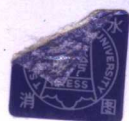
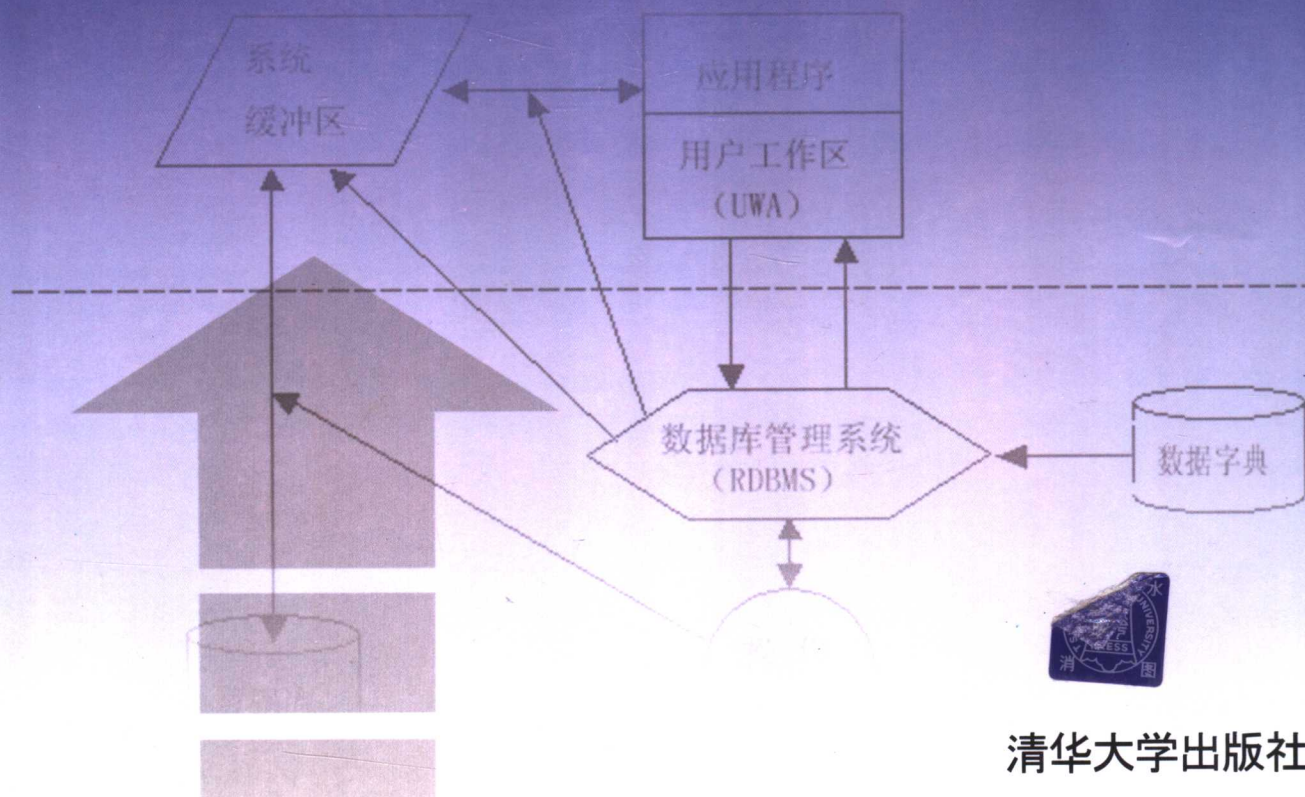


