

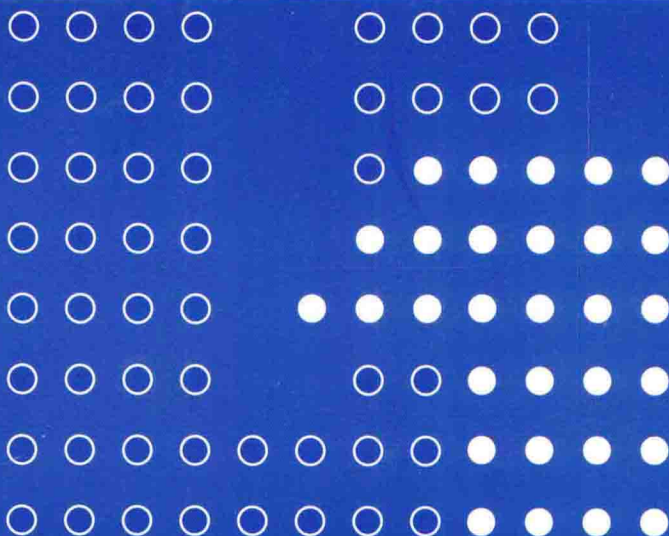


普通高等教育“十一五”国家级规划教材 计算机系列教材

数据库原理与应用

(第2版)

宋金玉 陈萍 陈刚 编著



清华大学出版社



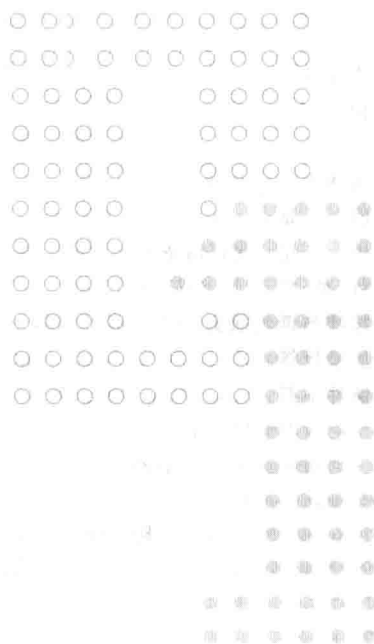


普通高等教育“十一五”国家级规划教材 计算机系列教材

宋金玉 陈萍 陈刚 编著

数据库原理与应用

(第2版)



清华大学出版社
北京

内 容 简 介

本书第1~2章系统地阐述数据库的基本概念和基本理论,包括数据库的概念、数据库管理数据的特点、数据库系统体系结构、数据模型等;第3~5章从数据模型的三个组成要素的角度,介绍关系数据模型所采用的数据结构、关系操作语言和完整性约束、SQL查询语言的使用以及指导关系模式设计的关系规范化理论;第6~8章全面地分析数据库管理系统的数据存储管理、查询优化和事务管理等核心技术;第9~10章介绍数据库应用系统中,数据库设计和在应用编程中访问与管理数据库中数据的方法;第11章介绍数据库应用领域发展相对成熟的、应用较广泛的一些技术。

本书以《高等学校计算机科学与技术专业公共核心知识体系与课程》和《高等学校计算机科学与技术专业核心课程教学实施方案》为指导,适合作为高等院校计算机及其相关专业数据库课程的教材,也可作为从事数据库理论研究、数据库应用系统的设计与开发,以及数据库管理与维护等工作人员的参考用书。

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 13701121933

图书在版编目(CIP)数据

数据库原理与应用/宋金玉,陈萍,陈刚编著.--2版.--北京:清华大学出版社,2014

计算机系列教材

ISBN 978-7-302-37242-4

I. ①数… II. ①宋… ②陈… ③陈… III. ①数据库系统—教材 IV. ①TP311.13

中国版本图书馆CIP数据核字(2014)第153408号

责任编辑:张瑞庆

封面设计:常雪影

责任校对:李建庄

责任印制:何 芊

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 刷 者:三河市君旺印务有限公司

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:22.75 字 数:565千字

版 次:2011年6月第1版 2014年9月第2版 印 次:2014年9月第1次印刷

印 数:1~2000

定 价:39.00元

产品编号:058883-01

随着社会信息化进程的推进,尤其是大数据时代的到来,数据库管理技术已经成为现代计算环境中数据管理的基础技术。数据库理论的研究一直未曾间断,而这其中绝大多数的研究仍以关系数据模型为基础。因此,本书中数据库原理的内容仍以关系数据模型为基础,以关系数据库理论、关系数据库管理系统核心技术、关系数据库设计与编程为主来组织,共11章内容,可分为如下5个部分。

