



ORACLE®

Oracle 技术系列丛书

Oracle 8

数据仓库分析、构建 实用指南

Oracle 8 Data Warehousing
A Practical Guide to Successful
Data Warehouse Analysis, Build, and Roll-Out

Oracle Press 授权出版

(美) Michael J. Corey, Michael Abbey 著
Ian Abramson, Ben Taub

陈越 郭渊博 张红旗 等译



机械工业出版社
China Machine Press

OSBORNE

Oracle 技术系列丛书

Oracle 8 数据仓库分析、 构建实用指南

Michael J. Corey

(美) Michael Abbey 著
Ian Abramson

Ben Taub

陈越 郭渊博 张红旗 等译
王亚弟 审校



机械工业出版社
China Machine Press

本书全面阐述了 Oracle 8 独特的数据仓库化能力, 全面提供数据仓库规划、构建、生产、维护和性能优化方面的信息。

本书适合于数据库管理人员、信息技术管理人员、数据仓库化团队成员使用。

Michael J. Corey, Michael Abbey, Ian Abramson, Ben Taub: Oracle 8 Data Warehousing, A Practical Guide to Successful Data Warehouse Analysis, Build, and Roll-Out.

Original edition copyright © 1998 by The McGraw-Hill Companies. All rights reserved.

Chinese edition copyright © 2000 by China Machine Press. All rights reserved.

本书中文简体字版由美国麦格劳-希尔公司授权机械工业出版社独家出版。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。

版权所有, 侵权必究。

本书版权登记号: 图字: 01-1999-3476

图书在版编目 (CIP) 数据

Oracle 8 数据仓库分析、构建实用指南/(美)苏瑞 (Corey, M. J.) 等著; 陈越等译.
—北京: 机械工业出版社, 2000.2

(Oracle 技术系列丛书)

书名原文: Oracle 8 Data Warehousing, A Practical Guide to Successful Data Warehouse Analysis, Build, and Roll-Out.

ISBN 7-111-07749-0

I. O... II. ①苏...②陈... III. 关系数据库-数据库管理系统, Oracle 8-指南
IV. TP311.132.3-62

中国版本图书馆 CIP 数据核字 (1999) 第 55742 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: 郭东青

北京昌平第二印刷厂印刷·新华书店北京发行所发行

2000年1月第1版·2000年6月第2次印刷

787mm×1092mm 1/16·25印张

印数: 6 001-9 000册

定价: 50.00元

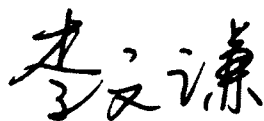
凡购本书, 如有倒页、脱页、缺页, 由本社发行部调换

贺 辞

祝贺中文版 ORACLE 8 系列丛书的出版。希望她能为中国的广大 ORACLE 用户和对数据库技术感兴趣的读者提供最先进的 ORACLE 技术知识。

ORACLE 软件系统有限公司希望通过中文版 ORACLE 8 系列丛书的出版，更好地建立起与广大用户和数据库工作者之间技术交流的桥梁。

Oracle 中国有限公司总经理

A handwritten signature in black ink, appearing to read '李文谦' (Li Wenqian), written in a cursive style.

1998.6

序

数据库技术的发展，使它已经成为现代信息技术的重要组成部分，成为现代计算机信息系统和计算机应用系统的基础和核心。可以说，如果没有数据库技术的发展，没有优秀的数据库产品的推出和应用，社会信息化的进程将是难以实现的。因此，在衡量一个国家信息化的程度时，其数据库的建设规模、数据库信息量的大小和使用程度也就成为重要的标志之一。

我国引进数据库技术始于70年代末，从微型计算机上运行的数据库到当前的大型数据库系统的引入和应用，已经有20多年的历史。20多年来，虽然在微型计算机数据库知识的普及和应用上取得了很大的进展和成绩，在大型数据库系统的开发和应用上也取得了进步，甚至还有了国产化的数据库软件，但如果从对数据库系统的应用效果和对数据库技术的掌握上来比较，则与发达国家之间仍然存在较大的差距，特别是在大型数据库系统的开发、建设和应用水平方面差距更大。这种差距主要表现在两个方面：第一是数据库的数量及所收集信息的数量和质量与发达国家相比存在很大的差距；第二是对数据库技术和产品的掌握和应用上更有待于提高和加强，应用人才急需培养，经验有待积累和总结。前者的改善应依靠于对数据库应用基础工作的加强，如重视基础数据的收集和整理，即重视数据库工程的建设，并制定相应的数据政策；而后者的改善则会更多地依赖于人们对数据库技术和对数据库产品的掌握。为此，我们必须首先占有充足的资料并加以消化。如果数据库厂家和出版机构能提供较完整的、质量较高的技术资料 and 书籍，并为较多的数据库技术人员和应用人员所掌握，无疑将会促进国内数据库技术人才的成长并推动数据库应用水平的提高。

