

数据模型-概述

Smartbi的数据模型实现将所有查询结果归集后，基于CUBE模型重新构建数据结构：以“维度”和“度量”进行构建，同时增加了“成员”和“命名集”的定义，实现了数据模型构建的灵活性及应用广泛性。

Smartbi的数据模型基于成熟的建模理论和方法，总体而言主要体现在两方面：模型架构和模型表关系。

模型架构

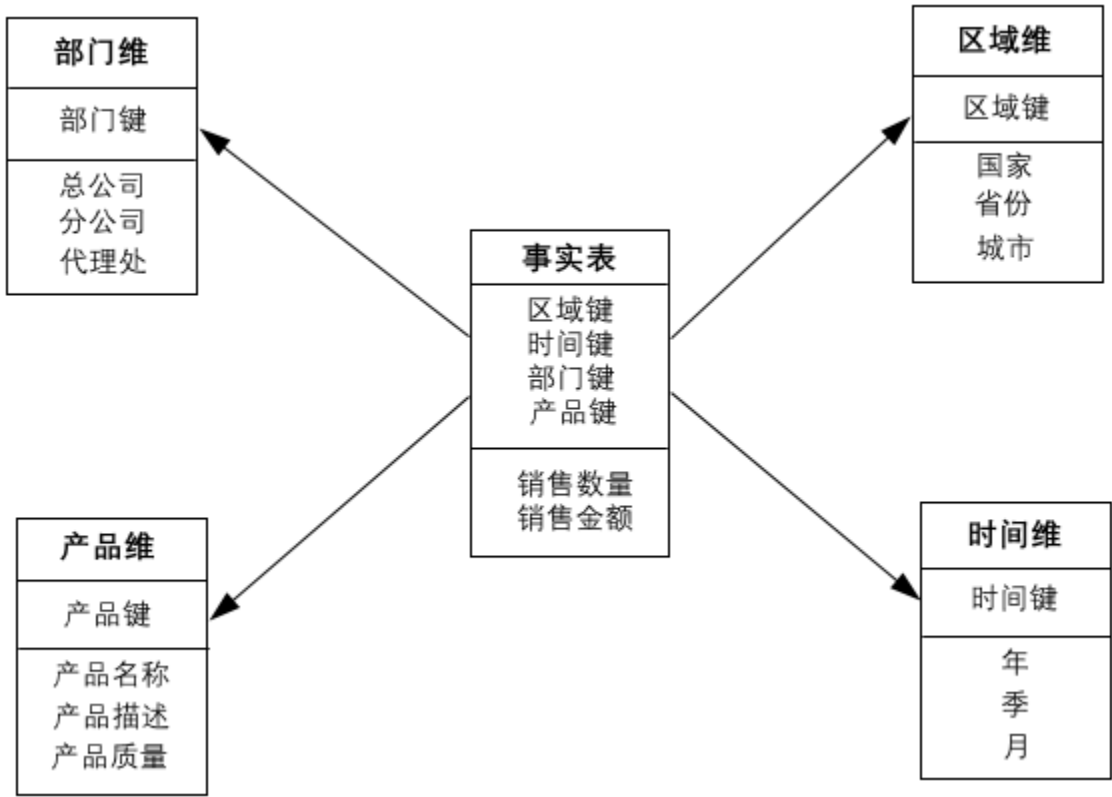
数据模型是围绕事实表和维度表的关系而进行模型的构建：

- 事实表：是数据模型中的中央表，它包含联系事实与维度表的数字度量值和键。事实数据表包含描述业务（例如产品销售）内特定事件的数据。
- 维度表：是维度属性的集合，是分析问题的一个窗口，是人们观察数据的特定角度，是考虑问题时的一类属性，属性的集合构成一个维。

