



普通高等教育“十一五”国家级规划教材



高等学校计算机科学与技术教材

数据库技术与应用

COMPUTER Science and Technology

□ 冯凤娟 编著

- 原理与技术的完美结合
- 教学与科研的最新成果
- 语言精练，实例丰富
- 可操作性强，实用性突出



清华大学出版社

● 北京交通大学出版社

普通高等教育“十一五”国家级规划教材
高等学校计算机科学与技术教材

数据库技术与应用

冯凤娟 编著

清华大学出版社
北京交通大学出版社
· 北京 ·

内 容 简 介

本书详细介绍了数据库系统的组成、数据库系统的发展历程、关系模型、关系代数运算、关系数据库标准语言 SQL、规范化理论、数据库设计,以及数据库四种保护措施(安全性、完整性、并发控制和数据库恢复)的原理和实现技术。本书用一个应用系统(教学管理系统)贯穿全书,在各个章节中以该应用系统的实际问题为例,详细讲解各章节的原理、技术、方法和语法,以及开发应用系统的步骤和各步所要完成的工作。

同时以 Oracle11G 数据库管理系统为例,介绍了在 OEM 环境下如何管理数据库各对象,如何使用 Oracle 开发语言 PL/SQL 编写存储过程、存储函数、包、数据库触发器等程序结构,以及在 Oracle11G 系统中保证数据库安全性、完整性、并发控制、备份与恢复的具体命令和方法。

本书内容全面,深入浅出,概念清晰,条理清楚,案例丰富,不仅适合于教学,也适合于自学。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13501256678 13801310933

图书在版编目(CIP)数据

数据库技术与应用/冯凤娟编著. —北京:清华大学出版社;北京交通大学出版社,2011.3
(高等学校计算机科学与技术教材)

ISBN 978-7-5121-0523-2

I. ①数… II. ①冯… III. ①数据库系统—高等学校—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2011)第 031073 号

责任编辑:谭文芳 特邀编辑:宋林静

出版发行:清华大学出版社 邮编:100084 电话:010-62776969 <http://www.tup.com.cn>

北京交通大学出版社 邮编:100044 电话:010-51686414 <http://press.bjtu.edu.cn>

印刷者:北京瑞达方舟印务有限公司

经 销:全国新华书店

开 本:185×260 印张:28.5 字数:730 千字

版 次:2011 年 4 月第 1 版 2011 年 4 月第 1 次印刷

书 号:ISBN 978-7-5121-0523-2/TP·637

印 数:1~3 000 册 定价:45.00 元

本书如有质量问题,请向北京交通大学出版社质监组反映。对您的意见和批评,我们表示欢迎和感谢。

投诉电话:010-51686043, 51686008; 传真:010-62225406; E-mail: press@bjtu.edu.cn。

前 言

数据库系统是计算机科学的重要分支，由于数据库具有数据结构化、较低的冗余度、较高的程序与数据独立性、易于扩充和易于编制应用程序等优点，大型的信息系统都是建立在数据库之上的。数据库技术成为目前最活跃、应用最广泛的计算机领域之一，几乎所有的大型应用系统都涉及数据库，以数据库方式存储数据。

由于数据库应用的广泛性，数据库管理系统已发展为通用的系统软件，目前数据库管理系统有很多种，如 Oracle、Sybase、SQL Server、DB2 等都是优秀的大型数据库管理系统。

要想很好地使用这些数据库管理系统进行应用系统的开发，必须全面地掌握数据库系统的原理以及这些原理在某一数据库管理系统（如 Oracle）中的应用。数据库原理是数据库应用的理论基础，数据库应用是数据库原理的实践过程，两者缺一不可。

本书详细介绍了数据库系统的组成、数据库系统的发展历程、关系模型、关系代数运算、关系数据库标准语言 SQL、规范化理论、数据库设计，以及数据库四种保护措施（安全性、完整性、并发控制和数据库恢复）的原理和实现技术。本书一个应用系统（教学管理系统）贯穿全书，在各个章节中以该应用系统的实际问题为例，详细讲解各章节的原理、技术、方法和语法，以及开发应用系统的步骤和各步所要完成的工作。

同时以 Oracle 11G 数据库管理系统为例，介绍了在 OEM 环境下如何管理数据库各对象，如何使用 Oracle 开发语言 PL/SQL 编写存储过程、存储函数、包、数据库触发器等程序结构，以及在 Oracle 11G 系统中保证数据库安全性、完整性、并发控制、备份与恢复的具体命令和方法。