第1部分包括2章(第1~2章)。介绍数据库管理技术的发展历史,突出数据库技术相对于人工管理和文件系统这两种数据管理方式所具有的优点,优势在于采用数据模型表示数据,数据模型的发展是数据管理技术发展的一条主线。随着信息社会海量数据的产生,以及数据呈现多态性等,数据模型必将得到相应的发展。阐明数据库系统的三级模式结构以及二级映像机制的作用,帮助学生了解虽然目前实际的数据库管理系统产品种类很多,但它们通常遵循ANSI/SPARC体系结构,该体系结构提供了数据独立性,减轻了用户使用系统的负担,使DBMS产品得以广泛使用。

第2部分包括3章(第3~5章)。从数据模型的三个组成要素的角度,介绍关系数据模型所采用的数据结构、关系操作能力的表达方法、关系模型对于存储在数据库中的数据具有的约束能力。帮助学生认识关系模型之所以被广泛的应用和发展,在于其可用单一的数据结构(关系)来表达实体以及实体间的联系,能比较真实地模拟现实世界,容易被人所理解且便于在计算机上实现。介绍SQL语言对关系模型的支持,以及数据库管理系统软件对SQL语言标准的支持。介绍不良的关系模式设计可能带来的数据冗余和更新异常问题,以及解决此类问题的关系规范化理论的相关概念和方法。

第3部分包括3章(第6~8章)。介绍关系数据库管理系统的实现技术,包括DBMS如何在存储器上组织数据字典、用户数据、数据之间的联系,以及实现数据快速存取所采用的存储结构和存取方式。此外,RDBMS还必须对SQL查询进行优化,来提高查询效率和系统性能。而DBMS要得到广泛的应用必须保证数据一致性,保证系统中一切事务的原子性、一致性、隔离性和持久性,因此DBMS必须对各类故障进行恢复,对并发事务操作进行控制。

第4部分包括2章(第9~10章)。介绍数据库应用系统中数据库设计以及在应用编程中访问和管理数据库中数据的方法。其中,数据库设计是基于应用系统需求分析中对数据的需求,解决数据的抽象、数据的表达和数据的存储等问题,其目标是设计出一个满足应用要求,简洁、高效、规范合理的数据库。第9章基于E-R模型的规范设计方法,介

绍数据库设计各个阶段所做的工作。第 10 章基于一个简单的图书管理系统实例,介绍目前在应用系统中采用编程方式来访问和管理数据库中数据的主要方法。

第 5 部分包括 1 章(第 11 章)。简要分析数据库系统的研究与发展趋势,介绍数据库应用领域发展相对成熟的、应用较广泛的一些技术,包括分布式数据库、数据仓库及数据挖掘等。

本版的内容结构与第 1 版大体一致,修订的主要工作是把第 1 版教材在使用过程中产生的问题和对教材内容的新认识加以体现,包括如下几个主要方面:

(1) 对第 3 章中一些关系操作的内涵进行了更有助于教学的阐述。对第 4 章在细节上有较大的修订,更换了一些例题,更注重介绍 SQL 对关系模型的支持,同时也侧重描述 DBMS 对 SQL 标准的支持程度及相关实现策略。

(2) 本书在介绍关系数据库管理系统的实现技术时,是基于当今的商用或试验用的 DBMS 中所采用的概念和策略,为了更好地理解相关内容,基于特定的 DBMS 产品实例进行了补充介绍。

(3) 对数据库设计及数据库编程部分进行了较大的修订,更改扩充了 3 个章节的内容,并对 5 个章节的文字内容重新进行阐述。

教材的修订工作由宋金玉主持完成,陈萍和陈刚参与了部分章节的修订,所在大学和学院的有关方面给予了大力的支持和保障,在此表示感谢!也向使用第 1 版教材和提供宝贵意见的师生表示感谢!

