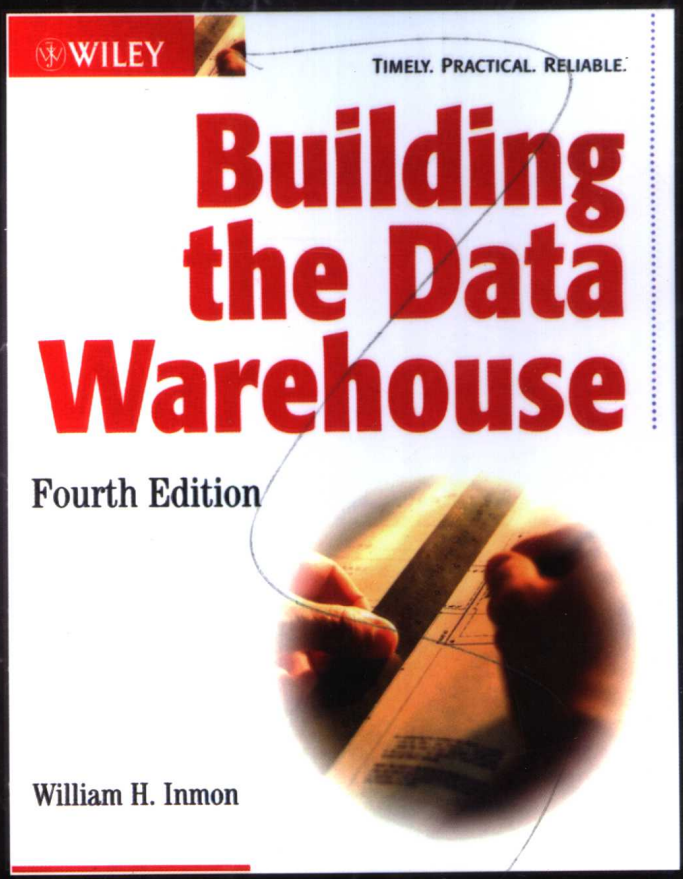
HZ BOOKS WILEY

计 算 机 科 学 丛 书

原书第4版

数据仓库

(美) William H. Inmon 著 王志海 等译 黄厚宽 田盛丰 审校



Building the Data Warehouse
Fourth Edition

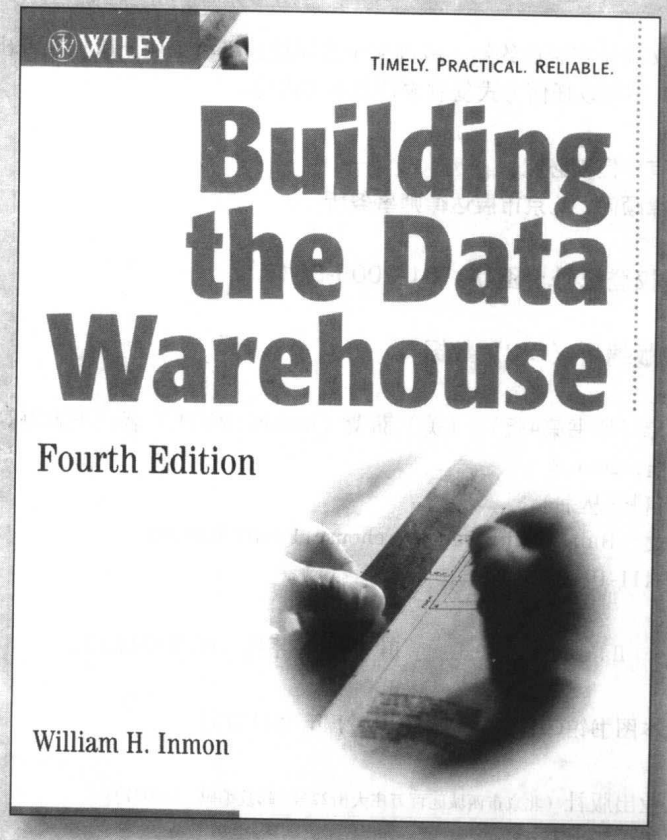
机械工业出版社
China Machine Press

计 算 机 科 书

原书第4版

数据仓库

(美) William H. Inmon 著 王志海 等译 黄厚宽 田盛丰 审校



Building the Data Warehouse

Fourth Edition

机械工业出版社
China Machine Press

本书系统讲述数据仓库的基本概念、基本原理以及建立数据仓库的方法和过程。主要内容包括：决策支持系统的发展、数据仓库环境结构、数据仓库设计、数据仓库粒度划分、数据仓库技术、分布式数据仓库、EIS系统和数据仓库的关系、外部和非结构化数据与数据仓库的关系、数据装载问题、数据仓库与Web、ERP与数据仓库以及数据仓库设计的复查要目。

本书是数据仓库之父撰写的关于数据仓库的最权威著作，既可作为相关专业的研究生教材，也是数据仓库的研究、开发和管理人员的必备指南。

William H. Inmon: Building the Data Warehouse, Fourth Edition (ISBN: 0-7645-9944-5)

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

本书版权登记号：图字：01-2005-5618

图书在版编目（CIP）数据

数据仓库（原书第4版）/（美）茵蒙（Inmon, W. H.）著；王志海等译. －北京：机械工业出版社，2006.8

（计算机科学丛书）

书名原文：Building the Data Warehouse, Fourth Edition

ISBN 7-111-19194-3

Ⅰ. 数…　Ⅱ. ①茵…　②王…　Ⅲ. 数据库系统　Ⅳ. TP311.13

中国版本图书馆CIP数据核字（2006）第051127号

机械工业出版社（北京市西城区百万庄大街22号　邮政编码　100037）

责任编辑：王　玉

北京慧美印刷有限公司印刷 · 新华书店北京发行所发行

2006年8月第1版第1次印刷

184mm × 260mm · 20.75印张

定价：39.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：（010）68326294

出版者的话