喜闻机械工业出版社华章公司与ORACLE出版公司合作，为配合ORACLE 8在中国的发行，由机械工业出版社买断了ORACLE出版公司出版的ORACLE 8系列丛书的中文简体字版的出版权，并组织国内从事ORACLE应用开发的科技人员和教学人员进行翻译出版，还邀请国内数据库专家对译稿进行了审定，以保证丛书在技术

上的权威性。无疑，这对大型数据库系统特别是ORACLE数据库系统的开发和应用将起到很好的推动作用。

众所周知，ORACLE公司推出的ORACLE 8是一种面向网络计算的数据库(the database for Network Computing)并支持对象关系模型的数据库产品。该系列丛书全面地介绍了ORACLE 8的功能和技术，具体书目见封底。

该丛书内容丰富，涵盖了大型数据库应用开发中的全部技术内容，有的资料，如《Oracle 8数据仓库分析、构建实用指南》、《新版Oracle 8故障解决手册》等都是在国内首次面世，很值得数据库技术工作者参考和阅读。

无疑，这套丛书应该有广泛的读者，它可供大型数据库系统，特别是ORACLE系统应用的系统管理员、应用程序员、系统分析员和设计人员以及广大用户学习和参考，也可做为ORACLE系统培训的教材和高等学校本科和研究生的学习参考书，同时也是掌握大型数据库系统理论和实践的好材料。

应感谢机械工业出版社的领导和组织翻译丛书的先生们和女士们，辛勤的劳动将为我国信息化事业的发展增加动力。ORACLE中国有限公司对丛书的出版给予极大的关注与支持，李文谦总经理并为丛书的出版题词。

预祝丛书的出版和发行获得成功，并得到读者的欢迎。

中国科学院研究生院 教授
ORACLE大学(中国) 校长

罗晓沛

译者序

随着市场竞争的不断加剧和信息系统的广泛应用,人们不再满足于仅能提供日常事务处理的系统,他们要求信息系统能对企业、机构的决策提供支持。在当今信息时代,决策者们越来越意识到他们的宝贵数据资源对其经营策略是何等重要,如何把它们转变为信息并正确理解其含义将有助于管理层对其业务有更深刻的了解并制定出更为有效的商业决策。于是,数据仓库技术应运而生了,它是一个为支持机构的决策处理过程而设计的信息仓储,其目的是为智能化事务决策提供必要的信息管理。企业根据其需求,适当地应用数据仓库技术,就能为其带来空前的投资回报。

本书是由数据仓库的“做者”所写的,不同于一般纯技术性书籍,它从满足商业需求的数据仓库战略着眼,从成功实现数据仓库工程的战术入手,全面地阐述了数据仓库的分析、构建、生产和使用的每一重要环节,其中包含了作者大量的商业运作、工程管理、实现技术等各方面的实际经验和指导原则,堪称数据仓库化的“孙子兵法”。

本书介绍了数据仓库的基本概念、基本原理和基本结构;深入研究了数据仓库的工程管理;基于 Oracle 8 技术着重讨论了数据仓库的逻辑设计、物理设计、聚居数据、分区、备份与恢复、安全、性能优化及对其实现提供支持的 RAID 技术;讲述了用户如何利用数据仓库进行数据获取,介绍了 OLAP (联机分析处理)、数据挖掘、Internet 等技术,并以 Oracle 为核心用形象生动的实例介绍了新一代的查询、OLAP、聚集导航工具。从而形成了以数据仓库为中心,以 OLAP、数据挖掘、模型库、Internet 为数据访问方式的综合决策支持系统体系。

数据仓库技术是一种方兴未艾的技术,特别是在我国还未被广泛地应用,我们为能将本书介绍给国内的企业决策者和高级管理人员、信息系统管理人员与开发者、商务专家和顾问、以及广大当前和未来数据仓库的用户而感到由衷的欣慰。

参加全书译制工作的有陈越、郭渊博、张红旗、谢辉、贺小川、杨英杰、王鲁、韦大伟、王光霞、席建民、牛光、曹守见、王新昌、杜学绘、杨向荣、范新峰、鹤荣育、易晓峰、李景峰、张彦杰、费晓飞、卢贤玲、宋震、何熹、王国正、麻信洛、季晓中、徐文军、张景生、牟书贞等,由陈越负责统稿,由王亚弟审校。