本书分 8 章，每章后面配有大量习题，帮助学生复习掌握课程内容。本书各章节组织如下。

第 1 章介绍数据库系统概述。介绍数据库系统的组成、数据库系统的发展历程、数据模型、数据库系统体系结构。

第 2 章介绍关系数据库。介绍关系模型的基本概念、关系代数的各种运算，并举例说明。

第 3 章介绍关系数据库标准语言 SQL。介绍了 SQL 语言的特点，各种 SQL 命令——数据查询命令、数据定义 DDL 命令、数据操纵 DML 命令、数据控制 DCL 命令的语法和功能，介绍 Oracle 11G 的各种函数和嵌入式 SQL 语言的语法，并举例说明。Oracle 11G 数据库的 SQL 语言是在标准 SQL 语言的基础上进行了扩充。

第 4 章介绍关系数据库规范化理论。介绍了函数依赖、函数依赖的逻辑蕴涵、Armstrong 公理、关键字求解理论和算法、关系模式的分解和关系模式的规范化过程，并举例说明。

第 5 章介绍数据库设计。介绍了开发数据库应用系统的步骤，以及每一步需要完成的工作，并举例说明。

第 6 章介绍数据库的保护。分别介绍了安全性、完整性、并发控制、数据库恢复这四种

数据库保护措施的原理和实现技术，同时详细地介绍了在 Oracle 11G 系统中这四种保护措施的具体实现方法和操作步骤，并举例说明。

第 7 章介绍 Oracle PL/SQL 高级应用。介绍了 Oracle 开发语言 PL/SQL 的组成和各种语法结构，以及如何使用 PL/SQL 开发和测试存储过程、存储函数、包、数据库触发器，并举例说明。

第 8 章介绍数据库系统的发展方向。介绍了分布式数据库系统和并行数据库系统的特点、体系结构和数据库管理系统的功能。

作者深入调查了目前许多高校讲授数据库课程的详细情况，同时参阅和借鉴了国内外许多优秀教材的内容和优点，结合教学过程中的体会和经验，对本书内容进行了合理的取舍。使得本书更加符合初学者学习数据库课程的认知规律，进一步体现概念讲解的条理性、逻辑性，使内容讲解循序渐进、深入浅出，易于读者学习和掌握。

本书提供了丰富的例题，并将所学内容和相关知识点组织到相应例题中，使读者能很容易地加深对知识点的理解和掌握。本书所有的例题和操作过程都是作者亲自实践过的，因此都是可以重复的，建议读者在阅读本书时，跟随本书内容进行具体的操作实践，这样会收到事半功倍的效果。本书提供了大量的习题，从不同的角度进一步帮助读者理解和掌握所学的知识点。另外，书后还附有综合练习和习题答案，帮助读者检验学习效果。

本书结合 Oracle 数据库的最新版本 Oracle 11G，讲解了 Oracle 11G 的高级应用知识，使读者很容易学会利用 Oracle 11G 的 OEM 环境进行数据库的管理工作，真正做到学以致用。

本书内容全面，深入浅出，概念清晰，条理清楚，案例丰富，不仅适合于教学，也适合于自学。如果作为数据库原理与应用课程的教材，建议总学时为 64 学时，其中主讲学时 48 学时。由于课程学时的限制，实验学时各学校可适当调整。另外，除了实验学时外，应该安排学生自由上机的时间，以加强学生的实际动手能力。

首先要感谢北京交通大学软件学院芦苇院长、陈旭东老师和赵宏老师，更要感谢北京交通大学出版社的谭文芳编辑，要是没有他们的坚持和鼓励，恐怕也就没有本书的出版。感谢为此书审稿的老师、所有帮助过我的同仁及我的家人。

