

数据库
专业人员
的宝库

SQL Server 7.0 数据库开发系列丛书

SQL SERVER 7

Data Warehousing 数据仓库

[美] Michael Corey Michael Abbey 等著
希望图书创作室 译

- 掌握市场上最复杂的数据库工具
- 学会使用 SQL Server 7 设计、建立、管理和优化数据仓库
- 利用 FREE 内建 OLAP 服务、Microsoft Decision Support Server、Plato 和数据挖掘工具

本书配套光盘内容包括:

1. 与本书配套电子书
2. SQL Server 7 评测版软件



北京希望电子出版社
Beijing Hope Electronic Press
www.bhpe.com.cn

SQL Server 7 数据库开发丛书

SQL Server 7 Data Warehousing

数 据 仓 库

[美] Michael Corey Michael Abbey 等著

希望图书创作室 译

本书随书光盘内容包括:

1. 与本书配套电子书
2. SQL Server 7 评测版软件



北京希望电子出版社

Beijing Hope Electronic Press

www.bhp.com.cn

内 容 简 介

本书是“SQL Server 7 数据库开发丛书”之一。作者是数据库和数据仓库技术方面的权威。

本书首先介绍了数据仓库,以及在规划决策支持解决方案的部署时要考虑的许多新事务,如小组成员、作用域、风险承担者、分析工具、数据库等。一旦掌握了这些知识,就会了解 Microsoft 划时代的数据仓库版本 SQL Server 7。SQL Server 7 提供了全部集成化的数据仓库解决方案,并引导读者进行设计、管理和安装 Microsoft 决策支持系统。功能全面和价格低廉的成功结合将数据仓库的蓝图改变为现在我们所看到的样子。本书讨论的许多关于数据仓库的理论和技術是最基本的,无论选择什么样的后端(数据库的选择)、迁移和转换工具及前端(分析工具),都应采纳这些有益的建议。

本书内容实用、结构合理、可操作性强,对数据库和数据仓库进行了深刻的分析,并将这些概念深入浅出地表达出来,适用于数据仓库一般用户学习 Microsoft SQL Server 7,也是广大数据仓库设计、管理和维护人员、数据仓库工程技术人员的重要参考资料。可作为相关专业大学、社会培训、自学的参考教材。

本书光盘内容包括与本书配套的电子图书和 SQL Server 7 评测版软件。

版 权 声 明

本书英文版名为“SQL Server 7 Data Warehousing”,由 McGraw-Hill 出版公司出版,版权归 McGraw-Hill 出版公司所有。本书中文版由 McGraw-Hill 出版公司授权出版。未经出版者书面许可,本书的任何部分不得以任何形式或任何手段复制或传播。

系 列 书 名: SQL Server 7 数据库开发系列丛书

书 名: SQL Server 7 数据仓库

文本著作者: (美) Michael Corey Michael Abbey 等著 希望图书创作室 译

审 校 者: 李航海 刘桂兰

责 任 编 辑: 刘晓融

CD 制 作 者: 希望多媒体创作中心

CD 测 试 者: 希望多媒体测试部

出版、发行者: 北京希望电脑公司 北京希望电子出版社

地 址: 北京海淀路 82 号, 100080

网址: www.bhp.com.cn

E-mail: lwm@hope.com.cn

电话: 010-62562329, 62541992, 62637101, 62637102(图书发行, 技术支持)

010-62633308, 62633309(多媒体发行, 技术支持)

010-62613322-215(门市) 010-62531267(编辑部)

经 销: 各地新华书店、软件连锁店

排 版: 希望图书输出中心

CD 生 产 者: 文录激光科技有限公司

文本印刷者: 北京双青印刷厂

开本 / 规格: 787×1092 16 开本 20.25 印张 458 千字

版次 / 印次: 2000 年 1 月第 1 版 2000 年 4 月第 2 次印刷

印 数: 5 000-10 000 册

本 版 号: ISBN 7-900031-00-6/TP·00

定 价: 40.00 元(1CD, 含配套书)

说明: 凡我社光盘配套图书若有缺页、倒页、脱页、自然破损, 本社发行部负责调换

译者的话

SQL Server 7 的发布是本行业的一件大事，该产品在过去的几年中取得了巨大的进步，SQL Server 已成为关系数据库技术中易于使用和管理的标准。

SQL Server 7 版本的发布也是 Microsoft 的一件大事，它的确是本行业中最优秀的数据库专家通力合作的结晶。约 5 年前，Microsoft 认识到，为真正成为企业级数据库的玩家，需要为跨入 21 世纪建立一个坚固的基础。为此，Microsoft 招聘了一些本行业中最优秀的专家，这个小组在最后的 5 年中主要考虑一个目标：SQL Server 7。SQL Server 7 的一个重点就是数据仓库。SQL Server 7 提供的集成解决方案不仅包括一个极其易用、可伸缩的关系数据库和一个性能最佳的并行查询处理器，还包括一个最好的 OLAP 服务器，它可用于多维分析、数据转换服务，一个强大而灵活的数据提取和加载工具，一个用于数据仓库的设计和管理的知识库，用于捕获和重放工作负载的配置器和索引分析工具，以及性能优化的自然语言查询界面。这项技术的所有功能都可在一个集成的产品中使用。