在此,还要感谢同事们、家人为本书完成所提供的支持和帮助。感谢我们这个团结的集体。

由于译者水平有限,时间仓促,错误之处在所难免,望广大读者不吝指正。

译者

1999 年 10 月于郑州信息工程大学

引言

当 Mike Corey 打来电话，请求我为他与 Michael Abbey 合著的一本书写篇引言时，我并不惊奇。的确，作为全世界 Oracle 系统开发人员不可缺少的资料，他们关于数据仓库的著作非常流行，实用指南尤其适用于高级管理人员和技术员工。我马上意识到本书的市场前景必将十分明朗。因而，我很快同意为这本书写一篇引言。

我以前写过关于数据仓库在一个有线电视公司的使用和开发的文稿，我曾担任该公司的信息主管。这个公司正在逐步把它自己重新定义为一个“宽带”公司，一个所谓的“宽带”公司就是它可以超越传统有线电视的限制来传递信息。这样的公司通过它的数据线路或光纤网络，提供包括数据、语音、视频及其他新型的交互式服务。在公司转型过程中，关于顾客、商业趋势、商业竞争及无数其他信息要素的深入准确的认识是确保成功转型的一个关键因素。要有效管理如此庞大的信息量，建立相应的数据仓库是成功完成这种转型的重要一环。现在，在这个公司的战略核心中仍保持着数据仓库的投资。

最近，我在天气频道公司工作。就像被闪电照亮两次一样，我再次为数据仓库的广阔前景所吸引。可以肯定地讲，天气频道公司在日常工作的基础上获得了大量的气象数据，这些数据对实时预报是有用的，同时也可以作为历史数据而供日后天气趋势分析使用。目标用户（如飞行员和大型农场主），可以从由数据挖掘得到的准确知识中获利颇丰。这样一种数据挖掘过程可以带来与天气情况相关的大量产品。当前，公司的高级经理已开始意识到这种数据仓库的潜力，一个全方位数据仓库工程的开发工作正在进行中。

在这样两个位置上（对于有线电视公司来讲作为一个信息分发者，对于天气频道公司来讲作为一个内容创建者），我把数据仓库的重要性和影响排在 IT 业投资的第三位。在哈佛商学院的实例库中充满着各种研究案例，其中许多都是关于企业由于使用数据仓库而成功崛起的。例如，Lexus 公司可以使用数据仓库的分析工具来分析在消费品市场上所有顾客的消费行为，同时可以避免研究得到过时企业业绩信息的陷阱。WAL_MART 同样使用复杂的数据库技术以保证库存管理准确反映用户需求，而不必臆测用户的需求。

但是数据仓库系统并不便宜。每个 Lexus 经销商必须购买自己的 IBM AS400 电脑和一个卫星天线，将销售信息和服务信息传送到美国加利福尼亚州特朗丝地区的总部。除了这些技术性投资之外，许多服务器、数据库、开发者的预算都需要被分配相应款项。控制数据仓库工程的成本（包括硬件、软件、开发费用）是件困难的任务。仔细研究商业案例，并且在具体的顾客知识上取得巨大进展，对于开发数据仓库工程极其重要。当然，翻一翻关于那些“神像”（数据仓库）的实用指南，也可以对那些工程人员有所帮助。

这是一部关于“怎样做”的书，它是“做者”写的。我希望它可以成为一个行动的召唤。它将首次挑战读者：“在我的公司中，数据仓库可以起到什么重大作用？它是如何工作

的？我们应该收集什么样的数据？我们应该挖掘什么样的数据？”我保证它将会激励每个读者去行动。

Robert Strickland

Weather Channel 公司高级副总裁、信息技术主管

Robert Strickland 先生住在马萨诸塞州波士顿地区附近，他是天气频道公司的高级副总裁、信息技术主管。天气频道公司的总部位于乔治亚州的亚特兰大城，它是 LandMark 公司附属公司。Robert Strickland 先生以前曾是 MediaOne 公司主管信息系统的高级副总裁。

前 言

数据仓库——这个名词使人想起这样一种情形：“饥渴”的用户正在访问大量的数据。那么，所有这些用户来自哪里？他们想做什么？他们为什么在这儿？系统专业人员能够做哪些让用户满意的事？我们中的许多人都还记得 Kevin Costner 的电影——“梦之田园”，讲述的是一个美国家庭在他们的前院建立了一个空想的篮球场，希望那些篮球大师们将会魔术般地出现，结果他们真的出现了！电影的主题是：“如果我们建立它，他们就会来！”猜猜数据仓库会怎样？数据仓库恰恰相反，即我们最好叫它“田园之梦”，归结到电影主题上就应该是“如果他们来了，我们应该建立它！”。在数据仓库的体育馆中，最重要的球员应该是用户。当他们需要时，我们就提供他们想要的。我们拥有技术，那么障碍是什么呢？

