

# 《Star Schema完全参考手册》

## 图书基本信息

书名：《Star Schema完全参考手册》

13位ISBN编号：9787302291527

10位ISBN编号：7302291527

出版时间：2012-7

出版社：清华大学出版社

作者：亚当森

页数：389

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：[www.tushu000.com](http://www.tushu000.com)

## 内容概要

《Star Schema完全参考手册:数据仓库维度设计权威指南》内容简介：学习维度设计的最佳实践。

《Star Schema完全参考手册:数据仓库维度设计权威指南》全面、深入地介绍了设计原理及其理论基础。全书围绕设计概念组织并辅以详实的示例，可作为初学者逐步深入学习星型模式的参考资料，同时也为专家提供了内容丰富的参考资源。《Star Schema完全参考手册:数据仓库维度设计权威指南》首先介绍维度设计的基础知识，向读者展示如何将这些基础应用于不同的数据仓库体系结构，包括W.H.Inmon并HRalph Kimball倡导的体系结构。《Star Schema完全参考手册:数据仓库维度设计权威指南》通过介绍一系列高级技术来帮助读者解决现实世界中存在的复杂性，获得最佳性能并适应商业智能和ETL软件产品的需求。读者可以将书中的设计任务和设计产品运用于所有项目中，而无论项目采用什么方法和架构。

## 作者简介

作者：（美国）亚当森（Christopher Adamson）译者：王红滨 王念滨 初妍 Christopher Adamson是星型模式设计专家，并且是经验丰富的数据仓库实践者。他在世界各地演讲和授课，撰写数据仓库相关内容，并通过Oakton软件有限责任公司提供数据仓库咨询服务。作为一名咨询专家，Christopher与各个行业的客户一起工作，一开发数据仓库战略，识别并确定项目优先顺序，设计并修改解决方案，以及管理实现。他服务的客户包括财富500强的公司、政府机关和数据仓库软件提供商。Christopher在世界范围内为成千上万学生讲授多维设计。他是The Data Warehousing Institute（TDWI）的教员，为重要会议提供现场讲座，并且提供网络教学。Christopher撰写了许多有关数据仓库的文章和书籍。他是Mastering Data Warehouse Aggregates（Wiley，2006）一书的作者，并与人合著了Data Warehouse Design Solutions（Wiley,1998）一书。他撰写有关数据仓库的文章并为读者解答问题的博客网址为[www.StarSchemaCentral.com](http://www.StarSchemaCentral.com)。

## 书籍目录

### 第1部分 基础

#### 第1章 分析型数据库与维度设计

- 1.1 维度设计
- 1.2 星型模式
- 1.3 星型模式的应用
- 1.4 指导原则
- 1.5 本章小结
- 1.6 补充阅读

#### 第2章 数据仓库体系结构

- 2.1 inmon的企业信息化工厂
- 2.2 kimball的维度数据仓库
- 2.3 独立型数据集市
- 2.4 体系结构和维度设计
- 2.5 本章小结
- 2.6 补充阅读

#### 第3章 星型模式与多维数据集

- 3.1 维度表特性
- 3.2 事实表特性
- 3.3 缓慢变化维
- 3.4 多维数据集
- 3.5 本章小结
- 3.6 补充阅读

### 第2部分 多种星型模式

#### 第4章 过程处理中的事实表

- 4.1 事实表与业务过程
- 4.2 具有不同时间的事实
- 4.3 具有不同粒度的事实
- 4.4 从多个事实表中分析事实
- 4.5 本章小结
- 4.6 补充阅读

#### 第5章 维度一致性问题

- 5.1 星型模式之间的协同
- 5.2 维度与横向钻取
- 5.3 一致性维度
- 5.4 结构与一致性
- 5.5 本章小结
- 5.6 补充阅读

### 第3部分 维度设计

#### 第6章 深入学习维度表

- 6.1 将维度分组到表中
- 6.2 分解大型维度
- 6.3 角色维度和别名使用
- 6.4 避免空值
- 6.5 行为维度
- 6.6 本章小结
- 6.7 补充阅读