望学术同仁不吝赐教,教师和学生继续给予支持,并在使用过程中多提宝贵意见,我们将不胜感激。另外,为了方便教学,本书配套出版了习题解答和实验指导;配备了电子课件,教材示例数据库等,如需要可与作者联系(Email: sjyhello@163.com)或者到清华大学出版社网站(www.tup.com.cn)下载。

作 者

2014 年 6 月于南京解放军理工大学

F O R E W O R D

第1章 数据库系统概论 /1

1.1 数据管理技术的发展 /1

1.2 数据库的基本概念 /5

1.2.1 数据库 /5

1.2.2 数据库管理系统 /9

1.2.3 数据库系统 /10

1.3 数据库系统体系结构 /12

1.3.1 数据库系统的三级模式结构 /13

1.3.2 二级映射与数据独立性 /17

1.3.3 DBMS 的模块组成 /19

1.3.4 DBMS 的客户/服务器体系结构 /20

1.4 小结 /22

习题 /23

第2章 数据模型 /26

2.1 现实世界客观对象的抽象过程 /26

2.2 概念模型 /27

2.3 数据模型概述 /29

2.3.1 关系模型 /31

2.3.2 半结构化数据模型 /32

2.3.3 面向对象数据模型 /35

2.4 小结 /36

习题 /36

第3章 关系数据库理论 /38

3.1 关系模型概述 /38

3.1.1 关系的数据结构 /38

3.1.2 关系的完整性约束 /43

3.1.3 关系操作 /46

3.2 关系代数 /47

- 3.2.1 传统的集合运算 /48
- 3.2.2 专门的关系运算 /50
- 3.2.3 用关系代数表达式实现关系查询 /55

3.3 关系演算 /56

- 3.3.1 关系演算中关系的表示 /56
- 3.3.2 元组关系演算 /57
- 3.3.3 域关系演算 /59

3.4 小结 /61

习题 /62

第4章 关系数据库标准查询语言 SQL /67

4.1 SQL 语言概述 /67

4.2 SQL 中的数据定义 /70

- 4.2.1 基本表的定义 /70
- 4.2.2 基本表的修改 /74
- 4.2.3 基本表的删除 /75

4.3 SQL 中的数据查询 /75

- 4.3.1 SQL 的单表查询 /76
- 4.3.2 SQL 的连接查询 /83
- 4.3.3 SQL 的嵌套查询 /86
- 4.3.4 SQL 的集合查询 /94

4.4 SQL 中的数据更新 /95

- 4.4.1 SQL 插入数据语句 /95
- 4.4.2 SQL 修改数据语句 /96
- 4.4.3 SQL 删除数据语句 /97
- 4.4.4 更新操作的完整性检查 /98
- 4.4.5 触发器 /100

4.5 SQL 中的视图 /105

- 4.5.1 视图的概念 /105
- 4.5.2 视图的定义 /106
- 4.5.3 视图删除 /107
- 4.5.4 视图的查询 /108

4.5.5	视图的更新	/109
4.5.6	视图的作用	/111
4.6	SQL 中的授权控制	/113
4.7	小结	/119
	习题	/119
第5章 关系模式的规范化设计 /124		
5.1	关系模式的设计问题	/124
5.2	关系模式的规范化	/125
5.2.1	函数依赖	/125
5.2.2	基于函数依赖的范式	/127
5.2.3	多值依赖与 4NF	/131
5.2.4	关系模式的规范化	/133
5.3	函数依赖的理论	/134
5.3.1	函数依赖集的逻辑蕴含	/134
5.3.2	Armstrong 公理	/135
5.3.3	属性集闭包	/136
5.3.4	函数依赖集等价和最小函数 依赖集	/138
5.3.5	候选键及其求解方法	/140
5.4	模式分解	/142
5.4.1	模式分解的概念	/143
5.4.2	无损连接分解和保持函数依赖 分解	/144
5.4.3	模式分解算法	/147
5.5	小结	/150
	习题	/151
第6章 数据库的存储管理 /155		
6.1	数据库存储管理的数据	/155
6.2	磁盘上数据的存储	/157
6.2.1	磁盘的物理特性	/158

