

Practical...

# 数据库

## 系统原理与应用

# Server Administration

文家焱 施平安 编著

冶金工业出版社

# 数据库系统原理与应用

文家焱 施平安 编著

北 京

冶金工业出版社

2002

## 内 容 简 介

本书是一本讲述数据库系统原理与应用技术的书籍，所阐述的内容主要包括数据库系统概述、关系数据库系统、结构化查询语言——SQL 语言、嵌入式 SQL 语言、查询优化、关系规范化、数据库设计、数据库事务处理技术（如数据库的恢复策略和并发控制等）、数据库安全性和完整性、面向对象数据库、并行数据库系统、分布式数据库以及数据库的新应用、SQL Server 2000、数据库设计/建模工具的使用等。

本书内容丰富，语言简明易懂，书中还配有大量的例题和练习，相信读者认真学习完本书后一定会获益非浅。

本书适合于数据库技术的初、中级学者，也可作为大、中专院校相关专业师生的学习和参考用书。

### 图书在版编目 (CIP) 数据

数据库系统原理与应用 / 文家焱等编著. —北京：  
冶金工业出版社，2002. 11  
ISBN 7-5024-3119-5

I. 数... II. 文... III. 数据库系统  
IV. TP311.13

中国版本图书馆 CIP 数据核字 (2002) 第 075165 号

出版人 曹胜利 (北京沙滩嵩祝院北巷 39 号, 邮编 100009)

责任编辑 程志宏

广东省肇庆新华印刷有限公司印刷; 冶金工业出版社发行; 各地新华书店经销

2002 年 11 月第 1 版, 2002 年 11 月第 1 次印刷

787mm × 1092mm 1/16; 22.5 印张; 516 千字; 350 页; 1-2500 册

35.00 元

冶金工业出版社发行部 电话: (010) 64044283 传真: (010) 64027893

冶金书店 地址: 北京东四西大街 46 号 (100711) 电话: (010) 65289081

(本社图书如有印装质量问题, 本社发行部负责退换)

# 前 言

## 1. 关于数据库技术

尽管数据库技术的出现和应用只有短短的几十年，但是数据库技术的发展是突飞猛进的。当今的数据库系统已经成为一个大家族，丰富多样的数据模型、层出不穷的新兴技术、不断扩展的应用领域使得现代数据库正朝着多维度、多方面迅速发展。现代信息技术的发展已经不能离开数据库技术的支持，可以说，没有数据库技术的发展，没有相关优秀数据库产品的推出和应用，社会的信息化就不可能实现。

现在，衡量一个国家信息化程度时，数据库的建设规模和数据库信息量的大小和应用程度已经成为重要的检验内容之一。因此，不管是计算机专业的本、专科学生和研究生，还是对计算机有着强烈爱好的人士，数据库原理和应用技术都是不但要学，而且要学好的重要内容。

## 2. 本书结构

本书共有 16 章，具体结构如下：

第 1 章：数据库系统概述。主要讲述数据库管理系统、数据抽象与数据独立性、数据模型、数据库语言等内容。

第 2 章：关系数据库系统。主要讲述关系数据模型的数据结构、数据完整性等内容。

第 3 章：SQL 语言。主要讲述 SQL 的产生与发展、数据定义、SQL 查询、数据更新、SQL 数据控制等内容。

第 4 章：嵌入式 SQL 语言。主要讲述嵌入式 SQL 语言要解决的问题、嵌入式 SQL 语言的语法格式、动态 SQL 语言介绍、静态 SQL 与动态 SQL 的比较等内容。

第 5 章：查询优化。主要讲述查询优化的一般策略、代数优化等内容。

第 6 章：关系规范化。主要讲述关系规范化理论、函数依赖、关系模式的规范化、关系模式规范化的原则、关系模式分解等内容。

第 7 章：数据库设计。主要讲述需求分析、概念结构设计、逻辑结构设计、数据库物理设计、数据库的实施等内容。

第 8 章：数据库事务处理技术。主要讲述数据库故障分类、恢复的实现技术、具有检查点的恢复技术、Oracle 的事务处理技术等内容。

第 9 章：并发控制。主要讲述锁协议、两段锁协议、死锁及其处理方法、Oracle 的封锁机制等内容。

第 10 章：数据库安全性与完整性。主要讲述用户标识与鉴别、访问控制、数据加密、统计数据库的安全性、完整性约束类型、完整性约束机制等内容。

第 11 章：面向对象数据库。主要讲述新的数据库应用及特点、面向对象数据模型、面向对象数据库管理系统、对象/关系数据库等内容。

第 12 章：并行数据库系统。主要讲述并行数据库的体系结构、并行数据库的查询优化技术等内容。

第 13 章：分布式数据库系统。主要讲述分布式事务、分布式数据库系统中的并发控

制、分布式查询处理和优化等内容。

第 14 章：数据库的新应用。主要讲述决策支持系统、数据仓库、数据挖掘等内容。

第 15 章：关系数据库管理系统——SQL Server 2000。主要讲述 SQL Server 2000 的装配置、SQL Server 2000 的基本使用、SQL Server 2000 的备份与恢复等内容。