Smartbi支持星型模型、雪花模型和星座模型、单表模型。

星型模型

星型模型：所有维表都直接连接到事实表上，整个视图就像星星一样，视图如下：



星型模型强调的是对维度进行预处理，将多个维度集合到一个事实表，形成一个宽表。

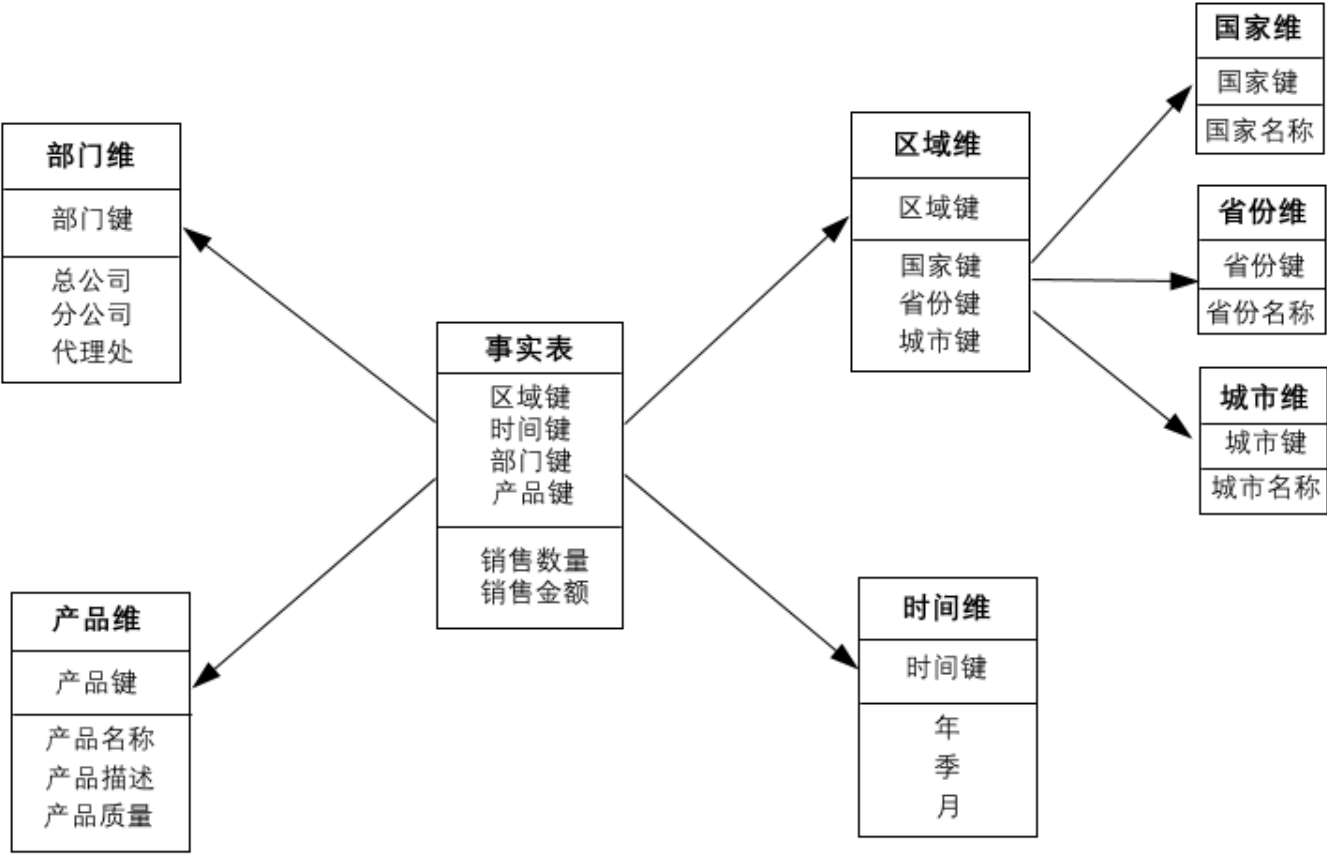
该模型类型在Smartbi数据模型设计中呈如下视图：



模型表左侧绿底表示事实表，蓝底表示维表

雪花模型

雪花模型：一个或多个维表没有直接连接到事实表上，而是通过其他维度表连接到事实表，视图就像多个雪花连接在一起，视图如下：



该模型类型在Smartbi数据模型设计中呈如下视图：

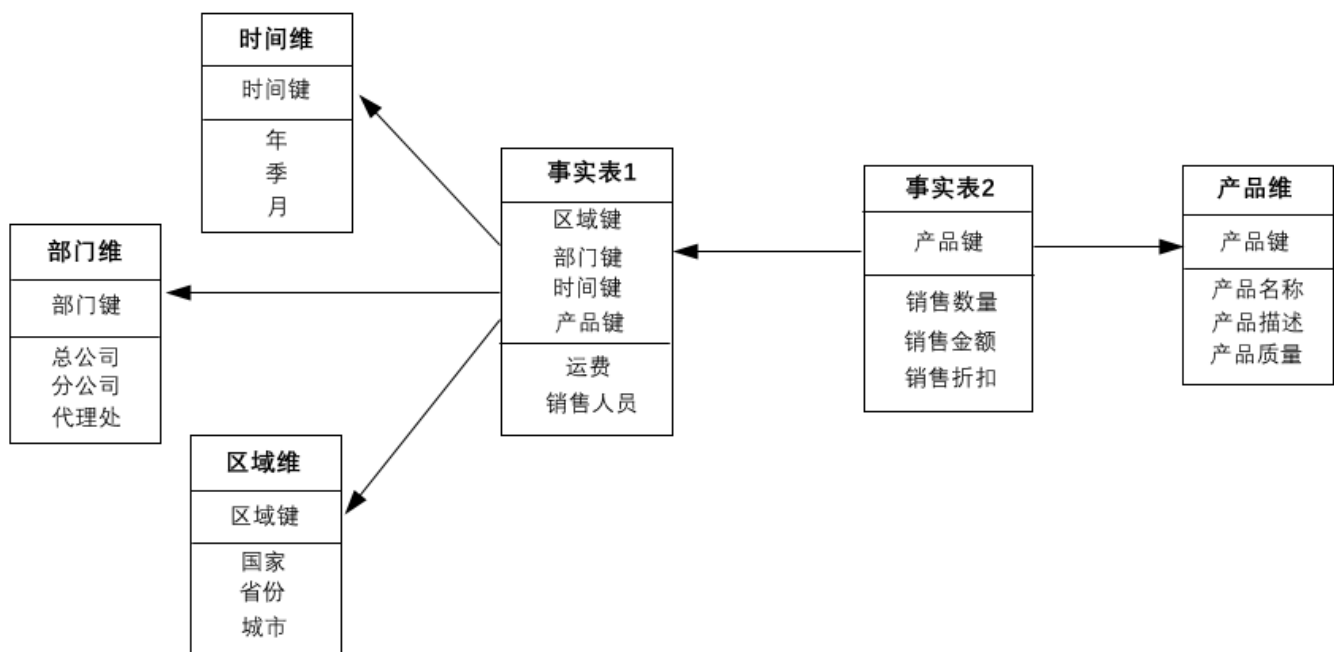


模型表左侧绿底表示事实表，蓝底表示维表

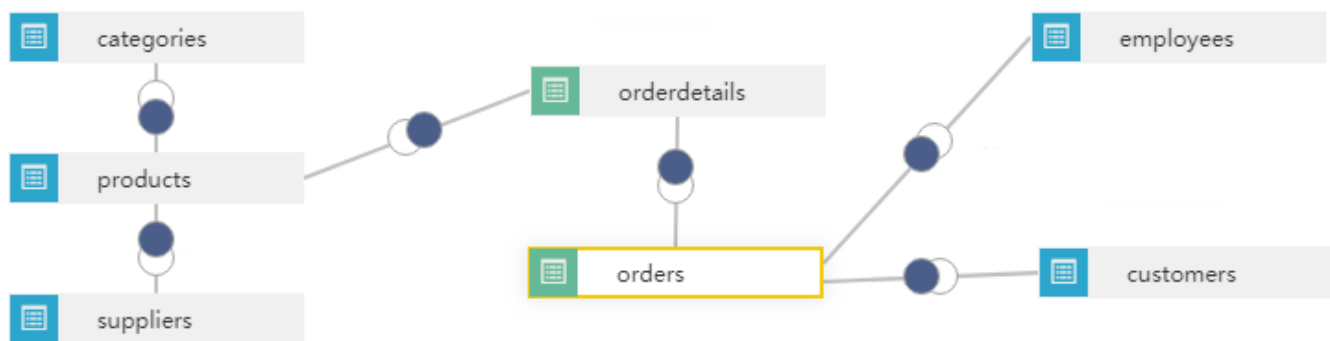
雪花模型是对星型模型的扩展，它对星型模型的维表进一步层次化，原有的各维表可能被扩展为小的事实表，形成一些局部的“层次 ”区域，这些被分解的表都连接到主维度表而不是事实表。如将产品维表分解为产品类别维表。

星座模型

星座模型：包含多个事实表，维表是公共的，可以共享，视图如下：