我们经常使用数据仓库这样一种新的方式来做陈旧的事，通俗讲就是，“新瓶装旧酒，换汤不换药”。在 20 世纪 80 年代，我们中的许多人都用电子表格软件来清除工作障碍，整天盯着那些表示商业销售额、收据、借方、贷方、待付款项、可付款项及其他内容的数据。想到着手处理 AB67 单元，数字从 12 390 ~ 24 780，看到这些变动引起屏幕上曲线的微微起伏，我们非常高兴。这是使用“假设分析”工具的首次探险。从那时起，接连不断地出现了许多种决策支持方案——从最基本的到现在最复杂的产品。

数据的仓库化可将数据转变成信息。在大范围的数据源中系统地收集相关的商业数据、数据的表示、数据访问方式的描述都是这一过程的组成部分。简单点讲，我们使用数据仓库得到信息，而且把信息传送到大范围的用户——那些在公司底层的人和那些在公司会议室中的实权人物。

本书总结了数据仓库在四个商业应用的使用经验，除了 Oracle 8 系统的实现细节之外，通过本书可以熟悉数据仓库的设计、构建和生产。我们在本书中谈及了数据仓库相关的“是什么”和“为什么”的问题，讨论了某些如何把数据仓库从理论变成现实的问题，这些问题必将使您获益匪浅，同时本书还深入研究了任何工程项目最令人头疼的领域——办公室政治。很多工程项目就只是因为参与者不能良好地合作而失败的，它们要么踌躇不前，要么勉强进行，要么干脆下马。

数据仓库的设计是一项多律性的设计任务。“设计”不仅仅意味着设计一个数据库，它还意味着你需要设计你的数据抽取、数据加载以及终端用户的数据访问方式。现在许多数据仓库的书解释了设计数据仓库的基本问题——规范化 (Normalization)、逆规范化 (denormalization) 及星形模式 (star schema) 等。本书同样也包括上述问题，而且提供一种被证明是有效的方法，来设计数据仓库。我们阐述了怎样进行数据仓库的设计，解答了你的疑问以及那些在设计数据仓库时可能对你有所帮助的各个方面的问题。我们提供的方法将引导您完成某些需要完成的任务，这些任务构成了数据仓库的设计基础。Oracle 8 在下列方面也是独一无二的：

- 我们的解决方案是面向用户的——现在的许多数据仓库不能满足用户的需求。
- 我们的解决方案展示用户怎样参与设计过程——用户应该广泛参与数据仓库的设计

过程，以保证数据仓库的成功实现。

- 我们的解决方案将引导您走过设计的过渡阶段，并且帮助您成功地建立数据仓库。它的设计方法与当前操作型系统（Operational System，又称为 OLTP 数据库，它是支持商业事务操作的数据库系统——译者注）的设计方法是有差异的，它是另一种设计范例。

我们将探讨把在制图板上所设计的数据仓库变成实际的、大量用户可使用的数据仓库过程中的各个方面的问题。当许多企业正在市场上寻找可以解决他们的数据存储和检索问题的数据库系统时，Oracle 公司前期投入市场的 Oracle 服务器产品给用户带来很大效益，从而使 Oracle 公司更加靠近市场前沿。在本书中，我们涉及了 Oracle 8 的细节，并且集中从软件角度来加以讨论。我们还看到了在拓展数据仓库整套解决方案时会遇到的普遍性问题。把工程项目按时按预算成本完成并获利的策略，将与 Oracle 8 的技术问题一起讨论。在设计工作的早期，以及在其之后的基础性工作阶段，关注这些技术问题是明智的。数据的仓库化——您应该决定接受的任务，它将会改变您的生活。请继续读吧。