高等院校计算机教材

数据库原理及应用 (Oracle) 实用教程

刘甫迎 王道学 党晋蓉 编著



清华大学出版社

73.967
857

数据库原理及应用 (Oracle) 实用教程

刘甫迎 王道学 党晋蓉 编著

清华大学出版社

北 京

内 容 简 介

本书从实用的角度对数据库原理、创建和管理数据库、数据库管理员的岗位工作进行了详细的阐述。

本书有 12 章和两个附录, 详细介绍了数据库的意义, 数据库的由来和发展, 数据模型, 实体联系模型 (E-R), 关系数据库, 函数依赖、范式, 数据库的设计与维护, 数据库的安全性、完整性、并发控制与恢复, 分布式数据库等基本概念、原理和理论。本书打破了其他数据库原理的书忌讳写某一具体的数据库的惯例, 并叙述了后端大型数据库管理系统的工业标准——Oracle, 且将之作为上述基本理论的具体例子贯穿全书。特别是书中还著述了 Oracle 的 SQL、PL/SQL 和 Oracle 的开发工具 Oracle Developer (包含 Forms Builder、Reports Builder 等)、数据库管理工具 Enterprise Manager 以及 JDeveloper 和 Oracle9i J2EE 高级技术等内容, 所以本书又是一本学习 Oracle 的基于 C/S 模式和 B/S 编程以及提高 DBA (数据库管理员) 能力的教材。

本书有实例、实验指导书、习题、教学大纲, 便于学习与教学, 可作为高等院校及软件学院的教材, 也适合作为从事数据库软件开发和应用的人员的参考用书。

版权所有, 翻印必究。举报电话: 010-62782989 13901104297 13801310933

本书封面贴有清华大学出版社激光防伪标签, 无标签者不得销售。

本书防伪标签采用清华大学核研院专有核径迹膜防伪技术, 用户可通过在图案表面涂抹清水, 图案消失, 水干后图案复现; 或将表面膜揭下, 放在白纸上用彩笔涂抹, 图案在白纸上再现的方法识别真伪。

图书在版编目 (CIP) 数据

数据库原理及应用 (Oracle) 实用教程/刘甫迎, 王道学, 党晋蓉编著. —北京: 清华大学出版社, 2004.10
ISBN 7-302-09733-X

I. 数… II. ①刘… ②王… ③党… III. 关系数据库-数据库管理系统, Oracle-高等学校-教材
IV. TP311.138

中国版本图书馆 CIP 数据核字 (2004) 第 105434 号

出 版 者: 清华大学出版社 地 址: 北京清华大学学研大厦

<http://www.tup.com.cn> 邮 编: 100084

社 总 机: 010-62770175 客 户 服 务: 010-62776969

责任编辑: 刘利民

文稿编辑: 钟志芳

封面设计: 姜凌娜

版式设计: 杨 洋

印 刷 者: 北京密云胶印厂

装 订 者: 北京鑫海金澳胶印有限公司

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印 张: 22.5 字 数: 520 千字

版 次: 2004 年 10 月第 1 版 2004 年 10 月第 1 次印刷

书 号: ISBN 7-302-09733-X/TP·6730

印 数: 1~5000

定 价: 29.00 元

本书如存在文字不清、漏印以及缺页、倒页、脱页等印装质量问题, 请与清华大学出版社出版部联系调换。联系电话: (010)62770175-3103 或 (010)62795704

前 言

早在 20 世纪 60 年代, 数据库作为现代信息系统基础的一门软件学科便应运而生了。现在, 数据库技术已成为计算机领域中最重要技术之一, 它是软件学科中一个独立的分支。它的出现使得计算机应用渗透到工农业生产、商业、行政、教育、科学研究、工程技术和国防军事的各个部门。管理信息系统 (MIS)、办公自动化系统 (OA)、决策支持系统等都是使用了数据库管理系统或数据库技术的计算机应用系统。越来越多的人希望学习和了解数据库。

Oracle 是世界上第一个以 SQL 语言为基础的关系数据库。从 1979 年问世, 到 1997 年对象关系数据库 Oracle8 的发表, 再到现在的 Oracle10g, 美国的 Oracle 公司一直致力于信息管理现代化技术及产品的研究与开发, 使 Oracle 在数据库技术领域创造了无数的第一, 它的销量和普及程度名列世界前茅。Oracle 实际上成了大型数据库管理系统的工业标准。

本书有 12 章和两个附录, 详细介绍了数据库的意义, 数据库的由来和发展, 数据模型, 实体联系模型 (E-R), 关系数据库, 函数依赖、范式, 数据库的设计与维护, 数据库的安全性、完整性、并发控制与恢复, 分布式数据库等数据库的基本概念、原理和理论。本书打破了其他数据库原理的书忌讳写某一具体的数据库的惯例, 并叙述了后端大型数据库管理系统的工业标准——Oracle, 且将之作为上述基本理论的具体例子贯穿全书。特别是书中还著述了 Oracle 的 SQL、PL/SQL 和 Oracle 的开发工具 Oracle Developer (包含 Forms Builder、Reports Builder 等)、数据库管理工具 Enterprise Manager 以及 JDeveloper 和 Oracle9i J2EE 高级技术等内容, 所以本书又是一本学习 Oracle 的基于 C/S 模式和 B/S 编程以及提高 DBA (数据库管理员) 能力的教材。

