



“十二五”普通高等教育本科国家级规划教材
普通高等学校计算机教育“十二五”规划教材

数据库原理及 应用教程

(第3版)

*DATABASE PRINCIPLES AND
APPLICATIONS
(3rd edition)*

陈志泊 ◆ 主编
王春玲 许福 范春梅 ◆ 编著



人民邮电出版社
POSTS & TELECOM PRESS



“十二五”普通高等教育本科国家级规划教材
普通高等学校计算机教育“十二五”规划教材

数据库原理及 应用教程

(第3版)

*DATABASE PRINCIPLES AND
APPLICATIONS
(3rd edition)*

陈志泊 ◆ 主编
王春玲 许福 范春梅 ◆ 编著

人 民 邮 电 出 版 社
北 京

图书在版编目 (C I P) 数据

数据库原理及应用教程 / 陈志泊主编. -- 3版. --
北京: 人民邮电出版社, 2014. 2

“十二五”普通高等教育本科国家级规划教材 普通
高等学校计算机教育“十二五”规划教材
ISBN 978-7-115-33927-0

I. ①数… II. ①陈… 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(2014)第004347号

内 容 提 要

本书是第2版的修订版。在修订过程中,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,并进一步吸取了第2版教材的优点和教学过程中的体会和经验,全面系统地讲述了数据库技术的基本原理和应用,内容取舍合理,重点突出,符合教学和读者认识规律。

全书共七章,主要内容包括:数据库系统概述、关系数据库、关系数据库标准语言SQL、关系数据库理论、数据库安全保护、数据库设计和SQL Server 2008高级应用。本书除介绍数据库技术的基本原理外,还以SQL Server 2008为背景介绍了数据库技术的实现,包括数据库和数据表的维护、查询与统计、视图管理、存储过程和触发器的管理、用户管理、约束和默认管理、数据库的备份和还原、Transact-SQL程序设计等内容,使读者可以充分利用SQL Server 2008平台深刻理解数据库技术的原理,达到理论和实践的紧密结合。

本书内容循序渐进、深入浅出、概念清晰、条理性强,每一章节都给出了大量的实例,并进行解释说明。同时,每章后面都附有大量的习题,包括选择题、填空题、简答题等,从各种不同的侧面进一步帮助读者对所学知识的锻炼和掌握;另外,书后还附有习题答案,帮助读者检验学习和练习效果。

为方便读者学习和教师讲课,本书还提供了PowerPoint电子讲稿,读者可到人民邮电出版社网站(www.ptpress.com.cn)下载。

本书可作为高等院校计算机及相关专业的本、专科教材,也可供从事计算机软件工作的科技人员、工程技术人员以及其他有关人员参阅。

-
- ◆ 主 编 陈志泊
 - ◆ 编 著 王春玲 许 福 范春梅
责任编辑 滑 玉
责任印制 彭志环 杨林杰
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京天宇星印刷厂印刷
 - ◆ 开本: 787×1092 1/16
印张: 17.75 2014年2月第3版
字数: 465千字 2014年2月北京第1次印刷
-

定价: 39.80 元

读者服务热线: (010)81055256 印装质量热线: (010)81055316
反盗版热线: (010)81055315

编者的话 (修订版)

本书自第1版出版以来,受到了广大读者和计算机爱好者的热烈欢迎,并于2005年被评为北京市高等教育精品教材,2006年列为“十一·五”国家规划教材建设行列,2008年进行了第2版的修订工作,2012年被评为第一批“十二·五”普通高等教育本科国家级规划教材。但是,随着数据库技术的飞速发展,新技术、新知识层出不穷,一本教材的内容不可能一成不变,也不可能包罗万象。同时,在本书第1、2版的使用过程中,很多读者也给我们提出了很多有益的建议和意见,在这种情况下,我们经过认真分析、讨论,将第1、2版中的部分内容进行了合理的取舍,使之更加完善。

本书是对原来第2版的进一步修订。在修订过程中,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,并进一步吸取了第1版教材的优点和教学过程中的体会和经验,与上一版相比,本书具有以下几方面的特点:

(1)更加符合初学者学习数据库课程的认识规律,进一步提高了概念讲解准确性、逻辑性,内容讲解循序渐进,深入浅出,易于读者学习和掌握。

(2)进一步完善和扩充了关系数据库系统的模式结构、数据模型的有关概念和理论,使概念更加清晰,条理性更强。

(3)进一步完善和扩充了关系数据库的基本理论知识,主要包括函数依赖、函数依赖的逻辑蕴涵、Armstrong 公理、关键字的求解理论与算法、关系模式分解及其有关算法、查询优化等。

(4)进一步丰富了例题,将所学内容和相关知识点组织到相应例题中,使读者能很容易地进一步加深对知识点学习的理解和掌握。

(5)进一步丰富了习题的类型和数量,除了简答题之外,还增加了选择题、填空题等,从各种不同的侧面进一步帮助读者对所学知识的锻炼和掌握;另外,书后还附有习题答案,帮助读者检验学习和练习效果。

(6)采用了较新的 SQL Server 2008 管理库管理系统平台,并扩充了有关的高级应用知识,使读者很容易学会利用 SQL Server 2008 环境进行数据库的管理工作,真正做到学以致用。