目 录

贺辞	
序	
译者序	
引言	
前言	
第1章 数据仓库概论	1
1.1 特别糟糕的一天	1
1.2 当前报表构架的问题所在	2
1.3 目标：决策支持	3
1.4 一个数据自动出纳机	4
1.5 什么是数据仓库？	5
1.5.1 面向主题	5
1.5.2 完整性	6
1.5.3 非易失性	7
1.5.4 居于不同时间	8
1.6 决策支持不同于事务处理	8
1.7 数据仓库：一个肮脏的小秘密	9
1.8 投资回报	10
1.9 关于本书	12
第2章 需要考虑的问题	13
2.1 不要轻信我们——应该客观点	13
2.2 文章和书籍只包含观点，而不包含事实	13
2.3 从商业需求开始——而不是技术	13
2.4 哪些数据进入数据仓库？	14
2.5 数据集市还是数据仓库？	15
2.6 大笨钟与小迭代——开发阶段	16
2.7 参照完整性（RI）永存	17
2.8 查询工具——多多益善	18
2.9 数据仓库的不同之处——开发者的观点	18
2.10 数据仓库的不同之处——用户的观点	19
2.11 授权用户	20
2.12 Oracle 数据仓库为什么独占鳌头	21
2.12.1 全面解决方案	21
2.12.2 OLAP 和通用服务器	21
2.12.3 一个仓储/多种解决方案	22
2.12.4 现存的关系	22
2.13 投资回报	22
第3章 组建开发团队	24
3.1 数据仓库开发团队角色清单	24
3.2 数据仓库工程所需要的角色	25
3.2.1 数据仓库工程导演	26
3.2.2 数据仓库工程经理	27
3.2.3 数据准备专家/商务分析专家	29
3.2.4 数据仓库工程师	29
3.2.5 数据库管理员	30
3.2.6 系统管理员	31
3.2.7 数据迁移专家	32
3.2.8 数据转换/整饰专家	33
3.2.9 数据集市开发负责人	33
3.2.10 质量保证/测试专家	34
3.2.11 基础结构专家	35
3.2.12 权威用户（旧有系统专家）	35
3.2.13 培训者	36
3.2.14 技术书记员	36
3.2.15 公共关系人员	36
3.2.16 企业中的盟友	37
3.2.17 桌面帮助功能设计人员	37
3.2.18 工具专家	37
3.2.19 经销商关系人	38
3.2.20 Web 大师	38
3.3 顾问	39
第4章 数据仓库工程管理	40
4.1 什么是工程管理？	40
4.2 什么是一个工程	42
4.3 工程范围说明书	43
4.4 工作细目结构	45
4.5 工程估算	47
4.6 工程管理总结	51
第5章 数据仓库设计	53
5.1 设计——下一逻辑步骤	53
5.2 深入与用户交流	53

5.3 数据仓库数据库与操作型数据库	55	7.3.4 并行度	93
5.4 “规范的”操作型设计	56	7.3.5 在实例级设置并行查询	94
5.4.1 第一范式	57	7.4 为基于代价的优化器收集统计信息	95
5.4.2 第二范式	57	7.4.1 SQL 的 Analyze 命令	96
5.4.3 第三范式	58	7.4.2 粗估与计算的比较	97
5.5 维数据仓库设计	59	7.4.3 如何分析对象	97
5.6 星型模式	60	7.4.4 何时分析对象	99
5.6.1 事实表	61	7.4.5 在何处查看统计数据	100
5.6.2 维表	61	7.4.6 过时的统计数据	100
5.7 粒度	63	7.4.7 统计数据收集的一致性	101
5.8 数据仓库设计与指南	64	7.5 分割表空间	102
5.8.1 工程分析与计划	64	7.5.1 系统支持的表空间	102
5.8.2 高层设计	66	7.5.2 应用表空间	103
5.8.3 中层设计	67	7.5.3 为用户指定表空间	104
5.8.4 低层设计	68	7.6 从理论到实践	104
第 6 章 数据集市	70	7.6.1 系统的逆规范化	104
6.1 数据集市	71	7.6.2 星型模式的实现	108
6.1.1 独立数据集市	71	7.7 双向模式存取方法	111
6.1.2 维数据库	74	7.7.1 双向模式的属主	112
6.1.3 影响设计的维	74	7.7.2 存取模式	112
6.1.4 下钻需求对设计的影响	76	7.7.3 激活和挂起一个模式	113
6.2 查询数据集市	77	7.7.4 双向模式存取方法的威力	115
6.2.1 查询条件的类型	77	7.8 物理实现核查清单	116
6.2.2 聚集	78	第 8 章 将数据移入数据仓库	118
6.3 数据仓库与数据集市	78	8.1 数据库对象	119
6.4 参照完整性	80	8.2 SQL * Loader	119
6.4.1 主键	80	8.2.1 SQL * Loader 的特征	120
6.4.2 外键	81	8.2.2 激活 SQL * Loader	120
6.5 最好的数据集市工具软件	81	8.2.3 SQL * Loader 的输入和输出	122
6.6 数据集市还是数据仓库?	83	8.2.4 控制文件	123
第 7 章 物理数据仓库	85	8.2.5 并行和直接装载	133
7.1 超大型数据库	85	8.2.6 抑制产生重演日志和恢复信息	134
7.1.1 时机窗	85	8.3 导入和导出	136
7.1.2 精心管理和填充超大型数据库	86	8.3.1 操作方式	137
7.2 民族语言支持 (NLS)	87	8.3.2 操作方法	137
7.2.1 客户端民族语言支持	87	8.3.3 导出	137
7.2.2 服务器端民族语言支持	88	8.3.4 导出程序的参数	138
7.2.3 初始化参数文件	89	8.3.5 关于导出分区表的一点说明	139
7.3 并行性	90	8.3.6 导出样例 1	139
7.3.1 大规模并行处理和对称式多处理	90	8.3.7 导出样例 2	142
7.3.2 并行查询	90	8.3.8 导出样例 3	142
7.3.3 并行化与数据仓库	92	8.3.9 导入	143
		8.3.10 导入程序的参数	144

