate		DATETIME	默认值	👁		demo数据源.default.default																																																																									
📅 ShippedDate		DATETIME	默认值	👁		demo数据源.default.default																																																																									
A ShipName		STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
A ShipAddress		STRING	默认值	👁	请选择	demo数据源.default.default																																																																									
A ShipCity	发货城市	STRING	默认值	👁	请选择	demo数据源.default.default																																																																									

标记地理维

用于以维层次的形式标识地理信息字段。

入口：地理信息字段的右键菜单选择 **标记地理维度**。

已标记地理维度的字段必须存储在层次目录下，才能保证地理维层次字段的下钻效果实现。

应用详情请参考 [功能示例-地理层次维下钻](#) 章节。

删除

用于删除表。

立即刷新

显示隐藏字段

显示别名

100

行

名称	别名	数据类型	数据格式	可见性	脱敏规则	元数据
A ₀ ShipAddress	ShipAddress	STRING	默认值		请选择	demo数据源.default.default
A ₀ ShipCity	发货城市	STRING	默认值		请选择	demo数据源.default.default
A ₀ ShipProvince	新建目录	STRING	默认值		请选择	demo数据源.default.default
A ₀ ShipRegion	新建层次	STRING	默认值		请选择	demo数据源.default.default
A ₀ ShipPostalCode	新建计算字段	STRING	默认值		请选择	demo数据源.default.default
A ₀ ShipCountry	标记地理维度	STRING	默认值		请选择	demo数据源.default.default
orderdetails	订单明细表	表示地理信息字段的右键菜单选择 “标记地理维度”				
customers	顾客表					
employees	雇员表					

保存自助数据集

保存当前自助数据集到指定目录。

数据抽取

将当前自助数据集的数据从原始数据库中抽到到高速缓存库。详情请参见 [数据抽取](#) 章节。