全书内容相互衔接,成为一个逻辑整体。每章除提供大量的实例进行讲解外,还附有一定数量的习题,便于读者练习。为方便读者学习和教师讲课,本书还提供了 PowerPoint 电子讲稿,读者可到人民邮电出版社的网站(www.ptpedu.com.cn)下载。

本书内容全面,深入浅出,概念清晰,条理清楚,不仅适合于教学,也适合读者自学。如果作为教材,建议总学时为 60 学时,其中主讲学时 40 学时。由于课程学时的限制,实验学时各学校可适当调整,一般为 20 学时左右;另外,除实验学时外,最好安排学生自由上机的时间,加强学生的实际动手能力。

本书由陈志泊担任主编，全书内容由陈志泊修改定稿，全书由陈志泊、王春玲、许福、范春梅编写，其中，第1、4章由陈志泊编写，第5、7章主要由王春玲编写，第3、6章主要由许福老师编写，范春梅负责第2章的编写以及习题的提供与验证。另外，黄金燕、张惠萍、王鹏等参加了文字编辑、Powerpoint讲稿制作工作。

由于时间仓促，加之作者水平有限，书中不当之处在所难免，敬请广大读者批评指正。

编 者

2014年1月

目 录

第 1 章 数据库系统概述	1
1.1 信息、数据、数据处理与数据管理	1
1.1.1 数据与信息	1
1.1.2 数据处理与数据管理	2
1.2 数据库技术的产生、发展	3
1.3 数据库系统的组成	8
1.4 数据库系统的内部体系结构	10
1.4.1 数据库系统的三级模式结构	10
1.4.2 数据库系统的二级映像与数据独立性	12
1.4.3 数据库系统的三级模式与二级映像的优点	12
1.5 数据库系统的外部体系结构	13
1.6 数据库管理系统	15
1.6.1 DBMS 的主要功能	15
1.6.2 DBMS 的组成	16
1.6.3 DBMS 的数据存取的过程	17
1.7 数据模型	17
1.7.1 数据模型的概念及分类	17
1.7.2 数据模型的组成要素	18
1.8 三个世界及其有关概念	19
1.8.1 现实世界	19
1.8.2 信息世界	19
1.8.3 计算机世界	21
1.8.4 概念模型的 E-R 表示方法	21
1.9 四种数据模型	23
1.9.1 层次模型	23
1.9.2 网状模型	25
1.9.3 关系模型	26
1.9.4 面向对象模型	29
1.10 数据库系统的发展	30
1.10.1 第一代数据库系统	30

1.10.2 第二代数据库系统	30
1.10.3 第三代数据库系统	32
1.11 数据库技术与其他相关技术的结合	32
1.11.1 分布式数据库	33
1.11.2 数据仓库与数据挖掘技术	34
1.11.3 多媒体数据库	37
1.11.4 数据库技术的研究领域	38
1.12 小结	38
习题	39

第 2 章 关系数据库	42
2.1 关系模型的数据结构及其形式化定义	42
2.1.1 关系的形式化定义及其有关概念	42
2.1.2 关系的性质	44
2.1.3 关系模式	46
2.1.4 关系数据库与关系数据库模式	46
2.2 关系的码与关系的完整性	47
2.2.1 候选码与主码	47
2.2.2 外码	48
2.2.3 关系的完整性	48
2.3 关系代数	49
2.3.1 关系代数的分类及其运算符	50
2.3.2 传统的集合运算	50
2.3.3 专门的关系运算	52
2.4 关系演算*	56
2.4.1 元组关系演算语言	56
2.4.2 域关系演算语言 QBE	61
2.5 小结	65
习题	65

第 3 章 关系数据库标准语言——SQL	68
3.1 SQL 的基本概念与特点	68
3.1.1 SQL 的发展及标准化	68

3.1.2 SQL 的基本概念	69	3.7.3 删除数据	116
3.1.3 SQL 的主要特点	69	3.8 视图	117
3.2 SQL Server 2008 R2 简介	70	3.8.1 创建视图	117
3.2.1 SQL Server 的发展与版本	70	3.8.2 修改视图	119
3.2.2 SQL Server 2008 R2 的主要组件	71	3.8.3 删除视图	120
3.2.3 Management Studio	71	3.8.4 查询视图	120
3.3 创建与使用数据库	72	3.8.5 更新视图	121
3.3.1 数据库的结构	73	3.9 数据控制	121
3.3.2 SQL Server 的系统数据库	73	3.9.1 权限与角色	121
3.3.3 SQL Server 的示例数据库	74	3.9.2 系统权限与角色的授予与收回	122
3.3.4 创建用户数据库	74	3.9.3 对象权限与角色的授予与收回	122
3.3.5 修改用户数据库	77	3.10 小结	123
3.3.6 删除用户数据库	80	习题	124
3.3.7 查看数据库信息	80		
3.4 创建与使用数据表	81	第4章 关系数据库理论	128
3.4.1 数据类型	81	4.1 规范化问题的提出	128
3.4.2 创建数据表	83	4.1.1 规范化理论的主要内容	128
3.4.3 定义数据表的约束	85	4.1.2 不合理的关系模式存在的存储异常 问题	128
3.4.4 修改数据表	88	4.2 函数依赖	131
3.4.5 删除基本表	89	4.2.1 函数依赖的定义	131
3.4.6 查看数据表	90	4.2.2 函数依赖的逻辑蕴涵定义	132
3.5 创建与使用索引	91	4.2.3 函数依赖的推理规则及正确性	133
3.5.1 索引概述	91	4.2.4 完全函数依赖与部分函数依赖	134
3.5.2 索引的类型	92	4.2.5 传递函数依赖	136
3.5.3 索引的设计	93	4.2.6 属性集的闭包及其算法	135
3.5.4 索引的使用	94	4.2.7 候选键的求解理论和算法	136
3.5.5 使用数据库引擎优化顾问	96	4.2.8 函数依赖推理规则的完备性	137
3.6 数据查询	99	4.2.9 函数依赖集的等价、覆盖和最小函 数依赖集	138
3.6.1 SELECT 命令的格式与基本使用	99	4.3 关系模式的分解*	140
3.6.2 条件查询	101	4.3.1 模式分解问题	141
3.6.3 常用库函数及统计汇总查询	103	4.3.2 无损连接分解	141
3.6.4 分组查询	104	4.3.3 无损分解的测试算法	142
3.6.5 查询的排序	105	4.3.4 保持函数依赖的分解	142
3.6.6 数据表连接及连接查询	106	4.4 关系模式的范式	145
3.6.7 子查询	108	4.4.1 第一范式	145
3.6.8 合并查询	112	4.4.2 第二范式	146
3.6.9 存储查询结果到表中	112	4.4.3 第三范式	148
3.7 数据操纵	113	4.4.4 BC 范式	152
3.7.1 修改数据	113		
3.7.2 添加数据	115		