第 16 章：数据库设计/建模工具的使用。主要讲述 PowerDesigner 和 Erwin。

此外，为了巩固读者对知识的掌握，本书每章后面都附有习题，书的后面还给出了参考答案，以方便对照练习。

### 3. 本书特点

考虑到数据库系统是一门理论性和应用性都很强的课程。因此，为了便于教师对本门课的教学和学生对知识的掌握，特别是为了鼓励学生努力学习和勤于思考，编者总结了多年来从事数据库系统理论与实践教学的经验，力图通过一个新颖的角度、合适的切入点数据库系统各方面的知识进行介绍，由浅入深、循序渐进地探讨数据库的基本原理和应用技术。

本书结构紧凑、风格统一，注重以简明易懂的笔调阐述深刻内容，再配以大量经过精心筛选的例题和习题，不仅方便了老师教学，也便于学生自学。即使基础较差的学生，相信通过本书的学习也能够尽快掌握数据库系统的理论和技术，进入数据库管理系统的应用和开发的高级阶段。

### 4. 学习指导

本书的参考学时为 50~80 学时，可以分配 10~20 学时用于学生上机实践。根据不同层次的学生和课程设置的要求，书中有些理论性较强、难度较大的内容可以选学或者不学。

本书末尾提供了大量参考文献，有些参考文献对数据库某一方面的内容（如数据库设计、关系规范化、数据库安全性、分布式数据库、并发数据库技术等）有更深入的阐述。读者如果想对某一方面的问题有更深入的了解，可以在学完本书的基本理论的基础上，进一步查阅相关的文献进行学习。

### 5. 适用对象

本书适用于数据库技术的初、中级学者，也可作为大、中专院校相关专业师生的学习和参考用书。

本书主要由文家焱执笔，施平安审定。本书在编写过程中得到了许多同事和家人的支持和鼓励，一些同行提出了许多宝贵的意见和建议，并提供了许多资料，在此对他们表示衷心的感谢。

由于作者水平有限，时间仓促，书中错误之处在所难免，恳请读者给予批评指正。

编者

2002 年 8 月

# 目 录

<b>第 1 章 数据库系统概述</b> .....	1
1.1 数据库技术简介 .....	1
1.1.1 数据库技术的应用 .....	1
1.1.2 数据库系统及其特点 .....	2
1.2 数据库管理系统 .....	3
1.2.1 DBMS 的主要功能 .....	4
1.2.2 DBMS 的系统结构 .....	5
1.3 数据抽象与数据独立性 .....	7
1.3.1 数据抽象 .....	7
1.3.2 数据独立性 .....	8
1.4 数据模型 .....	8
1.4.1 实体-联系模型 .....	12
1.4.2 网状数据模型 .....	13
1.4.3 层次结构模型 .....	19
1.4.4 关系数据模型 .....	26
1.4.5 面向对象数据模型 .....	31
1.5 数据库语言 .....	31
1.5.1 数据定义语言 (DDL) .....	31
1.5.2 数据操纵语言 (DML) .....	31
小结 .....	32
综合练习题一 .....	32
一、选择题 .....	32
二、填空题 .....	33
三、思考题 .....	33
<b>第 2 章 关系数据库系统</b> .....	34
2.1 关系数据模型的数据结构 .....	34
2.1.1 基本概念 .....	35
2.1.2 基本关系术语 .....	35
2.1.3 对关系的数学定义 .....	37
2.2 数据完整性 .....	38
2.3 关系代数 .....	39
2.3.1 集合操作 .....	40
2.3.2 专门的关系运算 .....	42
小结 .....	47
综合练习题二 .....	48
一、选择题 .....	48
二、填空题 .....	49
三、思考题 .....	50
<b>第 3 章 SQL 语言</b> .....	51
3.1 SQL 的产生与发展 .....	51
3.2 数据定义 .....	52
3.2.1 基本表的定义、修改和删除 .....	52
3.2.2 建立和删除索引 .....	55
3.3 SQL 查询 .....	55
3.3.1 SQL 的基本结构 .....	55
3.3.2 简单查询 .....	58
3.3.3 连接查询 .....	60
3.3.4 嵌套查询 .....	62
3.3.5 使用库函数进行查询 .....	64
3.3.6 集合运算 .....	68
3.4 数据更新 .....	69
3.4.1 插入数据 .....	69
3.4.2 更新数据 .....	70
3.4.3 删除数据 .....	70
3.5 视图 .....	71
3.6 SQL 数据控制 .....	72
小结 .....	73
综合练习题三 .....	73
一、选择题 .....	73
二、填空题 .....	74
三、思考题 .....	74
<b>第 4 章 嵌入式 SQL 语言</b> .....	76
4.1 嵌入式 SQL 语言要解决的问题 .....	76
4.2 嵌入式 SQL 语言的语法格式 .....	76
4.3 数据库与程序工作单元之间的通信 .....	77
4.3.1 SQL 通信区 (SQLCA) .....	77
4.3.2 宿主变量 .....	78