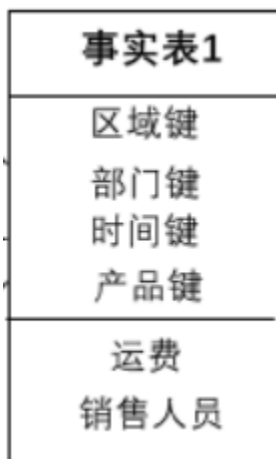
该模型类型在Smartbi数据模型设计中呈如下视图：



模型表左侧绿底表示事实表，蓝底表示维表

单表模型

指标、维度、属性关联在一起的一张表，如下图：



订单 Id	订...	发货日期	邮寄...	...	细分	城市	省...	国...	...	类别	子...	产品名称	销售额	数量	折扣
US-20...	2019...	2019/4/29	二级	管惠	公司	杭州	浙江	中国	华东	办公用品	用品	Fiskars 剪刀, ...	¥ 130	2	40%
CN-20...	2019...	2019/6/19	标准级	许安	消费者	内江	四川	中国	西南	办公用品	信封	GlobeWeis 搭...	¥ 125	2	0%
CN-20...	2019...	2019/6/19	标准级	许安	消费者	内江	四川	中国	西南	办公用品	装订机	Cardinal 孔加...	¥ 32	2	40%
US-20...	2019...	2019/12/13	标准级	宋良	公司	镇江	江苏	中国	华东	办公用品	用品	Kleencut 开信...	¥ 321	4	40%