本书的特点是: 其一, 内容新, 体现了最新版对象关系数据库 Oracle9i 版本以上的新技术; 其二, 本书作者长期从事 Oracle 数据库教材的编著和教学, 有丰富的教学经验, 并将其教学经验融入了此书; 其三, 在选材上“以必需够用为度”, 减轻学生负担 (例如, 去掉了传统数据库原理书中网状、层次数据库等现在用得较少的内容等); 其四, 突出实践动手能力和实用性 (有实例、案例、实验指导书、习题、教学大纲, 便于学习与教学), 力图使学生学习本书后便基本上可以编制和管理 Oracle 的程序和系统。**本书的电子教案可在 <http://computer.cec.deu.cn> 处的“课件发布”处下载。**

本书可作为高等院校及软件学院的教材, 也可作为从事数据库软件开发和应用人员的参考用书。

刘甫迎教授编著第 2 章、第 4 章、第 5 章、第 9 章、第 12 章和附录; 王道学副教授编著第 8 章、第 10 章、第 11 章; 党晋蓉副教授编著第 1 章、第 3 章、第 6 章、第 7 章; 全书由刘甫迎教授统稿, 杨绍华研究员主审本书并提供了关系代数除法的例子。在本书的编

著和出版的过程中，我们的学生提出了好的意见并做了一些工作，清华大学出版社的编辑也给予了我们很大的帮助，在此一并表示感谢！

由于编者水平有限，疏漏之处难免，敬请读者批评指正。

刘甫迎

2004年5月

目 录

第 1 章 数据库的概念	1
1.1 数据库的意义	1
1.1.1 信息处理及数据处理	1
1.1.2 数据模型和数据库技术	2
1.2 数据库的由来和发展	3
1.2.1 人工管理阶段 (20 世纪 50 年代中期以前)	3
1.2.2 文件系统阶段 (20 世纪 50 年代后期至 60 年代中后期)	4
1.2.3 数据库阶段 (20 世纪 60 年代末开始)	5
1.2.4 高级数据库阶段 (20 世纪 70 年代后期开始)	7
1.3 数据库的体系结构	11
1.4 数据库系统	16
1.5 数据库管理系统	19
习题	22
第 2 章 实体联系模型 (E-R Model)	23
2.1 实体和实体集合	23
2.2 联系和联系集合	24
2.3 属性、映射限制和关键字	25
2.3.1 属性	25
2.3.2 映射限制	26
2.3.3 关键字	28
2.4 实体联系 E-R 图解和将之归纳为表	29
2.4.1 实体联系图	29
2.4.2 把 E-R 图转变为表	31
2.5 概括和聚集	34
2.5.1 概括	34
2.5.2 聚集	36
2.6 E-R 数据库模式设计	37
2.6.1 映射基本集	37
2.6.2 实体集和联系集的使用	38
2.6.3 扩展 E-R 特征的使用	38

习题	39
第 3 章 关系模型及 Oracle 对象关系数据库	40
3.1 关系模型和基本概念	40
3.1.1 关系的定义	40
3.1.2 关系模型	41
3.1.3 关系数据语言概述	43
3.2 关系代数	43
3.2.1 传统的集合运算	43
3.2.2 专门的关系运算	44
3.3 关系演算	49
3.3.1 元组关系演算	49
3.3.2 域关系演算	51
3.4 关系数据库标准语言——SQL	52
3.5 Oracle 对象关系数据库系统	55
3.5.1 Oracle 系统的特点、产品结构及组成	56
3.5.2 Oracle 的体系结构 (实例的进程结构和内存结构)	58
3.5.3 Oracle 的配置方案	66
3.5.4 Oracle 运行过程	70
习题	70
第 4 章 Oracle 的 SQL 和 PL/SQL	72
4.1 Oracle 的 SQL 介绍	72
4.2 定义、修改、删除表	74
4.2.1 创建表 (CREATE TABLE)	74
4.2.2 修改表结构	81
4.2.3 索引的定义	83
4.2.4 删除表、索引	84
4.3 模式对象、直接量、函数和表达式	85
4.3.1 模式对象	85
4.3.2 对象及成分的命名	85
4.3.3 直接量	87
4.3.4 数据类型	88
4.3.5 空值	88
4.3.6 伪列	88
4.3.7 注释	89
4.3.8 算符	89

