

开发专家  
之 Oracle9i

# Oracle9i

## 数据仓库构建技术

飞思科技产品研发中心

编著



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: <http://www.phei.com.cn>

开发专家之 Oracle9i

# Oracle9i 数据仓库构建技术

飞思科技产品研发中心 编著

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

本书作者总结多年数据库开发经验,全面阐述了 Oracle9i 和 Oracle Warehouse Builder 独特的数据仓库化能力,提供了数据仓库规划、构建、生产、维护和性能优化等方面的知识和技能。本书主要内容包括数据仓库概念、Oracle Warehouse Builder 概述、OWB 安装与配置、使用 OWB 创建工程、数据仓库的逻辑设计、数据仓库物理设计、数据仓库的硬件和 I/O 考虑、数据装载和转换、实体化视图、数据仓库并行处理和分区、数据仓库中用于聚集的 SQL、在数据仓库中用于分析的 SQL、使用 XML 工具集、查询重写和联机分析处理。

本书适合于数据库管理人员、开发人员,以及广大 Oracle 数据库的爱好者,同时也适合大专院校计算机、信息系统、管理专业的师生作为教学参考用书。

未经许可,不得以任何方式复制或抄袭本书的部分或全部内容。

版权所有,侵权必究。

## 图书在版编目(CIP)数据

Oracle 9i 数据仓库构建技术/飞思科技产品研发中心编著. —北京:电子工业出版社,2003.1

(开发专家之 Oracle9i)

ISBN 7-5053-8117-2

I .O... II.飞... III.关系数据库—数据库管理系统, Oracle9i IV.TP311.138

中国版本图书馆 CIP 数据核字(2002)第 088954 号

责任编辑:郭 晶 赵红梅

印 刷:北京市增富印刷有限责任公司

出版发行:电子工业出版社 <http://www.phei.com.cn>

北京海淀区万寿路 173 信箱 邮编:100036

经 销:各地新华书店

开 本:787×1092 1/16 印张:27.75 字数:710.4 千字

版 次:2003 年 1 月第 1 版 2003 年 1 月第 1 次印刷

印 数:5000 册 定价:39.00 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

## 出版说明

“开发专家”是电子工业出版社计算机研发部长期以来精心培育的计算机科学技术类本版品牌。这个品牌是由多个专题系列组成的横向大系列，涵盖了计算机技术的各个方面，特别是一直受到极大关注的程序开发类系列，例如《开发专家之数据库》、《开发专家之网络编程》、《开发专家之 Delphi》、《开发专家之 Sun ONE》，以及《开发专家之 Oracle9i》等。这些专题系列基于各自的角度，从纵向上包含了该专题的所有内容。因此，整个“开发专家”的品牌架构纵横交错，囊括了所有的计算机技术和所有的技术层面，海纳百川而又极具可扩展性。

“开发专家”的作者队伍主要依托于“飞思科技产品研发中心”。“飞思科技产品研发中心”由专业的策划人员、权威的技术专家和资深的作者队伍共同组成。在图书的出版上，形成了以研发为基础、以出版为中心、以服务为支持的专业化出版框架和流程。通过深入的市场调查和技术跟踪，在综合了技术需求和读者焦点等因素的基础上，形成各系列丛书的写作重点和大纲，然后聘请业界的最前沿学者进行写作。同时，策划工作全程介入写作进程，严格控制写作质量，用最专业的技术背景、最深刻的理论基础、最具代表性的案例、最能为专业读者接受的形式，为读者提供品质最佳的图书产品，体现了出版者和著作者的完美结合。

多年来，计算机研发部始终把创造社会效益摆在首位，秉承一切为国内计算机技术专业读者服务的精神，为推动国内信息技术的发展、为体现国内技术的原创水平，穷尽所有的创意与努力，将出版者的命运与读者的支持紧紧地连在了一起。

在此，我们临出版之残酷竞争而不惧，旌旗猎猎而异军突起，这与广大读者的支持是分不开的。为使我们的脚步更坚实、使我们的队伍永远保持活力和创造力，我们期待着您能为我们的前进贡献出您的意见和建议。同时，我们也在等待着您的加入。

我们的联系方式：

电 话：(010) 68134545      68131648

电子邮件：support@fecit.com.cn

飞思在线：<http://www.fecit.com.cn>      <http://www.fecit.net>

答疑网址：<http://www.fecit.com.cn/question.htm>

通用网址：计算机图书、飞思、飞思教育、飞思科技、FECIT

电子工业出版社计算机研发部

# 前 言

## 关于本丛书

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,是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,衡量一个国家信息化的程度,数据库的建设规模、数据库信息量的大小和使用程度已经成为一个重要标志。

对于一个企业来说,如何在竞争日益激烈的市场环境中保持优势,很大程度上依赖于对信息的收集、处理和分析,而这些都依赖于数据库技术。作为全球最大的数据库产品提供商,Oracle 公司在推出 Oracle8i 产品以后,又将其 Oracle9i 数据库产品推向市场,Oracle9i 同 Oracle 的前期产品一样,将受到更加广泛的注意,成为计算机领域一颗耀眼的明星。

本套丛书立足于 Oracle9i 的系列产品,全面介绍了 Oracle9i 提供的强大功能特性。丛书分为 Oracle9i 基础、应用服务器、数据库高级管理、企业管理器、数据仓库构建技术和 J2EE 高级开发指南七个部分,内容涵盖了 Oracle9i 的核心产品。

《Oracle9i 基础与提高》一书介绍了 Oracle9i 数据库的基本概念、基础知识、基本操作,帮助广大初学者顺利掌握 Oracle9i 数据库产品提供的强大功能特性,为深入学习 Oracle9i 奠定基础。

《Oracle9i 应用服务器详解》一书从基础、管理和开发三个方面详细介绍了 Oracle9i 应用服务器为企业构建信息系统基础平台提供了强大支持,为构建典型企业应用提供了技术支撑。

《Oracle9i 数据库高级管理》为企业数据库管理员创建、维护高可用的大型企业数据库提供解决方案,帮助管理员在了解 Oracle9i 数据库体系的基础上,利用 Oracle9i 提供的新的、强大的管理特性执行管理操作。