#### 第7章 层次结构与雪花模式

- 7.1 钻取
- 7.2 属性层次和钻取
- 7.3 雪花模式
- 7.4 支架表
- 7.5 本章小结
- 7.6 补充阅读
- 第8章 深入学习缓慢变化技术
  - 8.1 时间戳维度
  - 8.2 类型3变化
  - 8.3 混合缓慢变化
  - 8.4 本章小结
  - 8.5 补充阅读
- 第9章 多值维度与桥接表
  - 9.1 标准的一对多关系
  - 9.2 多值维度
  - 9.3 多值属性
  - 9.4 本章小结
  - 9.5 补充阅读
- 第10章 递归层次与桥接表
  - 10.1 递归层次
  - 10.2 递归层次的扁平化
  - 10.3 层次桥接表
  - 10.4 变化与层次桥接表
  - 10.5 关于层次桥接表的相关讨论
  - 10.6 本章小结
  - 10.7 补充阅读
- 第4部分 事实表设计
- 第11章 事务、快照和累积快照
  - 11.1 事务事实表
  - 11.2 事实表快照
  - 11.3 累积快照事实表
  - 11.4 本章小结
  - 11.5 补充阅读
- 第12章 无事实的事实表
  - 12.1 无事实的事件
  - 12.2 条件、范围或资格
  - 12.3 本章小结
  - 12.4 补充阅读
- 第13章 特殊类型的星型模式
  - 13.1 特殊类型属性
  - 13.2 核心和自定义星型模式
  - 13.3 使用通用属性
  - 13.4 本章小结
  - 13.5 补充阅读
- 第5部分 性能
- 第14章 导出模式
  - 14.1 重构维度数据
  - 14.2 融合事实表
  - 14.3 旋转事实表

- 14.4 切片事实表
- 14.5 集合操作事实表
- 14.6 本章小结
- 14.7 补充阅读
- 第15章 聚集
- 15.1 聚集的基础知识
- 15.2 使聚集不可见
- 15.3 可选的汇总设计方法
- 15.4 本章小结
- 15.5 补充阅读
- 第6部分 工具及文档
- 第16章 设计与商业智能
- 16.1 商业智能与sql生成
- 16.2 适用于语义层的指导原则
- 16.3 使用商业智能工具的sql生成器
- 16.4 使用基于多维数据集的商业智能
- 16.5 本章小结
- 16.6 补充阅读
- 第17章 设计与etl
- 17.1 etl过程
- 17.2 加载星型模式
- 17.3 优化加载
- 17.4 清理数据
- 17.5 内部处理列
- 17.6 本章小结
- 17.7 补充阅读
- 第18章 如何设计并文档化维度模型
- 18.1 维度设计与数据仓库生命周期
- 18.2 设计活动
- 18.3 文档化维度模型
- 18.4 本章小结
- 18.5 补充阅读

## 章节摘录

版权页：插图：由于非可加的事实并未存储在事实表中，需要注意不要丢失这些事实。在许多业务过程中，比率是非常重要的度量，没有它们，系统会存在许多缺陷。作为模式设计的一部分，非可加事实应该被文档化，在第18章“如何设计并文档化维度模型”中将详细讨论。注意：关于非可加性的讨论还没有结束。除了非可加及可加的事实之外，还存在半可加的事实。关于半可加事实将在第11章“事务、快照和累积快照”中讨论。