4.3.9 函数	92
4.3.10 表达式与条件	94
4.4 数据操纵语言 (Insert、Update 和 Delete)	95
4.4.1 将新行插入 (INSERT) 表	95
4.4.2 修改 (UPDATE) 表的行	96
4.4.3 从表删除 (DELETE) 行	97
4.5 视图 (View)	97
4.5.1 视图定义	97
4.5.2 视图的查询	98
4.5.3 视图修改	99
4.6 Oracle 的 PL/SQL	100
4.6.1 什么是 PL/SQL	100
4.6.2 PL/SQL 的结构	101
4.6.3 控制结构	109
4.6.4 游标 (CURSOR)	113
习题	117
第 5 章 关系数据库设计理论	121
5.1 引言	121
5.2 函数依赖	122
5.3 范式	124
5.3.1 第一范式 (1NF)	125
5.3.2 第二范式 (2NF)	125
5.3.3 第三范式 (3NF)	126
5.3.4 BCNF	126
5.4 多值依赖和 4NF	128
习题	130
第 6 章 数据库的设计与维护	132
6.1 概述	132
6.2 需求分析	134
6.2.1 需求分析的基本步骤	134
6.2.2 数据字典	136
6.3 概念结构设计	136
6.3.1 概念模型	137
6.3.2 概念设计的主要步骤	137
6.4 逻辑结构设计	142

6.4.1 逻辑设计环境	142
6.4.2 关系数据库的逻辑设计	143
6.5 物理设计	145
6.5.1 物理设计的步骤	145
6.5.2 物理设计环境	146
6.5.3 物理设计性能的测量	147
6.6 实现与维护	148
6.6.1 数据库的实现	148
6.6.2 其他设计工作	149
6.6.3 运行与维护	150
6.7 用 Oracle 建立数据库	151
6.7.1 数据库结构和空间管理 (数据文件、日志文件、表空间、段、模式和模式对象)	151
6.7.2 用 Oracle 建立和修改数据库	159
6.7.3 初始化参数文件	169
6.8 数据字典	171
6.8.1 数据字典的结构	172
6.8.2 数据字典的使用	172
习题	173
第 7 章 数据库保护	174
7.1 数据库的安全性 (用户鉴别、特权、角色、审计)	174
7.1.1 数据库的存取控制和用户的建立	175
7.1.2 特权和角色	179
7.1.3 审计	185
7.2 数据完整性 (数据库触发器)	186
7.2.1 完整性约束	186
7.2.2 数据库触发器	187
7.3 并发控制	190
7.3.1 数据库不一致的类型	191
7.3.2 封锁	191
7.3.3 Oracle 多种一致性模型	192
7.3.4 封锁机制	192
7.3.5 手工的数据封锁	194
7.4 数据库后备和恢复	195
7.4.1 数据库恢复所使用的结构	195
7.4.2 在线日志	196

