

数据仓库
最权威指南！
维度建模

The Data Warehouse Toolkit:
the Complete Guide to Dimensional Modeling

数据仓库工具箱：

维度建模的完全指南（第二版）

Kimball 经典之作

[美]Ralph Kimball, Margy Ross 著

谭明金 译

数据仓库与数据挖掘技术应用丛书

数据仓库工具箱：

维度建模的完全指南（第二版）

[美] Ralph Kimball, Margy Ross 著

谭明金 译

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书从维度建模的基本知识入手,通过给出零售、库存、采购、订单管理、客户关系管理、账目、人力资源管理、财经服务、电信与公用事业、交通、教育、卫生护理、电子商务以及保险等方面的实例研究,全面、深入、透彻与具体地介绍了利用维度模型设计与开发数据仓库的综合技术知识。同时,本书对数据仓库在当前社会与技术背景下的发展前景进行了很有远见的探索。本书先介绍基本知识,然后逐个讨论具体实例内容,最后进行综合总体分析,在内容的结构方面很有特色。

本书主要是为数据仓库设计人员、技术人员以及管理人员而写的。此外,对于积极参与数据仓库初期规划工作的业务分析人员来说,本书也有很高的学习价值。

Copyright © 2002 by Ralph Kimball and Margy Ross.

All rights reserved. Authorized translation from the English language edition published by John Wiley & Sons, Inc.

本书简体中文专有翻译出版权由 John Wiley & Sons Inc. 授予电子工业出版社。未经许可,不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号:图字:01-2002-5359

图书在版编目(CIP)数据

数据仓库工具箱:维度建模的完全指南(第二版)/(美)金博尔(Kimball,R.)著;谭明金译. —北京:电子工业出版社,2003.10

(数据仓库与数据挖掘技术应用丛书)

书名原文: The Data Warehouse Toolkit: the Complete Guide to Dimensional Modeling
ISBN 7-5053-8931-9

I. 数… II. ①数… ②谭… III. 数据库系统 IV. TP311.13

中国版本图书馆CIP数据核字(2003)第062767号

责任编辑:朱沐红 zsh@phei.com.cn

印刷:北京增富印刷有限公司

出版发行:电子工业出版社 <http://www.phei.com.cn>

北京市海淀区万寿路173信箱 邮编 100036

经销:各地新华书店

开本:787×980 1/16 印张:27 字数:467千字

版次:2003年10月第1版 2003年10月第1次印刷

印数:6000册 定价:58.00元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。联系电话:(010)68279077。质量投诉请发邮件至 zlts@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

出版说明

如果没有对海量数据进行科学分析的能力,沃尔玛的老板再精明,也绝对想不到“啤酒与尿布”这两个风马牛不相及的东西之间还有着千丝万缕的联系。而将它们放在一起,竟然增加了啤酒销量,可见数据分析的巨大威力。

信息系统数年中收集了海量数据,且数据还正以指数级增长,企业迫切地需要高效、精确、科学地分析数据,以找出其背后的寓意,进而了解企业的经营状况和外部环境,做出科学的决断,在现代激烈的竞争中胜出。所以,如何将数据点石成金,更是摆在我们面前很现实也很诱人的一个问题。

现在,很多人已经意识到数据中潜在的大量商机,并踏踏实实地进行着从数据中沙里淘金的工作。特别是在信息化的大潮中,上至政府,下到企业,从银行到电信,再到网站、超市,人们都希望用数据分析这根魔杖赢得先机。与此同时,人们也在期盼着相关书籍,以便工作中学习参考。在广泛征询专家和用户的基础上,秉着选题全面、内容经典、译者严谨的原则,我们适时地推出了这套《数据仓库与数据挖掘技术应用丛书》,以飨读者。本丛书有如下几本:

- 数据仓库基础
- OLAP 解决方案: 多维信息系统的构建技术
- 数据仓库工具箱: 维度建模的完全指南(第二版)
- 数据仓库生命周期工具箱: 设计、开发和部署数据仓库的专家方法
- 数据仓库及其在电信领域中的应用
- 疑难数据仓库专家解决方案
- IBM 数据仓库和商业智能工具
- 可视化数据挖掘: 数据可视化和挖掘的技术和工具
- 点击流数据仓库
- Web 数据挖掘: 将客户数据转化为客户价值
- 企业信息工厂
- 机器学习与数据挖掘: 方法和应用

本丛书既包括商业智能(BI)的基础——数据仓库(DW),也包括数据仓库上的两类不同目的的数据增值操作——联机分析处理(OLAP)和数据挖掘(DM);既覆盖基础理论,如数据仓库基础,又提供不同领域的解决方案,如数据仓库在电信、银行、保险等领域的应用。

本丛书来自国外数据库领域一些著名作者的畅销书,以及国内第一线实施

者的精心总结。如一直位居 AMAZON 畅销书榜的数据仓库领域的畅销书作家 Ralph Kimball 的《数据仓库工具箱：维度建模的完全指南（第二版）》、《数据仓库生命周期工具箱：设计、开发和部署数据仓库的专家方法》，数据仓库之父 William H.Inmon 的《企业信息工厂》（Corporate Information Factory）等。

丛书的译者均来自工作在该领域一线的人员，既有该领域的理论和实践经验，又具备中英文翻译的功底。且多位译者先前均已读过原著，所以，自感翻译的过程不再枯燥，而是情趣盎然，乐在其中。

出版高品位、高品质的图书是博文视点的努力目标。希望您对我们的工作多提宝贵意见。您的意见是我们创造精品的动力源泉。

如果您希望将您的工作经验感悟等总结成书，我们将为您提供一流的服务，共创精品图书。

我们的联系方式如下：

地址：北京复兴路 47 号天行建商务大厦 604

邮编：100036

电话：010-51922832，68216158

传真：010-51922823

E-mail: jsj@phei.com.cn; zsh@phei.com.cn

博文视点资讯有限公司

2003 年 10 月

博文视点资讯有限公司 (BROADVIEW Information Co.,Ltd.) 于 2003 年 6 月 18 日正式成立, 是信息产业部直属的中央一级科技与教育出版社——电子工业出版社 (PHEI) 与国内最大的 IT 技术网站 CSDN.NET 和最具专业水准的 IT 杂志社《程序员》合资成立的以 IT 图书出版为主业、开展相关信息和知识增值服务的资讯公司。

我们的理念是: 创新专业出版体制; 培养职业出版队伍; 打造精品出版品牌; 完善全面出版服务。

秉承博文视点的理念, 博文视点的产品线为面向 IT 专业人员的出版物和相关服务。博文视点将重点做好以下工作:

- (1) 在技术领域开发专业作(译)者群体和高质量的原创图书
- (2) 在图书领域建立专业的选题策划和审读机制
- (3) 在市场领域开创有效的宣传手段和营销渠道

博文视点有效地综合了电子工业出版社、《程序员》杂志社和 CSDN.NET 的资源和人才, 建立全新专业的立体出版机制, 确立独特的出版特色和优势, 将打造 IT 出版领域的著名品牌, 并力争成为中国最具影响力的专业 IT 出版和服务提供商。

作为合资公司, 博文视点的团队融合了各方面的精英力量: 原电子工业出版社 IT 图书专业出版实力的代表部门——计算机图书事业部的团队; 《程序员》杂志和 CSDN 网站的主创人员; 著名 IT 专业图书策划人周筠女士及其创作群。这是一个整合专业技术人员和专业出版人员的团队; 这是一个充满创新意识和创作激情的团队; 这是一个不断进取, 追求卓越的团队。