最后，SQL Server 7 对整个行业带来了巨大的冲击，特别是数据仓库行业。SQL Server 7 中健全的功能、易于使用和低成本为市场注入了新的活力。数据仓库的商业利润和传统上只有资金雄厚的公司才能选择的 Online Analytical Processing (OLAP)，现在可用于各种公司。

为成功所付出的所有努力——包括数据仓库——是增强资源的可利用性，并描述如何成功地部署解决方案，最理想的是在想要的实现平台上使用特定的信息。用户不仅需要某个特定产品的信息，还需要成功实现过程中的信息和使数据仓库工程获得成功所需的工作小组，也需要从同一条战壕中的人、成功实现数据仓库的人和讲求实践而不单纯研究理论的人那里获得这些信息。本书提供了这两方面的内容。

本书内容实用、结构合理、可操作性强，对数据库和数据仓库进行了深刻的分析，并将这些概念深入浅出地表达出来，适用于数据仓库一般用户以及数据仓库爱好者学习、掌握 Microsoft SQL Server 7，也是广大数据仓库管理和维护专业人员、数据库仓库工程人员等日常使用的优秀参考资料。

参加本书翻译工作的主要有李增民、陈彬、仵春光、张怀亮、高小平、耿俊燕、赵志勇、杨国清、李文宣、付林、胡仁喜、陈公博、冯丽、陈凯、邓宪等，由杨皓担任主要的翻译和校对工作。严谨、求实、高品味、高质量是我们在策划和创作中孜孜以求的目标。尽管倾心相注，精心而为，但错误和不足在所难免，恳请读者不吝赐教和指正，我们定会全力改进，在后续的工作中提高。

NI564/02

译者

1999 年 10 月

作者简介

Michael Corey 是数据库技术公司的执行总管，数据库技术公司是位于美国 Wellesley, Mass 的 LLC.数据库技术公司的 PKS 系统集成辅助公司，LLC.数据库技术公司是在关系数据库、数据仓库和 Internet 解决方案供应商方面经验最丰富的第一家公司，也是 Microsoft 解决方案的合作伙伴。Mike 是部署关系数据库和数据仓库解决方案方面的技术和商业专家，他的网址是 mcorey@dbtinc.com。

Michael Abbey 生活工作在 Ontario 区域的 Ottawa。在过去的 12 年中，他一直致力于关系数据库技术领域。目前是加拿大、美国和欧洲的技术顾问，他的网址是 masint@istar.ca。

Ian Abramson 是位于 Toronto, Ontario 的 IAS 公司总经理。他在加拿大、美国和欧洲积累了超过 12 年的数据库专业知识，并与人合作编写了两本关于数据库和数据仓库的书。Ian 最精通数据仓库、系统开发和数据库管理，也常出任数据库和数据仓库顾问。他的网址是 ias@magi.com。

Larry Barnes 是 Microsoft 新西兰地域的工程师。他在该行业已有 20 多年的经验，主要从事客户机/服务器数据库开发和数据仓库。Larry 是 ODBC 标准协会的奠基人之一，出席过 Tech-Ed，是一位 Microsoft 认证系统工程师，解决方案开发人员和培训师，他还是“Bob 和 Larry 开发产品系列”的负责人之一，该系列是主要构建可伸缩的 Web 应用的流行研究系列产品。

Benjamin Taub 是中西部最大的数据仓库咨询中心之一 Dataspace, Incorporated（办公地点在芝加哥的 Ann Arbor 和堪萨斯市）的创始人和 CEO，他还是 International Oracle User Group-Americas (IOUS-A) 委员会的数据仓库集中区域经理，也是 McGraw-Hill 出版公司 Oracle Press 系列《Oracle 8 数据仓库》一书的作者之一。在他的职业生涯中，Taub 先生在 Andersen Consulting 和 Microstrategy 联合公司从事信息技术工作，是认证的公共会计师（CPA），拥有 Lehigh 大学的科学学位和 Michigan 大学的工商管理硕士学位，他的网址是 btaub@dspace.com。