4.4.5 多值依赖与第四范式	155	6.1.2 数据库设计方法简述	201
4.5 关系模式的规范化	158	6.1.3 数据库设计的步骤	202
4.5.1 关系模式规范化的目的和原则	158	6.2 系统需求分析	204
4.5.2 关系模式规范化的步骤	159	6.2.1 需求分析的任务	204
4.5.3 关系模式规范化的要求	159	6.2.2 需求分析的方法	205
4.6 小结	161	6.3 概念结构设计	207
习题	162	6.3.1 概念结构设计的必要性	207
第 5 章 数据库安全保护	166	6.3.2 概念模型的特点	208
5.1 数据库的安全性	166	6.3.3 概念结构设计的方法与步骤	208
5.1.1 数据库安全性的含义	166	6.4 逻辑结构设计	216
5.1.2 安全性控制的一般方法	166	6.4.1 逻辑结构设计任务和步骤	216
5.1.3 SQL Server 2008 的数据安全性 机制	170	6.4.2 初始关系模式设计	217
5.1.4 SQL Server 2008 的身份验证模式	171	6.4.3 关系模式规范化	218
5.1.5 SQL Server 2008 的登录账号和服务 器角色	172	6.4.4 模式评价与改进	218
5.1.6 SQL Server 2008 的数据库用户账号 和数据库角色	174	6.5 物理结构设计	220
5.2 完整性控制	179	6.5.1 确定物理结构	220
5.2.1 数据库完整性的含义	179	6.5.2 评价物理结构	222
5.2.2 完整性规则的组成	179	6.6 数据库实施	222
5.2.3 完整性约束条件的分类	180	6.6.1 建立实际数据库结构	222
5.2.4 数据完整性的实施	181	6.6.2 装入数据	222
5.2.5 规则	182	6.6.3 应用程序编码与调试	223
5.2.6 默认	183	6.6.4 数据库试运行	223
5.3 并发控制与封锁	184	6.6.5 整理文档	223
5.3.1 数据库并发性的含义	184	6.7 数据库运行和维护	224
5.3.2 事务 (Transaction)	185	6.7.1 维护数据库的安全性与完整性	224
5.3.3 并发操作与数据的不一致性	186	6.7.2 监测并改善数据库性能	224
5.3.4 封锁	188	6.7.3 重新组织和构造数据库	224
5.4 数据库的恢复	193	6.8 小结	225
5.4.1 数据库恢复的含义	193	习题	225
5.4.2 数据库恢复的原理及其实现技术	193	第 7 章 SQL Server 2008 高级 应用	229
5.4.3 数据库的故障和恢复的策略	194	7.1 Transact-SQL 程序设计	229
5.5 小 结	196	7.1.1 变量	229
习题	197	7.1.2 运算符	230
第 6 章 数据库设计	199	7.1.3 批处理	233
6.1 数据库设计概述	199	7.1.4 流程控制命令	233
6.1.1 数据库设计的任务、内容和特点	199	7.1.5 常用命令	237
		7.1.6 常用函数	240
		7.2 存储过程	249

7.2.1 存储过程的概念、优点及分类	249	7.3.5 修改触发器	260
7.2.2 创建存储过程	250	7.3.6 删除触发器	261
7.2.3 查看存储过程	253	7.4 备份和还原	261
7.2.4 重新命名存储过程	253	7.4.1 备份和还原概述	261
7.2.5 删除存储过程	253	7.4.2 创建备份设备	264
7.2.6 执行存储过程	253	7.4.3 备份数据库	265
7.2.7 修改存储过程	254	7.4.4 还原数据库	267
7.3 触发器	254	7.4.5 备份和还原系统数据库	268
7.3.1 触发器概述	254	7.5 小结	269
7.3.2 触发器的工作原理	255	习题	269
7.3.3 创建触发器	257	SQL Server 2008 综合练习	271
7.3.4 查看触发器	259	参考文献	276