4.3.3 游标 .....	79	6.3 关系模式的规范化 .....	108
4.4 程序实例 .....	80	6.3.1 第一范式 .....	108
4.5 动态 SQL 语言介绍 .....	83	6.3.2 第二范式 .....	108
4.5.1 嵌入式 SQL 实现动态 SQL 的方式 .....	83	6.3.3 第三范式 .....	109
4.5.2 动态游标的使用 .....	84	6.3.4 改进的第三范式—— BCNF 范式 .....	110
4.6 静态 SQL 与动态 SQL 的比较 .....	86	6.3.5 多值依赖与第四范式 .....	113
4.6.1 安全性方面的比较 .....	86	6.4 关系模式规范化的原则 .....	117
4.6.2 代码方面的比较 .....	89	6.5 数据依赖的公理系统 .....	117
小结 .....	90	6.5.1 Armstrong 公理系统 .....	117
综合练习题四 .....	91	6.5.2 Armstrong 公理系统的 完备性 .....	121
一、选择题 .....	91	6.5.3 函数依赖集等价与最小 依赖集问题 .....	122
二、填空题 .....	91	6.6 关系模式分解 .....	123
三、思考题 .....	91	6.6.1 无损连接模式分解 .....	124
<b>第 5 章 查询优化 .....</b>	<b>92</b>	6.6.2 保持函数依赖的模式分解 .....	127
5.1 查询优化概述 .....	92	6.6.3 模式分解的算法 .....	128
5.1.1 实例分析 .....	93	小结 .....	130
5.1.2 查询优化的一般策略 .....	95	综合练习题六 .....	131
5.2 代数优化 .....	95	一、选择题 .....	131
5.2.1 关系代数等价变换规则 .....	95	二、填空题 .....	131
5.2.2 查询的内部表示形式 .....	97	三、思考题 .....	132
5.2.3 启发式关系代数表达式 优化算法 .....	97	<b>第 7 章 数据库设计 .....</b>	<b>133</b>
5.2.4 查询优化实例 .....	98	7.1 数据库设计概述 .....	133
小结 .....	101	7.1.1 数据库设计的任务、特征与 目标 .....	133
综合练习题五 .....	101	7.1.2 数据库设计方法——规范 设计法 .....	134
一、选择题 .....	101	7.2 需求分析 .....	136
二、填空题 .....	101	7.2.1 收集资料 .....	137
三、思考题 .....	101	7.2.2 分析和整理资料 .....	137
<b>第 6 章 关系规范化 .....</b>	<b>102</b>	7.2.3 数据字典 .....	138
6.1 关系规范化理论 .....	102	7.2.4 需求分析实例 .....	138
6.1.1 关系规范化的定义 .....	102	7.3 概念结构设计 .....	144
6.1.2 一个不好的关系模式 .....	102	7.3.1 设计局部 E-R 图 .....	144
6.2 函数依赖 .....	104	7.3.2 集成局部 E-R 图为初步 E-R 图 .....	146
6.2.1 函数依赖的定义 .....	104		
6.2.2 完全函数依赖 .....	106		
6.2.3 传递函数依赖 .....	106		
6.2.4 键 (key) .....	107		

7.4 逻辑结构设计 .....	148	9.4 死锁及其处理方法 .....	182
7.4.1 E-R 图向关系模式的转换 .....	148	9.4.1 死锁的概念 .....	182
7.4.2 规范化处理 .....	152	9.4.2 死锁的预防 .....	182
7.4.3 设计外模式 .....	152	9.4.3 死锁的检测与处理 .....	184
7.5 数据库物理设计 .....	153	9.5 多粒度封锁和意向锁 .....	184
7.5.1 设计物理结构 .....	153	9.6 Oracle 的封锁机制 .....	186
7.5.2 物理设计实例 .....	154	小结 .....	188
7.6 数据库的实施 .....	155	综合练习题九 .....	188
小结 .....	156	一、选择题 .....	188
综合练习题七 .....	157	二、填空题 .....	188
一、选择题 .....	157	三、思考题 .....	188
二、填空题 .....	157	<b>第 10 章 数据库安全性与完整性 .....</b>	<b>190</b>
三、思考题 .....	157	10.1 概述 .....	190
<b>第 8 章 数据库事务处理技术 .....</b>	<b>158</b>	10.2 用户标识与鉴别 .....	191
8.1 事务的基本概念 .....	158	10.3 访问控制 .....	192
8.1.1 事务及其性质 .....	158	10.3.1 自由访问控制 .....	192
8.1.2 事务的状态 .....	159	10.3.2 强制访问控制方法 .....	194
8.1.3 SQL 中的事务定义 .....	160	10.3.3 基于视图的访问控制 .....	196
8.2 数据库故障分类 .....	161	10.3.4 推断控制 .....	197
8.3 恢复的实现技术 .....	163	10.3.5 审计 .....	197
8.3.1 数据转储 .....	163	10.4 数据加密 .....	198
8.3.2 基于日志的恢复 .....	165	10.5 统计数据库的安全性 .....	198
8.4 具有检查点的恢复技术 .....	170	10.6 数据库完整性 .....	199
8.5 介质故障的恢复与数据库镜像 .....	172	10.7 完整性约束类型 .....	200
8.6 Oracle 的事务处理技术 .....	172	10.7.1 列值非空完整性约束 .....	200
小结 .....	177	10.7.2 键值惟一完整性约束 .....	201
综合练习题八 .....	177	10.7.3 主键完整性约束 .....	201
一、选择题 .....	177	10.7.4 外键(参照)完整性约束 .....	201
二、填空题 .....	177	10.7.5 检查完整性约束 .....	202
三、思考题 .....	177	10.8 完整性约束机制 .....	202
<b>第 9 章 并发控制 .....</b>	<b>178</b>	10.8.1 精心构造的事务 .....	203
9.1 并发控制简介 .....	178	10.8.2 特权最小化 .....	203
9.1.1 并发的目的 .....	178	小结 .....	204
9.1.2 并发所引起的问题 .....	178	综合练习题十 .....	205
9.2 锁协议 .....	179	一、选择题 .....	205
9.2.1 锁的类型 .....	180	二、填空题 .....	205
9.2.2 锁的授予 .....	180	三、思考题 .....	205
9.3 两段锁协议 .....	181	<b>第 11 章 面向对象数据库 .....</b>	<b>207</b>