《Oracle9i 企业管理器详解》分析了 Oracle 的集成管理平台,详细介绍了如何利用企业管理器执行日常管理操作。

《Oracle9i 数据仓库构建技术》一书分析了 Oracle 提供的数据库解决方案,以及如何利用 Oracle 推出的数据库产品构建高可用的数据仓库,如何维护数据仓库,以及如何挖掘数据仓库数据。

《Oracle9i J2EE 高级开发指南》一书详细介绍了 J2EE 企业应用开发和部署的体系结构,以及 Oracle 如何为 J2EE 提供全面而强大的支持。

本丛书内容丰富,涵盖了大型数据库应用开发中的关键技术,很多内容在国内首次面世,如《Oracle9i J2EE 高级开发指南》、《Oracle9i 应用服务器详解》,很值得数据库技术工作者参考和阅读。本丛书具有广泛的读者,可供 Oracle 系统管理员、应用程序员、系统分析人员和设计人员等广大用户学习和参考。

## 关于本书

20 世纪 80 年代中期,“数据仓库之父” William H.Inmon 先生在其《建立数据仓库》一书中定义了数据仓库的概念,随后又给出了更为精确的定义:数据仓库是在企业管理和决策中面向主题的、集成的、与时间相关的、不可修改的数据集合。与其他数据库应用不同的是,数据仓库更像一种过程,是对分布在企业内部各处的业务数据的整合、加工和分析的过程,而不是一种可以购买的产品。

从目前的形势看,数据仓库技术紧跟 Internet 技术发展,成为信息社会中获得企业竞争优势的又一个关键。据美国 Meta Group 市场调查机构的资料表明,《幸福》杂志所列的全球 2000 家大公司中已有 90% 将 Internet 网络和数据仓库这两项技术列入企业计划,而且有很多企业为使自己在竞争中处于优势已经率先采用数据仓库技术。

本书在 Oracle9i 系列丛书中,主要面向企业数据仓库系统的构建与维护。本书全面阐述了 Oracle9i 和 Oracle Warehouse Builder 独特的数据仓库化能力,并提供了数据仓库规划、构建、生产、维护和性能优化方面的知识和技能。本书主要内容包括数据仓库概念、Oracle Warehouse Builder 概述、OWB 安装与配置、使用 OWB 创建工程、数据仓库的逻辑设计、数据仓库物理设计、数据仓库的硬件和 I/O 考虑、装载和转换、实体化视图、数据仓库并行和分区、数据仓库中用于聚集的 SQL、在数据仓库中用于分析的 SQL、使用 XML 工具集、查询重写和 OLAP。附录部分还介绍了 OWB 的批服务和 OWB 公共视图表的知识。

本书由飞思科技产品研发中心策划并组织编写,余金山、龚洵、汤义、陶钧、柯锋涛、林慧、周敏龙、舒振、赵青松、蒋平、王丽艳、张耀宏、马满好、刘程、吴永波、任意广、何晓晔、刘亮、未明亮、陈晓等人参与了本书的编写工作,在此一并致谢!

由于作者水平有限,疏漏之处在所难免,敬请读者批评指正。

我们的联系方式:

电 话: (010) 68134545      68131648

电子邮件: support@fecit.com.cn

飞思在线: <http://www.fecit.com.cn>      <http://www.fecit.net>

答疑网址: <http://www.fecit.com.cn/question.htm>

通用网址: 计算机图书、飞思、飞思教育、飞思科技、FECIT

飞思科技产品研发中心

# 目 录

|                                         |  |
|-----------------------------------------|--|
| 第 1 章 数据仓库概述..... 1                     |  |
| 1.1 数据仓库概述..... 1                       |  |
| 1.1.1 从数据库到数据仓库..... 2                  |  |
| 1.1.2 数据的仓库化过程..... 4                   |  |
| 1.2 数据仓库的体系结构..... 5                    |  |
| 1.2.1 数据库体系化环境..... 5                   |  |
| 1.2.2 数据仓库体系结构..... 7                   |  |
| 1.3 数据仓库的组成部分..... 9                    |  |
| 1.3.1 数据集市..... 9                       |  |
| 1.3.2 关系数据库..... 9                      |  |
| 1.3.3 数据源..... 10                       |  |
| 1.3.4 数据准备区..... 10                     |  |
| 1.3.5 显示服务..... 11                      |  |
| 1.4 Oracle9i 数据仓库解决方案..... 12           |  |
| 第 2 章 Oracle Warehouse Builder          |  |
| 概述..... 13                              |  |
| 2.1 Oracle Warehouse Builder 特性..... 13 |  |
| 2.2 OWB 的生命周期管理..... 15                 |  |
| 2.3 OWB 开发阶段简介..... 17                  |  |
| 2.4 OWB 定义阶段简介..... 18                  |  |
| 2.5 OWB 生成阶段简介..... 19                  |  |
| 2.6 OWB 装载和管理阶段简介..... 19               |  |
| 第 3 章 OWB 安装与配置..... 21                 |  |
| 3.1 OWB 的安装要求..... 21                   |  |
| 3.2 安装 OWB..... 23                      |  |
| 3.3 配置 Oracle Warehouse Builder         |  |
| Repository..... 26                      |  |
| 3.4 配置 OWB Runtime..... 30              |  |
| 3.5 配置 Oracle Warehouse Builder         |  |
| Browser..... 32                         |  |
| 第 4 章 使用 OWB 创建工程..... 35               |  |
| 4.1 理解 OWB 工程..... 35                   |  |
| 4.2 创建 OWB 工程..... 39                   |  |
| 4.3 OWB 控制台..... 41                     |  |
| 4.3.1 操作环境..... 41                      |  |
| 4.3.2 工具按钮..... 44                      |  |
| 4.3.3 控制台工具栏..... 46                    |  |
| 4.4 命名策略..... 47                        |  |
| 4.5 使用 Warehouse Builder 向导..... 48     |  |
| 4.6 搜索导航树..... 49                       |  |
| 第 5 章 数据仓库的逻辑设计..... 51                 |  |
| 5.1 数据仓库设计方法概述..... 51                  |  |
| 5.2 创建数据仓库的基本步骤..... 54                 |  |
| 5.2.1 收集运营环境文档..... 55                  |  |
| 5.2.2 选择数据仓库的                           |  |
| 实现技术..... 56                            |  |
| 5.2.3 设计数据仓库模型..... 56                  |  |
| 5.2.4 创建数据准备区..... 59                   |  |
| 5.2.5 创建数据仓库数据库..... 59                 |  |
| 5.2.6 从可操作系统中                           |  |
| 析取数据..... 60                            |  |
| 5.2.7 清理和转换数据..... 60                   |  |
| 5.2.8 将数据装入数据仓库                         |  |
| 数据库..... 61                             |  |
| 5.2.9 准备显示信息..... 62                    |  |
| 5.2.10 将数据分发到数据集市..... 62               |  |
| 5.3 逻辑设计概述..... 62                      |  |
| 5.4 创建逻辑设计..... 63                      |  |
| 5.5 数据仓库框架..... 63                      |  |
| 5.6 数据仓库对象..... 64                      |  |
| 5.6.1 事实表..... 64                       |  |
| 5.6.2 维表..... 65                        |  |
| 5.6.3 层次结构..... 65                      |  |
| 5.6.4 级别..... 66                        |  |
| 5.6.5 惟一标识符..... 67                     |  |
| 5.6.6 关系..... 67                        |  |
| 5.6.7 数据仓库对象和它们之间的                      |  |
| 关系实例..... 67                            |  |
| 5.7 使用 OWB 创建数据仓库模块..... 68             |  |
| 5.7.1 维对象的定义规则..... 68                  |  |