原 PHEI 计算机图书事业部以及武汉的创作团队, 在合资之前就已经各自具备了很强的策划和出版实力, 并创造出了优秀的出版业绩。而《程序员》杂志社则是国内最有影响力的 IT 专业杂志, 《程序员》杂志的年发行量超过 80 万份。CSDN 网站是国内最大的 IT 技术网站, 到 2003 年 5 月底, CSDN 网站的注册会员已超过了 54 万, 这两者已成为中国 IT 技术交流与推广的最佳平台, 积累了丰富的作者和读者资源。

电子工业出版社与《程序员》杂志和 CSDN 网站的合作以最有效率的方式形成了出版资源、媒体资源、网络资源的整合和互动, 成为 2003 年 IT 出版界倍受瞩目的事件。

“技术凝聚实力, 专业创新出版”, BROADVIEW 与您携手共迎信息时代的机遇与挑战!

译者序

数据仓库是近年来适应利用数据支持决策分析的强烈需求而发展起来的数据库应用技术。诚然，数据仓库以数据库为基础，但它在需求、客户、体系结构与运行机制等方面与数据库存在重大的不同。Ralph Kimball 说：“我们花了二十年的时间往数据库中加入数据，现在该是拿出来使用的时候了。”

不过，正如许多人错误地认为“数据仓库只是数据库在企业某个地方的拷贝”一样，有人认为，以传统数据库单一数据资源为中心并主要按记录形式进行的插入、更新与查询等数据库日常事务操作是必不可少的，而数据仓库分析型应用则是可选的。这种看法似乎有道理，但是随着时代与技术的高速发展，谁愿意仅仅停留在数据本身的操作之上并依靠人力做出分析决策，而不通过精确、高效、快速与全面的数据分析去获取更大的效益呢？从这个意义上讲，数据仓库将不仅显得必要，而且是数据处理的发展方向。这便是数据仓库能掀起数据库热的重要原因。

Ralph Kimball 是数据仓库界最知名的演讲者、咨询师与授课人之一。他致力于数据仓库的应用与教学已有二十多年的历史。几年来，Ralph Kimball 独立或者与他人合作编写了多部数据仓库方面的专著。其中，1996 年《数据仓库工具箱》一书（本书的第一版）的出版被认为是数据仓库方面具有里程碑意义的事件，该书也成为畅销书。目前，Ralph Kimball 是数据仓库方面的权威。他提出的许多维度建模概念被广泛应用于数据仓库的设计与开发之中，相关书籍逐渐成为这方面的经典著作。

阅读 Ralph Kimball 的书，你会从各个方面感受到一位老师，或者说大师正在给你授课。当然，Ralph Kimball 的语气似乎略显主观，但不妨认为，正是他的造诣才使其见解显得鲜明而直接。

参与本书翻译的还有王璐、沈鑫剡、胡勇强、李志、白志勇、谭晓辉、王光金、唐捷、陈林等，在此一并致谢。

数据仓库是一门新兴技术，参考资料不多，术语使用很不统一，并且本书涉及其他行业的较多专业知识，加之译者水平有限，译文中必定存在错误和不当之处，恳请读者不吝赐教。Ralph Kimball 说：“我们从你那里学到的与我们教授给你的一样多。”译者更深信，“从你那里学到的与我翻译给你的一样多！”

译者

2003 年 3 月

致 谢

首先，我们要感谢成千上万阅读本书的读者、接受培训的学员以及进行项目咨询的客户。我们从你那里学到的与我们教授给你的一样多。总的来说，你对数据仓库行业产生了深远的积极影响。祝贺你！

如果没有业务伙伴的帮助，本书是不可能写成的。我们要感谢 Ralph Kimball 协会的 Julie Kimball 在发起这项工程时所表现出的远见与果断。在 Julie 作为催化剂而开始活动的同时，DecisionWorks 咨询机构的 Bob Becker 承担了保持工程得以运转的设计、复查与宣传等一揽子工作。我们谨对他们的巨大帮助表示诚挚的谢意。

本书的写作还得到了朋友们的帮助。他们为书中的某些章节提供了宝贵意见或反馈信息。我们还要感谢 DecisionWorks 的 Bill Schmarzo、Attachmate 公司的 Charles Hagensen 与 InfoDynamics 的 Warren Thornthwaite 分别为本书的第 6、7 与 16 章提出的建议。

