

WILEY

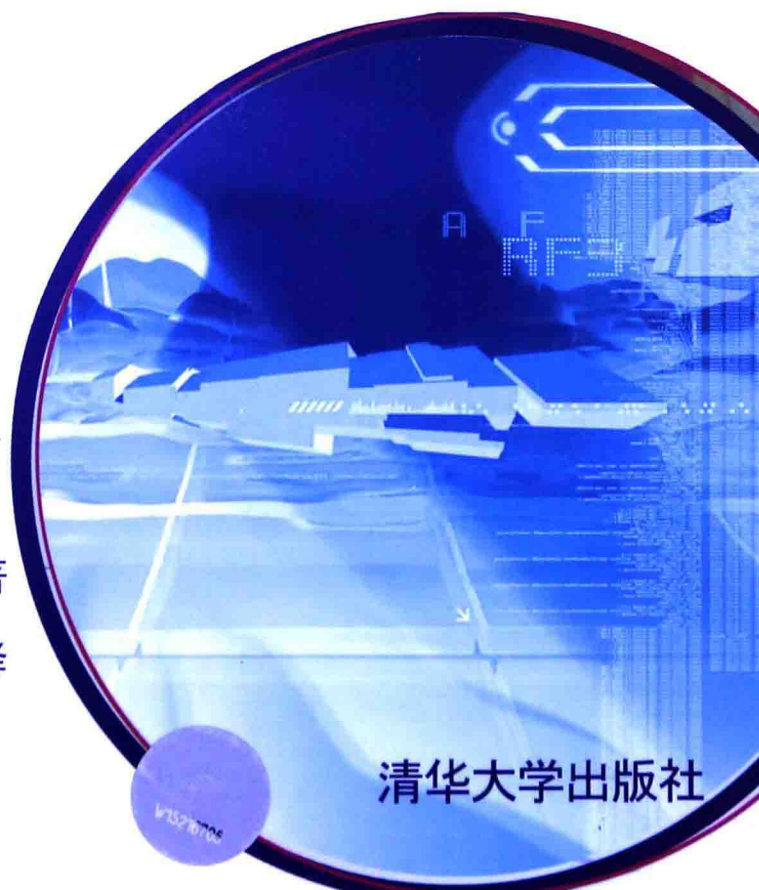
大数据应用与技术丛书

The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition

数据仓库工具箱(第3版)

——维度建模权威指南

[美] Ralph Kimball
Margy Ross 著
王念滨 周连科 韦正现 译



清华大学出版社

V328706

大数据应用与技术丛书

数据仓库工具箱(第3版)

——维度建模权威指南

[美] Ralph Kimball 著
Margy Ross

王念滨 周连科 韦正现 译



清华大学出版社

北 京

Ralph Kimball, Margy Ross

The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition

EISBN: 978-1-118-53080-1

Copyright © 2013 by Ralph Kimball and Margy Ross

All Rights Reserved. This translation published under license.

本书中文简体字版由 Wiley Publishing, Inc. 授权清华大学出版社出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

北京市版权局著作权合同登记号 图字: 01-2013-9333

本书封面贴有 Wiley 公司防伪标签, 无标签者不得销售。

版权所有, 侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

数据仓库工具箱(第3版)——维度建模权威指南 / (美)金博尔(Kimball, R.), (美)罗斯(Ross, M.) 著; 王念滨, 周连科, 韦正现 译. —北京: 清华大学出版社, 2015

(大数据应用与技术丛书)

书名原文: The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition

ISBN 978-7-302-38553-0

I. ①数… II. ①金… ②罗… ③王… ④周… ⑤韦… III. ①数据库系统 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2014)第 273663 号

责任编辑: 王 军 刘伟琴

装帧设计: 牛艳敏

责任校对: 邱晓玉

责任印制: 何 芊

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175 邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 清华大学印刷厂

经 销: 全国新华书店

开 本: 185mm×260mm 印 张: 25.5 字 数: 621 千字

版 次: 2015 年 1 月第 1 版 印 次: 2015 年 1 月第 1 次印刷

印 数: 1~4000

定 价: 69.80 元

产品编号: 056621-01

前 言

自 Ralph Kimball 于 1996 年首次出版 *The Data Warehouse Toolkit*(Wiley)一书以来，数据仓库和商业智能(Data Warehousing and Business Intelligence, DW/BI)行业渐趋成熟。尽管初期仅有部分大型公司采用，但从那时起，DW/BI 逐渐为各种规模的公司所青睐。业界已建立了数以千计的 DW/BI 系统。随着数据仓库原子数据的不断增加以及更新越来越频繁，数据容量不断增长。在我们的职业生涯中，我们见证了数据库容量从 MB 到 GB 再到 TB 甚至 PB 的发展过程，但是，DW/BI 系统面临的基本挑战并未发生重大变化。我们的工作就是管理组织中的数据并将其用于业务用户的决策制定过程中。总的来说，您必须实现这一目标，确保商务人士制定更好的决策，并从他们的 DW/BI 投资中获得回报。

