



高等院校计算机应用技术规划教材

数据库原理及技术应用教程 (Oracle)



刘甫迎 饶 斌 刘 焱 编著



中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE



高等院校计算机应用技术规划教材

数据库原理及技术应用教程 (Oracle)

刘甫迎 饶 斌 刘 焱 编著

中国铁道出版社
CHINA RAILWAY PUBLISHING HOUSE

内 容 简 介

本书包括 12 章和 3 个附录,详细介绍了数据库基础、数据库的由来和发展、数据模型、关系数据库、函数依赖、范式、PD CASE 工具、数据库的设计与维护(CDM、PDM 数据模型)、分布式数据库、安全性与完整性、并发控制与恢复等数据库的基本原理和技术。讲述了后端大型数据库的工业标准——Oracle,且将之贯穿全书。Oracle 的网格计算、Oracle RAC、PL/SQL、闪回技术、企业管理器 OEM、应用开发工具(Forms、Reports、JDeveloper)以及 J2EE 和互联网服务器应用等内容,使本书同时又成为一本学习 Oracle(以新版本 11g 为主,兼顾 9i、10g 版本)的基于 C/S、B/S 模式编程和 DBA(数据库管理员)的教科书。

本书提供实例、实验和实习、习题、教学大纲等,便于学习与教学。本书适合作为高等院校及软件学院的教材,还可作为从事数据库软件开发和应用者的参考书。

图书在版编目(CIP)数据

数据库原理及技术应用教程: Oracle/刘甫迎, 饶斌,
刘焱编著. —北京: 中国铁道出版社, 2009. 8
高等院校计算机应用技术规划教材
ISBN 978-7-113-10426-9

I. 数… II. ①刘…②饶…③刘… 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(2009)第 142304 号

书 名: 数据库原理及技术应用教程(Oracle)
作 者: 刘甫迎 饶 斌 刘 焱 编著

策划编辑: 翟玉峰 沈 洁

责任编辑: 王占清

特邀编辑: 李红玉

封面设计: 付 巍

版式设计: 于 洋

编辑部电话: (010) 63583215

编辑助理: 邱雪姣

封面制作: 白 雪

责任印制: 李 佳

出版发行: 中国铁道出版社(北京市宣武区右安门西街 8 号 邮政编码: 100054)

印 刷: 河北省遵化市胶印厂

版 次: 2009 年 9 月第 1 版 2009 年 9 月第 1 次印刷

开 本: 787mm×1092mm 1/16 印张: 26.5 字数: 654 千

印 数: 4 000 册

书 号: ISBN 978-7-113-10426-9/TP·3521

定 价: 39.00 元

版权所有 侵权必究

本书封面贴有中国铁道出版社激光防伪标签,无标签者不得销售

凡购买铁道版的图书,如有缺页、倒页、脱页者,请与本社计算机图书批销部调换。

诞生于 20 世纪 60 年代的数据库,是人类进入 21 世纪知识经济时代不可或缺的“武器”,其建设规模和使用水平已成为衡量一个国家信息化程度的重要标志。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,是高等学校计算机类专业和信息管理类专业的一门专业基础课,越来越多的人希望学习数据库。

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,到现在的 Oracle 11g,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。

笔者较早打破了其他介绍数据库原理的书忌讳写某一具体的数据库的惯例,本书除详细介绍数据库的基本原理和技术外,继续将 Oracle 作为实例贯穿全书,使之同时又成为一本学习 Oracle 的基于 C/S 模式和 B/S 编程以及 DBA (数据库管理员) 的教科书。

本书的主要特点是:

- 全书以新版本 Oracle 11g 为基础,同时,重要章节(例如,Oracle 的安装及目录结构、企业管理器 OEM 等)又对比 Oracle 9i、Oracle 10g 的内容进行讲述。体现了最新技术的先进性,同时兼顾了较低的两个版本用户还较多的情况,以便满足各类用户的需要。
- 突出了 Oracle 11g 的网格计算,将之贯穿全书(见第 2 章的网格计算基础、第 8 章的 Oracle RAC 技术、第 9 章的 OEM 管理 Oracle RAC、附录 C 的 RAC 的安装与配置指导书)。
- 介绍了 Oracle 11g 数据库高可用性的内容: LogMiner、数据卫士、RMAN、Oracle 闪回技术、Oracle RAC 技术和自动存储管理(ASM)等(见第 8 章)。
- 按 Oracle 11g 的要求,本书编写了企业管理器 OEM 的相关内容,详细介绍了 Oracle 11g 的 OEM 提供的两种独立的控制方式:数据库控制和网格控制(见第 9 章)。
- 介绍了 PowerDesigner CASE 技术,用之建立数据库 CDM、PDM 数据模型(见第 6 章、第 7 章)。
- 本书第 11 章介绍了 Oracle Developer Suite,讲述了其他相关书籍较少涉及的应用开发工具 Forms Developer、Reports Developer, JDeveloper 以及 J2EE 和互联网服务器应用等内容,使本书成为一本不多见的、内容较全的 Oracle 教科书。
- 本书注重理论与实践相结合,突出实践动手能力和实用性。有实例、实验指导书,便于读者参考、使用,力图使学生学习本书后便基本可以编制 Oracle 程序和应用系统。
- 本书附有教学大纲(见附录 A)、专业实习(见附录 C)、习题,图文并茂,便于学习与教学。
- 本书作者长期从事 Oracle 数据库教材的编著和教学、科研工作,有此课程的教学经验,在此书的材料组织过程中将其经验融入了进去。