7.4.3 归档日志	197
7.4.4 数据库后备	198
7.4.5 数据库恢复	199
习题	200
第 8 章 Oracle 的企业管理器 (Enterprise Manager)	202
8.1 企业管理器的三层框架结构	202
8.1.1 客户层	203
8.1.2 管理服务层	204
8.1.3 节点层	204
8.1.4 独立启动 (standalone)	205
8.2 企业管理器的配置	206
8.2.1 进入 OMS (Oracle Management Server)	206
8.2.2 OEM、OMS 功能简介	210
8.3 Oracle DBA (数据库管理员) 的工作	212
8.3.1 使用 OEM 管理数据库	212
8.3.2 用户管理	224
8.3.3 管理表和表空间	226
8.3.4 管理视图、索引和触发器	233
8.3.5 管理其他数据库对象	238
8.4 客户端配置	239
8.4.1 配置连接数据库	239
8.4.2 客户端测试	244
习题	244
第 9 章 分布式数据库系统	246
9.1 概述 (客户/服务器结构、服务器_服务器结构)	246
9.1.1 分布式数据库的特征	247
9.1.2 全功能分布式数据库的规则和目标	248
9.2 分布式数据库的连接	250
9.2.1 分布式数据库全局名和数据库链	251
9.2.2 连接客户 (Clients) 和服务器 (Servers)	253
9.2.3 连接服务器 (Servers) 到其他服务器 (Servers)	254
9.3 分布式查询处理及其他	255
9.3.1 建立分布式查询	255
9.3.2 在分布式事务 (Distributed Transaction) 中的其他语句	256
9.4 事务管理	256

9.4.1 事务	257
9.4.2 Oracle 的事务管理	257
9.5 多协议信息交换	259
9.5.1 SQL*Net	259
9.5.2 SQL*Net 连接组成成分	261
9.5.3 Multi Protocol Interchange (多协议交换)	262
9.5.4 配置 SQL*Net V2 客户机 (Client)	264
9.5.5 配置 listener	265
9.5.6 配置 SQL*Net V2 服务器 (Server)	266
9.5.7 启动 listener	267
9.6 表快照与复制	268
9.6.1 人工复制表	268
9.6.2 用触发器 (TRIGGERS) 复制表	269
9.6.3 自动复制和更新表	269
习题	269
第 10 章 数据库应用程序开发工具——Oracle Developer	271
10.1 Forms Builder	271
10.1.1 Form 应用基础	271
10.1.2 Form 应用设计	277
10.1.3 创建主从型 Form	280
10.1.4 项属性	282
10.1.5 创建 Form 应用程序	284
10.1.6 触发器	288
10.2 Reports Builder	296
10.2.1 Reports Builder 基础	296
10.2.2 创建一个简单的报表	297
10.2.3 创建图形 (Graphics)	302
10.2.4 在主界面中调用报表	305
习题	306
第 11 章 Oracle J2EE 高级技术与 JDeveloper	307
11.1 Java2 企业版 (J2EE)	307
11.1.1 J2EE 体系结构	307
11.1.2 J2EE 应用程序构成	308
11.2 Oracle9i 的 J2EE 技术与 OC4J	310
11.3 Oracle9i Java 商业组件 (BC4J)	314

11.3.1 BC4J	314
11.3.2 BC4J 应用结构	315
11.3.3 BC4J 的组件	316
11.4 JDeveloper	317
11.4.1 创建基于 BC4J 的应用工程	318
11.4.2 BC4J JSP 应用工程	322
习题	329
第 12 章 Oracle 程序实例——图书管理信息系统	330
12.1 一个简单的图书管理系统	330
12.2 图书数据库上 Form 的建立	333
习题	340
附录 A 《数据库原理及应用（Oracle）实用教程》教学大纲	341
附录 B 实验指导书	343
参考文献	348

第 1 章 数据库的概念

本章主要介绍数据库的意义、数据库的发展史、数据库的体系结构、数据库系统以及数据库管理系统等内容，以便读者对数据库的概念有一个基本的了解。

1.1 数据库的意义

1.1.1 信息处理及数据处理

诞生于 20 世纪中叶的计算机科学较之其他现代科学技术的发展更迅速，在 21 世纪到来之际，它几乎可以称为“知识爆炸”了。21 世纪是信息和知识的社会，如何组织和利用这些庞大的信息和知识已成为衡量一个国家科学技术水平高低的重要标志。

早在 20 世纪 60 年代，数据库技术作为现代信息系统基础的一门软件学科便应运而生了。现在，数据库技术已成为计算机领域中最重要技术之一，它是软件学科中一个独立的分支。它的出现使得计算机应用渗透到工农业生产、商业、行政、教育、科学研究、工程技术和国防军事的各个部门。管理信息系统（MIS）、办公自动化系统（OA）、决策支持系统等都是使用了数据库管理系统或数据库技术的计算机应用系统。