自 *The Data Warehouse Toolkit* 第 1 版出版以来，维度建模作为一种主要的 DW/BI 展现技术受到广泛认可。从业者与学者都认识到数据展现要获得成功，就必须建立在简单性的基础之上。简单性是使用户能够方便地理解数据库，使软件能够方便地访问数据库的基础性的关键要素。许多情况下，维度建模就是时刻考虑如何能够提供简单性。坚定不移地回到业务驱动的场景，坚持以用户的可理解性和查询性能为目标，才能建立始终如一地服务于组织的分析需求的设计。维度建模框架将成为 BI 的平台。基于我们多年来积累的经验以及大量实践者的反馈，我们相信维度建模是 DW/BI 项目成功的关键。

维度建模还是建立集成化的 DW/BI 系统的主导结构。当您使用维度模型的一致性维度和一致性事实时，可以增量式地建立具有可实践的、可预测的、分布式的复杂 DW/BI 系统的框架。

尽管业界的一切始终在变化，但 Ralph Kimball 于 17 年前提出的核心维度建模技术经受住了时间的考验。诸如一致性维度、缓慢变化维度、异构产品、无事实的事实表以及企业数据仓库总线矩阵等概念仍然是全球范围内设计论坛所讨论的问题。最初的概念通过新的和互补的技术被逐渐完善并强化。我们决定对 Kimball 的书籍发行第 3 版，因为我们感到有必要将我们所收集到的维度建模经验汇集到一书中。我们每个人都具有 30 年以上的关注决策支持、数据仓库和业务智能的经历。我们希望分享在职业生涯中反复利用的维度建模模式。本书还包含基于现实场景的特定的实践性的设计建议。

本书的目标是提供维度建模技术的一站式商店。正如书名所体现的那样，本书是一本维度设计原则和技术的工具箱。本书既能满足那些刚刚进入维度 DW/BI 行业的新手的需要，也描述了许多高级概念以满足那些长期战斗在这一行业的老手的需要。我们相信本书

在维度建模主题方面所涵盖内容的深度是独一无二的。本书是权威性的指南。

预期读者

本书面向数据仓库和商业智能设计人员、实践人员和管理人员。此外,积极参与 DW/BI 项目的业务分析人员和数据管理者也会发现本书内容对他们来说是非常有益的。

即使您并未直接负责维度模型的开发工作,但我们相信熟悉维度建模的概念对项目组所有成员都是非常重要的。维度模型对 DW/BI 实现的许多方面都有影响,从业务需求的转换开始,通过获取、转换和加载(ETL)过程,最后到通过商业智能应用发布数据仓库的整个过程。由于涉及内容的广泛性,无论您是主要负责项目管理、业务分析、数据结构、数据库设计、ETL 和 BI 应用,还是教育和支持,都需要熟悉维度建模。本书适合于方方面面的读者。

对那些已经阅读过本书前期版本的读者来说,在本书中将发现一些熟悉的案例研究,然而,这些案例都被更新了,增加了更丰富的内容,几乎每个案例都包括样例企业数据仓库总线矩阵。我们为新的主题区域(包括大数据分析)提供了相应内容。

本书内容偏向对技术的讨论。主要从关系数据库环境出发讨论维度建模,这一环境与联机分析处理(OLAP)存在的细微差别在适当之处都进行了说明。本书假定读者对关系数据库概念有一定的了解,例如表、行、键和连接等。鉴于我们在讨论维度模型时不采用某一特定的方法,所以不会就某一特定数据库管理系统具体的物理设计和调整指导展开深入的讨论。

各章预览

本书将围绕一系列商业场景或案例研究进行组织。我们相信通过实例来研究设计技术是最有效的方法,因为这样做可以使我们分享非常实际的指导以及现实世界的适用经验。尽管未提供完整的应用或业界解决方案,但这些案例可用来讨论出现在维度建模中的模式。据我们的经验来看,通过远离自己所熟悉的复杂问题,更容易抓住设计技术的要素。阅读过本书以前各版本的读者对这一方法的反应非常积极。