CN-20...	2018...	2018/6/2	二级	万兰	消费者	汕头	广东	中国	中南	办公用品	器具	KitchenAid 搅...	¥ 1,376	3	0%
CN-20...	2017...	2017/10/31	标准级	俞明	消费者	景德镇	江西	中国	华东	技术	设备	柯尼卡 打印机...	¥ 11,130	9	0%
CN-20...	2017...	2017/10/31	标准级	俞明	消费者	景德镇	江西	中国	华东	办公用品	装订机	Ibico 订书机, ...	¥ 480	2	0%

维度
度量

模型相关概念

我们在Smartbi设计数据模型架构前，需要对模型设计紧密相关的概念有清晰的认识和理解：

度量值

度量值是存储要汇总的值的的事实数据表列。

Smartbi数据模型中的度量值表达式通常利用 聚合函数（如 SUM、MIN、MAX、AVERAGE 等）在查询时生成标量值结果（值永远不会存储在模型中）。

度量值表达式涵盖广泛，从简单的列聚合到更复杂的公式应有尽有，这为模型开发者提供了便利，因为在很多情况下都无需写MDX表达式创建度量值：如“订单明细”表中的“销售量”列可以通过多种方式进行汇总（包括 SUM、COUNT、AVERAGE、MEDIAN、MIN、MAX 等）。