|       |                                              |     |
|-------|----------------------------------------------|-----|
| 5.7.2 | 创建一个维 .....                                  | 69  |
| 5.7.3 | 使用 OWB 创建维 .....                             | 70  |
| 5.7.4 | 使用 Oracle Enterprise Manager<br>维向导创建维 ..... | 73  |
| 5.7.5 | 维的多个层次结构 .....                               | 77  |
| 5.7.6 | 使用规范维表 .....                                 | 78  |
| 5.8   | 查看维 .....                                    | 79  |
| 5.9   | 使用带有约束的维 .....                               | 79  |
| 5.9.1 | 修改维定义 .....                                  | 80  |
| 5.9.2 | 使用 New Time Dimension<br>Wizard .....        | 81  |
| 5.10  | 使用 OWB 创建事实表定义 .....                         | 83  |
| 第 6 章 | 数据仓库的物理设计 .....                              | 87  |
| 6.1   | 从逻辑设计到物理设计的转换 .....                          | 87  |
| 6.2   | 物理设计 .....                                   | 87  |
| 第 7 章 | 数据仓库中的硬件和 I/O<br>设计 .....                    | 91  |
| 7.1   | 数据仓库中的硬件和 I/O<br>设计概述 .....                  | 91  |
| 7.2   | RAID 配置 .....                                | 92  |
| 7.2.1 | RAID Level 0 .....                           | 93  |
| 7.2.2 | RAID Level 1 .....                           | 93  |
| 7.2.3 | RAID Level 5 .....                           | 93  |
| 第 8 章 | 数据装载和转换 .....                                | 95  |
| 8.1   | 数据仓库中的装载和转换概述 .....                          | 95  |
| 8.1.1 | 多阶段数据转换 .....                                | 95  |
| 8.1.2 | 管道数据转换 .....                                 | 96  |
| 8.2   | 装载机制 .....                                   | 96  |
| 8.2.1 | SQL*Loader .....                             | 97  |
| 8.2.2 | 外部表 .....                                    | 99  |
| 8.2.3 | OCI 和直接路径 API .....                          | 101 |
| 8.2.4 | 导入/导出 .....                                  | 101 |
| 8.3   | 转换机制 .....                                   | 101 |
| 8.3.1 | 使用 SQL 进行的转换 .....                           | 101 |
| 8.3.2 | 使用 PL/SQL 进行的转换 .....                        | 106 |
| 8.3.3 | 使用表函数进行的转换 .....                             | 106 |
| 8.4   | 装载和转换语法 .....                                | 113 |
| 8.4.1 | 并行装载语法 .....                                 | 113 |
| 8.4.2 | 主搜索语法 .....                                  | 116 |

|        |                                             |     |
|--------|---------------------------------------------|-----|
| 8.4.3  | 异常处理语法 .....                                | 117 |
| 8.4.4  | 透视语法 .....                                  | 118 |
| 8.5    | 添加转换 .....                                  | 120 |
| 第 9 章  | 使用 Oracle 透明网关获取<br>SQL Server 数据 .....     | 127 |
| 9.1    | 透明网关概述 .....                                | 127 |
| 9.1.1  | Oracle 异类服务 .....                           | 127 |
| 9.1.2  | Oracle 透明网关访问 Microsoft<br>SQL Server ..... | 127 |
| 9.1.3  | 网关的体系结构 .....                               | 128 |
| 9.1.4  | 网关的处理过程 .....                               | 129 |
| 9.2    | 配置网关 .....                                  | 130 |
| 9.2.1  | 网关的配置 .....                                 | 130 |
| 9.2.2  | 配置网关的 Net8 Listener .....                   | 131 |
| 9.2.3  | 配置网关访问的 Oracle<br>数据库服务器 .....              | 133 |
| 9.2.4  | 完成可选的配置任务 .....                             | 133 |
| 9.2.5  | 创建数据库连接 .....                               | 134 |
| 9.2.6  | 为多重 Microsoft SQL Server<br>数据库配置网关 .....   | 135 |
| 9.3    | 使用网关 .....                                  | 137 |
| 9.3.1  | 通过网关访问 Microsoft SQL<br>Server .....        | 137 |
| 9.3.2  | Oracle 数据库服务器 SQL 构造<br>处理 .....            | 138 |
| 9.3.3  | 数据类型转换 .....                                | 138 |
| 9.3.4  | 使用同义词 .....                                 | 139 |
| 9.3.5  | 使用 PASS-THROUGH<br>特征 .....                 | 139 |
| 9.3.6  | 执行存储的过程和函数 .....                            | 139 |
| 9.3.7  | 完成分布式的查询 .....                              | 140 |
| 9.3.8  | 复制数据到 Microsoft SQL<br>Server .....         | 141 |
| 9.3.9  | 复制数据到 Oracle 数据库 .....                      | 142 |
| 9.3.10 | 监测网关会话 .....                                | 142 |
| 9.4    | Oracle 透明网关实例 .....                         | 143 |
| 9.4.1  | 创建实例表 .....                                 | 143 |
| 9.4.2  | 简单查询实例 .....                                | 145 |
| 9.4.3  | 复杂查询实例 .....                                | 146 |