请注意我们在第 2 章未采用案例研究方法。鉴于 Kimball 集团所发明的维度建模技术得到行业的广泛认可,我们整理出了这些技术,并简短地进行了描述。尽管并不指望读者会像阅读其他章节那样从头到尾仔细阅读,但我们觉得这一以技术为中心的章节对读者来说是一种有益的参考。

除第 2 章以外,本书其他章节相互关联。我们以基本概念开始,随着内容的展开,介绍了更高级的内容。读者应该顺序阅读各章。例如,除非您阅读了第 16 章之前有关零售、采购、订单管理和客户关系管理的那几章的内容,否则很难理解第 16 章的相关内容。

对那些已经阅读过本书以前版本的读者来说,可能会忽略前面几章。尽管对前面的事实和维度比较熟悉,但不希望读者跳过太多的章节,否则可能会错过一些已经更新的基本概念。

第 1 章：数据仓库、商业智能及维度建模初步

本书以数据仓库、商业智能及维度建模入门开始，探讨了整个 DW/BI 结构的所有组件并建立了本书其他章节所用到的核心词汇。消除了一些有关维度建模的神化和误解。

第 2 章：Kimball 维度建模技术概述

本章描述了超过 75 个维度建模技术以及模式。

第 3 章：零售业务

零售是用于描述维度建模的经典实例。我们之所以从该行业开始讨论是因为该行业为大众所熟悉。并不需要大家都对该行业有非常深入的了解，因为我们主要是希望通过该章的学习使大家能够关注核心的维度建模概念。该章以设计维度模型常用的 4 步过程开始，对维度表开展了深入的研究，包括贯穿全书反复使用的日期维度。同时，我们还讨论了退化维度、雪花维度以及代理键。即使您并不从事零售行业的工作，也需要仔细阅读这一章，因为该章是其他各章的基础。

第 4 章：库存

该章是对第 3 章零售业讨论的延伸，讨论了零售业的另一个案例，但请将注意力转移到零售业的另一个业务过程上。该章介绍了企业数据仓库总线架构以及具有一致性维度的总线矩阵。这些概念对那些希望建立集成的、可扩展的 DW/BI 架构的人来说是非常关键的一章。我们还讨论了三种基础类型的事实表：事务、周期快照和累积快照。

第 5 章：采购

该章强调了在构思 DW/BI 环境时，企业组织的价值链的重要性。我们还探讨了用于处理缓慢变化维度属性的一系列基础和高级的技术；讨论了基本的类型 1(重写)、类型 2(增加行)和类型 3(增加列)，并在此基础上介绍了类型 0，以及类型 4~类型 7。

第 6 章：订单管理

在研究该案例时，我们考察了在 DW/BI 系统中常常需要首先考虑实现的业务过程，因为这些过程支持核心业务性能度量——我们将哪些商品以何种价格卖给哪些顾客？讨论了在模式中扮演多种角色的维度。还讨论了在处理订单管理信息时，建模人员将会面对的常见挑战，例如，表头/列表项考虑、多币种或多种度量单位，以及五花八门的事务标识符的杂项维度等。

第 7 章：会计

该章主要讨论了建模数据仓库中的总账信息。描述了处理年度-日期(year-to-date)事实和多种财政日历，以及将多个业务过程中的数据合并到事实表的适当方法。还对维度属性

层次提供了详细的指导，从简单的规范的固定深度层次到包含参差不齐的可变深度层次的桥接表。

第8章：客户关系管理

大量的DW/BI系统建立在需要更好地理解客户并向其提供服务的前提下。该章讨论了客户维度，包括标准化地址和处理多值维度属性的桥接表。该章还讨论了对复杂的客户行为建模的模式，以及如何从多个数据源中合并客户数据的方法。

第9章：人力资源管理

该章讨论了人力资源维度模型具有的几种特性，包括那些维度表行为类似事实表的情况。该章讨论了分析方案软件包，以及对递归管理层次及调查问卷的处理方法。对几种处理多值技能关键词属性的方法进行了比较。

第10章：金融服务

银行案例研究探讨了那些每个业务列表项具有特定描述性属性和性能度量的异构产品的超类和子类模式的概念。显然，并不是只有金融服务行业需要处理异构产品。该章还讨论了账户、客户和家庭之间所存在的复杂关系。