Rajan Venkitachalam 是数据库技术公司的总管，数据库技术公司是位于美国 Wellesley, Mass 的 LLC.数据库技术公司的 PKS 系统集成辅助公司，LLC.数据库技术公司是在关系数据库、数据仓库和 Internet 解决方案供应商方面经验最丰富的第一家公司。Rajan 现住 Massachusetts 的 Acton，他的网址是 rajanv@dbtinc.com。

前 言

当 Michael Corey 与我联系，说想与他的长期合作伙伴 Michael Abbey 合作编写《SQL Server 数据仓库》这本书时，我非常高兴。这有两个原因：首先，Michael Corey 和 Michael Abbey 是两位本行业中的优秀作者，并曾合作编写了六本数据库出版物。他们对数据库和数据仓库有深刻的理解，并能将这些概念深入浅出地表达出来。

其次，这是他们编写的第一本有关 SQL Server 的书，这表明随着 SQL Server 7 的 Fall Comdex 1998 的发布，数据库的优秀作者现在认为 SQL Server 是一个企业级的数据库。Michael Corey 和 Michael Abbey 的丰富经验，以及 Larry Barnes（数据库行业的老手，Microsoft 中受人尊敬的 SQL Server 专家）的参与，正是我自豪地写这篇前言的原因。

SQL Server 7 的发布是本行业的一件大事，对于 Microsoft 和我本人都是如此。从个人的角度看，我已为使 SQL Server 7 成为现实奋斗了 4 年多。SQL Server 7 对于我有特殊的意义，因为我个人自从运行在 OS/2 的 SQL Server Version 1.1 开始就与 SQL Server 产品结下了不解之缘。该产品在过去的几年中取得了巨大的进步，SQL Server 已成为关系数据库技术中易于使用和管理的标准，我们用 SQL Server 来衡量现在能支持成千上万个用户和 TB 大小的数据库的水平。Microsoft Terraserver Web 站点 <http://www.terraserver.com> 是这方面的一个最好的例子。

SQL Server 7 版本也是 Microsoft 的一件大事，它的确是许多本行业中最优秀的数据库专家通力合作的结晶。大约 5 年前，Microsoft 认识到，为真正成为企业级数据库的玩家，需要为跨入 21 世纪建立一个坚固的基础。这个坚固的基础就从一个“伟大的”小组开始。Microsoft 招聘了一些本行业中最优秀的专家：Jim Gray，一个可伸缩事务处理系统的幻想家，他的履历上写有 IBM DB2、Tandem 和 Digital 的 Rdb；Goetz Bereson 和 Peter Spiro，本行业 20 多年的老手，他们分别是 Digital 的 Rdb 数据库的小组组长和体系结构设计人员；Casey Kiernan，Ingres 的专家和 SQL Server 管理工具的构建者；Bill Baker，富有 OLAP 行业 20 多年的丰富经验，是 IRI 软件规则的创立者；以及 Amir Netz，MS OLAP Services 的开发带头人。这个名单还可以列出一大串。

这个小组在最后的 5 年中主要考虑一个目标：SQL Server 7。SQL Server 7 的一个重点是数据仓库。SQL Server 7 提供的集成解决方案不仅包括一个及其易用、可伸缩的关系数据库及一个艺术级全并行查询处理器，还包括一个艺术级的 OLAP 服务器，可用于多维分析、数据转换服务、一个强大而灵活的数据提取和加载工具、一个用于数据仓库的设计和管理的知识库、用于捕获和重放工作负载的配置器和索引分析工具，以及性能优化的自然语言查询界面。这项技术的所有功能都可在一个集成的产品中使用。

最后，SQL Server 7 对整个行业带来了巨大的冲击，特别是数据仓库行业。SQL Server 7 健全的功能、易于使用和低成本为市场带来一个变化。数据仓库的商业利润和传统上只有能支付得起的公司才能选择的 Online Analytical Processing (OLAP)，现在可用于各种公司。

为成功所付出的所有努力——包括数据仓库——是资源的可利用性和描述如何成功部署解决方案的指导，最理想的是在想要的部署平台上使用特定的信息。用户不仅需要某个特定产品，还需要成功过程的信息和使数据仓库工程获得成功所需的工作小组的信息。需要从同一条战壕中的人、成功实现数据仓库的人和讲求实践而不单纯研究理论的人那里获得这些信息。本书提供了这两方面的内容。