|        |                                    |     |
|--------|------------------------------------|-----|
| 9.4.4  | 连接 Microsoft SQL Server 表 .....    | 146 |
| 9.4.5  | 写数据的能力 .....                       | 147 |
| 9.4.6  | 数据字典查询 .....                       | 149 |
| 9.4.7  | PASS-THROUGH 特征 .....              | 150 |
| 9.4.8  | 执行存储过程 .....                       | 153 |
| 第 10 章 | 实体化视图 .....                        | 155 |
| 10.1   | 实体化视图概述 .....                      | 156 |
| 10.1.1 | 数据仓库环境下的实体化视图 .....                | 156 |
| 10.1.2 | 分布式计算的实体化视图 .....                  | 156 |
| 10.1.3 | 移动计算的实体化视图 .....                   | 156 |
| 10.1.4 | 实体化视图需求 .....                      | 156 |
| 10.1.5 | 总结管理的组件 .....                      | 157 |
| 10.1.6 | 总结设计建议 .....                       | 158 |
| 10.2   | 实体化视图类型 .....                      | 159 |
| 10.2.1 | 带有聚集的实体化视图 .....                   | 159 |
| 10.2.2 | 只包含连接的实体化视图 .....                  | 162 |
| 10.2.3 | 嵌套实体化视图 .....                      | 163 |
| 10.3   | 创建实体化视图 .....                      | 165 |
| 10.3.1 | 命名 .....                           | 170 |
| 10.3.2 | 存储特性 .....                         | 171 |
| 10.3.3 | 创建方法 .....                         | 171 |
| 10.3.4 | 启用查询重写 .....                       | 171 |
| 10.3.5 | 刷新选项 .....                         | 171 |
| 10.3.6 | ORDER BY 子句 .....                  | 172 |
| 10.3.7 | 实体化视图日志 .....                      | 172 |
| 10.3.8 | 使用 Oracle Enterprise Manager ..... | 172 |
| 10.4   | 使用 OWB 创建实体化视图定义 .....             | 172 |
| 10.5   | 注册已有的实体化视图 .....                   | 176 |
| 10.6   | 分区和实体化视图 .....                     | 176 |
| 10.6.1 | 跟踪分区修改情况 .....                     | 177 |
| 10.6.2 | 实体化视图分区 .....                      | 178 |
| 10.6.3 | 预建表分区 .....                        | 179 |
| 10.7   | 为实体化视图选择索引 .....                   | 179 |
| 10.8   | 修改实体化视图 .....                      | 180 |

|        |                             |     |
|--------|-----------------------------|-----|
| 10.9   | 使用 OWB 修改实体化视图 .....        | 180 |
| 10.10  | 分析实体化视图 .....               | 182 |
| 第 11 章 | 数据仓库的并行处理和分区 .....          | 185 |
| 11.1   | 并行处理概述 .....                | 185 |
| 11.2   | 并行粒度 .....                  | 186 |
| 11.3   | 分区设计 .....                  | 187 |
| 11.3.1 | 分区类型 .....                  | 187 |
| 11.3.2 | 分区修剪 .....                  | 192 |
| 11.3.3 | Partition-wise 连接 .....     | 193 |
| 第 12 章 | 数据仓库中用于聚集的 SQL .....        | 195 |
| 12.1   | 数据仓库中用于聚集的 SQL 概述 .....     | 195 |
| 12.1.1 | 多维数据分析 .....                | 195 |
| 12.1.2 | 优化性能 .....                  | 196 |
| 12.1.3 | 一个聚集语法 .....                | 197 |
| 12.2   | ROLLUP 对 GROUP BY 的扩展 ..... | 198 |
| 12.3   | CUBE 对 GROUP BY 的扩展 .....   | 199 |
| 12.4   | GROUPING 函数 .....           | 202 |
| 12.5   | GROUPING SETS 表达式 .....     | 205 |
| 12.6   | 合成列 .....                   | 207 |
| 12.7   | 连接分组 .....                  | 208 |
| 12.8   | 使用聚集所需考虑的事项 .....           | 211 |
| 12.9   | 使用 WITH 子句计算 .....          | 212 |
| 第 13 章 | 在数据仓库中用于分析的 SQL .....       | 215 |
| 13.1   | 在数据仓库中用于分析的 SQL 概述 .....    | 215 |
| 13.2   | 求秩函数 .....                  | 217 |
| 13.3   | 窗口聚集函数 .....                | 226 |
| 13.4   | 报表聚集函数 .....                | 229 |
| 13.5   | LAG/LEAD 函数 .....           | 232 |
| 13.6   | FIRST/LAST 函数 .....         | 232 |
| 13.7   | 线性回归函数 .....                | 234 |
| 13.8   | 反百分比函数 .....                | 235 |
| 13.9   | 假定分阶和分布函数 .....             | 238 |
| 13.10  | WIDTH_BUCKET 函数 .....       | 239 |

|        |                                   |     |
|--------|-----------------------------------|-----|
| 13.11  | CASE 表达式.....                     | 240 |
| 第 14 章 | 使用 XML 工具集.....                   | 243 |
| 14.1   | 从源提取数据.....                       | 243 |
| 14.2   | 将数据存储到目标中.....                    | 244 |
| 14.3   | 使用运行控制.....                       | 244 |
| 14.4   | 调用 XML 工具集.....                   | 244 |
| 14.5   | 典型控制文件.....                       | 245 |
| 14.6   | 存储在文件中的 XML 文档.....               | 245 |
| 14.7   | 作为其他对象的 XML 文档.....               | 248 |
| 14.8   | 为控制文件的文档类型定义.....                 | 250 |
| 第 15 章 | 查询重写.....                         | 253 |
| 15.1   | 查询重写概述.....                       | 253 |
| 15.2   | 使用查询重写.....                       | 256 |
| 15.2.1 | 为查询重写初始化参数.....                   | 257 |
| 15.2.2 | 控制查询重写.....                       | 257 |
| 15.2.3 | 使能查询重写所需<br>的权限.....              | 258 |
| 15.2.4 | 查询重写的准确性.....                     | 259 |
| 15.3   | Oracle 如何重写查询.....                | 259 |
| 15.3.1 | 文本匹配查询重写方法.....                   | 260 |
| 15.3.2 | 一般的查询重写方法.....                    | 261 |
| 15.3.3 | 何时需要约束和维.....                     | 261 |
| 15.3.4 | 视图约束.....                         | 262 |
| 15.3.5 | 表达式匹配.....                        | 263 |
| 15.3.6 | 日期分段.....                         | 264 |
| 15.3.7 | 选择兼容性.....                        | 265 |
| 15.3.8 | 连接兼容性检查.....                      | 269 |
| 15.4   | 查询重写特殊案例.....                     | 277 |
| 15.4.1 | 使用部分陈旧的实体化视图<br>的查询重写.....        | 277 |
| 15.4.2 | 使用复杂实体化视图<br>的查询重写.....           | 279 |
| 15.4.3 | 使用嵌套实体化视图<br>的查询重写.....           | 279 |
| 15.4.4 | 带有 CUBE、ROLLUP 和<br>分组集的查询重写..... | 280 |
| 15.5   | 验证查询重写.....                       | 284 |
| 第 16 章 | 联机分析处理.....                       | 289 |
| 16.1   | Oracle OLAP.....                  | 289 |