第11章：电信

该章从结构上来看与前几章有一些差别，主要是为了鼓励读者在执行维度模型设计评审时辩证地考虑问题。该章从乍看似乎是合理的维度设计开始。您能够从中发现什么问题？此外，该章还讨论了地理位置维度的特性。

第12章：交通运输

该章的案例考察了不同粒度级别的相关事实表，指出描述旅程或网络中区段的事实表的特性。进一步深入考察了日期和时间维度，包括特定国家日历和跨多个时区的同步问题。

第13章：教育

该章考察了几类无事实的事实表。此外，探讨了处理学生应用和研究基金申请流水线的累积快照事实表。该章为读者提供了了解教育团体中各种各样业务过程的机会。

第14章：医疗卫生

我们所遇见的最复杂的模型来自医疗卫生行业。该章描述了处理此类复杂性的方法，包括使用桥接表建模多项诊断以及与病人治疗事件相关的提供商。

第 15 章：电子商务

该章主要关注点击流 Web 数据的细节，包括其独有的多维性。该章还介绍了用于更好地理解包含顺序步骤的任何过程的步骤维度。

第 16 章：保险业务

作为本书提供的最后一个案例研究，我们将书中前述的各个模式关联到一起。可将该章看成是对以前各个章节的总结，因为这些建模技术按层次划分。

第 17 章：Kimball DW/BI 生命周期概述

通过前述章节的介绍，您已经熟悉了维度模型的设计方法，该章概述了在典型的 DW/BI 项目生命周期中将会遇到的活动。该章是对由我们与 Bob Becker、Joy Mundy 和 Warren Thornthwaite 共同编写的 *The Data Warehouse Lifecycle Toolkit, Second Edition*(Wiley, 2008)一书的简短概述。

第 18 章：维度建模过程与任务

该章对处理 Kimball 生命周期中的维度建模任务提出了具体的建议。本书的前 16 章包含维度建模技术和设计模式；该章描述责任、操作方式，以及维度建模设计活动的发布物。

第 19 章：ETL 子系统与技术

在构建 DW/BI 环境时，ETL(获取、转换、加载)系统将会消耗大量的时间和精力，与其他部分工作比较，几乎不成比例。仔细考虑的最佳实践揭示了在几乎所有维度数据仓库后端都将发现的 34 个子系统。该章首先讨论了在设计 ETL 系统之前必须考虑的需求和约束，然后描述了 34 个与获取、清洗、一致性、发布和管理有关的子系统。

第 20 章：ETL 系统设计与开发过程和任务

该章深入探讨了与 ETL 设计和开发活动有关或无关的具体技术。那些对 ETL 负有责任的读者都应该阅读本章。

第 21 章：大数据分析

最后一章主要关注大数据这一正在流行的主题。我们认为，大数据是对 DW/BI 系统的自然扩展。首先讨论了几种可选的结构，包括 MapReduce 和 Hadoop，描述了这些可选系统如何与当前的 DW/BI 结构共存的问题。接着讨论了针对大数据的管理、结构、数据建模和数据治理的最佳实践。

Web 资源

Kimball 集团网站 www.kimballgroup.com 包含许多补充的维度建模内容和资源:

- 注册 Kimball Design Tips 可收到有关维度建模和 DW/BI 主题的实践指导。
- 访问目录可获得 300 个设计技巧和文章。
- 通过学习 Kimball 大学公开的和现场的课程,可获得高质量、独立于提供商的教育,并分享我们的经验和文章。
- 获得 Kimball 集团咨询服务以利用我们几十年积累的有关 DW/BI 的宝贵经验。
- 向 Kimball 论坛的其他维度设计参与者提问。

小结

本书的写作目标是基于作者 60 多年来从实际业务环境中获得的经验和来之不易的教训,为读者提供正式的维度设计和开发技术。DW/BI 系统必须以业务用户的需求来驱动,如此才能真正从维度角度设计和展现。我们坚信,如果您能够接受这一前提,将会朝建立成功的 DW/BI 系统迈出巨大的一步。

既然知道从何开始,请开始仔细阅读本书。在第 1 章中将讨论 DW/BI 和维度建模的基本内容,确保每个人对关键术语和结构性概念具有统一的认识。

致 谢

首先，请让我对成千上万阅读过我们出版的工具箱系列书籍的读者、参加过我们开设的课程的学员以及参与我们项目咨询的客户表示感谢。我们为你们教授的与从你们那里学到的一样多。一句话，你们对数据仓库和商业智能行业的发展产生了深远的积极影响，祝贺大家！