11.1 新的数据库应用及特点 .....	207	13.1.1 什么是分布式数据库系统 ..	234
11.2 面向对象数据模型 .....	209	13.1.2 同构与异构 .....	235
11.2.1 对象与封装 .....	209	13.1.3 分布式数据存储 .....	235
11.2.2 类与实例 .....	210	13.1.4 数据透明度的问题 .....	237
11.2.3 类层次结构 .....	211	13.2 分布式事务 .....	238
11.2.4 继承与多重继承 .....	211	13.2.1 系统失效模式 .....	238
11.2.5 对象标识 .....	213	13.2.2 提交协议 .....	239
11.2.6 重载、过载和迟绑定 .....	214	13.2.3 三阶段提交协议 .....	241
11.2.7 复杂对象与对象包含 .....	214	13.3 分布式数据库系统中的	
11.3 面向对象数据库管理系统 .....	216	并发控制 .....	242
11.3.1 关系与参照完整性 .....	216	13.3.1 单锁管理器方法 .....	242
11.3.2 组合对象 .....	217	13.3.2 分布式锁管理器方法 .....	242
11.3.3 位置透明性 .....	218	13.3.3 死锁处理 .....	244
11.3.4 对象版本的应用 .....	218	13.4 分布式查询处理和优化 .....	244
11.3.5 模式演变 .....	218	13.5 分布式查询处理的一般过程 .....	245
11.3.6 OODB 语言 .....	220	13.5.1 查询变换 .....	245
11.4 对象/关系数据库 .....	220	13.5.2 简单连接处理 ( Simple Join	
11.5 OODB 的现状与发展 .....	223	Processing ) .....	246
小结 .....	224	13.5.3 半连接策略 .....	246
综合练习题十一 .....	224	小结 .....	247
一、选择题 .....	224	综合练习题十三 .....	248
二、填空题 .....	224	一、选择题 .....	248
三、思考题 .....	224	二、填空题 .....	248
<b>第 12 章 并行数据库系统 .....</b>	<b>226</b>	三、思考题 .....	248
12.1 并行数据库系统简介 .....	226	<b>第 14 章 数据库的新应用 .....</b>	<b>249</b>
12.2 并行数据库的体系结构 .....	227	14.1 决策支持系统 .....	249
12.2.1 共享内存 .....	228	14.1.1 决策支持系统的定义 .....	249
12.2.2 共享磁盘结构 .....	228	14.1.2 DSS 的性能目标 .....	250
12.2.3 无共享结构 .....	229	14.1.3 DSS 的结构 .....	250
12.3 数据划分方法 .....	230	14.2 数据仓库 .....	251
12.4 并行数据库的查询优化技术 .....	232	14.2.1 数据仓库的概念与结构 .....	251
小结 .....	232	14.2.2 数据集市 .....	253
综合练习题十二 .....	233	14.2.3 数据仓库系统 .....	254
一、选择题 .....	233	14.2.4 数据仓库的数据获取 .....	255
二、填空题 .....	233	14.2.5 数据仓库的数据组织 .....	256
三、思考题 .....	233	14.2.6 联机分析处理 .....	257
<b>第 13 章 分布式数据库系统 .....</b>	<b>234</b>	14.3 数据挖掘 .....	258
13.1 概述 .....	234	14.3.1 数据挖掘的基本概念 .....	258

14.3.2 数据挖掘的任务 .....	259	15.7 SQL Server 2000 的事务 处理技术 .....	308
14.3.3 数据挖掘的对象、方法与 技术 .....	260	15.7.1 SQL Server 2000 中 事务的执行 .....	308
小结 .....	260	15.7.2 事务日志 .....	310
综合练习题十四 .....	261	15.7.3 SQL Server 2000 的锁机制 .....	310
一、选择题 .....	261	15.8 SQL Server 2000 的安全管理 .....	311
二、填空题 .....	261	15.8.1 身份验证模式 .....	312
三、思考题 .....	261	15.8.2 用户账户及管理 .....	313
<b>第 15 章 关系数据库管理系统—— SQL Server 2000.....</b>	<b>262</b>	15.8.3 用户权限的管理 .....	316
15.1 SQL Server 2000 简介 .....	262	小结 .....	318
15.2 SQL Server 2000 的安装配置 .....	263	综合练习题十五 .....	318
15.2.1 硬件和操作系统要求 .....	264	一、选择题 .....	318
15.2.2 安装过程 .....	264	二、填空题 .....	319
15.3 SQL Server 2000 的基本使用 .....	270	三、思考题 .....	319
15.3.1 SQL Server 提供的管理 工具及作用 .....	270	<b>第 16 章 数据库设计/建模工具的使用 .320</b>	
15.3.2 SQL Server 2000 数据库的 基本操作 .....	273	16.1 PowerDesigner .....	320
15.3.3 索引 .....	281	16.1.1 PowerDesigner 简介 .....	320
15.3.4 视图 .....	283	16.1.2 建立概念数据模型 .....	322
15.4 SQL Server 2000 的备份与恢复 .....	289	16.1.3 建立商业处理模型 .....	329
15.4.1 备份 .....	289	16.1.4 建立物理数据模型 .....	332
15.4.2 恢复 .....	290	16.2 ERwin .....	336
15.5 数据查询 .....	291	16.2.1 ERwin 简介 .....	336
15.6 SQL Server 2000 的完整性 约束 .....	296	16.2.2 ERwin 基本操作 .....	337
15.6.1 规则 .....	296	小结 .....	341
15.6.2 默认值 .....	299	综合练习题十六 .....	342
15.6.3 约束 .....	301	一、选择题 .....	342
15.6.4 存储过程 .....	302	二、填空题 .....	342
15.6.5 触发器 .....	305	三、思考题 .....	342
		<b>附录 参考答案 .....</b>	<b>343</b>
		<b>参考文献 .....</b>	<b>349</b>