6.2.2	磁盘上数据的缓冲存取	/159
6.3	文件的组织结构	/161
6.3.1	定长记录	/162
6.3.2	变长数据和记录	/164
6.3.3	列存储	/165
6.4	文件的存储结构	/166
6.4.1	堆文件	/166
6.4.2	顺序文件	/166
6.4.3	聚集文件	/167
6.4.4	散列文件(直接文件)	/169
6.4.5	SQL Server 的文件存储	/170
6.5	索引	/173
6.5.1	索引的概念	/174
6.5.2	聚集索引和非聚集索引	/175
6.5.3	稠密索引和稀疏索引	/179
6.5.4	多级索引	/182
6.5.5	倒排索引和文档检索	/184
6.5.6	位图索引	/185
6.6	索引文件的结构	/187
6.6.1	B ⁺ 树	/187
6.6.2	散列索引	/193
6.7	小结	/194
	习题	/195

第7章 关系查询与优化 /198

7.1	数据库系统的查询处理步骤	/198
7.2	查询分析与预处理	/200
7.3	代数优化	/202
7.3.1	代数优化的必要性	/202
7.3.2	基于代数等价的启发式优化	/205
7.4	物理优化	/209
7.4.1	操作符的实现算法	/210

7.4.2	基于代价的物理优化方法	/215
7.5	小结	/218
	习题	/219

第8章 事务管理 /222

8.1	事务的概念	/222
8.1.1	概念的引入	/222
8.1.2	事务的定义	/223
8.1.3	事务的 ACID 特性	/225
8.1.4	事务的管理	/226
8.2	事务的恢复	/227
8.2.1	故障及其错误状态	/227
8.2.2	恢复的实现技术	/229
8.2.3	恢复的策略	/232
8.3	并发控制	/238
8.3.1	并发控制的必要性	/238
8.3.2	封锁技术	/243
8.4	小结	/263
	习题	/264

第9章 数据库设计 /267

9.1	数据库设计概述	/267
9.1.1	数据库设计的内容	/267
9.1.2	数据库设计的方法	/267
9.1.3	数据库设计的阶段	/268
9.2	需求分析	/270
9.2.1	需求分析的任务	/270
9.2.2	需求分析的方法	/271
9.3	概念设计	/275
9.3.1	E-R 模型	/276
9.3.2	概念模型设计方法	/277
9.4	逻辑结构设计	/280

9.4.1	E-R图向关系模型的转换	/280
9.4.2	关系模式的优化	/284
9.4.3	用户外模式的设计	/284
9.5	物理设计	/285
9.5.1	确定数据库的存储结构	/286
9.5.2	确定数据库的存取方法	/286
9.6	数据库的实施与维护	/287
9.6.1	数据库数据的装入	/287
9.6.2	数据库的运行和维护	/288
9.7	小结	/289
	习题	/289
第10章	数据库编程	/292
10.1	嵌入式SQL	/292
10.1.1	嵌入式SQL的处理过程	/292
10.1.2	嵌入式SQL语句与主语言之间的通信	/293
10.1.3	建立和关闭数据库连接	/294
10.1.4	游标的使用	/295
10.2	存储过程	/297
10.2.1	存储过程的类型	/297
10.2.2	存储过程的创建和执行	/299
10.2.3	存储过程的优点	/302
10.3	函数	/303
10.3.1	系统函数	/303
10.3.2	用户自定义函数	/306
10.4	ODBC编程	/309
10.4.1	ODBC工作原理	/310
10.4.2	ODBC的工作流程	/313
10.5	ADO.NET	/317
10.5.1	ADO.NET的起源	/318
10.5.2	ADO.NET的对象模型	/319

10.5.3	ADO.NET 的数据访问	/320
10.6	小结	/325
	习题	/326
第 11 章 数据库技术的新发展 /327		
11.1	数据库技术的发展趋势	/327
11.1.1	关系数据库的缺陷	/327
11.1.2	第三代数据库系统	/329
11.1.3	数据库技术的发展	/329
11.2	分布式数据库系统	/331
11.2.1	分布式数据库系统的概念	/332
11.2.2	分布式数据库系统面临的新问题	/333
11.2.3	分布式数据库系统的体系结构及功能组成	/338
11.3	数据仓库与联机分析处理	/338
11.3.1	数据仓库	/338
11.3.2	联机分析处理	/342
11.4	数据挖掘技术	/343
11.4.1	数据挖掘的概念	/344
11.4.2	频繁项集挖掘	/344
11.4.3	发现相似的项	/345
11.4.4	大规模数据的聚簇	/346
11.5	小结	/348
	习题	/348
参考文献 /349		