现在让我们从头至尾阅读本书，作为参考指南，让数据仓库帮助理解你去过哪里，现在在哪里，但最重要的是，想去哪里。

Doug Leland

Group Product 经理

SQL Server 市场部

Microsoft 公司

致 谢

这是我的第七本书。当我回想是什么使之成为现实时，就觉得应感谢许多人。首先要感谢我的家庭。在整个过程中，他们牺牲得最多。我要感谢我的妻子 Julie 和孩子 John, Annmarie 和 Michael 在本书编写过程中对我的爱、理解和支持。对我的好友和七本书的合作作者 Michael Abbey，感谢你一直与我并肩作战。这些书送给我的最大礼物就是友谊。对于 Ian Abramson，我只能说“YATFG”。感谢您在本书和其他书的编写过程中所给予的友谊和支持。在他加入前，我一直不知道应如何编写一本书。对于这个小组的最新成员 Larry “Microsoft 先生” Barnes，感谢您在幕后的努力，例如让我们使用 SQL Server 的最新版，您在 SQL Server 数据库方面的渊博知识给我流下了难以磨灭的印象。对于其他合作作者 Rajan, Evan, Jeremy, Ben，感谢你们的努力，没有你们，工作是不可能完成的。对于 Microsoft 公司在波士顿分部的 Stev Shelgrin：你是一个精英，对您所作的工作无论怎样感谢都不过分。对 Doug Leland, Steve Murchie, Thomas Gruver, Jim Ewel，非常感谢你们为本书的出版所作的幕后工作，我们需要你们的帮助时，你们总是毫不吝惜地提供给我们。对 Osborne/McGraw-Hill 的 Scott Rogers 和 Wendy Rinaldi：感谢你们的工作。我知道与我们这样的一组作者一起工作能培养耐心。对这里没有提及的许多人在此说一句：非常感谢。

Michael J. Corey

感谢 Mike Corey, Ian Abramson, Ben Taub, Larry Barnes, Evan Ross, Rajan Venkitachalam 和 Jeremy Fitzgerald 为完成本书所作的工作，还要感谢 Osborne 公司的工作人员，特别是 Scott, Wendy, Mark 和 Cynthia Dennis。Yossi Amor 在我整个 IT 生涯中提供了很大的帮助，特别是其 SQL Server 的知识。

Michael Abbey

感谢我的父亲 Joe，他总是鼓励我以最大的潜能完成工作。那么，爸爸，我做得怎么样？Susan, Balia Jillian 是我的灵感和定位点，感谢 Michael Abbey 引导我进入这个世界 YATFG。Evan Ross 很高兴能与像你这样优秀的人一起写作。Mike Corey 是一位有包容心的人，非常感谢。对于其他组员 Ben, Larry, Rajan 和 Jeremy Fitzgerald，感谢你们一同完成这本优秀的书，这是我引以为自豪的一个小组。

Ian Abramson

感谢 Stev Shelgren, Rick Green, Mike Kosek 和 Dianne Gregg 支持我的工作，感谢我的研究伙伴 Bob Familiar，他是一位特别有创造力的人。最重要的是，我要感谢我的妻子 Linda 容忍我全身心地投入计算机事业中。

Larry Barnes

感谢 Dataspace Incorporated 中的每个人对我的支持，这是我所合作的最杰出的数据仓库专家组。另外，在这个过程中，David Gregory 的支持也是极其有价值的。而且，要感谢与我合作的所有新旧作者，希望有机会与他们再次合作。非常感谢，这总是很有意

义的！

Benjamin Taub

感谢我的妻子 Ririka 和儿子 Daniel 的耐心的牺牲。感谢家庭中的其他成员。如果没有 Mike C，我不会有机会写作的。在这个工作中，我还认识了许多人 Michael A., Ian, Ben, Larry, Evan 和 Jeremy，没有他们本书是写不出来的。

Rajan Venkitachalam

引 言

要说决策支持系统中的突破可能有点保守。随着技术的日新月异，这种现象或许比以前任何时候都更接近现实。我们在大学中见到的一本数据库理论书宣称，在过去的 20 年中，如果汽车工业的进步与计算机工业的进步相同，现在就能用 10 磅购买最近 25 年的 Rolls Royce。

数据仓库是这个区域的新成员？或许……我们认为这个项目只是多年来所做工作的另一种方式——使用许多数据源中的数据的大杂烩来进行决策。记得早期微型计算机（或许 PC 的 T-Rex——时钟速度为 7MHz 的 8088）中的 Lotus1-2-3 吗？进入单元格 C84，并将其中的数字 32000 改为 40000，再把光标移到临近的单元格中，观察电子表格中的变化。此时会问一个问题：“滑雪蜡产品清单中 25% 的变化会如何影响专业滑雪设备销售的底线？”然后，移到单元格 B68 中，就会看到其中的数字魔术般地从 23.6% 变为 23.9%。这难道不奇怪吗？