本书由刘甫迎、饶斌、刘焱编著。刘甫迎编写了第 2 章、第 4 章、第 9 章、第 12 章和附录 A、附录 B;饶斌编写了第 1 章、第 5 章、第 8 章、第 10 章、第 11 章;刘焱编写了第 3 章、第 6 章、

第7章、附录C；全书由刘甫迎教授统稿。在本书的编写和出版过程中，中国铁道出版社给予了很大的帮助，党晋蓉教授等老师和曾克蓉、李朝蓉等学生也做了许多工作，在此一并表示感谢！

由于作者水平有限，不妥之处在所难免，敬请广大读者批评指正。

编者

2009年8月

第 1 章 数据库基础	1
1.1 数据库概念	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	2
1.2.1 人工管理阶段	3
1.2.2 文件系统阶段	4
1.2.3 数据库阶段	4
1.2.4 高级数据库阶段	6
1.3 数据库的体系结构	10
1.4 数据库系统和 DBA (数据库管理员)	13
1.5 数据库管理系统 (DBMS)	16
习题	18
第 2 章 关系模型及 Oracle 网络计算数据库	20
2.1 关系模型的基本概念	20
2.1.1 关系的定义	20
2.1.2 关系模型	21
2.1.3 关系数据语言概述	22
2.2 关系代数	23
2.2.1 传统的集合运算	23
2.2.2 专门的关系运算	24
2.3 关系演算	28
2.3.1 元组关系演算	28
2.3.2 域关系演算	30
2.4 关系数据库标准语言——SQL	30
2.5 Oracle 对象关系数据库系统	33
2.5.1 Oracle 系统的特点、产品结构及组成	34
2.5.2 Oracle 的体系结构 (实例的进程结构和内存结构)	36
2.5.3 Oracle 的配置方案	42
2.5.4 Oracle 运行过程	43
2.6 Oracle 的网格计算	43
2.6.1 网格计算基础	43
2.6.2 Oracle 11g 的网格基础架构	45

2.7 Oracle 的安装与使用环境	47
2.7.1 产品的获取	47
2.7.2 Oracle Database 的安装	47
2.7.3 Oracle 数据库文件目录结构	51
2.7.4 安装 Oracle Database 后的程序组与服务	53
习题	59
第 3 章 实体-联系模型 (E-R Model)	61
3.1 实体和实体集合	61
3.2 联系和联系集合	62
3.3 属性、映射限制和关键字	63
3.3.1 属性	63
3.3.2 映射限制	63
3.3.3 关键字	64
3.4 E-R 图和将之归纳为表	66
3.4.1 实体-联系图 (E-R 图)	66
3.4.2 把 E-R 图转变为表	68
3.5 概括和聚集	70
3.5.1 概括	70
3.5.2 聚集	71
3.6 E-R 数据库模式设计	72
3.6.1 映射基本集	72
3.6.2 实体集和联系集的使用	72
3.6.3 扩展 E-R 特征的使用	73
习题	73
第 4 章 Oracle 的 SQL 和 PL/SQL	74
4.1 Oracle 的 SQL 语言	74
4.1.1 Oracle 的 SQL 介绍	74
4.1.2 SQL 的简单查询 (SELECT)	74
4.2 创建、修改、删除表 (Table)	76
4.2.1 创建表 (CREATE TABLE)	76
4.2.2 修改表结构	82
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	84
4.3.1 模式对象	84
4.3.2 对象及成分的命名	85