|         |                                       |     |
|---------|---------------------------------------|-----|
| 16.2    | 管理 Oracle OLAP.....                   | 291 |
| 16.2.1  | 管理服务环境.....                           | 291 |
| 16.2.2  | OLAP 服务示例管理器.....                     | 292 |
| 16.2.3  | 启动 OLAP 服务.....                       | 294 |
| 16.2.4  | 停止 OLAP 服务.....                       | 295 |
| 16.2.5  | 获得系统权限.....                           | 295 |
| 16.2.6  | 修改配置设置.....                           | 296 |
| 16.2.7  | 解锁数据库标识.....                          | 297 |
| 16.2.8  | 管理会话.....                             | 298 |
| 16.2.9  | 查看状态信息.....                           | 298 |
| 16.2.10 | 创建新的服务.....                           | 299 |
| 16.3    | 配置数据仓库.....                           | 299 |
| 16.3.1  | 设置管理区域.....                           | 300 |
| 16.3.2  | 编辑数据库资源.....                          | 301 |
| 16.4    | 数据仓库本地化处理.....                        | 302 |
| 16.5    | 性能调整.....                             | 302 |
| 16.6    | OLAP API 概述.....                      | 303 |
| 16.7    | 理解 OLAP API 元数据.....                  | 307 |
| 16.7.1  | OLAP API 元数据概述.....                   | 307 |
| 16.7.2  | Oracle 企业管理器中的<br>OLAP API 元数据对象..... | 308 |
| 16.7.3  | OLAP API 中的 MDM 元<br>数据对象.....        | 308 |
| 16.7.4  | MdmSource 类.....                      | 309 |
| 16.7.5  | MdmDimension 类.....                   | 309 |
| 16.7.6  | MdmLevel 类.....                       | 310 |
| 16.7.7  | MdmHierarchy 类.....                   | 311 |
| 16.7.8  | MdmListDimension 类.....               | 313 |
| 16.7.9  | MdmMeasure 类.....                     | 314 |
| 16.7.10 | MdmAttribute 类.....                   | 315 |
| 16.8    | 连接到数据库.....                           | 315 |
| 16.8.1  | 连接进程概述.....                           | 316 |
| 16.8.2  | OLAP API 中的连接类.....                   | 316 |
| 16.8.3  | 建立一个连接.....                           | 318 |
| 16.9    | 发现可用的元数据.....                         | 320 |
| 16.10   | 创建查询.....                             | 335 |
| 16.11   | 选择数据.....                             | 337 |
| 16.11.1 | 基于键选择元素.....                          | 337 |
| 16.11.2 | 基于元素值选择元素.....                        | 338 |

|         |                                               |     |
|---------|-----------------------------------------------|-----|
| 16.11.3 | 基于阶选择元素 .....                                 | 339 |
| 16.11.4 | 基于层次结构位置<br>选择元素 .....                        | 341 |
| 16.12   | 执行计算 .....                                    | 343 |
| 16.13   | 使用 TransactionProvider .....                  | 346 |
| 16.14   | 提取查询结果 .....                                  | 351 |
| 16.14.1 | 提取查询结果 .....                                  | 351 |
| 16.14.2 | 为不同的数据显示使用<br>CompoundCursor .....            | 355 |
| 16.14.3 | 定义游标的行为 .....                                 | 358 |
| 16.14.4 | 计算范围 .....                                    | 359 |
| 16.14.5 | 定义提取块大小 .....                                 | 361 |
| 16.15   | 创建动态查询 .....                                  | 364 |
| 第 17 章  | Oracle9i Discoverer .....                     | 373 |
| 17.1    | Discoverer 概述 .....                           | 373 |
| 17.1.1  | Oracle9i Discoverer 和 Oracle9i<br>应用服务器 ..... | 374 |
| 17.1.2  | 启动和关闭 Discoverer<br>服务 .....                  | 375 |
| 17.2    | Oracle Discoverer Plus 概述 .....               | 376 |
| 17.3    | 打开 Workbook .....                             | 376 |
| 17.3.1  | 打开 Workbook 的操作 .....                         | 376 |
| 17.3.2  | 使用 Page Items 转换<br>视角 .....                  | 378 |
| 17.4    | 格式化数据显示 .....                                 | 379 |
| 17.4.1  | 编辑数据字体 .....                                  | 379 |
| 17.4.2  | 插入一个变量到标题中 .....                              | 379 |
| 17.4.3  | 插入图像到标题中 .....                                | 380 |
| 17.5    | 演练数据 .....                                    | 381 |