决策支持和 Microsoft——显然关联。找出全球最大的软件巨人，将它与各种软件需求结合起来。

《SQL Server 7 数据仓库》综合了两本书的内容。本书首先介绍了数据仓库，在规划决策支持解决方案的部署时要考虑许多新事务：小组成员、作用域、风险承担者、分析工具、数据库，这很容易使最富有雄心壮志的工程进入进退两难的境地。一旦掌握了这些知识，就会了解 Microsoft 划时代的数据库版本 SQL Server 7，它提供了全部集成化的数据仓库解决方案，且价格合理。功能全面和价格低廉的成功结合将数据仓库的蓝图改变为我们现在看到的样子。

SQL Server 7 的旅行是非常令人兴奋的。90 年代和不久的 20 世纪（看起来有点古怪，但我们愿意使用这两个数字，不会去除它们。）在向我们招手，本书讨论的许多理论是最基本的，无论选择什么样的后端（数据库的选择）、迁移和转换工具及前端（分析工具），都应采纳这些有益的建议。我们把不受控制的数据仓库比作快餐。商务按部就班地进行，生活是美好的，事情都按可管理的步伐前进。工作是超负荷的，最终期限在迫近，工程的作用域进展得很快、很疯狂，将完成日期都错过了。《SQL Server 7 数据仓库》将阻止这些事情的发生，并引导读者进行设计、管理和 Microsoft 决策支持系统的安装。读者将会喜欢这次冒险……

关于 CD

随书另附的 CD-ROM 包含 SQL Server 7 的评测版本，在安装 120 天后将过期。

系 统 需 求

- Pentium 或 Alpha 芯片(166MHZ 或更高)的个人电脑。
- Microsoft Windows 95, Windows 98 或 Windows NT Workstation 4.0, Windows NT Server 4.0 或 Windows NT Server 企业版(所有 Windows NT 版本均需安装 Service Pack 4 或更高版本)。
- 32MB 内存。
- Microsoft Internet Explorer 4.01, 含 Service Pack 或更高版本。
- 硬盘空间。

典型安装:185MB(含 20MB 的英文查询)

- 最小安装:80MB。
- 至多可使用两个处理器。
- 光驱。
- VGA 或更高分辨率的显示器，推荐 Super VGA。
- Microsoft 鼠标或兼容定点设备。

本软件由 Osborne / McGraw-Hill 公司在 Microsoft 公司特别授权后复制。因此, Osborne / McGraw-Hill 对该产品提供担保及支持, 如果您的 CD 有缺陷, 请将之送回 Osborne McGraw-Hill 公司更换, 请不要去 Microsoft 寻求技术支持, 该公司软件的最终用户不是 Microsoft 产品的已注册用户, 所以将不会拥有“已注册用户可得到的”改进、升级或其他服务。