Kimball 集团的同事 Bob Becker、Joy Mundy 和 Warren Thornthwaite 等，在将近 30 年的时间里，与我们一起将本书所描述的技术应用过差不多上千次。本书所讲述的每种技术都经历了实践的检验。非常感谢他们对本书的投入和反馈——更重要的是，与 Julie Kimball 一道，作为商业合作者所经历的这些岁月。

John Wiley & Sons 出版公司的执行编辑 Bob Elliott、项目编辑 Maureen Spears，以及 Wiley 公司出版小组的工作人员对本书的出版投入了极大的热情并开展了出色的工作。与你们一起工作非常愉快。

感谢我们的家庭，感谢你们对我们职业生涯一如既往的无条件的支持。亲爱的 Julie Kimball 与 Scott Ross 以及孩子们 Sara Hayden Smith、Brian Kimball 和 Katie Ross，你们都以各种方式为本书的出版做出了贡献。

作者简介

Ralph Kimball 是 Kimball 集团的创建者。从 20 世纪 80 年代中期以来，他一直是数据仓库和商业智能行业维度建模方法的思想开拓者。大量 IT 专业人士接受过其教育。自 1996 年以来，由他及其同事们所撰写的工具箱系列书籍一直是最受读者青睐的书籍。Ralph Kimball 曾就职于 Metaphor 并建立了 Red Brick 系统，他在施乐 Palo Alto 研究中心(PARC)工作期间，与他人一起共同发明了星型工作站，这是首个利用视窗、图标和鼠标的商业产品。Ralph Kimball 毕业于斯坦福大学电子工程系并获得博士学位。

Margy Ross 是 Kimball 集团总裁。自 1982 年以来，她主要关注数据仓库和商业智能，强调业务需求和维度建模的重要性。与 Ralph Kimball 一样，Margy Ross 也为许多学生讲授过维度设计最佳实践，她与 Ralph Kimball 合作，共同撰写了 5 本工具箱序列书籍。Margy Ross 曾工作于 Metaphor 并与其他人共同创立了 DecisionWorks 咨询公司。她毕业于美国西北大学工业工程系并获得硕士学位。

译者序

围绕 Kimball 与 Inmon 先生有关数据仓库体系架构的论战已持续多年, Kimball 数据仓库架构并非唯一选择。正如作者在书中所述:“我们承认许多成功构建数据仓库/商业智能系统的组织采用其他架构方法。”但论战的结果是提出了大量促使数据仓库研究和应用得以快速发展的新技术和新方法。

本书就是这些新技术、新方法的集大成者。书中对许多技术的探讨来源于实际需求。因此,本书的写作是以案例研究为本体,根据应用的需求提出解决方案。针对应用中反映出来的问题,适时提出相关技术和方法。许多问题对于应用开发者来说都似曾相识,例如对于缓慢变化维度的探讨。实际上,我们往往认为,数据仓库中的数据主要应用于分析工作,因此不太关注个别数据的变化对整体带来的影响。本书却从实际业务需求出发,提出了从类型 0~类型 7 等 8 种方法,给人以“于无声处听惊雷”的感觉。

Kimball 先生是数据仓库业界的翘楚之一。本书提出的许多技术贴近应用,对技术的讨论深入浅出,便于读者学习和应用。例如维度属性层次、桥接表、支架表、无事实的事实表等都是解决在实际工作中经常遇到的棘手问题的良好方法。

对于这本经典之作,译者本着“诚惶诚恐”的态度,在翻译过程中力求“信、达、雅”,但是鉴于译者水平有限,错误和不当之处在所难免,恳请读者批评指正。

参加本书翻译工作的还有博士研究生祝官文、王瑛琪、何鸣、张爽等,在此一并致谢。

译者

目 录

第 1 章 数据仓库、商业智能及 维度建模初步..... 1	
1.1 数据获取与数据分析的区别..... 1	
1.2 数据仓库与商业智能的目标..... 2	
1.3 维度建模简介..... 5	
1.3.1 星型模式与 OLAP 多维 数据库..... 6	
1.3.2 用于度量的事实表..... 7	
1.3.3 用于描述环境的维度表..... 9	
1.3.4 星型模式中维度与事实的 连接..... 11	
1.4 Kimball 的 DW/BI 架构..... 14	
1.4.1 操作型源系统..... 14	
1.4.2 获取-转换-加载(ETL)系统..... 14	
1.4.3 用于支持商业智能决策的 展现区..... 16	
1.4.4 商业智能应用..... 17	
1.4.5 以餐厅为例描述 Kimball 架构..... 17	
1.5 其他 DW/BI 架构..... 19	
1.5.1 独立数据集市架构..... 19	
1.5.2 辐射状企业信息工厂 Inmon 架构..... 20	
1.5.3 混合辐射状架构与 Kimball 架构..... 22	
1.6 维度建模神话..... 22	
1.6.1 神话 1: 维度模型仅包含 汇总数据..... 23	
1.6.2 神话 2: 维度模型是部门级而 不是企业级的..... 23	