3.2.3 粒度 事实表行所表达的细节层次被称为粒度。清楚地声明事实表的粒度是模式设计过程中的一个重要工作。明确的粒度能够确保对事实表中行的意思的理解不会产生混淆，保证所有的事实按照同样的细节层次记录。可以采用多种方式描述粒度。许多模式设计者采用简单地枚举相关维度的方式来描述粒度。例如，图3—1所示的订单事实表（ORDER FACTS）的粒度可以描述为“按照每天、销售人员、产品、客户排序”。以上方式听起来好像说清楚了，它揭示了星型模式的重要信息。在此方式下，粒度的声明包含以下隐含意思：对给定的某天，如果一个客户在同一销售商处，就同一产品提交了多个订单，那么这些订单将被组合到同一行中。这样的聚合可能抛弃了有用的信息。多数情况下，模式设计者会避免在将数据加载到事实表之前聚合数据。尽可能地保持最细粒度的数据，星型模式就能解决范围更加宽泛的分析型需求。初期的业务需求可能不需要详细数据，但分析型的需求往往具有变化的趋势，一旦构建了聚集模式，对详细信息的需求将就此展开。无论采用何种数据仓库结构，这一指导原则都普遍适用。对多维数据仓库来说，事实表包含粒度数据至关重要，因为多维数据仓库是细节数据的集中存储仓库。如果性能成为关注的焦点，可以在细粒度事实表中附带聚合信息，第15章中将会讨论该问题。针对企业信息化工厂体系结构来说，这一指导原则可以宽松一些，因为不同的仓库均包含粒度数据。在此环境下，数据集市的事实表可以聚集数据，而不用担心丢失信息的问题。不过，对研究粒度数据的新需求可能会导致重新开发数据集市。

技巧：尽可能将事实表的粒度设置为最详细的数据粒度级别。这一原则有助于确保分析型应用具有最大的灵活性。假如能够有不同的仓库保存不同粒度的数据，则可以降低要求，但这种方式可能会对未来的应用带来限制。

# 《Star Schema完全参考手册》

## 编辑推荐

《Star Schema完全参考手册:数据仓库维度设计权威指南》由清华大学出版社出版。



## 精彩短评

- 1、搞不懂，一个没有跟终端客户沟通机制的电子商务网站，还能存活多久？
- 2、正好有项目要用，还比较权威
- 3、内容太少，讲得太深。不容易理解。
- 4、很少看到这么深入浅出的书了，而且很多知识都是毫无保留的贡献出来。不像某些书只是藏着掖着的讲一个大概该书没有和具体的OLAP平台挂钩，但涵盖的内容非常详细，可惜最高只能给5分，否则绝对要打一个10分的！！
- 5、正版，挺有用的，快递很快，赞一个
- 6、收到书了，全新，不错！！纸质一般
- 7、书本摸着舒服，感觉很好，很满意
- 8、很好，大师级人家是不一样
- 9、非常不错，讲的很实用，正在细细深读~
- 10、内容不错，纸张不好。有电子的英文版，还没来得及判断翻译如何。
- 11、十分不错，强烈推荐！
- 12、讲得很详细，我所关心的问题基本都解释了。翻译的也可以。
- 13、简单看了一下，比较适合入门者。不错的资料！
- 14、翻译糟糕得令人发指。如果不会翻译，麻烦你就用英语原文术语，谢谢。还有一点，不知道是纸张不好或者是打印的问题，书里的图表看着模糊。
- 15、等读完再补充
- 16、师徒传承，数据仓库生命周期工具箱之后的又一力作，既有理论高度，又贴近操作。
- 17、没看完呢，看了一点，总体感觉讲的还是不错的，值得一看
- 18、在书店看到的，翻看两章，印象深刻，坚决的买回来了！！！！
- 19、本书侧重设计、理论层面，如何用传统关系型schema来实现Star Schema
- 20、本书对于基本概念都将清楚了.涉及了基本概念,维度建模,性能,ETL等.例子还蛮多的.不过中文版有些地方翻译的不好,部分概念还是比较别扭.不过对于了解还是OK的.
- 21、理论和实践上有差异，但是在数仓上这本算是必读了吧.....我反复读过前面几章，因为这些内容属于基础又重要，才能慢慢体会，但是转换成自己的语言仍不自信。数仓的内容真的有些抽象啊，作为一个原先受过前台编程训练的我。
- 22、看了，感觉还不错。
- 23、关于星型模式讲解得很详细，挺不错的一本书
- 24、纯理论，无实操
- 25、很适合初学者，有很好的参考价值。
- 26、书的质量很好，至于内容还没看，相信还不错送货的速度也比较快，
- 27、还可以吧，还没看，纸张不是很好，发黄。