第 1 章

数据库系统概述

数据库技术是数据管理的最新技术，是计算机科学技术中发展最快的领域之一，也是应用最广的技术之一，它是专门研究如何科学地组织和存储数据、如何高效地获取和处理数据的技术。它已成为各行各业存储数据、管理信息、共享资源的最先进最常用的技术。因此，数据库课程不仅是计算机科学与技术专业、信息管理与信息系统专业的必修课程，也是许多非计算机专业的必修课程。

随着计算机技术的飞速发展，信息已成为当今社会各种活动的核心资源，通过对这些信息资源的进一步开发利用，可有效降低相应活动的成本，使各种社会资源得到最大限度节约和合理运用，而这其中起着基础和核心作用的就是数据库。

数据库 (Database)，简单地说就是数据的仓库，即数据存放的地方。我们周围有许多数据库的例子，如通讯录是一个小数据库，图书馆则是一个典型的大型数据库。小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。在计算机三大主要应用领域 (科学计算、过程控制和数据处理) 中，数据处理所占比例约为 70%。20 世纪 60 年代末，数据库技术作为数据处理的最新技术应运而生。

本章首先回顾数据管理技术的三个发展阶段，然后介绍数据库中的相关概念。学习本章后，读者应了解数据库的发展阶段及各阶段的主要特点，掌握数据库中的有关的基本概念、数据库系统的组成及各部分的主要功能，重点掌握实体、属性定义和实体型之间的联系类型及特点，了解三种数据模型的特点及区别，为后续各章的学习打下基础。

1.1 信息、数据、数据处理与数据管理

1.1.1 数据与信息

在数据处理中，最常用到的基本概念就是数据和信息，二者之间既有区别又有联系。

1. 信息 (Information)

(1) 信息的定义。信息是人脑对现实世界事物的存在方式、运动状态以及事物之间联系的抽象反映。信息是客观存在的，人类有意识地对信息进行采集并加工、传递，从而形成了各种消息、情报、指令、数据及信号等。例如，对于学生基本情况来说，某同学的学号是“S1”，姓名是“赵亦”，性别是“女”，年龄是“17 岁”，所在系别是“计算机”等，这些都是关于某个同学的具体信息，是该同学当前存在状态的反映。

(2) 信息的特征。

① 信息源于物质和能量。信息不可能脱离物质而存在，信息的传递需要物质载体，信息的获

取和传递要消耗能量。如信息可以通过报纸、电台、电视和计算机网络进行传递。

② 信息是可以感知的。人类对客观事物的感知,可以通过感觉器官,也可以通过各种仪器仪表和传感器等,不同的信息源有不同的感知形式。如报纸上刊登的信息通过视觉器官感知,电台广播的信息通过听觉器官感知。

③ 信息是可存储、加工、传递和再生的。人们用大脑存储信息,叫做记忆。计算机存储器、录音、录像等技术的发展,进一步扩大了信息存储的范围。借助计算机,还可对收集到的信息进行整理。

2. 数据(Data)

(1) 数据的定义。数据是由用来记录信息的可识别的符号组合的,是信息的具体表现形式。例如,上面提到的描述某个同学的信息,可用一组数据“S1、赵亦、女、17、计算机”表示。由于这些符号在此已被赋予了特定的语义,因此,它们就具有传递信息的功能。

可见,数据和它的语义是不可分割的。例如,对于数据:(赵亦,计算机),我们可以赋予它一定的语义,它表示“赵亦”所在系为“计算机”系。如果不了解其语义,则无法对其进行正确解释,甚至解释为“赵亦”学习的课程为“计算机”。

(2) 数据的表现形式。可用多种不同的数据形式表示同一信息,而信息不随数据形式的不同而改变。如“2000年硕士研究生将扩招30%”,其中的数据可改为汉字形式“两千年”、“百分之三十”,而表达的信息是一致的。

由于早期的计算机系统主要用于科学计算,因此计算机中处理的数据主要是整数、浮点数等传统数学中的数字。但是,在现代计算机系统中,数据的概念是广义的,数据的概念内涵已被大大地拓宽了,其表现形式不仅包括数字,还包括文字、图形、图像和声音等,它们都可以经过数字化后存储到计算机中。

3. 数据与信息的联系

通过前面的分析可以看出,信息与数据之间存在着固有的联系:数据是信息的符号表示或载体,信息则是数据的内涵,是对数据的语义解释。如上例中的数据“2000”、“30%”被赋予了特定的语义,此处的2000表示的是“2000年”,30%表示的是“研究生将扩招30%”,因此,它们就具有了传递信息的功能。

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去区分它们。

1.1.2 数据处理与数据管理

数据处理是将数据转换成信息的过程,包括对数据的收集、管理、加工利用乃至信息输出的演变与推导等一系列活动。其目的之一是从大量的原始数据中抽取和推导出有价值的信息,作为决策的依据;目的之二是借助计算机科学地保存和管理大量复杂的数据,以便人们能够方便地充分利用这些信息资源。

在数据处理过程中,数据是原料,是输入,而信息是产出,是输出结果。“信息处理”的真正含义应该是为了产生信息而处理数据。

在数据处理中,通常计算比较简单,而数据的管理比较复杂。数据管理是指数据的收集、分类、组织、编码、存储、维护、检索和传输等操作,这些操作是数据处理业务的必不可少的基本环节。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,把数据有效地管理起来,以便最大限度地减轻程序员的负担;至于处理业务中的加工计算,因不同业务各不相同,要靠程序员根据业务情况编写应用程序加以解决。所以,数据处理是与数据管理相联系的,数据