第 1 章 数据库系统概论

数据库系统已经成为现代社会日常生活的重要组成部分,在每天的工作和生活中,人们经常与数据库系统打交道。例如,到银行存钱或取钱,预订宾馆房间或机票,在图书馆查找图书或者从网上购物,等等,所有这些活动都会涉及人或通过计算机程序访问数据库。在以上这些数据库应用中,大多数信息都是以文本或数字形式来存储和访问的。随着计算机技术的发展,数据库得到了更广泛的应用。多媒体数据库可以存储图片、视频片段和卫星图像;地理信息系统(GIS)可以存储和分析地图、气象数据和卫星图像;许多公司使用数据仓库(data warehouse)和联机分析处理(online analytical processing, OLAP)来提取、分析大型数据库中的有用信息以辅助决策;实时(real-time)和主动数据库(active database)技术则用于控制工业和制造业的生产过程;数据库搜索技术也已经应用到 WWW 上,以满足用户对信息搜索的需求。

数据库是信息化社会中信息资源开发与利用的基础。对于一个国家来说,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这个国家信息化程度的重要标志。

本章主要介绍数据管理技术的发展,数据库的基本概念和数据库体系结构,目的是使读者了解为什么使用数据库技术,数据库系统的功能,发挥作用的机理等。本章是全书的基础和导引。

1.1 数据管理技术的发展

数据库技术是应数据管理任务的需要而产生的,是数据管理的关键技术,是构成信息系统的核心和基础。

数据管理是指对数据进行分类、组织、编码、存储、检索和维护,它是数据处理的中心问题。而数据处理是指对各种数据进行收集、存储、加工和传播的一系列活动的总称。

在数据管理领域,数据与信息是分不开的。一般把信息理解为关于现实世界事物存在方式或运动状态的反映。而数据,通常指用符号记录下来的、可以识别的信息。可见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称为载体,信息则是数据的内涵,是对数据语义的解释。

在计算机问世以前,对数据的管理只能是手工和机械方式。在计算机问世后,在应用需求的推动下,在计算机硬件、软件发展的基础上,数据管理技术经历了人工管理、文件系统管理和数据库系统管理三个阶段。

1. 人工管理阶段

在人工管理阶段(20 世纪 50 年代中期以前),计算机主要用于科学计算。硬件方面的现状是,外部存储器只有磁带、卡片和纸带等,还没有磁盘等直接存取存储设备,所以数

据并不保存。软件方面还没有出现操作系统,尚无数据管理的软件,应用程序(用户)负责管理数据,对数据进行批处理。这个时期对于数据的管理是由用户自己完成的,所以称为人工管理。

人工管理数据具有如下特点。

1) 数据不保存

计算机主要用于计算,并不对数据进行其他操作,也没有磁盘等直接存取存储设备,数据并不单独保存在计算机系统中。程序中的数据,随着程序的运行完成,其所占用的内存空间同指令所占用的内存空间一起被释放,退出计算机系统。

2) 数据面向程序

数据需要由应用程序自己设计、说明(定义)和管理,程序员在编写程序时自己规定数据的存储结构、存取方法、输入方式等。

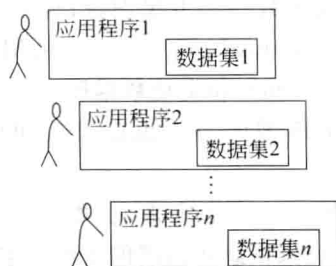


图 1-1 人工管理阶段应用程序与数据之间的对应关系

3) 数据不能共享

数据完全面向特定的应用程序,数据的产生和存储依赖于定义和使用数据的程序。多个程序使用相同数据时,也必须各自定义,重复存储,因此会产生冗余数据。但这是不可避免的,因为一个程序所使用的数据并不能为另一个程序所知,因此数据不能共享,如图 1-1 所示。