由于时间仓促，加之作者水平和经验有限，书中难免有不足或错误之处，敬请专家和读者批评指正。

编 者

2011 年 1 月

目 录

第 1 章 数据库系统概述	1
1.1 数据、数据库与数据库管理系统的基本概念	1
1.1.1 数据与信息	1
1.1.2 数据处理与数据管理	2
1.1.3 数据库	2
1.1.4 数据库管理系统	2
1.1.5 数据库系统	4
1.2 数据管理技术的发展历程	5
1.2.1 人工管理阶段	5
1.2.2 文件管理阶段	5
1.2.3 数据库管理阶段	6
1.3 数据模型	9
1.3.1 数据模型的概念和种类	9
1.3.2 概念模型	14
1.3.3 逻辑数据模型和物理数据模型	15
1.4 数据库系统的内部体系结构	24
1.4.1 数据库系统的三级模式结构	24
1.4.2 数据库的两层映像	25
1.4.3 两级数据独立性	26
1.4.4 数据库系统的三级模式与两层映像的优点	26
1.4.5 数据库操作过程	27
1.5 数据库系统的外部体系结构	28
1.5.1 单用户结构的数据库系统	28
1.5.2 主从式结构的数据库系统	28
1.5.3 分布式结构的数据库系统	28
1.5.4 客户-服务器结构的数据库系统	29
1.5.5 浏览器-服务器结构的数据库系统	29
1.6 数据库系统的组成	30
1.6.1 硬件	30
1.6.2 软件	30
1.6.3 数据库	31
1.6.4 人员	31
总结	32

习题	33
第2章 关系数据库	37
2.1 关系模型的数据结构及其形式化定义	37
2.1.1 关系的形式化定义及其有关概念	37
2.1.2 关系模式	41
2.1.3 关系数据库与关系数据库模式	42
2.2 关系操作	42
2.2.1 基本的关系操作	42
2.2.2 关系数据语言	43
2.3 关系的完整性	43
2.3.1 实体完整性	43
2.3.2 参照完整性	44
2.3.3 用户定义完整性	45
2.4 关系代数	46
2.4.1 传统的集合运算	46
2.4.2 专门的关系运算	48
2.4.3 扩充的关系代数运算	52
2.4.4 关系代数表达式示例	53
总结	57
习题	58
第3章 关系数据库标准语言——SQL	61
3.1 SQL 概述	61
3.1.1 SQL 发展历程	61
3.1.2 SQL 的特点	62
3.1.3 SQL 的基本概念	62
3.1.4 SQL 语言的分类	63
3.2 了解 Oracle 11G	64
3.2.1 Oracle 11G 的主要组件	64
3.2.2 企业管理器 OEM	65
3.3 创建与使用 Oracle 数据库	70
3.3.1 Oracle 数据库的物理存储结构	70
3.3.2 创建用户数据库	71
3.3.3 Oracle 数据库的模式对象	75
3.4 创建与管理基本表	77
3.4.1 示例数据库表结构	77
3.4.2 创建基本表	78
3.4.3 修改基本表结构	87
3.4.4 删除、重命名和截断基本表	92
3.4.5 查看基本表	94

3.5 创建与管理视图	96
3.5.1 视图的建立	96
3.5.2 视图的删除	100
3.6 创建与管理索引	100
3.6.1 索引的分类	101
3.6.2 建立索引的原则	101
3.6.3 建立索引	102
3.6.4 索引的删除	104
3.6.5 查看索引	105
3.7 数据查询	106
3.7.1 SELECT 语句的语法格式	106
3.7.2 单表查询	107
3.7.3 联结查询	118
3.7.4 嵌套查询——子查询	124
3.7.5 集合查询	138
3.7.6 视图的查询	140
3.7.7 用 OEM 实现查询	142
3.8 数据操纵	144
3.8.1 基本表的数据操纵	144
3.8.2 视图的数据操纵	152
3.9 数据控制	155
3.9.1 权限和角色	155
3.9.2 系统权限与角色的授予和回收	156
3.9.3 对象权限与角色的授予和回收	157
3.10 Oracle SQL 函数	160
3.10.1 单行函数	160
3.10.2 聚组函数	168
3.10.3 函数的嵌套	169
3.11 嵌入式 SQL	170
3.11.1 SQL 语言的运行环境	170
3.11.2 嵌入式 SQL 的使用规定	171
总结	175
习题	176
第4章 关系数据库规范化理论	181
4.1 规范化问题的提出	181
4.1.1 规范化理论的主要内容	181
4.1.2 不合理的关系模式存在的异常问题	181
4.2 函数依赖	184
4.2.1 函数依赖的定义	184