雪花维度

雪花维度是单个业务实体的一组规范化表。例如：“产品”维度经过了规范化并存储在两个相关表中：产品类别维度和产品维度。这些规范化表位于事实数据表之外，形成雪花形状的设计。

在 Smartbi数据模型 中，可以选择模仿雪花维度设计，也可以将源表（非规范化）通过私有查询的方式集成到单个模型表中。如下图所示：



一般而言，单个模型表的优点比多个模型表的优点更多，最理想的决策取决于数据量和模型的可用性要求。

在选择集成到一个模型表中时，还可以定义一个层次结构，其中包含维度的最高和最低粒度。冗余非规范化数据的存储可能会增加模型存储大小，尤其是在维度表很大的情况下。

角色扮演维度

角色扮演维度是能以不同方式筛选相关事实的维度。例如，在Northwind中，订单维度表与订单销售明细事实之间有三种关系：可以使用同一维度表中按订单日期、发货日期或交货日期筛选事实。

数据仓库中接受的设计方法是定义单个日期维度表。在查询时，日期维度的“角色”是通过用于联接表的事实列建立的。例如，在按订单日期分析销售情况时，表联接与订单日期列相关。

在Smartbi数据模型中，我们用一种常见的建模方法解决，就是为每个角色扮演实例创建维度类型表。该模型可以包含日期表、发货日期表和交货日期表，每个表与其对应的销售表列之间都存在单一且活动的关系。

杂项维度

如果有多个维度，特别是包含几个属性（可能只有一种）并且这些属性的值很少时，则杂项维度非常有用。

合适的候选项包括订单状态列或客户人口统计列（性别、年龄分组等）。

杂项维度的设计目标是将多个“小”维度合并为一个维度，以减少模型的存储大小，还可以通过呈现更少的模型表来缓解“字段”窗格的混乱情况。

杂项维度表通常是所有维度属性成员的笛卡尔积，具有唯一标识列。可以在数据仓库中构建维度，也可以通过使用Smartbi数据模型中的私有查询创建，并此查询作为维度类型表加载到模型中。

退化维度

退化维度指的是筛选所需的事实数据表的属性。我们的northwind中就有例子，就是销售订单号。在这种情况下，创建仅包含这一列的独立表对模型设计并没有什么积极作用，因为这会增加模型的存储大小，并导致“字段”窗格混乱。

在Smartbi模型中，可以将销售订单号列添加到事实类型表，以允许按销售订单号进行筛选或分组。

无事实事实数据表

无事实事实数据表不包含任何度量值列。它仅包含维度键。

无事实事实数据表可以存储由维度键定义的观察值。例如，在特定日期和时间，特定客户登录到你的网站。你可以定义一个度量值以对无事实事实数据表中的行进行计数，从而对客户的登录时间和数量进行计数分析。

无事实事实数据表的更有力的用途是存储维度之间的关系，例如，假设销售人员可以分配到一个或多个销售区域，则需要设计为包含两列的无事实事实数据表：销售人员键和区域键，这两列中可以存储重复的值。

模型表关系

模型设计过程中，模型表的关系也是很重要的一项内容。

关系的确立需要通过匹配键列中的数据（通常是两表中同名的列）。在大多数情况下，该关系会将一个表中的主键（它为每行提供了唯一标识）与另一个表的外部键中的某项相匹配。例如，通过创建 orderdetails表中的 orderid（主键）与orders表中的 orderid列（外部键）之间的关系，则销售量就与订单相关联了。