数据库（Database，即 DB）是存储在一起的相关数据的集合，是存储数据的“仓库”。因此，要理解数据库就需要先了解在数据处理领域中常遇到的两个基本概念：“信息”（information）和“数据”（data）。

信息是关于现实世界事物的存在方式或运动状态的反映的组合。例如，上课用的黑板，它的颜色是黑的，形状是矩形，尺寸是长 3.2m，高 1.4m，材料是木材，这些都是关于黑板的信息，都是关于黑板的存在状态的反映，从不同角度“反映”或“刻画”了黑板这个事物。信息源于物质和能量，一切事物，包括自然和人类都产生信息，信息是物质和能量形态的反应，它不可能脱离物质而存在。信息传递需要物质载体，信息的获取和传递要消耗能量。信息是可以感知的和存储的，并且可以加工、传递和再生。电子计算机是信息处理领域中最先进的工具之一，人类对收集到的信息可以进行取舍整理。几乎和信息同样广泛使用的另一个概念是“数据”。所谓数据，通常指用符号记录下来的可加以鉴别的信息。例如，为了描述黑板的信息，可以用一组数据“黑色、矩形、3.2m×1.4m”来表示，由于“黑色”、“矩形”、“3.2”、“m”……这些符号已经被人们赋予了特定的语义，所以它们就具有了传递信息功能。

从上面的例子中, 可以看到信息和数据之间的固有联系: 数据是信息的符号表示或称为载体, 信息则是数据的内涵, 是对数据的语义解释。但另一方面, 某一具体信息与表示它的数据的这种对应关系又因环境而异。同一信息可能有不同的符号表示, 同一数据也可能有不同的解释。数据处理领域中的数据概念较之科学计算领域中数据概念已经大大地拓宽了。定义中所说的符号, 不仅包含数字符号, 而且包含文字、图像和其他符号; 而所谓“记录下来”也不仅是指用笔写在纸上, 还包括磁记录、光刻等各种记录形式。

由于信息是现实世界中事物的存在方式和运动状态的反应, 而现实世界的事物常常是相互关联的, 这就使得人们在了解、掌握事物之间的固有联系和运动规律的基础上, 可以从一些已知的信息出发, 经过演绎推理, 导出新的信息, 为人类社会生活的各种需要服务, 这就是常说的“信息处理”。例如, 根据上述黑板的尺寸数据和木工定额标准, 可以推算出制作黑板所需的木材数量和制作费用。

一般, 人们将原始信息表示成源数据, 然后对这些源数据进行综合推导加工, 得出新的数据。这些结果数据表示了新的信息, 可以作为某种决策的依据(或用于新的推导加工)。这整个过程通常称为“数据处理”。

电子计算机使大规模的数据处理成为可能, 它和通信、网络技术的发展一起, 进一步推动了信息处理和利用社会化, 极大地增强了人类社会信息处理能力。

1.1.2 数据模型和数据库技术

数据库系统就是实现有组织地、动态地存储大量相关数据, 方便用户访问的计算机软硬件资源组成的系统。而数据库技术是研究数据库的结构、存储、设计和使用的一门软件学科。因此, 数据库技术主要是研究如何存储、使用和管理数据。在计算机应用中, 数据处理占的比重最大, 而数据库系统是数据处理的核心机构, 所以它的效能往往决定了整个计算机应用的经济效益。

数据库离不开数据模型。数据模型是对现实世界客观事物及其联系的描述, 它反映数据项之间和记录之间的联系, 在数据库技术中使用模型的概念描述数据库的结构与语义。常用的 3 种数据模型是: 层次模型(hierarchical model)、网状模型(network model)和关系模型(relational model)。此外, 还有面向对象模型(object_oriented model)等。

数据库这门学科与其他基础软件、系统软件、应用软件有着密切的联系。例如: 操作系统, 数据技术是在操作系统的文件系统的基础上发展起来的, 而且数据库系统本身就是在操作系统(例如 Windows)支持下才能工作。数据库与网络技术和多媒体技术的关系也很密切, 例如, 分布式数据库要用上网络, 数据库甚至可在国际互联网(Internet)上交换海外的数据、声音、图像、图片等多媒体信息。数据库技术与数据结构的关系也是不可分的, 数据库技术不仅要用到数据结构的知识, 而且丰富了数据结构的内容。程序设计是使用数据库系统的最基本方式, 因为数据库中大量的应用程序多是用高级语言加上数据库的操纵语言编写的。集合论、数理逻辑是关系数据库的理论基础, 其很多概念、术语、思想