本书的编辑 John Wiley & Sons 的 Bob Elliott 与整个 Wiley 团队从技能、鼓励与热情等方面为这项工程提供了支持。我们还要感谢《智能企业》(Intelligent Enterprise) 的主编 Justin Kestelyn 允许本书使用一些 Ralph 的论文材料。

感谢家人在需要的时候出现在我们身边，并谅解我们对时间的占用。爱人 Julie Kimball 与 Scott Ross 及其孩子 Sara Hayden Smith, Brain Kimball 与 Katie Ross 都为本书付出了许多，而且这经常是在无意中做出的。感谢你们无条件的支持。

前 言

数据仓库行业自 1996 年 Ralph Kimball 发表《数据仓库工具箱》第一版 (Wiley 出版) 以来, 已经逐步走向成熟。虽然最初只是一些大型公司涉足其间, 但从那时起, 数据仓库技术开始被各种规模的机构所使用。目前, 产业界已经建造了成千上万的数据仓库, 数据容量正随着日益增多的原子型数据的加入以及十分频繁的更新操作的进行而持续地增长。同时, 越来越多的用于帮助人们设计、开发与应用数据仓库的工具, 纷纷由厂家推向市场。最为重要的是, 业务人员通过使用数据仓库, 能够更好地做出决策并因数据仓库的投资而获取回报。

《数据仓库工具箱》第一版的发表, 使维度建模方法作为展示数据仓库的主导技术被广泛地接受。数据仓库方面的普通从业人员与专家一样都意识到, 数据仓库的展示只有立足于简单性才有望获得成功。简单性是使数据库能够容易被用户所理解, 以及使数据库软件能高效地运行的基本要件。在很大程度上, 简单性掌控着维度建模应用的命脉。注重以业务需求作为考虑问题的出发点, 并坚持将用户的可理解性与查询性能作为设计的目标, 就能够建立起满足机构分析需要的完备设计。根据我们自己的经验以及来自许多像你一样的公司从业人员的大量反馈来看, 我们深信, 维度建模对于发起一个成功的数据仓库应用是绝对重要的。

同时, 维度建模方法还在建立分布式数据仓库系统方面, 发挥着惟一配套体系的作用。通过在一组维度模型中使用一致的维度与一致的事实, 可以得到一个可预见的实用框架, 从而逐步、递增地建立没有中心环节的复杂数据仓库系统。

虽然数据仓库产业内的一切都在迅速地变化, 但 Ralph Kimball 于 6 年前发表的维度建模核心技术却经受住了时间的检验。诸如渐变维度、异构产品、非事实型事实表与体系结构数据中心这样的一些概念, 今天仍然在世界范围的数据仓库设计专题讨论会议上持续得到关注。某些原始概念已经随着新的补充方法的出现而得到修改与扩展。决定出版 Kimball 原著第二版的原因是, 我们认为, 将维度建模方面的集体思想通过成套丛书的形式组织在一起, 将

是非常有用的。20 多年来，本书的每位作者都一直致力于开展决策支持与数据仓库方面的工作。我们希望把数据仓库从业经历中反复出现的维度建模模式，拿出来与你一起分享。本书配有许多基于现实世界的应用情景给出的具体而可行的推荐方案。

本书的目标是为维度建模技术提供一个百货店。正如书名所体现的那样，本书是关于维度建模原理与技术的工具箱。我们在注重给出那些刚刚接触维度数据仓库的读者所需要的知识的同时，也对一些高级概念进行了描述，以满足那些在这方面具有一定基础的读者的需要。我们相信，本书在覆盖维度建模主题的内容深度方面是别具一格的。