# 第 1 章 数据库系统概述

20 世纪 60 年代末，数据库技术开始崭露头角，发展到现在，数据库技术已经成为计算机科学的重要分支并已经在各行各业得到广泛的应用，成为存储、使用、更新信息资源的主要手段。数据库技术已经能够成为数据处理的公用支撑技术，其广泛的实际应用已经产生了巨大的社会和经济效益。因此，了解、使用和研究数据库技术，不断地推广使用数据库，不断地利用现有数据库和开发数据库新技术为社会和经济发展服务，是每个计算机专业人员都必须做的工作。

## 1.1 数据库技术简介

### 1.1.1 数据库技术的应用

数据库是相互关联的数据的集合。即它不仅包含数据本身，还包含数据之间的联系。数据库中的数据具有一定的逻辑关系，能够表达确定的意义，及时反映现实生活的某个方面，数据库中的数据具有按照特定的格式存储，具有较小的冗余度，较高的数据独立性和扩展性，数据可为用户共享使用。

对数据库中的数据专门实施管理职责的软件称为数据库管理系统（Database Management System，简称 DBMS）。数据库管理系统一般在操作系统支持下运行。例如目前较流行的大型数据库管理系统包括 Oracle，Sybase，Informix，SQL Server 2000 等，微型机上运行的有 Access，Visual FoxPro 以及大型数据库系统的微机版本等。

数据库技术主要应用在一些数据密集型应用（Data Intensive Applications）的领域，如铁路定/售票系统、工资/人事管理系统、图书馆管理、企业生产/物资调配/销售管理、办公自动化系统、银行信息系统、地理信息系统、军事情报信息/指挥决策/军事调度，管理信息系统等。

这种数据密集型应用主要有以下的一些特点：

1) 涉及的数据量很大，数据一般需要存放在外存中，内存中只能暂时存储很小的一部分。

2) 数据不随程序的结束而消失，而必须长期保留在计算机系统中，如银行系统必须长久地保存储蓄用户的信息。

3) 数据要为多个应用程序所共享，或者要求在一个单位或更大范围内共享。

对这种大量、持久和共享数据的管理是一个很重要的问题，早期人们一般采用文件处理系统加以管理，即所需的数据被存储在多个不同的文件中，通过编写不同的应用程序来实现将记录从相应的文件中取出或将记录插入到相应的文件中。在数据库管理系统出现以前，人们一般只能采用这样的系统来存储信息。

随着时间的推移，这种利用文件处理系统来管理数据的方法已经越来越不适应数据处理的需要。采用文件系统处理数据有以下几个弊端：

1) 数据的冗余和不一致。由于文件和程序是很长一段时间内由不同的程序员创建的, 因此不同的文件可能采用不同的格式, 不同的程序可能采用不同的语言。此外, 相同的信息也可能在几个地方重复存储。这种冗余除了导致数据存储和访问开销的增大之外, 还可能导致数据的不一致, 即同一数据的不同副本不一致。

2) 数据访问困难。

3) 文件处理系统中应用程序对文件的过分依赖将使应用程序的维护量非常大。

4) 文件系统一般不支持对文件的并发访问。

5) 安全性问题突出。并非所有用户都可以访问所有数据, 在这方面, 文件系统难满足系统安全性要求。

文件系统的上述缺点使得它难以满足越来越高的数据处理的要求, 这也进一步促进了数据库系统的发展。数据库系统以统一管理和数据共享为主要特征。在数据库系统中, 数据不再仅仅服务于某个程序或者用户, 而是作为系统的共享资源, 由数据库管理系统来对数据进行统一管理。

### 1.1.2 数据库系统及其特点

数据库系统是指引入数据库技术后的计算机系统。其组成如图 1-1 所示, 狭义地讲, 数据库系统由数据库和数据库管理系统组成; 广义地讲, 数据库系统由数据库+数据库管理系统+数据库管理员+应用程序+用户组成。

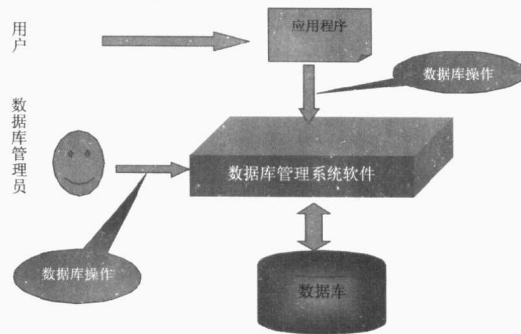


图 1-1 数据库系统的组成

数据库系统是一个有机结合的人机系统。数据库系统在整个计算机系统中的地位如图 1-2 所示。

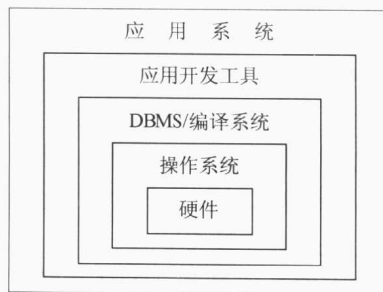


图 1-2 数据库在计算机系统中的地位

与人工管理和文件管理系统相比，数据库系统主要有以下一些特点：

1) 数据结构化。数据结构化是数据库与文件系统的根本区别。现在的数据库系统中的数据都必须按照某一特定的数据模型组织数据。例如在关系数据库系统里，数据库中的数据组织成二维表的形式，表由若干记录组成，每一记录又由若干属性项组成。每一记录都包含相同的属性项，但至少有一项属性值不相同，每一属性值必须满足所规定的完整性约束条件等。