4.2.2	函数依赖的逻辑蕴涵定义	186
4.2.3	函数依赖的推理规则	186
4.2.4	完全函数依赖与部分函数依赖	188
4.2.5	传递函数依赖	188
4.2.6	属性集的闭包及其算法	189
4.2.7	候选键的求解理论和算法	190
4.2.8	函数依赖推理规则的完备性	191
4.2.9	函数依赖集的等价、覆盖和最小函数依赖集	192
4.3	关系模式的分解*	193
4.3.1	模式分解的基本概念	194
4.3.2	无损联结分解	194
4.3.3	无损联结分解的测试算法	195
4.3.4	保持函数依赖的分解	198
4.4	关系模式的范式	198
4.4.1	第1范式	199
4.4.2	第2范式	200
4.4.3	第3范式	203
4.4.4	BC范式	204
4.4.5	关系模式规范化过程	207
总结		208
习题		209
第5章	数据库设计	213
5.1	数据库设计概述	213
5.1.1	数据库设计的任务	213
5.1.2	数据库设计的内容	213
5.1.3	数据库设计的特点	213
5.1.4	数据库设计的方法	214
5.1.5	数据库设计阶段	216
5.1.6	数据库设计过程的输入和输出	218
5.2	需求分析	219
5.2.1	需求分析的任务	219
5.2.2	需求分析的方法	220
5.2.3	评审	222
5.3	概念结构设计	223
5.3.1	概念模型的特点	223
5.3.2	概念结构设计方法	224
5.3.3	数据抽象与E-R模型设计	226
5.3.4	采用E-R方法的概念设计步骤	229
5.4	逻辑结构设计	238

5.4.1 关系模式的设计问题	238
5.4.2 逻辑结构设计的内容与步骤	239
5.4.3 E-R 图转换为关系模型	241
5.4.4 关系模式的规范化	242
5.4.5 关系模式的评价与优化	243
5.4.6 设计用户子模式	244
5.5 物理结构设计	245
5.5.1 设计物理结构	245
5.5.2 评价物理结构	247
5.6 数据库的实施和维护	247
5.6.1 数据库的实施	247
5.6.2 其他设计	249
5.6.3 数据库的运行与维护	250
总结	251
习题	251
第6章 数据库的保护	255
6.1 数据库的安全性管理	255
6.1.1 安全性管理的方法和原理	255
6.1.2 Oracle 系统的安全措施	259
6.2 数据库的完整性控制	281
6.2.1 完整性约束的概念	281
6.2.2 Oracle 中的完整性约束	289
6.2.3 OEM 工具管理约束条件	298
6.3 数据库的并发控制	301
6.3.1 并发控制的原理和方法	301
6.3.2 Oracle 系统中并发控制技术	314
6.4 数据库的恢复	319
6.4.1 数据库恢复的原理、方法和策略	319
6.4.2 Oracle 系统的备份与恢复技术	325
总结	344
习题	344
第7章 Oracle PL/SQL 高级应用	348
7.1 PL/SQL 程序设计	348
7.1.1 PL/SQL 的特点	349
7.1.2 PL/SQL 程序结构	349
7.1.3 PL/SQL 的流程控制结构	358
7.1.4 出错处理	366
7.1.5 游标	370
7.2 存储过程和函数	382

7.2.1 存储过程和函数与应用程序的区别	382
7.2.2 开发存储过程和函数	383
7.2.3 存储过程和函数的管理	389
7.2.4 存储过程和函数的调用和测试	391
7.2.5 存储过程和函数的安全性	394
7.2.6 存储过程和函数的优点	395
7.3 包	395
7.3.1 包的组成	396
7.3.2 开发包	396
7.3.3 包的管理	400
7.3.4 包的调用	401
7.3.5 系统包	402
7.3.6 包的优点	402
7.4 触发器	403
7.4.1 触发器的基本概念和类型	403
7.4.2 DML 触发器	405
7.4.3 数据库事件触发器和 DDL 触发器	420
7.4.4 INSTEAD OF 触发器（替代触发器）	424
总结	426
习题	426
第8章 数据库的发展方向	429
8.1 分布式数据库系统	429
8.1.1 分布式数据库系统的定义	429
8.1.2 分布式数据库系统的特点	429
8.1.3 分布式数据库的系统结构	431
8.1.4 分布式数据库管理系统	432
8.2 并行数据库系统	434
8.2.1 并行数据库系统的概述	434
8.2.2 并行数据库系统的功能	435
8.2.3 并行数据库的体系结构	435
8.2.4 并行数据库的并行处理技术	436
总结	438
习题	438
附录 A Oracle 11G 综合练习题	439
附录 B 部分习题参考答案	444
参考文献	446