4.3.3	直接量	86
4.3.4	LOB 数据类型	87
4.3.5	空值	87
4.3.6	伪列	88
4.3.7	注释	88
4.3.8	运算符	89
4.3.9	函数	91
4.3.10	表达式与条件	93
4.4	插入、修改和删除表数据	93
4.4.1	将新行插入 (INSERT) 表	93
4.4.2	修改 (UPDATE) 表的行	94
4.4.3	从表删除 (DELETE) 行	95
4.5	SQL 的数据查询	95
4.5.1	连接查询	95
4.5.2	嵌套查询	96
4.5.3	分组计算查询	97
4.5.4	集合的并运算	97
4.6	视图	98
4.6.1	视图定义	98
4.6.2	视图查询	98
4.6.3	视图修改	99
4.7	Oracle 的 PL/SQL	100
4.7.1	什么是 PL/SQL	100
4.7.2	PL/SQL 的结构	101
4.7.3	控制结构	108
4.7.4	游标 (CURSOR)	111
	习题	114
第 5 章	关系数据库设计理论	118
5.1	引言	118
5.2	函数依赖	119
5.3	范式	121
5.3.1	第一范式 (1NF)	121
5.3.2	第二范式 (2NF)	122
5.3.3	第三范式 (3NF)	122
5.3.4	BC 范式 (BCNF)	123
5.4	多值依赖和第四范式 (4NF)	124
5.5	连接依赖和第五范式 (5NF)	126

5.5.1 连接依赖	126
5.5.2 第五范式 (5NF)	127
习题	127
第 6 章 PowerDesigner CASE 技术	129
6.1 什么是 CASE 工具	129
6.2 PD 分析设计过程及若干级建模技术	132
6.2.1 PD 的分析设计过程	132
6.2.2 PD 的若干级建模功能	134
6.3 PD 的功能	135
6.3.1 PD 的一般功能	135
6.3.2 PD 12.5 的主要特征	136
6.4 PD 的分析设计环境	136
6.4.1 安装 PD	136
6.4.2 PD 用户界面	138
6.4.3 PD 的分析设计环境	140
6.4.4 PD 的公共资源	141
习题	142
第 7 章 数据库设计与维护	143
7.1 数据库设计概述	143
7.1.1 数据库设计步骤、生存期及其他	143
7.1.2 数据库设计的需求分析	145
7.2 用 PD 创建 CDM	147
7.2.1 概念数据模型 (CDM) 基础	147
7.2.2 用 PD 建立 CDM	149
7.2.3 定义业务规则	152
7.2.4 定义 CDM 中的域	154
7.2.5 定义数据项	155
7.2.6 定义实体	158
7.2.7 定义联系	162
7.2.8 定义继承	166
7.2.9 定义检查约束参数	168
7.2.10 由 CDM 生成 PDM	169
7.3 用 PD 创建物理数据模型 (PDM)	171
7.3.1 PDM 及其与 DBMS 的关系	171
7.3.2 建立物理数据模型 (PDM)	172
7.3.3 定义索引	174

7.3.4	定义视图	176
7.3.5	PDM 中的用户管理	177
7.3.6	检查 PDM 中的对象	177
7.3.7	逆向工程	177
7.4	数据库的实现与维护	179
7.4.1	数据库的实现与维护概述	179
7.4.2	由 PDM 生成数据库 SQL 脚本	181
7.5	用 Oracle SQL 建立数据库	182
7.5.1	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)	182
7.5.2	用 Oracle SQL 建立和修改数据库	189
7.5.3	初始化参数文件	198
7.5.4	数据字典 (Data Dictionary)	199
习题		201
第 8 章	数据库保护	203
8.1	数据库的安全性	203
8.1.1	数据库的存取控制和用户的建立	204
8.1.2	特权和角色	208
8.1.3	审计	213
8.2	数据完整性	213
8.2.1	完整性约束	214
8.2.2	数据库触发器	214
8.3	并发控制	217
8.3.1	数据不一致的类型	218
8.3.2	封锁	218
8.3.3	Oracle 多种一致性模型	218
8.3.4	封锁机制	219
8.3.5	手工的数据封锁	221
8.4	数据库后备和恢复以及高可用性	221
8.4.1	数据库恢复所使用的结构	221
8.4.2	在线及归档日志	222
8.4.3	数据库后备、Oracle 数据卫士和数据库恢复	227
8.4.4	Oracle 的闪回技术	238
8.4.5	Oracle RAC 技术	247
8.4.6	自动存储管理 (ASM)	249
习题		250

第 9 章 Oracle 企业管理器 (Oracle Enterprise Manager)	252
9.1 企业管理器的框架结构、安装及配置	252
9.1.1 OEM 的架构	252
9.1.2 OEM 的安装、配置及启动	256
9.2 使用企业管理器管理 Oracle 系统	263
9.2.1 建立和管理数据库 (数据库配置、数据文件管理)	263
9.2.2 数据库用户管理	271
9.2.3 方案和表管理	273
9.2.4 管理视图、索引和触发器	278
9.3 使用企业管理器维护 Oracle 系统	283
9.3.1 管理 (本地) 表空间	283
9.3.2 浏览预警文件信息	284
9.4 使用企业管理器进行性能调整	284
9.5 OEM Grid Control 网格控制管理	286
9.5.1 启动和访问 OEM 网格控制	286
9.5.2 使用网格控制监管全部 Oracle 环境	289
9.5.3 管理 Oracle RAC	295
习题	297
第 10 章 分布式数据库系统	298
10.1 概述	298
10.1.1 分布式数据库的特征	299
10.1.2 全功能分布式数据库的规则和目标	300
10.2 分布式数据库的连接	301
10.2.1 分布式数据库全局名和数据库链	302
10.2.2 连接客户 (client) 和服务器 (server)	304
10.2.3 连接服务器 (server) 到其他服务器 (server)	304
10.3 分布式查询处理及其他	305
10.3.1 建立分布式查询	305
10.3.2 在分布式事务 (distributed transaction) 中的其他语句	306
10.4 事务管理	307
10.4.1 事务	307
10.4.2 Oracle 的事务管理	307
10.5 多协议信息交换	309
10.5.1 Oracle Net	309
10.5.2 Oracle Net 连接组分	310
10.5.3 多协议交换	311
10.5.4 配置 Oracle Net 客户机 (client)	313