8.3.11 关于将数据导入分区的一点说明	146	10.8.1 建立索引时, 抑制产生恢复信息	196
8.3.12 导入样例 1	147	10.8.2 快速索引重建	197
8.3.13 导入样例 2	148	10.9 索引的创建和维护指南	197
8.3.14 导入样例 3	149	10.9.1 装载数据, 然后建立索引	198
8.4 使用 Oracle 企业管理器装载数据	150	10.9.2 为索引的创建分配大的排序空间	198
8.5 Oracle 数据集市套件	152	10.9.3 临时工作空间	198
8.6 基于引擎的工具	153	10.9.4 共享池和索引建立	198
8.6.1 ELT 工具特性	154	10.9.5 每个表中的索引数目	199
8.6.2 功能性测试	156	10.9.6 明确地设置存储参数	201
8.6.3 数据量测试	156	10.9.7 事务项参数	201
8.7 代码生成工具	157	第 11 章 备份与恢复数据仓库	202
8.7.1 ETI * EXTRACT	158	11.1 备份	202
8.7.2 Prism Solutions 公司的数据仓库软件	159	11.1.1 提供保护机制防止错误	202
8.8 Oracle 透明网关	162	11.1.2 数据备份能提供什么	204
8.9 在数据转换中运用网关技术	163	11.1.3 影响备份策略的因素	206
第 9 章 分区	164	11.1.4 最佳备份时机	209
9.1 什么是分区选项	164	11.1.5 何时进行备份	209
9.1.1 分区表与未分区表的区别	164	11.1.6 映像备份	210
9.1.2 创建分区对象	165	11.1.7 无人值守的 24 * 7 备份	211
9.1.3 选择分区键	166	11.1.8 备份技术的发展趋势	213
9.1.4 行放置	177	11.2 恢复	219
9.2 传统的分区	179	11.2.1 恢复的种类	219
9.2.1 分割	179	11.2.2 用映像备份进行恢复	223
9.2.2 分块	181	第 12 章 数据仓库安全	225
第 10 章 索引数据仓库	183	12.1 安全策略	225
10.1 索引哪些列	184	12.1.1 工作站安全	225
10.1.1 考虑列的可选度	184	12.1.2 窥探	226
10.1.2 考虑谓词中涉及的列	184	12.2 安全	226
10.1.3 数据仓库的独特性	185	12.2.1 查看数据仓库数据	226
10.2 单列索引和复合索引	185	12.2.2 如何在数据仓库中进行权限管理	227
10.3 Oracle 8 索引方法	187	12.2.3 在何处实现安全	229
10.3.1 B 树索引	187	12.3 基于工具的安全	229
10.3.2 位图索引	187	12.3.1 表的访问	230
10.4 分区表索引	190	12.3.2 商业域的访问	231
10.5 索引组织化表	191	12.4 Oracle 8 的口令安全	231
10.5.1 定义索引组织化表	191	12.4.1 帐号锁定	232
10.5.2 索引组织化表的优点	192	12.4.2 口令生存期与口令到期	232
10.5.3 何处使用索引组织化表	193	12.5 数据库用户管理	232
10.6 优化直方图	193	12.6 Oracle 企业管理器与安全	233
10.7 主键索引	195	12.6.1 Users	234
10.8 优化索引创建过程	195		