数据库专业人员的宝库

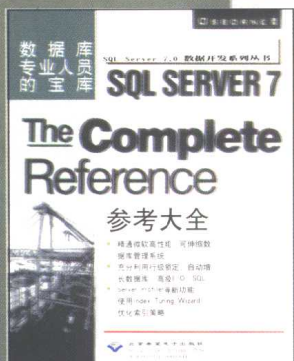
SQL Server 7.0 数据库开发系列丛书



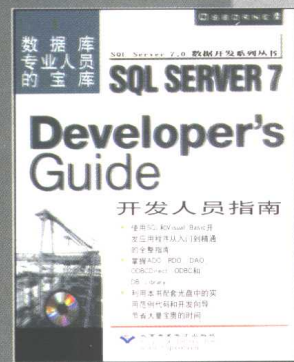
款款深情 倾情奉献



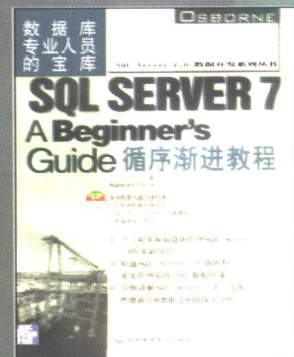
CX-2873
定价: 待订



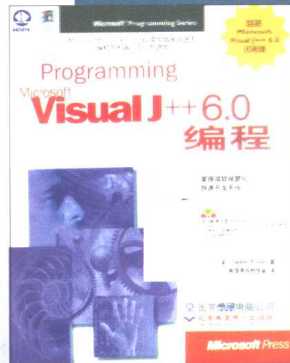
CX-2863
定价: 62.00元



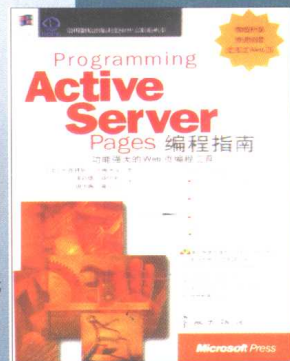
CX-2872
定价: 75.00元



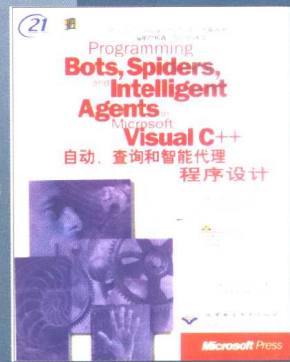
CX-2699
定价: 35.00元



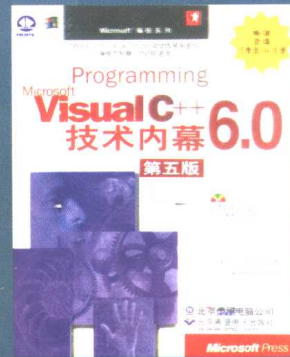
CX-2790
定价: 68.00元



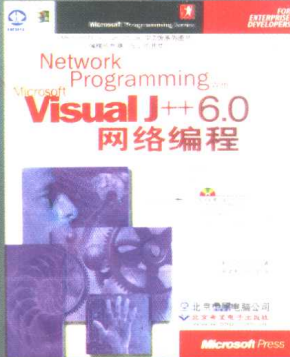
CX-2530
定价: 30.00元



CX-2852
定价: 66.00元



CX-2652
定价: 95.00元



北京希望电子出版社
Beijing Hope Electronic Press

地址: 海淀路82号(100080) 电话: 010-62562329, 62541992 传真: 010-62579874
网址: www.bhp.com.cn E-mail: qrh@hope.com.cn

美国微软 Office 权威系统编程书

精品推荐



CX-82881
定价:88.00 元



CX-82882
定价:55.00 元



CX-82854
定价:55.00 元



CX-82880
定价:68.00 元



CX-82891
定价:55.00 元



CX-82834
定价:50.00 元



CX-82829
定价:99.00 元



CX-82830
定价:30.00 元



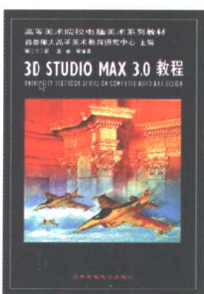
CX-82839
定价:120.00 元



CX-82837
定价:30.00 元



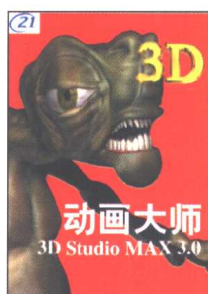
CX-82874
定价:30.00 元



CX-82826
定价:55.00 元



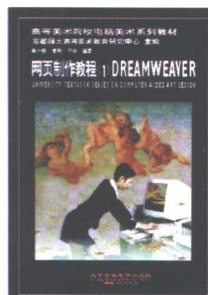
CX-82870
定价:35.00 元



CX-82859
定价:68.00 元



CX-82866
定价:35.00 元



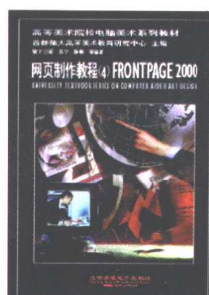
CX-82878
定价:55.00 元



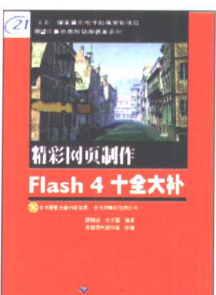
CX-82886
定价:50.00 元



CX-82909
定价:40.00 元



CX-82884
定价:50.00 元



CX-82887
定价:50.00 元



北京希望电子出版社

社址: 北京海淀路82号(黄庄路口东)

通讯: 北京中关村083信箱(100080)

电话: (010) 62562329 62541992

传真: (010) 62579874 62633308

网址: www.bhp.com.cn

目 录

第 1 章 什么是数据库仓库, 谁需要它, 为什么	1
1.1 令人沮丧的日子	1
1.2 目前报表体系结构中存在的问题	2
1.3 目标——决策支持	3
1.4 用于数据的自动柜员机 (ATM)	4
1.5 什么是数据仓库	5
1.6 决策支持与事务处理的区别	9
1.7 数据仓库: 一个不可告人的小秘密	9
1.8 投资收益率	10
1.9 关于本书	12
1.10 一个愉快的周末	12
第 2 章 应考虑的问题	13
2.1 不要只听本书所讲的——要从实际情况出发	13
2.2 质疑权威	13
2.3 从商业需求出发——不是技术需求	13
2.4 数据仓库中存储的是什么	14
2.5 数据市场还是数据仓库	15
2.6 开发阶段的大举措和小反复	17
2.7 参照完整性万岁	17
2.8 查询工具——可不止一个	18
2.9 数据仓库的差异——开发人员的展望	19
2.10 数据仓库的差异——用户的展望	20
2.11 给用户授权	21
2.12 为什么选择 Microsoft 建立数据仓库	21
第 3 章 筹建工作组	23
3.1 数据仓库工作组清单	23
3.2 数据仓库工作组所需的角色	25
第 4 章 数据仓库工程的管理	40
4.1 什么是工程管理	40
4.2 什么是工程	42
4.3 作用域说明	43
4.4 工作细目结构	45
4.5 工程估算	48
4.6 任务应直到工程停止前才完成	52