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭橥了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短、从业人员较少的现状下，美国等发达国家在其计算机科学发展的几十年间积淀的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章图文信息有限公司较早意识到“出版要为教育服务”。自1998年开始，华章公司就将工作重点放在了遴选、移译国外优秀教材上。经过几年的不懈努力，我们与Prentice Hall，Addison-Wesley，McGraw-Hill，Morgan Kaufmann等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出Tanenbaum，Stroustrup，Kernighan，Jim Gray等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及庋藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍，为进一步推广与发展打下了坚实的基础。

随着学科建设的初步完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都步入一个新的阶段。为此，华章公司将加大引进教材的力度，在“华章教育”的总规划之下出版三个系列的计算机教材：除“计算机科学丛书”之外，对影印版的教材，则单独开辟出“经典原版书库”；同时，引进全美通行的教学辅导书“Schaum's Outlines”系列组成“全美经典学习指导系列”。为了保证这三套丛书的权威性，同时也为了更好地为学校和老师们服务，华章公司聘请了中国科学院、北京大学、清华大学、国防科技大学、复旦大学、上海交通大学、南京大学、浙江大学、中国科技大学、哈尔滨工业大学、西安交通大学、中国人民大学、北京航空航天大学、北京邮电大学、中山大学、解放军理工大学、郑州大学、湖北工学院、中国国家信息安全测评认证中心等国内重点大学和科研机构在计算机的各个领域的著名学者组成“专家指导委员会”，为我们提供选题意见和出版监督。

这三套丛书是响应教育部提出的使用外版教材的号召，为国内高校的计算机及相关专业的教学度身订造的。其中许多教材均已为M. I. T.，Stanford，U.C. Berkeley，C. M. U. 等世界名牌大学所采用。不仅涵盖了程序设计、数据结构、操作系统、计算机体系结构、数据库、编译原理、软件工程、图形学、通信与网络、离散数学等国内大学计算机专业普遍开设的核心课程，而且各具特色——有的出自语言设计者之手、有的历经三十年而不衰、有的已被全世界的几百所高校采用。在这些圆熟通博的名师大作的指引之下，读者必将在计算机科学的宫殿中由登堂而入室。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证，但我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。教材的出版只是我们的后续服务的起点。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下:

电子邮件: hzjsj@hzbook.com
联系电话: (010) 68995264
联系地址: 北京市西城区百万庄南街1号
邮政编码: 100037

专家指导委员会

（按姓氏笔画顺序）

尤晋元	王　珊	冯博琴	史忠植	史美林
石教英	吕　建	孙玉芳	吴世忠	吴时霖
张立昂	李伟琴	李师贤	李建中	杨冬青
邵维忠	陆丽娜	陆鑫达	陈向群	周伯生
周克定	周傲英	孟小峰	岳丽华	范　明
郑国梁	施伯乐	钟玉琢	唐世渭	袁崇义
高传善	梅　宏	程　旭	程时端	谢希仁
裘宗燕	戴　葵			

译 者 序

计算机网络与数据库技术的迅速发展和广泛应用，使得企业管理进入一个崭新的时代。广大基层管理人员摆脱了繁重的制表业务和数据处理工作，管理工作进一步规范化，企业建立了各种在线事务处理信息系统，对各种日常业务处理提供了有效的支持。然而，面对当今竞争日趋激烈与瞬息万变的市场，各级管理人员迫切需要根据企业的现状和历史数据做出判断和决策。因此，各级管理人员希望能够从企业信息系统中获取有效的、一致的决策支持信息，及时准确地把握市场变化的脉搏，做出正确有效的判断和抉择。也就是说，数据处理的重点应该从传统的业务处理扩展到在线分析处理，并从中得到面向各种主题的统计信息和决策支持信息。随着企业事务处理系统的运行和建立，数据量越来越大，企业数据源越来越多。这种需求就比以往任何时候都更加迫切，也更加难于实现。

数据仓库技术就是针对上述问题而产生的一种技术解决方案，它是基于大规模数据库的决策支持系统环境的核心。正如本书作者W. H. Inmon所定义的，数据仓库是一个面向主题的、集成的、永久的且随时间不断变化的数据集合，用于支持管理层的决策。本书详尽地讲述了数据仓库的基本概念、基本原理，以及建立数据仓库的方法和过程。主要内容包括决策支持系统的发展、数据仓库环境结构、数据仓库设计、数据仓库粒度划分、数据仓库技术、分布式数据仓库、EIS系统和数据仓库的关系、外部和非结构化数据与数据仓库的关系、数据装载问题、数据仓库与Web、ERP与数据仓库以及数据仓库设计的复查要目。本书主要面向数据仓库的开发者、管理者、设计者、数据管理员、数据库管理员以及其他相关人员，对于计算机专业的本科生和研究生也有重要的参考价值。

我们研究小组对数据仓库技术和数据挖掘技术进行了很长时间的研究，并翻译了一些相关文献。1999年翻译并出版了本书的第2版，2003年翻译并出版了本书的第3版，都得到了社会各界的好评。为了反映数据仓库技术的进展，本书作者在不断地充实和修改其著作。应出版社的要求，我们承担了第4版的翻译工作，并推荐给读者。随着这几年我们研究的进展，对数据仓库技术和工程有了更为深入的理解。为此，我们对数据仓库所涉及的术语的译法重新进行了规范，在翻译了新增和修改内容的同时，将全部原有内容重新逐字校正了一遍，更正了以前译文中的一些错误，使语言更加准确、通顺，便于读者理解。本书的第1章和第2章由范亚琼负责，第3章和第4章由曹源负责，第5章和第6章由李广群负责，第7章至第13章由山丹负责，第14章至第19章以及词汇表由廉捷负责翻译，杨迪参加了第3章的部分翻译工作。本书最后的定稿与许多人先后的辛勤工作密切相关，他们是王琨、王继奎、董隽、刘森、林友芳、高思宇、王春花、宁云晖、李晓武、蔺永华、范星艳、高宏彬、贾旭光、李红松、秦远辉等。本书由王志海负责统一定稿，由黄厚宽教授和田盛丰教授共同审定全书。由于译者水平有限，错误之处望广大读者批评指正。