12.6.2	Roles	235
12.6.3	Profiles	235
12.6.4	OEM 的威力	236
第 13 章	冗余独立磁盘阵列	237
13.1	RAID 的产生与发展	238
13.1.1	革新阶段	240
13.1.2	仿制阶段	240
13.1.3	成熟阶段	240
13.1.4	巩固阶段	240
13.1.5	商品化阶段	241
13.2	为什么使用 RAID	241
13.3	RAID 1——磁盘镜像	242
13.4	RAID 2——不可兼容	243
13.5	RAID 3——有奇偶校验位的数据 分割	243
13.6	RAID 4——有奇偶校验位的数据 分块	244
13.7	RAID 5——独立的磁盘访问、无 奇偶校验的磁盘瓶颈	244
13.8	RAID——未来的选择, 为什么每 一个数据仓库都需要它	245
第 14 章	用户对数据的获取	247
14.1	查询工具的特性	248
14.1.1	易于使用	249
14.1.2	性能	250
14.1.3	多数据源	250
14.1.4	集中式管理	250
14.1.5	数据安全性	251
14.1.6	与 Web 的融合	252
14.1.7	集成化分析	253
14.2	使用 SQL* Plus 生成报表	254
14.3	Oracle Discoverer	255
14.3.1	Oracle Discoverer 管理器	255
14.3.2	Discoverer 终端用户工具	258
14.4	Cognos Impromptu	261
14.4.1	定义到数据库的连接	261
14.4.2	创建目录	262
14.4.3	添加表	262
14.4.4	定义所要求的连接	263
14.4.5	创建文件夹结构	264
14.4.6	增加参数配置文件	265
14.5	从终端用户的角度看 Impromptu	266
14.6	总结	268

第 15 章	关系型 OLAP 与聚集导航 器	269
15.1	OLAP: 仅仅是四个单词吗?	269
15.1.1	多维性	269
15.1.2	钻取	271
15.1.3	旋转或切片与切块	273
15.1.4	多视图模式	274
15.1.5	OLAP 小结	275
15.2	ROLAP 与 MOLAP	275
15.2.1	多维数据库	276
15.2.2	关系型数据库	277
15.2.3	回到特洛伊: ROLAP 与 MOLAP	280
15.2.4	ROLAP 与 MOLAP 的和平共处	284
15.3	为提高 ROLAP 的性能而对数据建 模	285
15.4	三种 ROLAP	286
15.5	结论	287
第 16 章	联机分析处理	288
16.1	什么是 MOLAP、ROLAP 和 DOLAP	290
16.2	MOLAP 还是 ROLAP?	290
16.3	我们为什么需要 OLAP?	291
16.4	OLAP 工具	292
16.4.1	OLAP 工具的评估	293
16.4.2	使用 Cognos Transformer 建立 OLAP 模型	294
16.4.3	Cognos 的 PowerPlay	299
16.5	Oracle 与 OLAP 的市场	304
第 17 章	Oracle Express 方案	305
17.1	为什么要用 Express	305
17.2	Express 的优势	306
17.2.1	Express 适合数据仓库	307
17.2.2	完全开放的连接	308
17.2.3	标准的操作环境	309
17.2.4	一步到位的方案	309
17.2.5	与数据仓库的集成	310
17.3	Oracle Express 服务器	310
17.3.1	概述	310
17.3.2	通信基础设施	311
17.4	Express 数据模型	311
17.4.1	维	312
17.4.2	变量	312
17.4.3	公式	312