▲ X.1 读者对象

本书主要是为数据仓库设计人员、技术人员以及管理人员而写的。此外，对于积极参与数据仓库初期规划工作的业务分析人员来说，本书也有很高的学习价值。

即使不直接负责维度模型的开发工作，但作为数据仓库项目组的成员，每个人对维度建模的概念有所了解也是非常重要的。维度模型对数据仓库实施的多个方面都有影响，这些方面涵盖从业务需求转换开始，经过数据转储，最后到通过分析型应用对数据仓库进行揭示等环节。由于这些工作涉及面十分广泛，因此项目团队需要对维度建模非常精通，而不管各自的主要职责内容是项目管理、业务需求分析、数据体系结构、数据库设计、数据转储、分析型应用还是培训与保障等。本书的写作就考虑了读者在这些方面的广泛需要。

对于已经阅读过本书第一版的读者来说，将会看到有些熟悉的案例研究重新出现在了本书中，不过，本书还是对它们进行了很大的更新，并增加了更为丰富的内容。书中为包括卫生护理、电信与电子商务等在内的新行业，开发了一些简单的应用。此外，本书也为诸如人力资源、账目管理、采购与客户关系管理这样的业务功能，引入了较多的可跨行业横向使用的案例研究。

本书略微涉及了一些技巧性的内容。维度建模主要是以关系型数据库作为背景来进行讨论的，并假定读者已经具备了诸如表、行、关键字与连接之类的关系型数据库概念方面的基本知识。这样做是考虑到如果不以一门特定技术为背景，而只是对维度模型进行泛泛的讨论，就不可能深入到具体的物

理设计内容中去，从而也就不可能为任何一种数据库管理系统提供统一的指导。

X.2 各章预览

本书是围绕一系列业务的简单应用或者案例研究来组织内容的。我们认为，对照实例来开发设计技术是特别有效的方法，因为这可以使人体体会到实实在在的启发。虽然并不打算将这些实例当做完整的应用或者行业方案，但它们也确实起到了作为讨论维度建模中出现的模式框架的作用。经验表明，通过从所有自己过于熟悉的应用复杂性中挣脱出来而考虑另外的业务，通常就更容易抓住设计技术的主要部分。第一版的读者已经对这种方法表示出浓厚的兴趣。

本书各章是相互照应的。翻开书本首先看到的是些基本概念，然后才逐步引入比较高级的内容。每位读者最好按顺序来阅读所有各章。比如说，除非已经阅读了前面介绍零售、采购、订单管理与客户关系管理等内容的几章，否则对于第 15 章描述保险的内容是很难理解的。

看过第一版的读者可能倾向于跳过本书的开头几章。虽然前面关于事实与维度的背景知识可能是你再熟悉不过的草根皮，但还是希望你不要跑得太过于靠前。例如，尽管第一个实例研究确实像第一版所做的那样，将注意力集中在零售行业方面，但是，该版本却推出一种全新的方法，并用它来给出一个致力于处理机构原子型基本数据的更为全面的实例。如果你向前跳得太快，就会错过这样的原理性内容以及针对基本概念做出的其他更新。

X.2.1 导读内容

本书用提示、关键概念与各章指南等内容对正文进行了修饰，以帮助现在的阅读和将来更容易的查阅。此外，本书还提供了一个经过扩展的术语表。

✱ 翻动各章的书页并寻找本行左边的那个灯泡图标，就可以看到分散在本书各处的提示内容。

🔑 每章都从用一个钥匙图标指示的关键概念列表开始。

X.3 各章目的

在开始叙述正文之前，我们想先在书的前面逐章预览一下各自所覆盖的概念。

第 1 章：维度建模初步

本书从维度建模的初步知识开始讨论。这一章从总体上开发了数据仓库体系结构的组成部分，并建立了将在本书其余部分使用的核心术语表。同时，对一些关于维度建模方面的神话与误解进行了疏导。此外，还讨论了规范化模型所承担的作用。