2) 数据共享。数据库中的数据是可以被多个用户所共享的，这和以前的文件系统有所不同，对文件系统来说，不同的应用程序需要自己独有的文件，文件之间基本上是不能共享的，这从另一方面导致了数据的大量存储，即冗余度太大，浪费太多的存储空间。而数据库中的数据由于可以为多个用户所共享，冗余度小，花费很少的存储空间，不同的用户可以同时通过数据库管理系统对数据库中的数据进行存取。

另一方面，数据库中的数据面向数据库系统整体而不是单个的应用，因此数据库更容易增加新的应用，数据库的扩展性好，因而能够更进一步地满足用户的要求。

3) 数据具有较高的独立性。关系数据库的数据库管理系统通过映像的功能，保证应用程序对数据结构和存取方法具有较高的数据独立性。这使得数据的物理存储结构和用户看到的逻辑结构有很大的差别。用户在操作数据时，并不需要了解数据库中的数据是如何存储的，只需要以简单的逻辑结构来操作数据。这大大简化了用户的操作，使得对数据库了解很少的用户也能迅速地掌握和使用数据库系统。

4) 数据由 DBMS 统一管理和控制。数据库作为多个用户和应用程序的共享资源，为了保证用户的数据的并发存取是即时和有效、不出错的，数据库必须有数据库管理系统提供并发控制功能。同时，数据库的安全性和完整性也是数据库的重要内容，这些都由数据库管理系统来加以保证。

## 1.2 数据库管理系统

数据库管理系统 (DBMS) 是管理数据的软件，利用它实现数据库系统的各种功能，图 1-3 表示了数据库管理系统在数据库系统的位置。

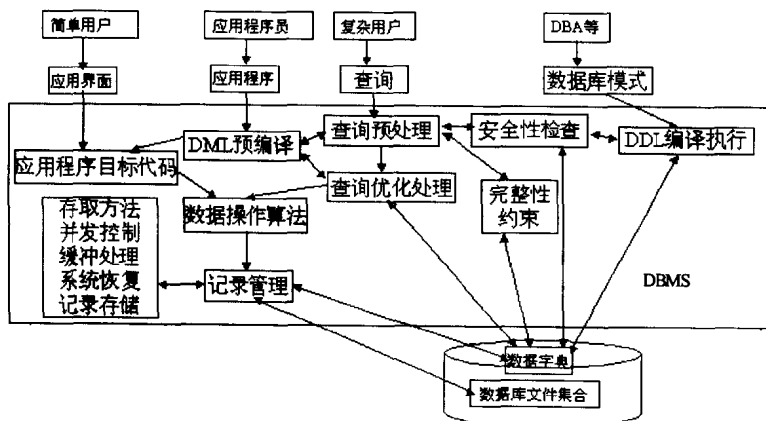


图 1-3 数据库管理系统结构

可以说,数据库管理系统是数据库系统的核心,是一种通用的软件。本节主要介绍数据库管理系统应具备的功能及当前数据库管理系统的主要体系结构。

### 1.2.1 DBMS 的主要功能

#### 1. 数据定义功能

提供数据的定义或者操作命令,如 FoxPro 系统中的 CREATE 命令和 MODIFY STRUCTURE 命令就属于数据定义功能。

#### 2. 数据操纵功能

数据库管理系统提供数据操纵语言 DML (Data Manipulation Language) 完成对数据库的查询、插入、追加、修改、删除等基本操作。不同的 DBMS 语言的语法格式也不相同,以其实现方法而言,一般可分为两种类型:一类 DML 可以独立使用,不依赖于其他程序设计语言,称为自主型语言;另一类称为宿主型 DML,嵌入到宿主语言中使用。如嵌入到 C 等程序设计语言中。在使用高级语言编写的应用程序中,如果需要调用数据库中的数据时,则要用宿主型 DML 语言来操纵数据。因此,DBMS 必须包含 DML 的编译或解释程序。

#### 3. 数据库的运行控制功能

数据库中的数据是可共享的宝贵资源,用户对数据的存取可能是并发的,当多个用户并发存取同一个数据库时,为了避免冲突和解决问题,DBMS 必须提供数据的完整性、安全性、并发控制以及发生故障之后的对数据库系统恢复等方面的功能。

数据库管理系统是数据库系统的一个重要组成部分。

从图 1-3 可以看出,数据库系统的用户一般有:

- 数据库管理员 (DBA)

数据库管理员 (Database Administrator, 简称 DBA) 的职责是监督管理其他用户对数据库和数据库管理系统软件的使用;负责系统安全和性能保障等。