管理技术的优劣，将直接影响数据处理的效率。数据库技术正是瞄准这一目标研究、发展并完善起来的专门技术。

1.2 数据库技术的产生、发展

通过前面的学习可知，数据处理的中心问题是数据管理。随着计算机硬件和软件的发展，数据管理经历了人工管理、文件系统和数据库系统三个发展阶段。数据库技术正是应数据管理任务的需要而产生、发展的。

1. 人工管理阶段

在 20 世纪 50 年代中期以前，计算机主要用于科学计算，当时硬件中的外存只有卡片、纸带和磁带，没有磁盘等直接存取设备。软件只有汇编语言，没有操作系统和管理数据的软件。所以，为了给程序提供完成科学计算和数据处理的数据，必须手工来完成。因此，称这样的数据管理方式为人工管理数据。

人工管理数据有如下几个特点。

(1) 数据不保存。因为还没有磁盘这样的可直接存取数据的存储设备，再者，当时计算机主要用于科学计算，对于数据保存的需求尚不迫切。只有在计算某一课题时才将原始数据随程序一起输入内存，运算处理结束后将结果数据输出，随着计算任务的完成，数据和程序一起从内存中被释放；若再计算同一课题时，还需要再次输入原始数据和程序。可见，不仅参加运算的原始数据不保存，运算的结果也不保存。

(2) 系统没有专用的软件对数据进行管理。数据需要由应用程序自己管理，没有相应的软件系统负责数据的管理工作。每个应用程序不仅要规定数据的逻辑结构，而且要设计物理结构，包括数据的存储结构、存取方法和输入方式等，因此，程序员的负担也很重。

(3) 数据不共享。数据是面向程序的，一组数据只能对应一个程序。即使多个应用程序涉及某些相同的数据时，也必须各自定义，无法互相利用、互相参照，因此，程序之间有大量的冗余数据。

(4) 数据不具有独立性。由于以上几个特点，以及没有专门对数据进行管理的软件系统，所以，这个时期的每个程序都要包括数据存取方法、输出/输出方式和数据组织方法等。因为程序是直接面向存储结构的，所以，如果数据的类型、格式或输入/输出方式等逻辑结构或物理结构发生变化，必须对应用程序做出相应的修改，因而，数据与程序不具有独立性，这也进一步加重了程序员的负担。

在人工管理阶段，程序与数据之间是一一对应的关系，其特点可用图 1-1 表示。

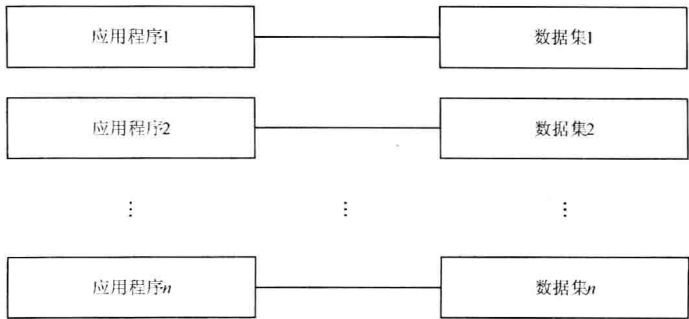


图 1-1 人工管理阶段应用程序与数据之间的对应关系

2. 文件系统阶段

在 20 世纪 50 年代后期至 60 年代中期, 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。随着数据量的增加, 数据的存储、检索和维护成为紧迫的需要。

此时, 在硬件方面, 已有了磁盘、磁鼓等直接存取设备; 在软件方面, 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。

文件系统阶段数据管理有如下特点。

(1) 数据以文件形式长期保存。数据以文件的组织方式, 长期保存在计算机的磁盘上, 可以被多次反复使用。应用程序可对文件进行查询、修改和增删等处理。

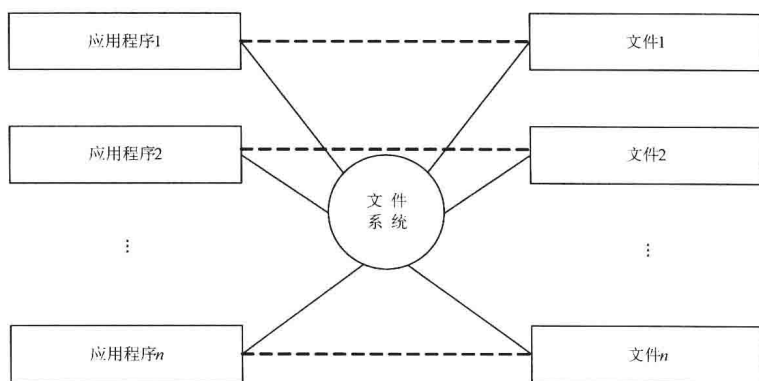
(2) 由文件系统管理数据。文件系统提供了文件管理功能和访问文件的存取方法。文件系统把数据组织成内部有一定结构的记录并以文件的形式存储在存储设备上, 这样, 程序只与存储设备上的文件名打交道, 不必关心数据的物理存储(存储位置、存储结构等), 而由文件系统提供的存取方法实现数据的存取, 从而实现了“按文件名访问, 按记录进行存取”的数据管理技术。