第 1 章 数据库系统概述

数据库是数据管理的最新技术，是计算机科学的重要分支。目前，信息资源已成为各个部门的重要财富和资源，建立一个满足各级、各部门信息处理要求的信息系统已成为一个企业生长和发展的重要条件。数据库的建设规模、数据库信息量的大小和使用频度已成为衡量国家信息化程度的重要标志。

数据库（DataBase）简单地说就是数据的仓库，即数据存放的地方。人们周围有许多数据库的例子，如通信录就是一个小型的数据库，而图书馆则是一个典型的大型数据库。小型数据库尚可用手工管理，而大型数据库必须由计算机进行管理。在计算机 3 大主要应用领域（科学计算、过程控制、数据处理）中，数据处理所占比例约为 70%。数据库技术作为数据处理的最新技术应运而生。

本章先介绍数据库中的主要概念，然后回顾数据管理技术的发展历程，接着介绍数据模型的种类和组成以及数据库的体系结构，最后介绍了数据库系统的几个组成部分。使读者对数据库技术的概况有所了解。

1.1 数据、数据库与数据库管理系统的基本概念

数据、数据库与数据库管理系统是与数据库技术密切相关的几个基本概念。

1.1.1 数据与信息

数据（data）是数据库中存储的基本对象，是指用符号记录下来的可以区别的信息；在计算机中可表示数据的种类很多，文字、图形、图像、声音都是数据。为了了解世界，交流信息，人们在计算机中需要描述、存储和处理这些表现形式和内容复杂的数据。可以对数据作如下定义：描述事物的符号记录称为数据。描写事物的符号可以是数字，也可以是文字、图形、图像、声音等，即有多种表现形式，但它们都是经过数字化后存入计算机的。

信息是关于现实世界中事物的存在方式或运动形态的反映，是人们进行各种活动所需的知识。在信息社会中，信息是一种资源，它与能源、材料一起构成客观世界的 3 大要素。信息是客观存在的，例如，对于学生的基本情况来说，学生的学号、姓名、性别、年龄、系号等都是学生的具体信息，是该学生当前状态的反映。

信息与数据之间存在着固有的联系；数据是信息的载体或是信息的符号表示，信息是被人们消化的数据，是数据的内涵，是对数据的语义解释。数据表示了信息，而信息只有通过数据形式表示出来才能被人们理解和接受。尽管两者在概念上不尽相同，但通常人们并不严格区分它们。

1.1.2 数据处理与数据管理

数据处理指将数据转换成信息的过程，广义地讲，数据处理包括对数据的收集、存储、传播、检索、分类、加工或计算、打印各类报表或输出各种需要的图形等一系列活动；其目的是从大量的原始数据中抽取和推导出有价值的信息。狭义地讲，数据处理是指对所输入的数据进行加工整理。

数据与数据处理之间的关系可以表示为：

$$\text{信息} = \text{数据} + \text{数据处理}$$

数据是原料，是输入；而信息是产出，是输出结果。

在数据处理的一系列活动中，数据收集、整理、组织、存储、维护、传播、检索和分类等操作是数据处理业务中必不可少的基本环节，这些基本环节统称为数据管理。对数据管理部分，应研制出一个通用、高效而又使用方便的管理软件（数据库管理系统），把数据有效地管理起来，并最大限度地减轻程序员的负担。数据处理是与数据管理相联系的，数据管理技术的优势将直接影响数据处理的效率。数据库技术正是为了实现这一目标而发展起来的专门技术。

1.1.3 数据库

数据库（Data Base, DB），可以直观地理解为存放数据的仓库，只不过这个仓库是在计算机的大容量存储器上，如硬盘就是一类最常见的计算机大容量存储设备。数据必须按一定的格式存放，因为它不仅需要存放，还要便于查找。所以数据库是长期存储在计算机内有组织、可共享的数据集合。数据库中的数据具有较小的数据冗余，可供多个用户共享，具有较高的数据独立性和扩展性。在数据库管理系统的支持下具有安全性控制机制，能够保证数据的安全、可靠、允许并发地使用数据库，能有效、及时地处理数据，并能保证数据的一致性和完整性。