10.5.5	配置 listener.....	314
10.5.6	配置 Oracle Net 服务器	315
10.5.7	启动 listener.....	316
10.6	表快照与复制.....	316
10.6.1	人工复制表.....	316
10.6.2	用触发器 (TRIGGERS) 复制表.....	317
10.6.3	自动复制和更新表	317
10.7	用 Oracle Net 配置助手工具配置客户端	317
10.7.1	配置连接数据库	317
10.7.2	客户端测试	322
	习题.....	322
第 11 章	数据库应用程序开发工具——Oracle Developer Suite	323
11.1	Oracle Developer Suite 概述.....	323
11.2	Forms Developer.....	326
11.2.1	Form 应用基础.....	326
11.2.2	Form 应用设计	331
11.2.3	创建主从型 Form.....	334
11.2.4	项属性	335
11.2.5	创建 Form 应用程序	337
11.2.6	触发器	341
11.3	Reports Developer.....	347
11.3.1	Reports Builder 基础	347
11.3.2	创建一个简单的报表	348
11.3.3	创建图形 (graphics)	353
11.3.4	在主界面中调用报表	356
11.4	Oracle J2EE 高级技术与 JDeveloper.....	357
11.4.1	Java2 企业版 (J2EE)	357
11.4.2	Oracle 的 J2EE 技术与 OC4J	360
11.4.3	Oracle Java 商业组件 (BC4J)	363
11.4.4	JDeveloper	366
11.5	Oracle 互联网服务器与应用	376
11.5.1	认识 Oracle 互联网服务器	376
11.5.2	Oracle 互联网服务器应用	378
	习题.....	379
第 12 章	综合应用实例——图书管理信息系统.....	381
12.1	一个简单的图书管理信息系统.....	381

12.2 图书数据库上 Form 的建立	383
习题	390
附录 A “数据库原理及技术 (Oracle)” 教学大纲	391
附录 B 实验指导书	393
附录 C 专业实习 (Oracle) —— 实时应用集群 (RAC) 安装与配置指导书	398
参考文献	410

第 1 章

数据库基础

本章介绍数据库的意义、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,以便对数据库的概念有一个基本的了解。

1.1 数据库概念

1.1.1 信息处理及数据处理

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等各个部门。管理信息系统(MIS)、办公自动化系统(OA)、决策支持系统(DSS)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, 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、宽 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,它可以存储,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。

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”表示,由于“黑色”、“矩形”、“3”、“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的功能。

从上面的例子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的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,而现实世界中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)。整个过程通常称为“数据处理”。

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,极大地增强了人类社会信息处理的能力。

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,它反映了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。常用的三种数据模型是: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。(关系数据模型将在第2章进行详细讲述。)

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只有在操作系统(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,例如分布式数据库需要网络的支持,数据库可在因特网(Internet)上交换数据、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割的,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过程都是用高级语言加上数据库操作语言编写的。集合论、数理逻辑是关系数据库的理论基础,它们的很多概念、术语、思想都可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 data 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。

从20世纪80年代起,关系数据库产品已相当成熟,抢占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 system)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 system)、主动数据库系统(active database system)、大型知识库系统(large knowledge base system)、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了以下四个阶段: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

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,因此程序之间会有重复。

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

文件系统阶段 (20 世纪 50 年代后期至 20 世纪 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。处理数据的方式有批处理, 也有联机实时处理。这一阶段的数据管理情况如下:

1. 特点

- 数据可长期保存在外存上。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。
- 数据的物理结构与逻辑结构有了区别, 但比较简单。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。
- 文件的形式已多样化, 有索引文件、链接文件和直接存取文件等, 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。
- 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。
- 确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, 修改数据的物理结构时, 仍然需要修改用户的应用程序。

文件系统有三大缺陷:

- 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。
- 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, 就可能导致同样的数据在不同的文件中不一样。
- 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

数据库阶段是从 20 世纪 60 年代末开始的。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。