译者简介

王志海　博士，特聘教授。1985年毕业于郑州大学计算机科学系，获理学学士学位，1987年毕业于哈尔滨船舶工程学院计算机与信息科学系，获工学硕士学位，1998年毕业于合肥工业大学计算机与信息学院，获博士学位。先后在澳大利亚Monash大学计算机科学与软件工程学院进行博士后研究工作，Deakin大学信息技术学院任研究员，Monash大学计算机科学与软件工程学院任高级研究员。曾任中国计算机学会人工智能与模式识别专业委员会委员，中国人工智能学会机器学习委员会委员，2003年国际软件工程大会数据挖掘技术在软件工程中应用研讨会（DMSE' 2003, USA）等程序委员会委员，2004~2007年历届亚太数据库知识发现与数据挖掘会议（PAKDD）程序委员会委员，2005年中国分类技术及其应用研讨会程序委员会委员等。在国际学术刊物，国际学术会议和国内学术刊物上发表论文约30多篇。

审校者简介

黄厚宽　教授，博士生导师。1940年9月生，1963年毕业于北京大学数力系六年制数学专业，1966年哈尔滨军事工程学院应用数学研究生毕业。1970～1980年参加我国首次洲际火箭发射落点水声测量系统研制，主持总体数学模型论证计算及专用计算机系统软件编制，获中央军委嘉奖及原国防科工委重大科技成果三等奖。1983～1985年先后在美国亚拉巴马大学和佛罗里达大学信息研究中心任访问教授。十多年来主持完成多个专家系统与工具及计算机应用系统，进行机器学习、数据挖掘、分布式人工智能的研究。获省部级科技进步奖6项，已发表论文150多篇，指导硕士与博士研究生80多人，俄罗斯高级访问学者1人。现任中国计算机学会人工智能与模式识别专委会副主任兼秘书长等。

田盛丰　教授，博士生导师。1944年11月生，1967年毕业于哈尔滨军事工程学院电子工程系，1968～1977年在七级部五院五零四研究所任实习研究员，1977年至今在北方交通大学计算机系任教。其中1982～1984年在美国纽约州立大学石溪分校作访问学者，主要研究人工智能；1997年在英国伦敦大学Royal Holloway学院计算机科学系合作研究人工智能项目。曾主持和参加了多项科研项目，包括国家自然科学基金项目“隧道工程预测专家系统”、“工程建设中知识系统的应用研究”、“断裂地质构造遥感图像判释专家系统”，教委博士点基金项目“隧道岩溶预测专家系统”，部委级项目“国防交通铁路工程保障指挥决策专家系统的改进与应用”等。发表论著2部及论文50多篇。

第2版前言

数据库及其理论已经出现好长时间了。早期的数据库主要是一些独立的数据库，应用于企业数据处理的各个方面——从事务处理到批处理，再到分析型处理。早期的大多数数据库系统主要集中于操作型的日常事务处理。近年来，出现了一种更高级的数据库观念，即一种数据库服务于操作型需求，而另一种数据库服务于信息型或分析型需求。从某种程度上讲，这种数据库的新颖思想是随着个人计算机技术、第四代程序设计语言（4GL）技术以及最终用户新需求的出现而产生的。

将操作型数据库和信息型数据库分离开，是出于以下原因:

- 服务于操作型需求的数据在物理上不同于服务于信息型或分析型需求的数据。
- 支持操作型处理的技术从根本上不同于支持信息型或分析型需求的技术。
- 操作型数据的用户群体不同于信息型或分析型数据所支持的用户群体。
- 操作型环境的处理特点与信息型环境的处理特点从根本上是不同的。