|         |                                    |     |
|---------|------------------------------------|-----|
| 17.5.1  | 完成展开演练 .....                       | 381 |
| 17.5.2  | 完成替换演练 .....                       | 382 |
| 17.6    | 为压缩和分析旋转数据 .....                   | 384 |
| 17.7    | 图形化数据 .....                        | 386 |
| 17.7.1  | 创建一个图形 .....                       | 387 |
| 17.7.2  | 编辑用户图形的序列 .....                    | 389 |
| 17.8    | 陈述用户数据中的例外 .....                   | 390 |
| 17.9    | 演练相关数据 .....                       | 392 |
| 17.10   | 创建 Workbook .....                  | 393 |
| 17.10.1 | 改变默认格式 .....                       | 394 |
| 17.10.2 | 使用 Workbook 向导创建<br>Workbook ..... | 394 |
| 17.11   | 完成计算 .....                         | 396 |
| 17.11.1 | 用户数据的排序 .....                      | 397 |
| 17.11.2 | 计算总利润 .....                        | 398 |
| 17.11.3 | 计算百分比 .....                        | 400 |
| 附录 A    | 批服务 .....                          | 403 |
| A.1     | 批服务器接口 .....                       | 403 |
| A.2     | 批服务命令行的特性 .....                    | 405 |
| 附录 B    | OWB 公共视图表 .....                    | 413 |
| B.1     | 数据模型 .....                         | 413 |
| B.2     | 实现模型 .....                         | 420 |
| B.3     | 平面文件/记录模型 .....                    | 421 |
| B.4     | 压缩分析/数据线性模型 .....                  | 423 |
| B.5     | 分类模型 .....                         | 426 |
| B.6     | 表达式和函数模型 .....                     | 426 |
| B.7     | 转换/映射模型 .....                      | 428 |
| B.8     | 运行公共视图 .....                       | 430 |

# 第 1 章 数据仓库概述

数据仓库（Data Warehouse）概念的形成是以 Prism Solutions 公司副总裁 W.H.Inmon 在 1992 年出版的《建立数据仓库》（Building the Data Warehouse）为标志的。数据仓库的提出是以关系数据库、并行处理技术和分布式技术的飞速发展为基础的，它是解决信息技术（IT）在发展过程中虽然拥有大量数据却缺乏有用信息（Data rich-Information poor）的综合方案。

从目前的形势看，数据仓库技术紧跟 Internet 技术发展，成为信息社会中获得企业竞争优势的又一关键。据美国 Meta Group 市场调查机构的资料表明，《幸福》杂志所列的全球 2000 家大公司中已有 90% 将 Internet 网络和数据仓库这两项技术列入企业计划，而且有很多企业为使自己在竞争中处于优势，已经率先采用数据仓库技术。

本章我们主要介绍数据仓库的基本概念、Microsoft 数据仓库的框架、数据传输层（OLE DB），如何设计数据仓库、数据仓库的数据库，如何向数据仓库中添加数据，如何分析和呈现数据仓库中的数据，最后介绍如何管理数据仓库。

本章主要内容：

- 数据仓库概述
- 数据仓库的体系结构
- 数据仓库的组成部分

## 1.1 数据仓库概述

随着计算机技术的飞速发展和企业界不断提出新的需要，数据仓库技术应运而生。传统数据库技术是以单一的数据资源，即以数据库为中心，进行从事务处理、批处理到决策分析等各种类型的数据处理工作。然而，不同类型的数据处理有着不同的处理特点，单一的数据组织方式进行组织的数据库并不能反映这种差异，满足不了数据处理多样化的要求。近年来，随着计算机的应用，特别是数据库技术的广泛应用，人们对数据处理的这种多层次特点有了更清晰的认识。

总结起来，当前的数据处理可以大致地划分为两大类：操作型处理和分析型处理。操作型处理也叫事务处理，是指对数据库联机的日常操作，通常是对一个或一组记录的查询和修改，主要是为企业特定应用服务的。人们关心的是响应时间，数据的安全性和完整性。分析型处理（又叫信息型处理）则用于管理人员的决策分析。例如，DSS、EIS 和多维分析等，经常要访问大量的历史数据。两者之间的巨大差异使得操作型处理和分析型处理的分离成为必然。这种分离划清了数据处理的分析型环境与操作型环境之间的界限，从而由原来的以单一数据库为中心的数据环境发展为一种新环境——体系化环境。

### 1.1.1 从数据库到数据仓库

数据库系统作为数据管理手段，主要用于事务处理。在这些数据库中已经保存了大量的日常业务数据。传统的 DSS 一般是直接建立在这种事务处理环境上的。数据库技术一直力图使自己能胜任从事务处理、批处理到分析处理的各种类型信息的处理任务。尽管数据库在事务处理方面获得了巨大成功，但它对分析处理的支持一直不能令人满意，尤其是当以业务处理为主的联机事务处理（On-line Transaction Processing, OLTP）应用与以分析处理为主的 DSS 应用共存于同一个数据库系统中时，这两种类型的处理发生了明显的冲突。人们逐渐认识到，事务处理和分析处理具有本质的不同，直接使用事务处理环境来支持 DSS 是行不通的。

具体来说，事务处理环境不适宜 DSS 应用的原因主要有以下五条。

#### 1. 用户的行为模式

在事务处理环境中，用户的行为特点是数据的存取操作频率高，但是每次操作处理的时间短。因此，系统可以允许多个用户按分时方式使用系统资源，同时保持较短的响应时间，OLAP 是这种环境下的典型应用。

在分析处理环境中，用户的行为模式与此完全不同，某个 DSS 应用程序可能需要连续运行几个小时，从而消耗大量的系统资源。

将具有如此不同处理性能的两种应用方式放在同一个环境中运行，这显然是不适当的。

#### 2. 数据的集成问题

DSS 需要集成的数据。全面且正确的数据是有效分析和决策的首要前提，相关数据收集得越完整，得到的结果也就越可靠。因此，DSS 不仅需要整个企业内部各部门的相关数据，还需要企业外部、竞争对手等处的相关数据。

事务处理的目的在于使业务处理自动化，一般只需要与本部门业务有关的当前数据。而对整个企业范围内的集成应用考虑得很少，当前绝大部分企业内数据的真正状况是以分散方式而非集成方式存在的。造成这种分散的原因有多种，主要有事务处理应用问题、“蜘蛛网”问题、数据不一致问题、外部数据和非结构化数据问题。

上述问题是事务处理环境所固有的，尽管每个单独的事务处理应用可能是高效的，能产生丰富的细节数据，但这些数据却不能成为一个统一的整体。对于需要集成数据的 DSS 应用来说，必须在应用程序中对这些纷杂的数据进行集成。可是，数据集成是一项十分繁杂的工作，都交给应用程序完成会大大增加程序员的负担。并且每做一次分析，都要进行一次这样的集成，将会导致极低的处理效率。DSS 对数据集成的迫切需要是数据仓库技术出现的重要原因之一。