1.6.3 神话 3: 维度模型是 不可扩展的..... 23	
1.6.4 神话 4: 维度模型仅用于 预测..... 23	
1.6.5 神话 5: 维度模型不能被 集成..... 24	
1.7 考虑使用维度模型的 更多理由..... 24	
1.8 本章小结..... 25	
第 2 章 Kimball 维度建模技术概述..... 27	
2.1 基本概念..... 27	
2.1.1 收集业务需求与数据实现..... 27	
2.1.2 协作维度建模研讨..... 27	
2.1.3 4 步骤维度设计过程..... 28	
2.1.4 业务过程..... 28	
2.1.5 粒度..... 28	
2.1.6 描述环境的维度..... 28	
2.1.7 用于度量的事实..... 29	
2.1.8 星型模式与 OLAP 多维 数据库..... 29	
2.1.9 方便地扩展到维度模型..... 29	
2.2 事实表技术基础..... 29	
2.2.1 事实表结构..... 29	
2.2.2 可加、半可加、不可加 事实..... 29	
2.2.3 事实表中的空值..... 30	
2.2.4 一致性事实..... 30	
2.2.5 事务事实表..... 30	

2.2.6	周期快照事实表	30	2.5.7	类型 6: 增加类型 1 属性到 类型 2 维度	36
2.2.7	累积快照事实表	30	2.5.8	类型 7: 双类型 1 和 类型 2 维度	36
2.2.8	无事实的事实表	31	2.6	处理维度层次关系	37
2.2.9	聚集事实表或 OLAP 多维数据库	31	2.6.1	固定深度位置的层次	37
2.2.10	合并事实表	31	2.6.2	轻微参差不齐/可变 深度层次	37
2.3	维度表技术基础	31	2.6.3	具有层次桥接表的参差不齐/ 可变深度层次	37
2.3.1	维度表结构	31	2.6.4	具有路径字符属性的可变 深度层次	37
2.3.2	维度代理键	32	2.7	高级事实表技术	37
2.3.3	自然键、持久键和 超自然键	32	2.7.1	事实表代理键	37
2.3.4	下钻	32	2.7.2	蜈蚣事实表	38
2.3.5	退化维度	32	2.7.3	属性或事实的数字值	38
2.3.6	非规范化扁平维度	32	2.7.4	日志/持续时间事实	38
2.3.7	多层次维度	32	2.7.5	头/行事实表	38
2.3.8	文档属性的标识与指示器	33	2.7.6	分配的事实	38
2.3.9	维度表中的空值属性	33	2.7.7	利用分配建立利润与 损失事实表	38
2.3.10	日历日期维度	33	2.7.8	多种货币事实	39
2.3.11	扮演角色的维度	33	2.7.9	多种度量事实单位	39
2.3.12	杂项维度	33	2.7.10	年-日事实	39
2.3.13	雪花维度	33	2.7.11	多遍 SQL 以避免事实表间的 连接	39
2.3.14	支架维度	34	2.7.12	针对事实表的时间跟踪	39
2.4	使用一致性维度集成	34	2.7.13	迟到的事实	40
2.4.1	一致性维度	34	2.8	高级维度技术	40
2.4.2	缩减维度	34	2.8.1	维度表连接	40
2.4.3	跨表钻取	34	2.8.2	多值维度与桥接表	40
2.4.4	价值链	34	2.8.3	随时间变化的多值桥接表	40
2.4.5	企业数据仓库总线架构	35	2.8.4	标签的时间序列行为	40
2.4.6	企业数据仓库总线矩阵	35	2.8.5	行为研究分组	40
2.4.7	总线矩阵实现细节	35	2.8.6	聚集事实作为维度属性	41
2.4.8	机会/利益相关方矩阵	35	2.8.7	动态值范围	41
2.5	处理缓慢变化维度属性	35	2.8.8	文本注释维度	41
2.5.1	类型 0: 原样保留	35	2.8.9	多时区	41
2.5.2	类型 1: 重写	35	2.8.10	度量类型维度	41
2.5.3	类型 2: 增加新行	36			
2.5.4	类型 3: 增加新属性	36			
2.5.5	类型 4: 增加微型维度	36			
2.5.6	类型 5: 增加微型维度及 类型 1 支架	36			