(3) 程序与数据间有一定独立性。由于文件系统在程序与数据文件之间的存取转换作用, 使得程序和数据之间具有“设备独立性”, 即当改变存储设备时, 不必改变应用程序。程序员也不必过多地考虑数据存储的物理细节, 而将精力集中于算法设计上, 从而大大减少了维护程序的工作量。

(4) 文件的形式已经多样化。由于有了磁盘这样的直接存取存储设备, 文件也就不再局限于顺序文件, 有了索引文件、链表文件等。因此, 对文件的访问既可以是顺序访问, 也可以是直接访问。但文件之间是独立的, 它们之间的联系需要通过程序去构造, 文件的共享性也还比较差。

(5) 数据具有一定的共享性。有了文件以后, 数据就不再仅仅属于某个特定的程序, 而可以由多个程序反复使用。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, 数据的存储结构和程序之间的依赖关系并未根本改变。

在文件系统阶段, 程序与数据之间的对应关系如图 1-2 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, 但一些根本性问题还没有彻底解决, 主要表现在以下几方面。

(1) 数据共享性差、冗余度大。一个文件基本上对应于一个应用程序, 即文件仍然是面向应用的。当不同的应用程序所使用的数据具有共同部分时, 也必须分别建立自己的数据文件, 数据不能共享。

(2) 数据不一致性。这通常是由数据冗余造成的。由于相同数据在不同文件中的重复存储、各自管理, 在对数据进行更新操作时, 不但浪费磁盘空间, 同时也容易造成数据的不一致性。

(3) 数据独立性差。在文件系统阶段, 尽管程序与数据之间有一定的独立性, 但是这种独立

性主要是指设备独立性, 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部存储器的物理结构要求。因此, 在文件系统中, 一旦改变数据的逻辑结构, 必须修改相应的应用程序, 修改文件结构的定义。而应用程序发生变化, 如改用另一种程序设计语言来编写程序, 也将引起文件的数据结构的改变。

(4) 数据间的联系弱。文件与文件之间是独立的, 文件间的联系必须通过程序来构造。因此, 文件系统只是一个没有弹性的、无结构的数据集合, 不能反映现实世界事物之间的内在联系。

3. 数据库系统阶段

从 20 世纪 60 年代后期开始, 计算机用于管理的规模更加庞大, 应用越来越广泛, 数据量急剧增加, 同时, 多种应用、多种语言互相覆盖地共享数据集合的要求也越来越强烈。

在这种情况下, 为了提高数据管理的效率, 人们开始对文件系统进行扩充, 但这样做没能解决问题。而此时, 硬件方面出现了大容量、存取快速的磁盘, 使计算机联机存取大量数据成为可能, 也为数据库技术的产生和发展提供了物质条件。同时, 硬件价格下降和软件价格上升, 使开发和维护系统软件的成本相对增加。因此, 文件系统的管理方法已无法适应各种应用的需要。于是为解决多用户、多个应用程序共享数据的需求, 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。

可见, 计算机技术的发展、数据管理的需求迫切性共同促进了数据库技术的诞生。

20 世纪 60 年代末期出现的对数据管理技术有着奠基作用的三件大事, 标志着以数据库系统为基本手段的数据管理新阶段的开始。

(1) 1968 年, 美国 IBM 公司推出了商品化的信息管理系统 (Information Management System, IMS)。该系统支持层次数据模型, 也就是使用层次结构来表示数据与数据之间的联系。

(2) 1969 年, 美国数据系统语言协会 (Conference On Data System Language, CODASYL) 下属的数据库任务组 (DataBase Task Group, DBTG) 发布了一系列研究数据库方法的 DBTG 报告, 该报告建立了数据库技术的很多概念、方法和技术。其中, DBTG 提出的网状数据模型, 可以表示任意数据之间的联系, 从而突破了已有层次模型的限制。

(3) 1970 年起, 美国 IBM 公司的研究员 E. F. Codd 连续发表文章, 提出了数据间的联系用数学上的“关系”来表示, 即提出了数据库的关系模型, 从而开创了数据库关系方法和关系数据理论的研究, 为关系数据库的发展和理论研究奠定了基础, 该模型一直沿用至今。

20 世纪 70 年代以来, 数据库技术得到迅速发展, 开发出了许多产品, 并投入运行。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, 与人工管理和文件系统相比, 数据库系统阶段管理数据的特点有如下几个方面。

(1) 结构化的数据及其联系的集合。在数据库系统中, 将各种应用的数据按一定的结构形式 (即数据模型) 组织到一个结构化的数据库中, 不仅考虑了某个应用的数据结构, 而且考虑了整个组织 (即多个应用) 的数据结构, 也就是说, 数据库中的数据不再仅仅针对某个应用, 而是面向全组织, 不仅数据内部是结构化的, 整体也是结构化的; 不仅描述了数据本身, 也描述了数据间的有机联系, 从而较好地反映了现实世界事物间的自然联系。

例如: 要建立学生成绩管理系统, 系统包含学生 (学号、姓名、性别、系别、年龄)、课程 (课程号、课程名)、成绩 (学号、课程号、成绩) 等数据, 分别对应三个文件。若采用文件处理方式, 因为文件系统只表示记录内部的联系, 而不涉及不同文件记录之间的联系, 要想查找某个学生的学号、姓名、所选课程的名称和成绩, 必须编写一段比较复杂的程序来实现, 即不同文件记录间的联系只能写在程序中, 在编写程序时考虑和实现。而采用数据库方式, 由于数据库系统不仅描