一对多（多对一）关系

一对多和多对一是相对关系，如下列的课程表与老师表：

Course表

C_id	C_name
101	Java
202	前端
303	英语

Teacher表

T_id	T_name
1	李老师
2	刘老师
3	郑老师

站在课程表的角度看是：多对一；站在老师表的角度看是：一对多。

在一对多或多对一的关系中，我们习惯把一的一方称为主表，把多的一方称为从表。

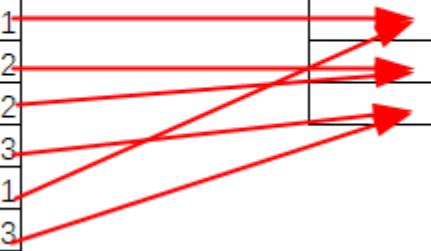
在数据库中建立一对多的关系，需要使用数据库的外键约束：

Course表

C_id	C_name	T_id
101	Java	1
202	前端	2
303	英语	2
304	python	3
305	Andriod	1
306	语文	3

Teacher表

T_id	T_name
1	李老师
2	刘老师
3	郑老师



Course表中的T_id字段建立外键约束，指向Teacher表中的T_id。

一对一关系

在一对一关系中，A 表中的一行最多只能匹配于 B 表中的一行，反之亦然。比如，一个人只能有一个身份证号，一个身份证号只能属于一个人。

Person表

P_id	P_name	P_address
1	张三	SH
2	李四	GZ
3	王五	BJ

IdCard表

C_id	C_number
1	100100100
2	200200200
3	300300300

如果相关列都是主键或都具有唯一约束，则可以创建一对一关系。

双方主键约束

Person表

P_id	P_name	P_address
1	张三	SH
2	李四	GZ
3	王五	BJ

IdCard表

C_id	C_number
1	100100100
2	200200200
3	300300300

Person表

P_id	P_name	P_address
1	张三	SH
2	李四	GZ
3	王五	BJ

IdCard表

C_id	C_number	P_id
1	100100100	1
2	200200200	2
3	300300300	3

unique:建立唯一约束

这种关系并不常见，因为一般来说，按照这种方式相关的信息都在一个表中。可以利用一对一关系来：

- 分割具有多列的表。
- 由于安全原因而隔离表的一部分。
- 保存临时数据，并且可以毫不费力地通过删除该表而删除这些数据。
- 保存只适用于主表的子集的信息。

多对多关系

多对多的关系，可以看成是两个多对一的关系，如下的课程表和学生表：

Course表

C_id	C_name
101	Java
202	前端
303	英语
304	python
305	Andriod
306	语文

Student表

S_id	S_name
1	张三
2	李四
3	王五

站在课程表的角度，一门课可以被多个学生选择，是多对一的关系；站在学生表的角度，一个学生可以选择多门课程，是多对一的关系。

这两个表的多对多关系，通过一张关联表来建立关联：

Course表

C_id	C_name
101	Java
202	前端
303	英语
304	python
305	Andriod
306	语文

Couse_Student表---中间表

C_id	S_id
101	1
101	2
202	1
202	3
303	1
303	2
303	3

Student表

S_id	S_name
1	张三
2	李四
3	王五

101课程被学号1、学号2的学生选择；学号1的学生同时选择了101、202、303的课程。

目前Smarbi数据模型构建仅支持一对多（多对一）和一对一关系。

模型选型

星型模型和雪花模型在架构体系中各有优劣，其对比如下：

区别项	星型模型	雪花模型
维度表	一级维度表	多级维度表（子维度表）
存储空间	多	少
数据冗余度	大	小
表宽度	宽	窄
扩展性	差	好
Join复杂度	低	高
查询逻辑	简单	难
查询速度	快	慢
可读性	简单	难
OLAP建模工具优化度	低	高
数据总量	多	少
可读性	容易	差
表个数	少	多

在实际业务场景中，往往是两种模型的综合应用。