第 2 章：零售营销

零售处理是用于图解维度建模的经典实例。之所以用这个经典例子作为第一个实例，是因为考虑到大家对它都比较熟悉。这样，你用不着对行业内容进行苦思冥想，相反，应该将注意力集中在维度建模的核心概念上。本章首先讨论了用于设计维度模型的四步处理过程。接着，从深度上探讨了维度表，这包括将在整本书中反复重用的日期维度。另外，还讨论了退化维度、雪花处理与代理关键字方面的内容。即使你不是一个零售从业人员，也需要阅读本章，因为其中充满了对各种基础内容的介绍。

第 3 章：库存

本章仍然在零售业范围内讨论本书的第二个实例研究，但将注意力转向了另外一个业务处理。这个实例研究为数据仓库总线结构以及一致的维度与数据的使用，提供了一个非常生动的例子。这些概念对于任何准备构造一个可扩展的集成数据仓库体系结构的人员来说，都是极为关键的。

第 4 章：采购

本章随着数据仓库建立的逐步深入，进一步强调了使用机构价值链的重要性。同时，开发了一系列处理渐变维度属性的基本与高级技术方法。

第5章：订单管理

在本实例研究中，简要地介绍了那些提供核心业务性能度量指数而经常要在数据仓库中首先实施的业务处理，这些度量指数包括“以何种价格向哪些顾客销售什么产品”等。其次，本章讨论了维度在单个设计结构中承担多种角色的情形。同时，还探讨了在处理订单管理信息时，建模人员要面临的一些常见挑战，比如标题/分列项目考虑因素、多种流通货币或者度量单位以及具有辅助事务指示符的杂项维度等。另外，对事务、周期快照与累积快照这三种基本类型的事实表进行了比较。本章最后为具有更多实时性要求的数据仓库处理提供了一些建议。

第6章：客户关系管理

大量的数据仓库都是在需要更好地了解与服务顾客的期望中建立起来的。本章覆盖了包括地址标准化、大容量维度的管理与针对不可预见的顾客层次体系进行建模等在内的有关顾客维度的主要考虑内容。同时，还讨论了多源顾客数据的合并。

第7章：账目

这是全新的一章，讨论了在数据仓库中对一般分类总账信息进行建模的知识。同时，还描述了本年度至今的事实与多种财政年度安排表的合适处理，以及对来自多个业务处理的数据进行组合的合并维度模型概念等。

第8章：人力资源管理

在这个新增的一章里，探讨了许多特定于人力资源维度模型的内容，比如维度表在什么情形下开始呈现出事实表的特点。另外，本章还引入了审计与关键词维度以及调查表数据的处理等。

第9章：财经服务

在本章的银行实例研究中，提出了其中的每项业务都具有惟一描述属性

与性能度量指数的异构产品的概念。显然，异构产品的处理需要并不仅仅局限于财经服务方面。其次，本章还讨论了账户、顾客与家庭方面的复杂关系。

第 10 章：电信与公用事业

这是新增的一章，在内容编排上具有某种程度的不同，其目的在于突出针对数据模型设计内容进行评估时要考虑的因素。此外，本章还探讨了地理位置维度的特性以及支持地理信息系统的可能性等内容。

第 11 章：交通

在本章的实例研究中，简单地描述了处在不同粒度层次上的相关事实表。同时，讨论了处理微型维度的另外一种方法，并比较详细地考察了涵盖诸如国别的历法与跨多个时区的时间同步之类的概念的日期与时间维度。

第 12 章：教育

本章对几个非事实型事实表进行了考察，并讨论了它们在分析没有出现的内容方面所具有的重要性。除此之外，还建立了一个学生应用流水线，这也是累积维度快照事实表方面的一个初步实例。

第 13 章：卫生护理

通常遇到的一些最为复杂的模型非卫生护理行业莫属。在这个新增的一章里，阐述了针对这些复杂问题进行处理的内容，比如，使用桥接表对与病人治疗有关的多项诊断与提供者进行建模。

第 14 章：电子商务

本章简单地介绍针对网上点击数据进行建模的知识。这里涉及的一些概念是从 Ralph Kimball 与 Richard Merz 合著的《数据网库综合开发技术》(The Data Webhouse Toolkit) 一书(由 Wiley 于 2000 年出版)中衍生而来的。