述数据本身，还描述数据之间的联系，记录之间的联系可以用参照完整性来表述，上述查询可以非常容易地联机查到。

此外，在数据库系统中，不仅数据是结构化的，而且存取数据的方式也很灵活，可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录。而在文件系统中，数据的最小存取单位是记录，不能细化到数据项。

（2）数据共享性高、冗余度低。所谓数据共享是指数据库中的一组数据集合可为多个应用和多个用户共同使用。

由于数据库系统从整体角度看待和描述数据，数据不再面向某个或某些应用，而是全盘考虑所有用户的数据需求、面向整个应用系统，所有用户的数据都包含在数据库中，因此，不同用户、不同应用可同时存取数据库中的数据，每个用户或应用只使用数据库中的一部分数据，同一数据可供多个用户或应用共享，从而减少了不必要的数据冗余，节约了存储空间，同时也避免了数据之间的不相容性与不一致性，即避免了同一数据在数据库中重复出现且具有不同值的现象。

同时，在数据库系统下，用户或应用不像在文件系统中那样各自建立自己对应的数据文件，而是从数据库中存取其中的数据子集，该数据子集是通过数据库管理系统（Data Base Management System, DBMS）从数据库中经过映射而形成的逻辑文件。同一个数据可能在物理存储上只存一次，但可以把它映射到不同的逻辑文件里，这就是数据库系统提高数据共享、减少数据冗余的根本所在，如图 1-3 所示。

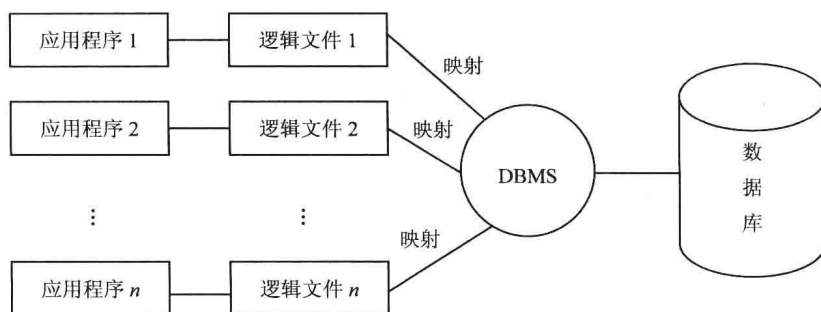


图 1-3 数据库系统中的数据共享机制示意图

（3）数据独立性高。所谓数据的独立性是指数据库中的数据与应用程序间相互独立，即数据的逻辑结构、存储结构以及存取方式的改变不影响应用程序。

在数据库系统中，整个数据库的结构可分成三级：用户的逻辑结构、整体逻辑结构和物理结构。数据独立性分两级：物理独立性和逻辑独立性，如图 1-4 所示。

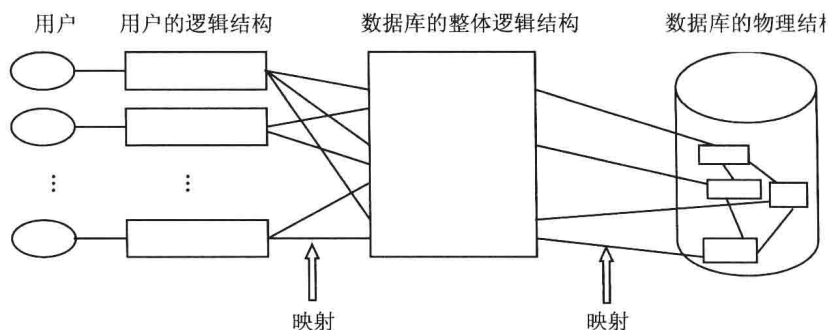


图 1-4 数据库的三级结构及其映射关系示意图

数据的物理独立性是指当数据的物理结构（如存储结构、存取方式、外部存储设备等）改变时，通过修改映射，使数据库整体逻辑结构不受影响，进而用户的逻辑结构以及应用程序不用改变。

数据的逻辑独立性是指当数据库的整体逻辑结构（如修改数据定义、增加新的数据类型、改变数据间的关系等）发生改变时，通过修改映射，使用户的逻辑结构以及应用程序不用改变。

数据独立性把数据的定义从程序中分离出去，加上数据的存取是由 DBMS 负责，从而简化了应用程序的编制，大大减少了应用程序的维护和修改。

（4）有统一的数据管理和控制功能。在数据库系统中，数据由数据库管理系统进行统一管理和控制。数据库可为多个用户和应用程序所共享，不同的应用需求，可以从整个数据库中选取所需要的数据子集。另外，对数据库中数据的存取往往是并发的，即多个用户可以同时存取数据库中的数据，甚至可以同时存取数据库中的同一个数据。为确保数据库数据的正确、有效和数据库系统的有效运行，数据库管理系统提供下述 4 个方面的数据控制功能。

① 数据的安全性（Security）控制：防止不合法使用数据库造成数据的泄露和破坏，使每个用户只能按规定对某些数据进行某种或某些操作和处理，保证数据的安全。例如，系统提供口令检查用户身份或用其他手段来验证用户身份，以防止非法用户使用系统。也可以对数据的存取权限进行限制，用户只能按所具有的权限对指定的数据进行相应的操作。