都直接用到关系数据库中。

1.2 数据库的由来和发展

数据库这个名词起源于 20 世纪 50 年代, 当时美国为了战争的需要, 把各种情报集中在一起, 存入计算机, 称为 Information Base 或 Database。1963 年美国 Honeywell 公司的 IDS (Integrated Data Store) 系统投入运行, 揭开了数据库技术的序幕。1965 年美国利用数据库帮助设计了阿波罗登月火箭, 推动了数据库技术的产生。当时社会上产生了许多形形色色的 Database 或 Databank, 但基本上都是文件系统的扩充。1968 年美国 IBM 公司推出了层次模型的 IMS 数据库系统, 并于 1969 年形成产品; 1969 年, 提出了 COBOL 语言的美国 CODASYL (Conference on Date System Language, 数据系统语言协会) 组织的数据库任务组 (DBTG) 发表了网状数据库系统的标准文本 (1971 年正式通过); 1970 年初, IBM 公司的高级研究员 E.F.Codd 发表论文提出了关系模型, 奠定了关系数据库的理论基础。

20 世纪 70 年代是数据库蓬勃发展的年代。网状系统和层次系统占领了市场, 关系系统开始处于实验阶段, IBM 公司研制出了原型关系语言 System R。1979 年关系软件 (Relational Software) 公司推出了第一个基于 SQL 的商用关系数据库产品 Oracle。

20 世纪 80 年代起, 关系数据库产品已相当成熟, 取代了网状系统和层次系统的市场。同时关系数据库理论也日趋完善, 走向更高级的阶段, 有了分布式数据库系统 (Distributed database Systems) 等。后来, 从不同的计算机应用领域提出了许多数据库的非传统应用课题, 诸如多媒体数据、空间数据、时序数据、科学数据、复杂对象、知识、超文本管理等。为了适应这类应用的需要, 提出了不少新的概念、新的数据模型和系统结构。经过几年的研究和实践, 逐步形成了面向对象数据库系统 (Object_Oriented Database Systems)、主动数据库系统 (Active Database Systems)、大型知识库系统 (Large Knowledge Base Systems)、数据库中的知识发现 (Knowledge Discovery in Database) 以及科学数据库 (Science Database) 等热点。21 世纪数据库技术必将获得更加长足的发展。了解数据库的由来和发展的历史以及数据管理技术各阶段的特点, 对学习好数据库显然十分必要。

综观数据管理技术的发展可知: 它与硬件 (主要是外部存储器)、软件以及计算机应用的范围有密切关系。数据管理技术大致经过以下 4 个阶段: 人工管理阶段、文件系统阶段、数据库阶段以及高级数据库阶段。

1.2.1 人工管理阶段 (20 世纪 50 年代中期以前)

这一阶段的计算机主要用于科学计算。硬件中的外存只有卡片、纸带等。软件只有汇编语言, 没有数据管理方面的软件。数据处理的方式基本上是批处理。这个时期的数据管理特点如下:

1. 数据不保存

进行某一课题计算时将原始数据随程序一起输入主存, 运算处理后将结果数据输出。任务完成后, 数据空间同程序空间一起释放。

2. 没有专用软件对数据进行管理

每个应用程序要包括存储结构、存取方法、输入/输出方式等, 数据结构与程序不具有独立性, 一旦存储结构改变, 就必须由应用程序员修改程序。由于程序直接面向存储结构, 因此不存在逻辑结构与物理结构的区别。

3. 只有程序 (PROGRAM) 的概念, 没有文件 (FILE) 的概念

即使有文件, 也大多是顺序文件, 其他组织方式必须由程序员自行设计与安排。

4. 数据面向应用

即一组数据对应于一个程序。由于各应用程序处理的数据不会毫无联系, 程序之间会有重复。

5. 对数据的存取以记录为单位

其灵活性差。

1.2.2 文件系统阶段 (20 世纪 50 年代后期至 60 年代中后期)