第 15 章：保险业

作为本书的最后一个实例研究，本章以单个互相关联的方案组的形式，叙述了在本书前面已经讨论过的多种技术方法。本章可以看做是对所有内容的一个汇总，因为各建模技术方法之间具有一定的层次关系，其情形就像高射投影仪透明胶片相互重叠一样。

第 16 章：建立数据仓库

在了解了维度模型的设计内容之后，本章从一个较高的视角层次，对典型数据仓库项目迭代设计的生命期内出现的行为内容进行了概括。这一章可看做是对由 Laura Reeves 与 Warren Thornthwaite 合著的《数据仓库生命周期模式工具箱》(The Data Webhouse Lifecycle Toolkit) 一书 (由 Wiley 于 1998 年出版) 的快速浏览。

第 17 章：相关知识与展望

作为本书的最后一章，本章对数据仓库应用的将来发展做出了预示。

术语表

本书最后给出了一个作为参考资源的详细术语表。它在你自己领域的一般业务理解与其他业务实例研究之间架起了一座进行沟通的桥梁。

X.4 配套的 Web 站点

你可以访问本书配套的 Web 站点：www.kimballuniversity.com。从网站可以得到如下资源：

- 注册“设计技巧”(Design Tips) 服务，从而可以通过电子邮件获取为维度建模与数据仓库设计定期发布的后续实用指南。
- 连接 Ralph Kimball 在《智能公司》(Intelligent Enterprise) 及其前身《DBMS 杂志》(DBMS Magazine) 上发表的所有文章。
- 学习 Kimball 大学根据作者的经历与著作开设的独立于厂家平台的高

X.5 总结

本书的目标是为维度数据仓库的设计介绍一套标准的技术方法。简略地说，作为读者，如果通过本书的阅读能建立起数据仓库必须由业务用户的需要来驱动，并因此从一个简单的维度视角来建立与展示数据仓库这样的理念，那么本书就达到目的了。我们深信，有了这些前提，你已向成功的数据仓库设计迈出了一大步。

到此为止，你应该知道我们要努力实现的目标了，当然该是深入讨论细节内容的时候了。第 1 章将从维度建模的初步知识开始介绍，以便统一给出一些关键的术语与有关体系结构的概念。随后，将从屡试不爽的零售业入手，讨论维度建模的基本技术方法。

目 录

| | |
|---------------------------|----|
| 第 1 章 维度建模初步 | 1 |
| 1.1 信息应用的不同领域 | 1 |
| 1.2 数据仓库的最终目标 | 2 |
| 1.2.1 出版业方面的比喻实例 | 4 |
| 1.3 数据仓库的组成 | 6 |
| 1.3.1 操作型源系统 | 7 |
| 1.3.2 数据聚集环节 | 7 |
| 1.3.3 数据展示 | 9 |
| 1.3.4 数据存取工具 | 12 |
| 1.3.5 其他要考虑的内容 | 13 |
| 1.4 维度建模词汇表 | 15 |
| 1.4.1 事实表 | 15 |
| 1.4.2 维度表 | 18 |
| 1.4.3 事实与维度的融合 | 20 |
| 1.5 关于维度建模的神话 | 22 |
| 1.5.1 应该避免的常见疏误 | 24 |
| 1.6 总结 | 25 |
| 第 2 章 零售营销 | 26 |
| 2.1 四步维度设计过程 | 27 |
| 2.2 零售实例的研究 | 28 |
| 2.2.1 第一步：选取业务处理 | 30 |
| 2.2.2 第二步：定义粒度 | 30 |
| 2.2.3 第三步：选定维度 | 31 |
| 2.2.4 第四步：确定事实 | 32 |
| 2.3 维度表属性 | 34 |
| 2.3.1 日期维度 | 35 |
| 2.3.2 产品维度 | 39 |