4.7 90 天——为成功进行规划	52
4.8 工程管理总结	52
第 5 章 数据仓库的设计	54
5.1 设计——下一个逻辑步骤	54
5.2 不断与用户交流	55
5.3 数据仓库中的数据库和操作系统中的数据库	56
5.4 “规范”的操作系统设计	58
5.5 维数数据仓库设计	60
5.6 星型模式	61
5.7 粒度	64
5.8 数据仓库设计的方法和指南	65
第 6 章 数据市场	71
6.1 数据市场	72
6.2 对数据市场的查询	78
6.3 数据仓库和数据市场	79
6.4 参照完整性	81
6.5 最佳的数据市场工具	83
6.6 数据市场还是数据仓库	84
第 7 章 物理数据仓库	86
7.1 巨型数据库	86
7.2 国家字符支持	87
7.3 SQL Server 7.0 体系结构	89
7.4 并行处理	90
7.5 查询处理器的优化程序和统计数据	92
7.6 数据库的建立和配置	93
7.7 理论用于实践	100
7.8 物理实现的检查列表	106
第 8 章 把数据移入数据仓库	108
8.1 数据交换服务概述	108
8.2 验证数据的有效性	109
8.3 清除数据	111
8.4 转换数据	112
8.5 移动数据	115
8.6 DTS 的输入和输出向导	115
8.7 为复制创建目标表	121

8.8 DTS 命令行——不是为了昏倒.....	121	13.9 用 NT 和 D SQL Server 7 实现 RAID	187
8.9 DTS 的 BCP.....	122	13.10 NT RAID 支持.....	188
8.10 DTS 程序包设计者.....	122	13.11 何时使用什么 NT RAID.....	190
8.11 用 DTS 进行开发.....	124	13.12 RAID——每个数据仓库都需要.....	190
第 9 章 查询处理器——数据仓库索引.....	125	第 14 章 为用户提供数据:查询技术简介...	192
9.1 用作索引的列.....	126	14.1 特定的访问.....	193
9.2 单列索引和复合索引.....	127	14.2 查询工具的特性.....	193
9.3 成簇索引.....	128	14.3 用 QUERY ANALYZER 生成报表.....	196
9.4 主键索引.....	129	14.4 Microsoft Excel.....	197
9.5 SQL Server 的查询处理器.....	130	14.5 Microsoft ACCESS.....	199
9.6 索引建立和维护的准则.....	134	14.6 COGNOS 的 IMPROMPTU.....	200
第 10 章 数据仓库高级技术.....	137	14.7 终端用户使用 IMPROMPTU.....	204
10.1 数据库文件.....	137	第 15 章 OLAP.....	208
10.2 文件和文件组的使用规则.....	143	15.1 决策支持.....	209
10.3 数据类型.....	144	15.2 ROLAP 和 MOLAP.....	212
10.4 设置列的数据类型.....	147	15.3 多维数据库 (MDDB).....	213
10.5 Unicode 数据的使用.....	149	15.4 关系数据库 (RDBMS).....	214
第 11 章 备份数据仓库.....	151	15.5 设计 OLAP 系统.....	222
11.1 SOL Server 7——如何存储数据.....	151	15.6 维数映射.....	229
11.2 数据仓库的备份.....	152	第 16 章 Microsoft 的决策支持服务器.....	231
11.3 如何恢复数据.....	162	16.1 SQL Server 7 OLAP 服务.....	231
11.4 硬件.....	165	16.2 终端用户访问.....	246
11.5 下一步——第三方实用程序.....	167	第 17 章 数据的采集.....	250
第 12 章 数据仓库的安全性.....	169	17.1 什么是数据的采集.....	250
12.1 安全策略.....	169	17.2 数据采集的优点.....	254
12.2 安全措施.....	169	17.3 数据采集协助决策制定过程.....	255
12.3 基于工具的安全机制.....	172	17.4 数据采集技术.....	256
12.4 SQL Server 7 的密码安全.....	174	17.5 数据采集解决方案检查列表.....	259
12.5 数据库用户管理.....	176	17.6 数据采集的未来.....	260
12.6 SQL Server 7 的 Enterprise Manager 和安全机制.....	176	第 18 章 数据仓库与万维网.....	263
第 13 章 RAID 和镜像.....	178	18.1 什么是 Internet.....	264
13.1 开发 RAID 的原因.....	179	18.2 什么是万维网.....	264
13.2 为什么要使用 RAID.....	182	18.3 Internet 在世界上的迅猛发展.....	265
13.3 RAID 0 级——磁盘条块化.....	182	18.4 什么是内部网络.....	266
13.4 RAID 1 级——磁盘镜像.....	182	18.5 内部网络的演变.....	266
13.5 RAID 2 级——不兼容.....	184	18.6 走进内部网络.....	267
13.6 RAID 3 级——带奇偶的数据条块化.....	184	18.7 内部网络和安全.....	268
13.7 RAID 4 级——带奇偶的数据 条块化 (一次一块).....	186	18.8 内部网络的优点.....	268
13.8 RAID 5 级——独立磁盘访问, 没有奇偶磁盘瓶颈.....	186	18.9 内部网络的缺点.....	271
		18.10 SQL Server Web Assistant Wizard.....	272
		18.11 运行 SQL Server Web Assistant Wizard.....	272