## 章节试读

### 1、《Star Schema完全参考手册》的笔记-第63页

为了便于分析研究，应该为每个过程建立一个事实表。

如何判断给定的两个事实是否属于同一个过程？

1，这些事实是否同时发生？

2，这些事实是否可以用在同一个粒度？

如果其中一条不满足，说明两个事实不属于同一个过程。

### 2、《Star Schema完全参考手册》的笔记-第107页

1，将维度分组到表中

需要确定维度属性是属于同一个表还是多个表，其实可以根据直觉判断，如果抽象成理论则是：对于有亲密关系的维度放到同一个表里，对于有同时浏览需求的维度放到同一个表里。

### 3、《Star Schema完全参考手册》的笔记-第52页

一个过程对应一个事实表

如果两个事实，不是同时发生或者粒度不同，就是两个过程

### 4、《Star Schema完全参考手册》的笔记-第16页

三类体系结构：

1，INMON，企业信息化工厂

2，KIMBALL，维度数据仓库

3，独立数据集市

INMON和KIMBALL体系结构相同点：

1，操作系统和分析系统是分离的

2，ETL过程整合了多种类型操作型系统的数据，将信息集中到一起，保存到一个单独的数据仓库中

区别：

1，KIMBALL是根据维度建模的原则设计的，而INMON是根据ER模型设计的。

2，KIMBALL能被分析型系统直接访问。

3，INMON里数据集市是为部门使用重新组织整理的，而KIMBALL的数据集市是包含在数据仓库里的

4，INMON里只有数据集市是按照维度设计的，而KIMBALL是全部按照维度设计的。

### 5、《Star Schema完全参考手册》的笔记-第35页

星型模型：事实表和维度表相关联，为维度表分配一个代理键，用于唯一地区分表中的每一条记录。

缓慢变化维：维度属性也会变化，但是相对于快速变化的事实表，维度变化比较缓慢，因此被称为缓慢变化维。

缓慢变化维的两种处理方法：

1，当一个维度值的源发生变化，并且不需要在星型模型中保留历史记录时，仅仅使用变化的值重写维度即可。

2，大部分的维度值变化需要记录历史。

多维数据集：

多维数据集预先计算好维度值和事实值来获取各种组合值。该方式的优势1是速度快，优势2是不受SQL限制。缺陷在于预先准备的数据量过大。

## 6、《Star Schema完全参考手册》的笔记-第3页

基本概念：

(1)操作型系统：直接支持业务系统的执行，就是通常所说的OLTP（联机事物处理系统），主要的操作类型包括插入、删除、查询、更新，使用基于第三范式的实体 - 关系设计

(2)分析型系统：支持对业务系统的评价，应该就是通常所说的OLAP（联机分析处理系统），主要的操作类型就是查询，使用维度设计（星型模型或者多维数据集）

(1)星型模型：针对关系型数据库的维度设计被称为星型模型，相关维度成为维度表中的列，事实则存储在事实表的各个列中。星型模型的示意图(2)雪花模型：星型模型中并不是严格遵守第三范式，比如在星星模型中商品维表中会带上品牌信息，按照第三范式，非关键字之间应该相互独立，品牌应该单独做一张表，这种设计被称为雪花模型，附加表通常被叫做支架表。

## 7、《Star Schema完全参考手册》的笔记-第83页

这一章内容不知道是理解问题还是翻译上的问题，没有很看明白到底是在讲什么，准备之后对照下英文版再看下，先跳过。

## 版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:[www.tushu000.com](http://www.tushu000.com)