概括地讲，数据库数据具有永久存储、有组织和可共享 3 个基本特点。

图书馆可视为一个数据库，书架对应表格，书籍对应原始数据，书名或作者名的目录卡片对应索引。

人们总是尽可能地收集各种各样的数据，然后对它们进行加工，目的是要从这些数据中得到有用的信息。在社会迅速发展的今天，人们接触的事物越来越多，反映这些事物的数据量也急剧增加，过去人们手工管理和处理数据，现在借助计算机存储和管理复杂的大量的数据，这样就可以方便而充分地利用这些宝贵的数据资源，数据库技术正是由于这一需求驱动而发展起来的一种计算机软件。

1.1.4 数据库管理系统

数据库管理系统（DataBase Management System, DBMS）作为数据库系统中管理数据的软件系统，数据库管理系统是数据库系统的核心组成部分。它可以科学地组织和存储数据，高效地获取和维护数据。对数据库的一切操作，包括定义、查询、更新及各种控制，都是通过数据库管理系统进行的。数据库管理系统总是基于某种数据模型，可以把数据库管理系统看成是某种数据模型在计算机系统上的具体实现。根据数据模型的不同，数据库管理系统可以分成层次型、网状型、关系型、面向对象型等。

数据库管理系统的主要目标是使数据成为方便用户使用的资源，易于为各类用户所共享，并增加数据的安全性、完整性和可用性。

数据库管理系统是位于用户与操作系统之间的一层数据管理软件。数据库管理系统与操作系统一样是计算机的基础软件，也是一个大型复杂的软件系统。它主要包括下列功能。

1. 数据定义功能

数据定义功能提供数据定义语言（Data Definition Language, DDL）或操作命令，用于对各级数据模式进行精确的描述。用户通过它可以方便地对数据库中的相关对象进行定义，如对数据库基本表、视图和索引等进行定义。

2. 数据操纵功能

数据操纵语言（Data Manipulation Language, DML），实现对数据库的基本操作，如对数据库中数据的查询、插入、删除和修改。

数据操纵语言可以分为两类：一类是 DML 可以独立交互式使用，不依赖于任何程序设计语言，称为自含或自主型语言。另一类是宿主型 DML，嵌入到宿主语言中使用。如嵌入到 C 程序设计语言中，在使用高级语言编写的应用程序中，需要调用数据库中的数据时，则要用宿主型 DML 语句来操纵数据。

3. 数据库的事务管理与运行

数据库的事务管理与运行控制功能是数据库管理的核心部分，数据库管理系统必须提供以下几方面的数据控制功能：并发控制功能、安全性控制功能、完整性控制功能、数据库的恢复功能、数据库的内部维护功能（如索引、数据字典的自动维护）等。

① 并发控制功能：是处理多个用户同时使用某些数据时可能产生的问题。数据库中的数据是提供给多个用户共享的，用户对数据的存取可能是并发的，即多个用户同时使用同一个数据。

② 安全性检查功能：是保护数据以防止不合法的使用造成数据的泄密和破坏，每个用户只能按规定对某些数据以某种方式进行使用和处理。

③ 完整性控制功能：是指保证数据的正确性、有效性和相容性，将数据控制在有效的范围内，并保证数据之间满足一定的关系。

④ 数据库恢复功能：是指在硬件故障、软件故障、操作员失误以及故意破坏而造成数据库部分数据或全部数据丢失时，将数据库从错误状态恢复到某一已知的正确状态的功能。

所有数据库的操作都要在这些控制程序的统一管理下进行，以保证数据安全性、完整性和多个用户对数据库的并发操作。

4. 数据库的建立和维护功能

数据库的建立和维护功能包括：数据库初始数据的输入、转换功能，数据库的转储、恢复功能，数据库的重新组织功能和性能监测、分析功能等。这些功能通常是由一些实用程序完成的，包括数据装载程序、备份程序、性能调整程序、文件重组程序等，它是数据库管理系统的一个重要组成部分。

5. 数据库的组织、存储和管理功能

DBMS 要分类组织、存储和管理各种数据的存取路径，包括数据字典的系统数据、用户数据。数据库的组织、存储和管理功能提供了数据库中数据和应用程序的一个界面，与磁盘

中数据打交道的是操作系统的文件子系统，该功能是把各种 DML 语句转换成底层的文件系统命令，起到数据的存储、检索和更新的作用。数据库管理系统的数据组织和存储管理的目标是提高空间利用率和方便存取，提供多种存取方法（如索引查找、Hash 查找、顺序查找）来提高存取效率。