2.8.11	步骤维度	41	3.8	抵制规范化的冲动	70
2.8.12	热交换维度	42	3.8.1	具有规范化维度的 雪花模式	70
2.8.13	抽象通用维度	42	3.8.2	支架表	72
2.8.14	审计维度	42	3.8.3	包含大量维度的蜈蚣 事实表	72
2.8.15	最后产生的维度	42	3.9	本章小结	74
2.9	特殊目的模式	42	第 4 章	库存	75
2.9.1	异构产品的超类与 子类模式	43	4.1	价值链简介	75
2.9.2	实时事实表	43	4.2	库存模型	76
2.9.3	错误事件模式	43	4.2.1	库存周期快照	76
第 3 章	零售业务	45	4.2.2	库存事务	79
3.1	维度模型设计的 4 步过程	46	4.2.3	库存累积快照	80
3.1.1	第 1 步: 选择业务过程	46	4.3	事实表类型	81
3.1.2	第 2 步: 声明粒度	46	4.3.1	事务事实表	81
3.1.3	第 3 步: 确定维度	47	4.3.2	周期快照事实表	82
3.1.4	第 4 步: 确定事实	47	4.3.3	累积快照事实表	82
3.2	零售业务案例研究	47	4.3.4	辅助事实表类型	83
3.2.1	第 1 步: 选择业务过程	49	4.4	价值链集成	83
3.2.2	第 2 步: 声明粒度	49	4.5	企业数据仓库总线架构	84
3.2.3	第 3 步: 确定维度	50	4.5.1	理解总线架构	84
3.2.4	第 4 步: 确定事实	50	4.5.2	企业数据仓库总线矩阵	85
3.3	维度表设计细节	53	4.6	一致性维度	89
3.3.1	日期维度	53	4.6.1	多事实表钻取	89
3.3.2	产品维度	56	4.6.2	相同的一致性维度	89
3.3.3	商店维度	59	4.6.3	包含属性子集的缩减上卷 一致性维度	90
3.3.4	促销维度	60	4.6.4	包含行子集的缩减 一致性维度	91
3.3.5	其他零售业维度	62	4.6.5	总线矩阵的缩减一致性 维度	91
3.3.6	事务号码的退化维度	63	4.6.6	有限一致性	92
3.4	实际的销售模式	63	4.6.7	数据治理与管理的重要性	92
3.5	零售模式的扩展能力	64	4.6.8	一致性维度与敏捷开发	94
3.6	无事实的事实表	65	4.7	一致性事实	94
3.7	维度与事实表键	66	4.8	本章小结	95
3.7.1	维度表代理键	66	第 5 章	采购	97
3.7.2	维度中自然和持久的 超自然键	68	5.1	采购案例研究	97
3.7.3	退化维度的代理键	68			
3.7.4	日期维度的智能键	68			
3.7.5	事实表的代理键	69			