17.4.4 关系	312	18.5.3 吸收大量数据的能力	349
17.5 使用 Oracle Express 服务器完成数 据操作	313	18.5.4 对数据质量的敏感性	349
17.6 Oracle Express 管理器	321	18.6 数据挖掘的前景	350
17.6.1 创建数据库	321	第 19 章 数据仓库与万维网	351
17.6.2 与现存数据库相关的操作	323	19.1 什么是 Internet	351
17.7 Express DBA 简介	324	19.2 什么是 World Wide Web	352
17.7.1 数据库	324	19.3 Internet 的飞速发展	353
17.7.2 导入/导出	325	19.4 什么是 Intranet	354
17.8 Oracle Express 对象	326	19.4.1 Intranet 的演化	354
17.9 Express 对象之旅	328	19.4.2 走近 Intranet	355
17.10 Oracle 财政分析器	331	19.4.3 Intranet 与安全	356
17.10.1 与 Oracle GL 的集成	331	19.4.4 Intranet 的优势	356
17.10.2 关键特性及功能	332	19.4.5 Intranet 的劣势	358
17.11 Oracle 销售分析程序	332	19.4.6 Intranet 小结	359
17.12 关系型访问管理器	333	第 20 章 数据仓库的协调	361
17.13 Oracle Web 代理	334	20.1 共享池	361
17.14 Express Web 应用程序的工作情况	336	20.1.1 三个共享池参数	362
第 18 章 数据挖掘	338	20.1.2 确定这些参数的值	363
18.1 什么是数据挖掘?	338	20.2 有关协调的各种特性	364
18.1.1 发现	339	20.2.1 只读表空间	364
18.1.2 关系	340	20.2.2 临时表空间	366
18.1.3 模式	340	20.2.3 并行感知优化器	367
18.2 数据挖掘带来的收益	342	20.2.4 读优先机制	367
18.2.1 欺诈检测	342	20.2.5 导出程序的直接路径存取选 项	368
18.2.2 投资回报	343	20.2.6 Hash 连接的处理	368
18.2.3 电子化方案的可伸缩性	343	20.3 主管程序	370
18.3 数据挖掘辅助决策	344	20.3.1 生成参数配置文件	370
18.4 数据挖掘技术	344	20.3.2 指派参数配置文件	371
18.4.1 神经网络	344	20.3.3 改变参数配置文件并重新设 置用户	372
18.4.2 关联发现	345	20.4 初始化参数文件	373
18.4.3 分类	346	20.4.1 参数项类型	373
18.4.4 分簇	347	20.4.2 重要参数的描述	373
18.4.5 顺序发现	348	20.5 使用不同的初始化参数文件启动 Oracle	378
18.5 数据挖掘方案核查表	348	20.6 OEM 实例管理程序	379
18.5.1 对 Oracle 数据库的直接访问	348		
18.5.2 可视化分析的能力	348		

第1章 数据仓库概论

在比特斯的电影——“黄色潜水艇”中，Jeremy 老是在抱怨：“如此少的时间，需要知道的又是如此之多”，这句话让我们来讲就是：“缩写词愈多，需要知道的愈多”。DSS、EIS、OLAP、ODBC 是不是缩写词？如果把这个缩写词清单列出来，你能列上好几页。许多年以来，计算机界一直热衷于缩写词和新技术。实际上，术语“数据仓库”仅仅只是另一个概念的重复而已，而这个概念已经使用过好些年了。记不记得 Lotus1-2-3？它第一次让你品尝到“假设分析（what-if）”桌面处理工具的滋味。这是一个数据仓库可以做到的全部——使用收集到的信息来帮助你的企业，让它的反应更准确、更灵敏、更快捷、更有效。

与流行的看法恰恰相反，大部分公司企业没有建立数据仓库。然而数据仓库却是决策支持系统能够正常运作的关键部件。本章将帮助理解用户需要从决策支持系统中得到什么，以及为什么只有数据仓库才能满足这些需要。我们将回答下列问题：

- 什么是决策支持系统？
- 决策支持系统的商业目的和技术目标是什么？
- 什么是数据仓库？
- 数据仓库的商业驱动力是什么？
- 数据仓库的技术驱动力是什么？

1.1 特别糟糕的一天

假设你最终已成为一名信息主管。18 年以前，当你作为一名会计师开始编写 COBOL 应用软件时，谁能想到你将会坐到今天这样一个重要位置？谁能想到你所在的企业会让你成为制定战略的角色？谁能想到你开的玩笑——例如信息主管们经常在经销商野餐联欢会和高尔夫球上消磨时间——几乎都是事实？谁知道呢？

昨天你的高尔夫打了 76 杆，（今天可能会打 74 杆），你在互联网上核对了你的新股票的价值。办公室外突然喧闹起来，你听到几个充满愤怒的声音：“嗨”，你在想：“那声音听起来很像 John Valjon 和 Norman Richelieu，他们分别是我们的销售副主管和财务主管”。你可以听得更清楚，附近有人在高声说话：“我根本不管他是不是很忙，我现在就要见他！Hen-chard 刚刚把我们赶了出来，我想知道些答案。”

你想：“现在 Richelieu 是有点傲慢，但我从不知道 Valjon 也是如此激动，我想知道到底发生了什么？”在这场暴风雨中，两位绅士由于愤怒脸涨得通红，额头也渗出了汗珠。

Richelieu 开始讲了，“我们仅去年就在计算机系统建设方面花掉 2700 万，但当 John 和我一同去见那个大人物时，为什么会这样——我们的报表根本不匹配？John 认为他在 10 月份卖出 45 000 宗货物，但我的报表上只显示了 42 500，怎能这样？你的雇员怎能犯这样简单的错误？”

John 接着说：“那仅仅只是我与你的人之间存在的问题之一。上个星期，我同 Ed Kram-den —— FutureChef 工业公司的采购主管会谈，在我们的目标帐目的前十位中，FutureChef 公