6. 数据字典

数据字典（Data Dictionary，DD）是存放数据库各级模式结构的定义，即对数据库结构的描述。这些数据是数据库系统中有关结构的定义信息，称为元数据。数据字典提供了对数据库数据描述的集中管理手段，对数据库的使用和操作都通过数据字典来进行。

在大型系统中，数据字典也可以单独成为一个系统。如 SQL Server 中的系统数据库就是系统的数据字典。

7. 数据通信功能

数据通信功能包括与网络中其他软件系统的通信功能，与操作系统（Operation System，OS）的联机处理、分时处理和远程作业传输的相应接口等，这一功能对分布式数据库系统尤为重要。

8. 其他功能

其他功能包括一个数据库管理系统与另一个数据库管理系统或文件系统的数据库转换功能；异构数据库之间的互访和互操作功能。

1.1.5 数据库系统

数据库系统（DataBase System，DBS）通常是指具有管理和控制数据库功能的计算机应用系统。DBS 是实现有组织地、动态地存储大量关联数据，方便多用户访问的计算机软件、硬件和数据资源组成的系统，即采用了数据库技术的计算机系统。因此，数据库系统一般由数据库、硬件、软件、人员等几部分组成，其中软件又包括操作系统、数据库管理系统、开发工具和开发语言、应用系统等软件，人员包括数据库管理员（DataBase Administrator，DBA）、系统分析员、应用程序员、数据库用户等各类人员。图 1-1 显示数据库系统的组成。图 1-2 显示了数据库系统在整个计算机系统中的地位。

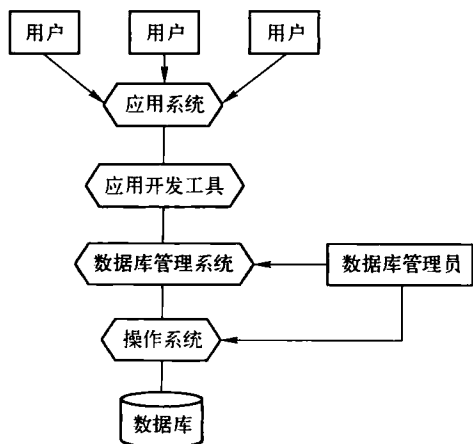


图 1-1 数据库系统的组成

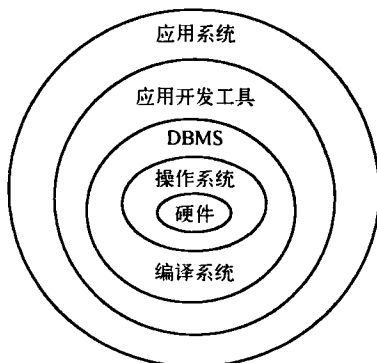


图 1-2 数据库系统在计算机系统中的地位

1.2 数据管理技术的发展历程

如前所述, 数据处理的中心问题是数据管理。数据管理技术的发展与硬件、软件、计算机应用的范围有密切联系, 数据管理大致经历了3个阶段: 人工管理阶段、文件管理阶段和数据库管理阶段。

1.2.1 人工管理阶段

20世纪50年代中期以前, 计算机主要用于科学计算。硬件方面, 外存储器只有卡片、纸带、磁带, 没有像磁盘这样可以随机访问、直接存取的外部存储设备。软件方面, 没有操作系统和专门管理数据的软件, 数据由计算机或处理它的程序自行携带。数据处理方式基本是批处理。人工管理数据有以下几个特点。