这一阶段的计算机不仅用于科学计算, 还大量用于信息管理。外存已有磁盘、磁鼓等直接存取设备。软件方面出现了高级语言和操作系统。操作系统中的文件系统 (有时也称为“信息处理模块”) 是专门处理外存的数据管理软件。处理数据方式有批处理, 也有联机实时处理。这一阶段数据管理情况如下:

1. 特点

(1) 数据可长期保存在外存的磁盘上。用户经常随时通过程序对文件进行查询、修改以及删除等处理。由于计算转向管理, 数据处理的工作量增大。

(2) 数据的物理结构与逻辑结构有了区别, 但较简单。程序与设备之间有设备独立性 (程序只需用文件名与数据打交道, 不必关心数据的物理位置), 由操作系统的文件系统提供存取方法 (读/写)。由存取方法实现数据的逻辑结构与物理结构之间的转换。

(3) 文件的形式已多样化, 有索引文件、链接文件和直接存取等, 因而对文件的记录可顺序访问, 也可随机访问。但文件之间是独立的, 联系要通过程序去构造, 文件的共享性差。

(4) 有了存储文件以后, 数据不再仅仅属于某个特定的程序, 而可以重复使用。但文件结构的设计仍然是基于特定的用途, 程序仍然是基于特定的物理结构和存取方法编制的, 因此, 数据结构与程序之间的依赖关系并未根本改变。

(5) 对数据的存取基本上还是以记录为单位。

2. 缺陷

在文件系统中, 改变存储设备, 不必改变应用程序。虽然文件系统提供了存取方法, 但这只是初级的数据管理。这种文件系统, 还未能彻底体现用户观点下的数据逻辑结构独立于数据在外存的物理结构要求。因此, 数据的物理结构修改时, 仍然需要修改用户的应用程序。

文件系统有 3 大缺陷:

(1) 数据冗余性 (redundancy), 由于文件之间缺乏联系, 造成每个应用程序都有对应的文件, 有可能同样的数据在多个文件中重复存储;

(2) 不一致性 (inconsistency), 这往往是由数据冗余造成的, 在进行更新操作时, 稍不谨慎, 就可能同样的数据在不同的文件中不一样;

(3) 数据联系弱 (poor data relationship), 这是文件之间独立, 缺乏联系造成的。

由于这些原因, 促使人们研究一种新的数据管理技术, 这就是 20 世纪 60 年代末产生的数据库技术。

1.2.3 数据库阶段 (20 世纪 60 年代末开始)

20 世纪 60 年代末, 磁盘技术取得了重大进展, 大容量 (数百兆字节以上) 和快速存取的磁盘陆续进入市场, 成本有了很大的下降, 为数据库技术的实现提供了物质条件。

60 年代中期, 出现的大多数系统 (database 或 databank) 还不能称为真正的数据库系统。数据管理技术进入数据库阶段的标志是前面讲到的 60 年代后期的 3 大事件, 即 1969 年 IBM 公司推出的 IMS 产品 (层次数据库系统) 和 CODASYL 研究和推荐的 DBTG 系统 (网状数据库系统), 以及 1970 年起, IBM 公司 E.F.Codd 连续发表的一系列论文, 奠定了关系数据库理论基础。

70 年代以来, 数据库技术得到迅速发展, 并投入使用。数据库系统阶段根据其支持的数据模型 (data model), 到目前为止经历了 3 代: 从上面讲到的层次网状代, 到关系模型代以及后面高级数据库阶段讲到的面向对象代。关系模型代从 70 年代初 E.F.Codd 奠定了关系数据库理论基础后, 70 年代末推出了一些试验系统, 80 年代初出现了一批商品化的关系数据库系统, 如 Oracle、SQL/DS、DB2、INGRES、INFORMIX、UNIFY 以及 dBASE、FoxBASE 等。SQL 语言在 1986 年被美 ANSI 和国际标准化组织 (ISO) 采纳为关系数据库语言的国际标准。

与文件系统相比, 数据库系统克服了文件系统的缺陷, 提供了对数据更高级、更有效的管理。概括起来, 数据库技术的管理方式具有以下特点:

1. 采用复杂的数据模型 (结构)

数据模型不仅要描述数据本身的特点, 还要描述数据之间的联系。这种联系是通过存