18.12 内部网络总结	276	19.7 Query Analyzer	293
第 19 章 数据仓库的调谐	278	19.8 索引分析器(Index Analyzer)	294
19.1 性能	278	第 20 章 数据仓库的维护	297
19.2 监视 SQL Server 7 的性能	282	20.1 管理任务自动化	297
19.3 调整 SQL Server 系统	285	20.2 响应事件	304
19.4 监视数据仓库的工作负荷	288	20.3 多服务器管理	305
19.5 创建一个配置器会话	288	20.4 定义多服务器作业	306
19.6 调节数据仓库	292	20.5 多服务器管理组成部件	307

第1章 什么是数据库仓库，谁需要它，为什么

在甲壳虫乐队主演的电影《黄色潜艇》中，Jeremy 不停的抱怨“时间太少了，而要学的东西太多了”。而现在对我们来讲，真是“缩写语词太多了，要学习的也太多了”。诸如 DSS、EIS、OLAP 以及 ODBC 等等这样的缩写词可以列出一大堆。多年来，计算机行业一直沉浸于无穷无尽的缩写词和新技术之中。实际上，术语“数据仓库”只不过是人们一直在讨论的概念的另一种重复而已。还记得以前使用的 LOUS-1-2-3 吗？这个软件首次尝试在台式机上处理“what if”。而这正是数据仓库所要解决的问题——使用企业收集到的信息，帮助用户工作得更好、更合理、更快、更有效率。

与通常的观念不同的是，大多数公司并不是仅为测试其技术人员的耐心才建立数据仓库的。事实上，数据仓库是一个优秀决策支持系统中的关键部件，本章中将介绍用户应从决策支持系统中得到什么，以及数据仓库为什么是满足这些需求所必须的。本章将回答以下几个问题：

- 什么是决策支持
- 决策支持的商业和技术目标是什么
- 什么是数据仓库
- 数据仓库的商业驱动是什么
- 数据仓库的技术驱动是什么

1.1 令人沮丧的日子

今天，这个人终于有幸成为信息主管(CIO)。有谁能想到他在 18 年前只不过是一名会计转行的 COBOL 程序员，而如今却位居如此重要的职位；有谁会想到公司有朝一日会让他担任如此重要的角色；有谁能想到 CIO 在研究价格昂贵的自动售货机时或在上高尔夫球课时，他的一句玩笑话却会如此引人注目，谁又能知道……

昨天，他购买的股票上涨了 76 点(或许今天会涨 74 点)令人非常兴奋，现正在 Internet 上查询着新股的行情。突然，一阵大吵大闹打破了宁静，窗外传来愤怒的吼声，他想：“噢，听起来好像是销售部副经理 John Valjon 和财务部主管 Norman Richelieu”。声音越来越近，“我不管他忙不忙，我要马上见他！Henchard 刚才训了我们一顿，我要知道为什么！”

他想，“Richelieu 太盛气凌人了，但我从未见过 Valjon 如此激动。到底出什么事了？”在争吵中，两位绅士怒容满面，额头汗水直流。

先是 Richelieu：“去年仅购买计算机一项就花了 2700 万美元，而结果呢？我和 John 一同去见那个大人物时，却发现我们报告上的数据居然不一致。John 说十月份共售出 4500 套，而我这儿的数据却是 42500 套。你的手下怎么会犯如此简单的错误？”

接着是 John：“这只是你我的问题之一。上周我见到了 FutureChef 行公司的采购主管 Ed kramdun 先生，我想用不着提醒你，在我们的十大客户中，Futurechef 名列第三。听说