- 系统分析员、系统设计员和应用程序员

这些用户主要的职责是根据用户需求设计数据库;与数据库管理员合作在数据库系统中建立数据库并加载数据;为最终用户编制应用程序,用以访问和管理数据。

- 最终用户

最终用户通过应用系统的用户接口使用数据库。

常用的接口方式有浏览器、菜单驱动、表格操作、图形显示、报表书写等,给用户提  
供简明直观的数据。

最终用户包括:

1) 偶然用户:访问数据库次数少,但每次访问数据库时往往需要不同的信息,需求不定,这类用户一般是企业或组织机构的高中级管理人员。

2) 简单用户:数据库的多数用户是简单用户。其主要工作是查询和修改数据库,一般都是通过应用程序员精心设计并具有友好界面的应用程序存取数据库。这类用户访问数据库次数多,需求稳定,多数业务人员。

3) 复杂用户:如工程师、科学家、经济学家、科学技术工作者等具有较高科学技术背景的人员。这类用户一般比较熟悉管理系统的各种功能,能够直接使用数据库语言访问

数据库,甚至能够基于数据库管理系统的 API 编制自己的应用程序,开发数据库等。

### 1.2.2 DBMS 的系统结构

数据库系统是数据密集型应用系统的核心。随着计算机硬件、软件的发展,计算机应用系统的支撑环境在不断地变化。为了适应这种变化,DBMS 的结构也在不断地演变。这里介绍 4 种主要的 DBMS 系统结构。

#### 1. 分时系统环境下的集中式数据库系统结构

数据库系统诞生于 20 世纪 60 年代中期,当时正是分时系统开始流行的时候,因此,早期的 DBMS 很自然地以分时系统作为 DBMS 首选的开发环境。从数据库的应用来看,数据是一个单位的共享资源,数据库系统的服务要面向全单位,而数据库的管理又要求集中,采用集中式数据库系统结构是当时的合理选择。在这种系统中,不但数据是集成的,数据的管理也是集中的。所有的程序,包括 DBMS,应用程序以及与用户终端进行通信的软件等都运行在一台宿主计算机上,所有的数据处理都是在宿主计算机中进行。宿主计算机一般是大型机、中型机或小型机。应用程序和 DBMS 之间通过操作系统管理的共享内存或应用任务区来进行通信,DBMS 利用操作系统提供的服务将数据写入磁盘,或从磁盘中读出数据。用户通过本地终端或远程终端来访问数据库。终端通常是非智能的,仅仅是人机交互的设备,不分担数据库系统的处理功能。具有这种系统结构的数据库系统现在还有很多在使用。目前大多数 RDBMS 也是从这种系统结构开始发展的。

集中式系统的主要优点是:具有集中的安全控制机制,以及处理大量数据和支持大量并发用户的能力。集中式系统的主要缺点是:购买和维持这样的系统一次性投资太大。

#### 2. 网络环境下的客户/服务器结构

20 世纪 70 年代后,由于微机的出现和迅速发展以及计算机网络的发展和广泛使用。出现了客户/服务器系统,客户/服务器是一种特殊的计算机,它们通过网络相连。客户机直接面向用户,接受并处理任务,并将任务中需要由服务器完成的部分委托给服务器完成。而服务器只接受客户机的委托,完成特定的任务。在客户/服务器系统中,处理是分布的。客户机与服务器功能划分的一般原则是:

- 1) 客户机提供多样化的用户接口、执行应用程序、对服务器提出服务请求等。
- 2) 服务器只完成客户机委托的公共服务。
- 3) 客户机与服务器的数据交换量应尽可能地少。
- 4) 消除瓶颈,提高全系统的功能。

这仅仅是原则,具体的划分与应用环境有关。因此,客户/服务器系统实际上是在微机局域网环境下,合理划分任务,进行分布式处理的一种应用系统结构,是解决微机大量使用而又无力承担所有处理任务这一矛盾的合理方案。

数据库是客户/服务器系统的一个重要应用领域。一般由客户机处理数据库的接口部分,如菜单管理、图形用户接口、嵌入式数据库语言的预处理、编辑、报表生成、二维数据库语言接口等。而 DBMS 的核心部分则由服务器来处理。甚至有些系统把查询处理和优化也放在客户机中,服务器只承担数据库的物理存取和事务管理。客户机在处理应用程序时,若遇访问数据库的要求,则以数据库语言(如 SQL)语句的形式提交给数据库服务器。

数据库服务器执行数据库语言语句，并返回结果给客户机。这与服务器仅仅向客户机传送原始文件，而不加任何处理是不同的。数据库服务器应在处理能力和资源配置上足以支持 DBMS，一般是大型机或小型机或高档微型机或并行处理的数据库机，这决定于数据库的规模。由于硬件价格的下降，用微处理器芯片构成的大规模并行的数据库成为数据库服务器的重要候选对象。

有些客户机需要经常访问或长期使用数据库中的部分数据，为了减少请求服务器服务的次数，常常将数据库中所用到的数据取到客户机的磁盘中，这叫取出 (CHCK OUT)；处理完后再送回 (CHECK IN)。客户机为了管理这些临时数据，可有自己的局部 DBMS。总之，客户机与服务器在功能上的划分决定于应用要求。