(1) 数据和程序不具有独立性

一组数据对应一组程序, 这就使得程序依赖于数据, 如果数据的类型、格式或者数据量、存取方法、输入/输出方式等改变了, 程序必须作相应的修改。

(2) 数据不能共享

由于数据是面向应用程序的, 在一个程序中定义的数据, 无法被其他程序利用, 因此, 程序与程序之间存在大量的重复数据。

(3) 系统中没有对数据进行管理的软件, 应用程序管理数据

数据管理任务包括物理结构、逻辑结构、存储结构、存取方法、输入/输出方式等完全由程序设计人员自负其责, 这就给应用程序设计人员增加了很大的负担。

(4) 数据不保存

当时计算机主要用于科学计算, 一般不需要将数据长期保存, 只是在计算某一课题时将数据输入, 运行结束后将结果输出, 这些数据也就没用了。

人工管理阶段, 应用程序与数据之间的一一对应关系如图1-3所示。

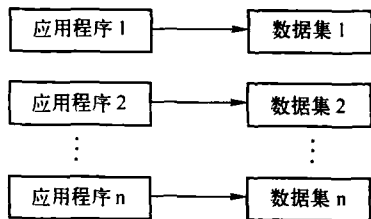


图1-3 人工管理阶段应用程序与数据之间的对应关系

1.2.2 文件管理阶段

20世纪50年代后期至60年代后期, 计算机应用范围逐步扩大, 不仅用于科学计算, 还大量用于信息管理。随着数据的增加, 数据的存储、检索和维护成为紧迫的需求。此时, 在硬件方面, 已经有了磁盘、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, 操作系统中有了专门管理数据的软件, 一般称为文件系统; 数据处理方式上不仅有批处理, 而且能够联机实时处理。文件系统管理数据具有下列特点。

(1) 由文件系统管理数据

由专门的软件即文件系统进行数据管理, 文件系统把数据组织成内部有一定结构的记录并以相互独立的数据文件的形式存储在存储设备上, 利用“按文件名访问, 按记录进行存取”的管理技术, 可以对文件进行修改、插入、删除等操作。程序与数据之间由文件系统

提供存取方法的转换，使应用程序与数据之间有一定的独立性。

(2) 数据可以长期保存

数据以文件为组织方式，长期保存在计算机的存储设备上，应用程序可以反复进行查询、修改、插入、删除等操作。

(3) 程序与数据有了一定的独立性

程序与数据分开存储，有了程序文件和数据文件的区别。数据的存取以记录为基本单

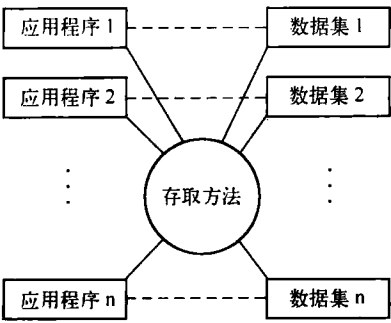


图 1-4 文件系统阶段应用程序与数据之间的对应关系

位，并进行了多种文件组织形式，如顺序文件、索引文件、随机文件等。在文件系统的支持下，数据的逻辑结构与物理结构之间可以有一定的差别，逻辑结构与物理结构之间的转换由文件系统的存取方法来实现。数据与程序之间有一定的独立性，程序只需用文件名访问数据，不必关心数据的物理位置。当改变保存文件的存储设备时，不必改变应用程序。这样，程序员可以集中精力在数据处理的算法上，而不必考虑数据存储的具体细节。

文件系统阶段，应用程序与数据之间的一一对应关系如图 1-4 所示。

与人工管理阶段相比，文件系统阶段对数据的管理有了很大的进步，但仍有以下缺点。

(1) 数据共享性差、冗余度大

数据冗余是指不必要的重复存储，同一数据项重复出现在多个文件中。一个文件基本上对应一个应用程序，不同的应用程序有相同的数据时，也必须建立各自的文件，因此数据不能共享；产生冗余，浪费存储空间。同时由于重复存储，各自管理，造成数据的不一致性，给数据的修改和维护带来困难。

(2) 数据独立性差

虽然程序与数据之间有一定的独立性，但这种独立性主要体现在存储设备的独立性上。文件系统中的数据文件是为某一特定的应用而设计的，数据与程序相互依赖，如果改变数据的逻辑结构或文件的组织方法，必须修改相应的应用程序，反之也是。系统不容易扩充。文件系统仍然是一个不具有弹性的、无结构的数据集合，文件之间是孤立的，不能反映现实世界之间的内在联系。

1.2.3 数据库管理阶段

从 20 世纪 60 年代后期开始，需要计算机管理的数据量急剧增长，并且对数据共享的需求日益增强。这时硬件已有大容量的硬盘，硬盘价格下降，软件价格上升，使得开发和维护系统软件的成本相对增加。在处理方式上联机实时处理要求更多，并提出分布式处理。在这种背景下，以文件系统作为数据管理手段已经不能满足应用的需求，于是为了解决数据的独立性问题，实现数据的统一管理，达到多应用程序共享数据的目的，数据库技术应运而生，出现了统一管理数据的专门软件系统，即数据库管理系统。

数据管理进入数据库管理阶段的标志是 20 世纪 60 年代末发生的三件大事。

1968 年，美国 IBM 公司推出层次模型的 IMS (Information Management System，信息管理