5.2	采购事务与总线矩阵	98	6.3.3	审计维度	133
5.2.1	单一事务事实表与多事务事实表	98	6.4	用于订单整个流水线的累积快照	134
5.2.2	辅助采购快照	101	6.4.1	延迟计算	136
5.3	缓慢变化维度(SCD)基础	101	6.4.2	多种度量单位	137
5.3.1	类型 0: 保留原始值	102	6.4.3	超越后视镜	138
5.3.2	类型 1: 重写	102	6.5	本章小结	138
5.3.3	类型 2: 增加新行	104	第 7 章	会计	139
5.3.4	类型 3: 增加新属性	106	7.1	会计案例研究与总线矩阵	139
5.3.5	类型 4: 增加微型维度	108	7.2	总账数据	141
5.4	混合缓慢变化维度技术	110	7.2.1	总账周期快照	141
5.4.1	类型 5: 微型维度与类型 1 支架表	110	7.2.2	会计科目表	141
5.4.2	类型 6: 将类型 1 属性增加到类型 2 维度	111	7.2.3	结账	141
5.4.3	类型 7: 双重类型 1 与类型 2 维度	112	7.2.4	年度-日期事实	143
5.5	缓慢变化维度总结	113	7.2.5	再次讨论多币种问题	143
5.6	本章小结	114	7.2.6	总账日记账事务	143
第 6 章	订单管理	115	7.2.7	多种财务会计日历	144
6.1	订单管理总线矩阵	116	7.2.8	多级别层次下钻	145
6.2	订单事务	116	7.2.9	财务报表	145
6.2.1	事实表规范化	117	7.3	预算编制过程	146
6.2.2	维度角色扮演	117	7.4	维度属性层次	148
6.2.3	重新审视产品维度	119	7.4.1	固定深度的位置层次	148
6.2.4	客户维度	120	7.4.2	具有轻微不整齐的可变深度层次	149
6.2.5	交易维度	122	7.4.3	不整齐可变深度层次	149
6.2.6	针对订单号的退化维度	123	7.4.4	不规则层次中的共享所有权	152
6.2.7	杂项维度	124	7.4.5	随时间变化的不规则层次	153
6.2.8	应该避免的表头/明细模式	125	7.4.6	修改不规则层次	153
6.2.9	多币种	126	7.4.7	其他不规则层次的建模方法	154
6.2.10	不同粒度的事务事实	128	7.4.8	应用于不规则层次的桥接表方法的优点	156
6.2.11	另外一种需要避免的表头/明细模式	129	7.5	合并事实表	156
6.3	发票事务	130	7.6	OLAP 角色及分析方案包	157
6.3.1	作为事实、维度或两者兼顾的服务级性能	131	7.7	本章小结	158
6.3.2	利润与损益事实	131	第 8 章	客户关系管理	159
			8.1	客户关系管理概述	160

8.2	客户维度属性	162	9.3	人力资源过程的总线矩阵	187
8.2.1	名字与地址的语法分析	162	9.4	分析解决方案软件包与 数据模型	188
8.2.2	国际姓名和地址的考虑	164	9.5	递归式雇员层次	189
8.2.3	客户为中心的日期	165	9.5.1	针对嵌入式经理主键 变化的跟踪	190
8.2.4	作为维度属性的聚集事实	166	9.5.2	上钻或下钻管理层次	190
8.2.5	分段属性与记分	166	9.6	多值技能关键字属性	191
8.2.6	包含类型 2 维度变化的 计算	169	9.6.1	技能关键字桥接表	191
8.2.7	低粒度属性集合的支架表	169	9.6.2	技能关键字文本字符串	192
8.2.8	客户层次的考虑	170	9.7	调查问卷数据	193
8.3	应用于多值维度的桥接表	171	9.8	本章小结	194
8.3.1	稀疏属性的桥接表	172	第 10 章	金融服务	195
8.3.2	应用于客户多种联系方式的 桥接表	173	10.1	银行案例研究与总线矩阵	195
8.4	复杂的客户行为	173	10.2	分类维度以避免出现维度 太少的情况	196
8.4.1	客户队列的行为研究分组	173	10.2.1	家庭维度	199
8.4.2	连续行为的步骤维度	175	10.2.2	多值维度与权重因子	199
8.4.3	时间范围事实表	176	10.2.3	再谈微型维度	200
8.4.4	使用满意度指标标记 事实表	177	10.2.4	在桥接表中增加 微型维度	202
8.4.5	使用异常情景指标标记 事实表	178	10.2.5	动态值范围事实	202
8.5	客户数据集成方法	178	10.3	异构产品的超类和 子类模式	203
8.5.1	建立单一客户维度的 主数据管理	179	10.4	热可交换维度	205
8.5.2	多客户维度的局部 一致性	180	10.5	本章小结	205
8.5.3	避免对应事实表的连接	180	第 11 章	电信	207
8.6	低延迟的实现检查	181	11.1	电信业案例研究与 总线矩阵	207
8.7	本章小结	182	11.2	设计评审的一般性考虑	209
第 9 章	人力资源管理	183	11.2.1	业务需求与实际可用 资源的权衡	209
9.1	雇员档案跟踪	183	11.2.2	关注业务过程	209
9.1.1	精确的有效和失效 时间范围	184	11.2.3	粒度	210
9.1.2	维度变化原因跟踪	185	11.2.4	统一的事实表粒度	210
9.1.3	作为类型 2 属性或事实 事件的档案变化	185	11.2.5	维度的粒度和层次	210
9.2	雇员总数周期快照	186	11.2.6	日期维度	211
			11.2.7	退化维度	211