具有客户机/服务器系统结构的数据库系统虽然在处理上是分布的，但是数据却是集中的，还是属于集中式数据库系统。即使系统中有多个数据库服务器，各分管一部分数据，如果它们是彼此独立的，则在用户眼中，它们是多个集中的数据库。用户在访问数据库时，必须知道所访问的数据在哪个数据库服务器上。如果所访问的数据涉及多个数据库服务器，如对不同服务器上的两个关系进行联接，则客户机只能分别从两个数据库服务器上取来两个关系中的数据，由客户机中的应用程序进行连接处理。因而，这种多数据库服务器系统仍属于集中式数据库系统的范畴。

由于微机性能价格比迅速提高，微机取代一部分大型机或者小型机的功能，客户/服务器数据库结构已经得到了广泛的应用。

### 3. 物理上分布、逻辑上集中的分布式数据库

一个单位规模扩大和地理分散后，其计算机一般通过联网来获得信息需求，但集中式数据库很难满足上述类型单位的要求。在这种情况下，集中式数据库存在通信开销大、性能差、可用性差、可扩充性差、难以管理等缺点。但分布式数据库能克服这些缺点。

分布式数据库的研究开始于 20 世纪 70 年代后期，最先研究的是物理上分布、逻辑上集中的分布式数据库。其基本思想是：把单位的数据模式（全局数据模式）按数据的来源和用途，合理地分布在系统的多个结点上，使大部分数据可以就地或就近存取。数据在物理上分布后，由系统统一管理，使用户感觉不到数据的分布。用户看到的似乎不是一个分布式数据库，而是一个数据模式为全局模式的集中式数据库，因而这种分布式数据库在逻辑上仍然是集中的。

目前这项技术在逻辑上已经成熟，并有商品化的产品。但它仍然没有摆脱集中式数据库的缺点。尤其当单位很大且分散时，全局数据模式很难设计、管理、扩充和修改。各个结点也受制于全局数据模式，缺少自治性。而且在实际应用中，一个单位很少从零开始设计一个分布式数据库系统，更多是集成现有的数据库。如果采用物理上分布、逻辑上集中的分布式数据库系统，则必须把原有各数据库的数据模式统一为全局数据模式。这无异于重新设计一个数据库。这种分布式数据库的用途是有限的。

### 4. 物理上分布、逻辑上分布的分布式数据库结构

这种分布式数据库的特点是结点自治和没有全局数据模式，每个结点所看到的数据模式仅仅限于该结点所用到的数据。它一般由本结点的数据模式和供本结点共享的其他结点上有关的数据模式两部分组成。结点间的数据共享由双边协商确定。如果一个新的结点要加入系统，它开始时可用本结点的数据，然后与有关结点协调，共享其他结点的有关数



据；本结点的数据也可供其他结点共享。这种扩展性完全是渐进的，不致影响原有系统的运行。由于每个结点所看到的数据模式是不一样的，好像系统中有多个逻辑数据库，这种分布式数据库因而在逻辑上是分布的。由于无全局数据模式，一个结点的数据模式的修改甚至一个结点的加入或撤离，仅仅影响有关的结点。这完全不像物理上分布、逻辑上集中的分布式数据库那样，影响所有的结点。一个结点在给数据对象命名时，只要在本结点的数据模式内惟一就行了，不必考虑与其他无关数据对象重名的问题。每个结点好像拥有一个满足自己需要的集中式数据库一样，而不受制于全局数据模式，甚至不必有“全局概念”，结点具有很高的自治性。

### 1.3 数据抽象与数据独立性

#### 1.3.1 数据抽象

数据库一般都具有复杂的数据结构。为便于理解，往往只提供数据的抽象视图，隐藏数据的存储结构和存储方法等细节来对用户屏蔽复杂性。一般将数据库系统抽象成一个具有三级模式的结构，即数据库的物理结构、数据库的逻辑结构和数据库的视图结构。其中物理结构也称为内模式，数据库的逻辑结构称为模式，而数据库的视图结构主要面向用户，称为数据库系统的外模式，如图 1-4 所示。

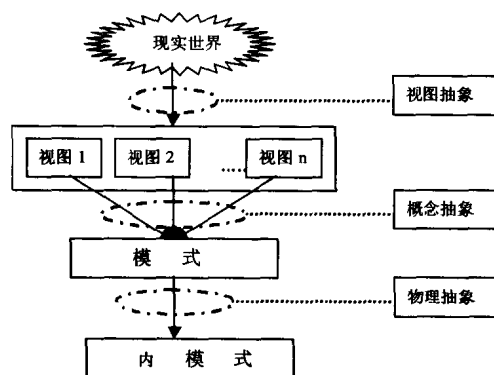


图 1-4 数据抽象的三个层次

1) 数据库的物理结构。数据库的物理结构是最低层次的抽象，主要用于描述数据到底是怎样存储的，即定义数据的存储方式和物理结构，一般也常被称为数据库的内模式、存储模式或者物理模式。

2) 数据库的逻辑抽象。数据库的逻辑抽象是比物理层稍高的抽象，也称数据库的逻辑模式或者概念模式。主要描述数据库中存储什么数据以及这些数据间存在什么关系，因而整个数据库通过少量相对简单的结构来描述。虽然简单的逻辑层结构的实现涉及到复杂的物理层结构，但逻辑层的用户一般不必知道这种复杂性，逻辑层抽象是由数据库管理员使用的，管理员必须确定数据库中应该保存哪些信息。

3) 数据库的视图结构。数据库的视图结构也称为数据库的外模式、子模式或用户模式。属于最高层次的抽象，视图结构一般只描述数据库的某个部分，反映的是用户关心的