由于这些原因（以及很多其他原因），当今建立系统的方法是将操作型处理及其数据与信息型或分析型处理及其数据分离开来。

本书讨论分析型的环境，或称为决策支持系统（DSS）环境，以及在这种环境中的数据结构问题。本书的重点是讨论信息型和决策支持系统处理的核心——“数据仓库”（或“信息仓库”）。

本书所讨论的问题是面向管理者和开发者的，在某些地方也涉及技术问题。但本书的大部分是关于数据仓库的问题和技术。本书旨在作为数据仓库设计者和开发者的一本指导性读物。

本书出第1版的时候，数据库的理论家们对数据仓库的概念大加嘲笑。有一个理论家说数据仓库技术将使信息技术倒退20年。另有人说不应该允许数据仓库技术的创建者在公共场合发表言论。另外一些学院派的研究人员宣称数据仓库技术根本就不是什么新技术，学术界早已经知道数据仓库技术，尽管那时没有出书、没有文章、没有课程、没有研讨会、没有学术会议、没有报告、没有参考文献、没有论文、也没有可用的术语或概念。

本书出第2版的时候，整个世界正在为互联网而疯狂。想要成功，就要在各种词之前加上字母“e”，如e-business，e-commerce，e-tailing等。记得一个风险投资家说过“我们现在有了互联网，为什么还要数据仓库呢？”

但是数据仓库技术已经远比那些想把所有数据放在一个数据库中的数据库理论家们期望的要好。数据仓库技术也挺过了由那些短视的风险投资家所带来的“.com”灾难。在技术常被华尔街和Main Street抛弃的这个时代里，数据仓库技术从来没有像现在这么活跃和强大。关于数据仓库技术，有着各种各样的学术会议、研讨会、书籍、文章、咨询等。更重要的是，现在有很多公司在做数据仓库。我们还可以发现，与大肆宣扬的所谓新经济不同，数据仓库技术确确实实在发挥着作用，尽管硅谷还在否认它。

第3版前言

本书的第3版预示着数据仓库技术更新、更强大的时代。当今，数据仓库技术已经不再是纯粹的理论，而是活生生的事实。新技术已经可以支持对数据仓库的各种新奇的需求。许多企业已经通过数据仓库运转它们的重要业务。由于有了数据仓库，获取信息的代价在急剧降低。对于混乱的遗留系统环境，管理人员最终有了一种可行的解决方案。企业第一次拥有了可用的企业范围内的历史数据“存储方式”。整个企业的数据集成真正成为可能，这在多数情况下还是第一次。许多企业正在学习如何从数据获取信息，以获得竞争优势。简而言之，数据仓库技术极大地冲破了技术的束缚。

数据仓库容易使人糊涂的地方在于它是一种体系结构，而不是一种技术。这一点使技术人员和风险投资家感到灰心，因为他们想买的是那些很好地打成了包的东西。但是，数据仓库本身不会将自己“封装”起来。体系结构和技术之间的差别就像是新墨西哥州圣达菲和砖块之间的差别一样。如果你在圣达菲的大街上开着车，你就会知道你是在圣达菲，而不是在别的什么地方。每一幢住宅、每一座办公楼、每一家饭馆都有显著的特征，提醒着我们“这里是圣达菲”。使圣达菲突显的外观和风格是建筑结构，而这种结构是由砖块和裸露的横梁构成的。当然，如果没有这些砖块和横梁就没有圣达菲的各种建筑。但是，砖块和横梁本身并不能构成结构。它们是独立的技术。就像你在美国西南部所有地方和世界的其他地方都能看到砖块，但它们并不是圣达菲。

因此，数据仓库和数据库及其他技术之间的关系，就像是体系结构和技术之间的关系。有了这种体系结构，就有相应的基础技术，两者之间有很大的差别。毫无疑问，数据仓库和数据库技术之间存在着关系，但是可以确定的是，它们不是同一种东西。数据仓库需要许多不同种类的技术支持。

有了本书的第3版，我们知道什么东西管用，什么东西不管用。在写第1版的时候，我们有一些开发和使用数据仓库的经验。但是说真的，当时的经验没有现在多。例如现在，我们可以确切地知道以下这些内容：