##### （1）事务处理应用的分散

当前企业内部各事务处理之间几乎是独立的，之所以出现这种现象有多种原因。有的原因是设计方面的，例如，系统设计人员为了减少系统开发费用和加快开发进度，总是采用简单而“有效”的设计方案，这种“有效”仅指对解决当前面临的问题有效，而不能保

证对以后新出现的问题继续有效。有的原因是经济方面的，当经费有限时，企业总是考虑先对关键的业务活动建立应用系统，然后再逐步建立其他业务的信息处理系统。还有的原因是历史、地理方面的，例如，某个大公司由分散在各地的多个子公司组成，企业的兼并等。

由于这种事务处理应用分散状况的存在，DSS 应用需要对分散在多个事务处理应用中的相关数据进行集成，为分析人员提供统一的数据视图。

### (2) “蜘蛛网”问题

在 DSS 应用中为了避免与其他用户的冲突，以及简化用户的数据视图，一种称为“抽取程序”的方法目前被广泛地应用，用户利用抽取程序从文件或数据库中查找有用的数据，然后这些数据被提取出来放入其他文件或数据库中供用户使用。这些经抽取得到的新文件或数据库又被某些用户再进行抽取，这种不加控制的连续抽取最终导致系统内的数据间形成了错综复杂的网状结构，人们形象地称为“蜘蛛网”。企业的规模越大，“蜘蛛网”问题就越复杂。

虽然网上的两个节点的数据可能归根结底是从原始库中抽取出来的同一节点数据，但其数据没有统一的时间基准，抽取算法各不相同，抽取级别也不相同，并且可能参考不同的外部数据。因而对同一问题的分析，不同节点会产生不同甚至截然相反的结果。这当然使决策者无从下手。

### (3) 数据不一致问题

前述的应用分散和“蜘蛛网”等问题导致了数据的不一致，这些数据不一致的形式是多种多样的。

- 同一字段在不同应用中具有不同的数据类型。例如，字段 Sex 在 A 应用中的值为“M/F”，在 B 应用中的值为“0/1”，在 C 应用中又为：“Male/female”。
- 同一字段在不同应用中具有不同的名字。例如，在 A 应用中的名称为“balance”，在 B 应用中名称为“bal”，在 C 应用中又变成了“currabal”。
- 同名字段，不同含义。例如，字段 weights 在 A 应用中表示人的体重，在 B 应用中表示汽车的重量等。

为了将这些不一致的数据集成起来，必须对它们进行转换后才进行分析。数据的不一致是多种多样的，对每种情况都必须专门处理，因此，这是一项很繁重的工作。

### (4) 外部数据和非结构化数据

在决策中经常用到外部数据，这部分数据不是由事务处理系统产生的，而是来自于其他外部数据源。例如，权威性刊物发布的统计数据、业界的技术报告、市场比较和分析报告、股票行情等，这些数据通常都是非结构化数据。在事务处理系统中，由于没有对外部数据进行统一管理，DSS 应用必须自行集成。

## 3. 数据动态集成问题

由于每次分析都进行数据集成以至于开销太大，所以一些应用仅在开始对所需数据进行了集成，以后就一直以这部分集成的数据作为分析的基础，不再与数据源发生联系，我们称这种方式的集成为静态集成。静态集成的最大缺点在于，如果在数据集成后数据源中的数据发生了改变，这些变化将不能反映给决策者，导致决策者使用的是过时的数据。对

于决策者来说，虽然并不要求随时准确地探知系统内的任何数据变化，但也不希望他所分析的是几个月以前的情况。因此，集成数据必须以一定的周期（例如 24 小时）进行刷新，我们称其为动态集成。显然，事务处理系统不具备动态集成的能力。

#### 4. 历史数据问题

事务处理一般只需要当前数据，在数据库中一般也只存储短期数据，且不同数据的保存期限也不一样，即使有一些历史数据保存下来了，也被束之高阁，未得到充分利用。但对于决策分析而言，历史数据是相当重要的，许多分析方法必须以大量的历史数据为依托。没有对历史数据的详细分析，是难以把握企业的发展趋势的。

通过上面所述可见，DSS 对数据在空间和时间的广度上都有了更高的要求，而事务处理环境难以满足这些要求。

#### 5. 数据的综合问题

在事务处理系统中积累了大量的细节数据，一般而言，DSS 并不对这些细节数据进行分析。这主要有两个原因，一是细节数据数量太大，会严重影响分析的效率；二是人多数细节数据不利于分析人员将注意力集中到有价值的信息上。因此，在分析前，往往需要对细节数据进行不同程度的综合。而事务处理系统不具备这种综合能力，根据规范化理论，这种综合往往因为产生某种数据冗余而被加以限制。

以上这些问题表明，在事务型环境中直接构建分析型应用是一种失败的尝试。数据仓库本质上是对这些存在问题的回答。但是建立数据仓库的主要驱动力并不是过去的缺点，而是市场商业经营行为的改变，市场竞争要求捕获和分析事务级的业务数据。建立在事务处理环境上的分析系统无法达到这一要求。要提高分析和决策的效率，分析型处理及其数据必须与操作型处理及其数据相分离。必须把分析型数据从事务处理环境中提取出来，按照 DSS 处理的需要进行重新组织，建立单独的分析处理环境。数据仓库正是为了构建这种新的分析处理环境而出现的一种数据存储和组织技术。

### 1.1.2 数据的仓库化过程

从信息的角度来看，数据仓库的目标就是在一个企业内将正确的信息发送给正确的人，这是一个动态的过程，并不是一时的解决方案，需要使用建立面向事务系统的不同开发方法。

数据仓库是一个数据集合，支持面向主题的、综合的、耗时的、非易失的管理决策。数据仓库集中于概念，而不是过程，包含了从多个处理系统中收集的关于一个概念的信息数据，搜集这些信息是相对比较稳定的，每隔一段时期进行一次。

开发数据仓库的第一步就是确定应当包含的主题，这需要同端用户、分析人员和执行人员交谈，了解信息需求的范围。在逻辑数据仓库转换成物理数据仓库之前，主题必须充分理解，因为建立数据仓库的目的就是面向主题的。