② 数据的完整性（Integrity）控制：系统通过设置一些完整性规则等约束条件，确保数据的正确性、有效性和相容性。

正确性是指数据的合法性，如年龄属于数值型数据，只能含 0, 1, …, 9，不能含字母或特殊符号。

有效性是指数据是否在其定义的有效范围，如月份只能用 1~12 的正整数表示。

相容性是指表示同一事实的两个数据应相同，否则就不相容，如一个人不能有两个性别。

③ 并发（Concurrency）控制：多个用户同时存取或修改数据库时，系统可防止由于相互干扰而提供给用户不正确的数据，并防止数据库受到破坏。

④ 数据恢复（Recovery）：由于计算机系统的硬件故障、软件故障、操作员的误操作及其他故意的破坏等原因，造成数据库中的数据不正确或数据丢失时，系统有能力将数据库从错误状态恢复到最近某一时刻的正确状态。

在数据库系统阶段，程序与数据之间的关系，如图 1-5 所示。

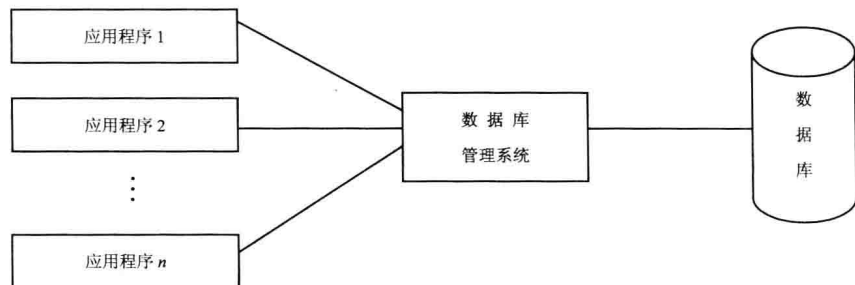


图 1-5 数据库系统阶段程序与数据之间的关系

从文件系统管理发展到数据库系统管理是信息处理领域的一个重大变化。在文件系统阶段，人们关注的是系统功能的设计，因此，程序设计处于主导地位，数据服从于程序设计；而在数据

库系统阶段，数据占据了中心位置，数据的结构设计成为信息系统首先关心的问题。

1.3 数据库系统的组成

数据库系统（DataBase System，DBS）是指在计算机系统中引入数据库后的系统。它主要由数据库、数据库用户、计算机硬件系统和计算机软件系统等几部分组成。有时人们常将数据库系统简称为数据库，它可用图 1-6 表示。

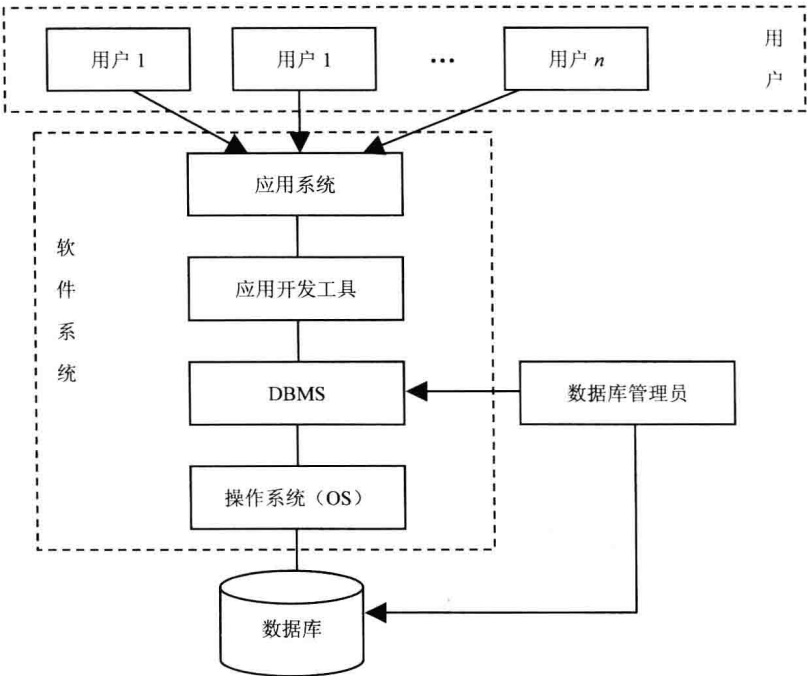


图 1-6 数据库系统的组成

1. 数据库

数据库（DataBase，DB）是存储在计算机内、有组织的、可共享的数据和数据对象（如表、视图、存储过程和触发器等）的集合，这种集合按一定的数据模型（或结构）组织、描述并长期存储，同时能以安全和可靠的方法进行数据的检索和存储。

数据库有如下两个特点。

（1）集成性。将某特定应用环境中的各种应用相关的数据及其数据之间的联系全部集中地并按照一定的结构形式进行存储，或者说，把数据库看成为若干个性质的数据文件的联合和统一的数据整体。

（2）共享性。数据库中的数据可为多个不同的用户所共享，即多个不同的用户可使用多种不同的语言，为了不同的应用目的，而同时存取数据库，甚至同时存取数据库中的同一数据。

2. 用户

用户是指使用数据库的人，他们可对数据库进行存储、维护和检索等操作。用户分为以下