- 数据仓库的建立要采用不同于应用程序的开发方法，不记住这点会带来很大的问题。
- 数据仓库在根本上不同于数据集市。两者不能混在一起，就像油和水一样。
- 数据仓库能够实现所承诺的功用，而不像许多被过分宣扬的、之后渐渐消逝的技术一样。
- 数据仓库中汇集了大量的数据，这样就需要有全新的技术来管理大规模的数据。

但是，或许数据仓库最吸引人的东西是数据仓库构成了许多其他各种形式处理的基础。可以改造和重复使用数据仓库中的各种粒度的数据。如果存在一个关于数据仓库永恒而深刻的真理，那就是：数据仓库为许多其他形式的信息处理提供了理想的基础。这个基础如此重要，有许多原因，比如：

- 真理只有单个版本。
- 如果需要，可以重新调整数据。
- 可以为新的、未知的应用随时提供数据。

最后，数据仓库技术降低了企业获取信息的代价。有了数据仓库，获取数据将不再昂贵，

数据访问也将更加快捷。

数据库及其理论已经出现好长时间了。早期的数据库主要是一些独立的数据库，应用于企业数据处理的各个方面——从事务处理到批处理，再到分析型处理。早期的大多数数据库系统主要集中于操作型的日常事务处理。近年来，出现了一种更高级的数据库观念，即一种数据库服务于操作型需求，而另一种数据库则服务于信息型或分析型需求。从某种程度上讲，这种数据库的新颖思想是随着个人计算机技术、第四代程序设计语言（4GL）技术以及最终用户新需求的出现而产生的。将操作型数据库和信息型数据库分离开，是出于以下原因:

- 服务于操作型需求的数据在物理上不同于服务于信息型或分析型需求的数据。
- 支持操作型处理的技术从根本上不同于支持信息型或分析型需求的技术。
- 操作型数据的用户群体不同于信息型或分析型数据所支持的用户群体。
- 操作型环境的处理特点与信息型环境的处理特点从根本上是不同的。

由于这些原因（以及很多其他原因），当今建立系统的方法是将操作型处理及数据与信息型或分析型处理及其数据分离开来。

本书讨论分析型的环境，或称为决策支持系统（DSS）环境，以及在这种环境中的数据结构问题。本书的重点是讨论信息型和决策支持系统处理的核心——数据仓库（或信息仓库）。

什么是分析型、信息型处理呢？这种处理服务于决策支持过程中的管理需求，一般称为DSS处理，要在大量的数据中分析处理探索趋势。不同于只查找1~2条数据记录（如操作型处理），当DSS分析人员进行分析型处理时，需要访问大量的数据记录。

DSS分析人员很少修改数据。而在操作型系统中，数据在个体记录层次上经常修改。在分析型处理中，需要经常访问记录，收集来的记录内容用于分析的需要，但很少或不需要对单个的记录进行更改。

相对于传统的操作型处理，在分析型处理中，响应时间的要求大大放宽。分析型处理的响应时间可以是30分钟到24小时。这样的响应时间标准对于操作型处理而言是一个巨大的灾难。

服务于分析型用户群体的网络比服务于操作型用户群体的网络的规模小得多。通常情况下，分析型网络的用户比操作型网络的用户少很多。

与应用于分析型环境的技术不同，操作型环境中的技术必须将技术本身与数据和事务锁定、数据争用、死锁等因素结合起来考虑。

这样，在操作型环境和分析型环境之间存在许多重大的区别。本书针对分析型的DSS环境进行讨论，并着重讨论以下问题:

- 数据的粒度。
- 数据分区。
- 元数据。
- 数据可信度的缺乏。
- DSS数据的集成。
- DSS数据的时间基准。
- 确定DSS数据的数据源——记录系统。
- 数据迁移及方法。