进行完物理设计以后，第二步就是事务性系统转移到数据仓库，由于事务数据库和数据仓库中数据的表示方式是不同的，转移数据到数据仓库中需要对数据进行的处理过程有：归纳、转换、解码、清除不合法的数据，分离、转换和移动源数据，满足数据仓库的

商业要求。这些过程需要自动执行，才能够满足数据仓库的并行要求。

在操作性系统中，数据是当前值，而且是准确的。例如，订单应用总是显示每个产品的当前情况，这些值可能因为查询的时间、地点等不同，显示的结果就不同。在数据仓库中，数据表示一个长期的过程所收集的信息，包含了商业主题的一个长期的快照。

第三步通过浏览、分析和报告，数据仓库中的信息可以呈现给用户。目前，已经有很多分析工具，从简单的书写报告到高级的数据开采。分析过程可以驱动数据仓库处理的交互，调整数据仓库的设计，提高仓库的数据存储能力，改善系统性能，支持更多的分析类型，如图 1-1 所示。

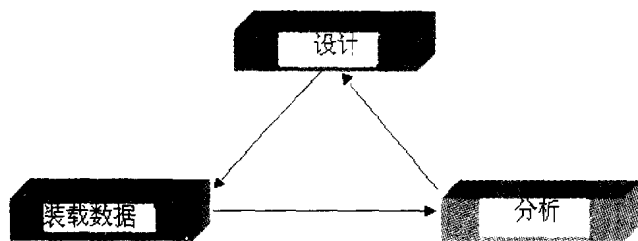


图 1-1 数据仓库的交互过程

## 1.2 数据仓库的体系结构

### 1.2.1 数据库体系化环境

数据库体系化环境是在一个企业或组织内，由面向应用的联机事务处理（On-Line Transaction Processing, OLTP）数据库及面向主题的数据仓库所组成的完整的数据环境，在这个数据环境内建立和进行一个企业或组织的从联机事务处理到企业管理决策的所有应用。这个数据环境分为两个部分，即操作型环境和分析型环境，分别为操作型处理和分析型处理这两类不同的数据处理服务。

数据库体系化环境的构成是广泛的，要建设一个企业的数据库体系化环境，不仅要求建立起各级数据库和数据仓库，还要求对面向应用的数据库之间、各级数据仓库之间及数据库与数据仓库之间的界限和联系做出合理划分和明确描述；对在不同的数据库或数据仓库上的数据处理和应用要进行明确的定义和划分；对软、硬件资源及其人员的配置做出明确规定。这些工作使体系化环境真正成为一个结构清晰、层次分明、联系明确、可有序运行的有机整体。

即使一个企业建立一个大而全的数据仓库，也有可能取得不是十分满意的效果。虽然数据仓库的数据组织是面向主题的，并为分析的需要保存了许多综合数据，但分析需求千变万化。我们不可能要求一个单一层次的数据仓库能完全符合各种各样的分析需求，数据仓库的主题需要不断完善、不断调整，综合数据的内容和形式也要灵活多变。另外，随着数据不断载入，数据仓库将越来越庞大，分析工作若完全基于单一层次的数据仓库，性能



将十分低下，因而需要建立起分层的数据仓库。如图 1-2 所示，根据管理层次的不同需要，在全局级数据仓库的基础上又导出了部门级和个人级数据仓库，全局级、部门级及个人级数据仓库三者浑然一体，适应了不同层次分析的要求，并与原有的操作型环境形成了一个多层次的体系化环境。

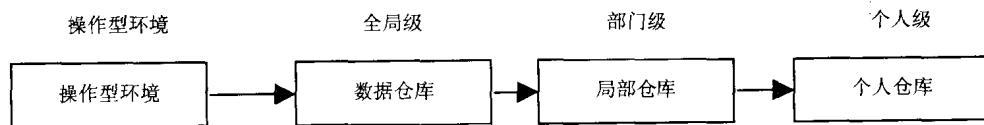


图 1-2 数据仓库在体系化环境中的位置

图 1-2 所示是对体系化环境的一个简单描述，它分为四个层次：操作型环境、全局级数据仓库、部门级的局部数据仓库和个人级的数据仓库。

在这个体系化环境中，操作型环境存放的是一些细节的操作型数据，服务于高性能事务处理。全局级数据仓库中除了存放细节数据外，还包含大量导出数据。部门级局部仓库中一般仅包含导出数据，而个人级数据仓库的数据都是暂时的，用于启发式分析。

数据从操作型环境经过综合整理进入全局数据仓库，企业中的有关部门再从全局数据仓库中组织适合自己分析需求的数据，建立自己的局部仓库。而个人不仅可以从全局数据仓库中提取数据，并且可以从部门级局部仓库中提取所需数据。这样，由于数据在全局数据仓库中都是集成的、一致的，所以部门和个人的抽取工作效率将会很高，并且不会出现本文开头提到的“蜘蛛网”现象。

四层体系化环境可以良好地与企业实际的部门组织结构对应起来。在实际环境中，一个企业的部门组织大都采用树状的结构，即层层负责、逐层管辖的分层结构。管理工作可分为高层、中层及基层三种。基层管理也称为操作管理，其主要任务是一般日常业务处理。操作型环境就是面向这一层次的，如进行联机事务处理。中层管理既包含一般业务处理，又需进行简单分析，做出一般的决策和控制。部门级的局部数据仓库面向这一层。高层管理的主要任务是进行战略决策，需要进行复杂的分析加工。个人级数据仓库可以说是面向这一层次的。

体系化环境的层次反映了数据与应用逻辑的抽象程度，如在操作型环境中的数据与事务处理逻辑常常是捆绑在一起的；在全局数据仓库中，数据组织的应用逻辑独立性最强，也就是说最不依赖于具体应用，其抽象程度最高；在部门级数据仓库和个人级数据仓库中，数据又与分析处理的逻辑联系在一起。我们不能简单地认为，在数据仓库中，数据最好是完全独立于任何应用逻辑的。比如在部门级数据仓库中，将数据组织与分析处理逻辑结合起来，将有利于提高分析效率。

数据处理的要求是多层次的，它理论上要求体系化环境应该有比较丰富的层次。层次越多，体系化环境也就越丰满，也就越能细致地反映企业中整个数据组织与处理。但是，体系化环境的层次越多，系统建设以及维护代价也就越大，而且层次间的界限不太分明，不利于建立结构清晰、层次分明、联系明确的体系化环境。