

## 目录

目录	1
建立模型	2
新建数据模型	2
模型设置	2
什么是维度、度量	2

# 建立模型

## 新建数据模型

连接完数据后，便可以将需要的多张数据表关联成一张表，并进行需要的数据处理（诸如字段重命名、空值处理、建立数据字典、添加计算字段、创建层级等操作），建立数据模型以进行后续的数据可视化分析工作，如下图所示：



下面我们分步骤介绍：

1、在“数据源”模块，添加数据模型  2、选择需要的数据连接，基于该连接建立数据模型

3、选择需要的一张或多张数据表，若选择多张数据表，则需要关联成一张宽表



拖入两张表时，若它们在原数据库中存在外键关联，则会自动进行关联；若无外键，系统会自动将两张表中相同名称的列设置为外键进行关联。用户也可以手动添加或修改“关联字段”。



完成关联后，下方会显示宽表中的所有字段，并将字段划分为维度、度量两种类型进行展示。



4、如果需要，可对字段进行处理，比如创建计算字段

5、保存后完成数据模型的建立



## 模型设置

用户可以对模型的连接方式，应用范围，是否同步复杂报表，缓存有效期进行设置。  对于连接方式为直连的模型，可以在更多内开启缓存有效期的设置  开启后缓存有效期设置后，可对缓存时间进行设置，时间单位包括时、分、秒。

缓存规则：首次访问后会对数据进行缓存，缓存有效期内，再次访问会直接访问缓存数据。当到达缓存失效时间点后，缓存被丢弃，之后再访问时，则需重新访问数据库获取数据并重新缓存。例如，若首次访问时间为“08时36分”，缓存有效期设置为“3小时”，则缓存的数据会在下一个三小时整数倍时间点失效，即“09时00分”失效。若缓存有效期设置为“6小时”，则缓存的数据会在下一个六小时整数倍时间点失效，即“12时00分”失效。若缓存有效期设置为“24小时”，则当日首次访问后缓存的数据，会在当日凌晨00时00分失效，这期间查看报表都会访问缓存数据。

## 什么是维度、度量

维度：观察数据时，使用的粒度

度量：汇总的统计值

聚合方式：汇总的方式，比如求和、求平均、最大值、最小值

怎么理解呢？假设我们有一份明细的订单交易数据，部分数据如下：



将这份数据导入有数后，我们可以用不同的粒度观察数据，有数会自动替我们进行汇总。

比如，观察“各地区的销售额”，“地区”是维度，“销售额”是度量。每个地区都对应成百上千行数据，有数对这些数据进行了求和汇总。如下图所示：



我们也可以观察“各省的销售额”，“省/自治区”是维度，“销售额”是度量。如下图所示：



数据导入有数后，默认会把字符型的字段归类为维度，数值型的字段归类为度量，用户也可以手动更改字段的类型。

表关联的说明 多张数据表进行关联时，有数提供4种关联类型：

举例说明这几种关联的区别：

1、内关联

关联前：

关联后：

2、左关联

关联前：

关联后：

3、右关联

关联前：

关联后： 创建数据模型时，支持“等于、不等于、小于、小于等于、大于、大于等于”关系的表关联。

跨数据连接关联表 有数支持将不同数据连接中的表进行关联，比如一张数据表来自MySQL数据库，一张数据表来自Excel文件，要将两张数据表关联成一张宽表后分析。此时需要将不同数据连接的表抽取至有数提供的MPP数据库中（此功能暂时只对付费用户开放）。

我们还可以通过“抽取设置”对抽取任务进行更灵活的设置，比如设置“每小时自动抽取一次”或“按照日期字段进行增量抽取”（关于抽取设置的更详细的内容，将在后续的项目管理章节进行介绍）。

添加自定义SQL视图 当基于数据库（比如MySQL、Oracle）类型的数据连接建立数据模型时，可以在有数内通过SQL语句建立自定义视图。

将数据库的Comment内容作为字段名 在数据模型编辑界面，我们可以进行批量编辑字段和批量设置数据格式的操作。

点击[批量编辑字段]，用户可以批量将别名修改为注释内容，也可以将字段说明修改为注释内容。

数据模型增加缓存有效期的设置 开启后，可对缓存时间进行设置，时间单位包括时、分、秒

缓存规则：首次访问后会对数据进行缓存，缓存有效期内，再次访问会直接访问缓存数据。当到达缓存失效时间点，缓存被丢弃，之后再访问时，则需重新访问数据库获取数据并重新缓存。

示例：若首次访问时间为“08时36分”，缓存有效期设置为“3小时”，则缓存的数据会在下一个三小时整数倍时间点失效，即“09时00分”失效。若缓存有效期设置为“6小时”，则缓存的数据会在下一个六小时整数倍时间点失效，即“12时00分”失效。若缓存有效期设置为“24小时”，则当日首次访问后缓存的数据，会在当日凌晨00时00分失效，这期间查看报表都会访问缓存数据。