本书适合开发人员、管理人员、设计人员、数据管理员、数据库管理员，以及其他在现代数据处理环境中进行系统建造的人员阅读。另外，本书也很适用于学习信息处理技术的学生。本书有些地方的讨论更具有技术性。但全书多数部分是关于数据仓库的问题和技术。本

书旨在作为数据仓库设计者和开发者的一本指导性读物。

本书是有关数据仓库的系列丛书中的第一本。第二本是《Using the Data Warehouse》(Wiley，1994)，着重阐述建立了数据仓库后所面临的一些问题。此外，还介绍了更大的体系结构的概念和操作型数据存储（ODS）的思想。操作型数据存储在体系结构上与数据仓库相似，两者的区别在于ODS仅适用于操作型系统，而不适用于信息型系统。该系列丛书的第三本是《Building the Operational Data Store》(Wiley，1999)，阐述什么是ODS以及如何建造ODS。

数据仓库系列丛书的第四本是《Corporate Information Factory, Third Edition》(Wiley，2002)。该书阐述了以数据仓库为中心的更大型的信息系统。在很多方面，有关CIF的书和有关DW的书是相辅相成的。有关CIF的书着眼点更高，而有关DW的书则做出了更为具体的讨论。该系列丛书还包括《Exploration Warehousing》(Wiley，2000)。该书阐述了使用统计技术对数据仓库中的数据所进行的一种特殊的处理模式分析。

无论如何，本书都是这一系列丛书的基石。数据仓库是其他所有DSS处理形式的基础。

也许本书结尾引用的参考文献最能雄辩地说明数据仓库和企业信息工厂所带来的进步。本书第1版出版时，除了少数论文外，没有其他书籍或白皮书可供参考引用。而这本第3版提到了许多书籍、论文和白皮书。确实，引用的参考文献只是揭示了大量重要工作中的一部分。

第4版前言

早期的数据库理论认为所有的数据都应该装载在一个公共的数据源中。这个想法不难得出。主文件是先于数据库而出现的，这些主文件存储在顺序介质上，为实现随之而来的各种应用而创建。在主文件之间根本没有数据集成。因此，将数据集成为单一的数据源——数据库的理念得到极大的认同。

数据仓库的诞生基于以上这些理念。数据仓库对于那些赞同传统数据库理论的人来说是一种智力上的威胁，因为数据仓库本身意味着应该建立不同种类的数据库。然而，建立不同种类的数据库的思想并不被数据库理论学家们所接受。

现在，数据仓库已经被认为是一种明智的选择。基于许多不同理由，人们相信数据仓库就是所想要的。近期的一项调查显示，公司用于数据仓库和商业智能方面的开销超过了事务处理和在线事务处理（OLTP）方面，这在几年前是不可想象的。

数据仓库的成熟期已经到来。

本书第4版的问世恰逢时宜，它掀起了数据仓库的新浪潮。

除了数据仓库中由来已久的概念外，本书第4版还囊括了数据仓库的基础知识，也包含了许多当今有关信息基础框架的主题。

本书中较为重要的新主题是：

- 依从准则（涉及Sarbanes Oxley, HIPAA, Basel II以及其他问题）
- 近线存储（扩展数据仓库使其无穷大）
- 多维数据库设计
- 非结构化数据
- 最终用户（他们是谁，他们需要什么）
- ODS和数据仓库

除了这些新主题外，本版还体现了更为庞大的围绕数据仓库所建立的体系结构。

技术伴随着数据仓库的发展而发展。在数据仓库发展的早期阶段，50GB~100GB的数据量被认为是一个庞大的数据仓库。现在，一些数据仓库已经达到千万亿字节的容量范围。其他技术包括多维技术——数据集市和星形连接方面的进展。此外，技术的进步也使得数据可以存储在非磁盘存储介质之上。

总而言之，技术的进步使今天的科技成果成为可能。没有现代技术的发展，就不会有数据仓库的出现。

本书可供数据仓库架构和系统设计师参阅。最终用户可能发现这本书的有用之处在于全面了解有关数据仓库的解释。管理者和学生们也将发现本书的有益之处。

目　录