4) 数据不具有独立性

数据独立性是指用户的应用程序与数据的逻辑结构和物理结构是相互独立的。在人工管理阶段,没有专门的软件对数据进行管理,程序直接面向存储结构(此阶段数据的逻辑结构与物理结构没有区别)。当数据的存储结构发生变化时,必须由应用程序做相应的修改,对数据进行重新定义。

2. 文件系统管理阶段

20 世纪 50 年代末到 60 年代中期,计算机不仅用于科学计算,还开始大量用于数据管理。这个时期硬件方面有了磁盘、磁鼓等直接存取设备,计算机所能处理的数据的量和速度得到了提高。软件方面出现了操作系统和高级语言,操作系统中有了专门管理数据的软件(称为文件系统)。数据处理方式上不仅能进行批处理,还能进行联机实时处理。这个时期对于数据的管理是由文件系统完成的,所以称为文件系统管理阶段。

文件系统管理数据具有如下的特点。

1) 由文件系统管理数据

文件系统是操作系统中负责存取和管理数据的模块,采用统一的方式管理用户和系统中数据的存储、检索、更新、共享和保护等。文件系统可把应用程序所管理的数据组织成相互独立的数据文件,利用“按文件名访问、按记录进行存取”的文件管理技术,实现对数据的修改、插入和删除等操作。

文件系统所管理的数据文件是一种有结构的文件,它包含若干逻辑记录,逻辑记录是文件中按信息在逻辑上的独立含义划分的一个信息单位,记录在文件中的排列可能有顺

序关系。此外,记录与记录之间不存在其他关系。数据文件之间更是相互独立、缺乏联系的,如图 1-2 所示。

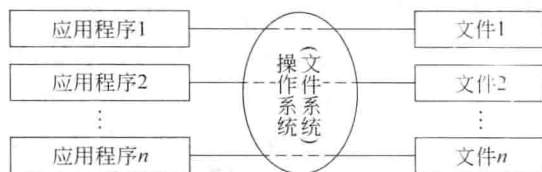


图 1-2 文件系统管理阶段应用程序与数据之间的对应关系

2) 数据可以长期保存

数据可以以“文件”的形式长期保存在磁盘等外部存储器上,应用程序可通过文件系统对磁盘上的文件中的数据进行管理。

3) 程序和数据之间具有“设备独立性”

利用文件系统,应用程序无需考虑如何将数据存放在存储介质上,只要知道文件名,给出有关操作要求便可存取数据,实现了“按名访问”。因此,数据文件可脱离应用程序单独存储,可以重复使用,程序和数据之间具有“设备独立性”。

4) 数据是面向应用的

在文件系统管理阶段中,文件的建立、存取、查询和更新等操作都要由应用程序来实现。数据需要由应用程序自己设计、说明(定义)和管理,程序员在编写程序时不仅要规定数据的逻辑结构,还要设计数据的物理结构,包括数据的存储结构、存取方法、输入方式等。即数据文件中只存储数据,不存储文件记录的结构描述信息。

数据的逻辑结构是用户可见的信息组织方式,可独立于物理环境加以构造;数据的物理结构是数据在物理存储空间中的存放方法和组织方式。比如,学生的姓名、性别、年龄等可用线性表这种逻辑结构组织,表中的数据元素称为一个记录,含有大量记录的线性表称为文件。而在物理存储空间可用一组地址连续的存储单元依次存储线性表的元素(称为顺序存储),也可用一组任意的存储单元存储线性表的数据元素(称为链式存储)。对线性表中的数据可采用后进先出的堆栈存取方法,也可采用先进先出的队列存取方法。数据的输入可采用在程序中进行简单赋值、批量赋值,或在程序运行时从键盘输入等方式。所有这些,都需要借助于某种计算机语言,采用适当的数据结构在应用程序中加以实现;应用程序在数据的存储结构基础上,采用一定的算法来实现对数据的存取、查询和更新等操作。

由此可见,文件系统管理阶段是数据管理技术发展中的一个重要阶段。但随着数据管理规模的扩大和数据量的急剧增加,利用文件系统进行数据管理仍存在以下缺点。

1) 数据的共享性差,冗余度大

此时,虽然程序和数据之间具有“设备独立性”,但数据仍然是面向应用的。数据文件的建立、存取等操作,都要由应用程序来实现,不同的应用程序所使用的相同数据,也必须存储在各自的数据文件中,数据文件之间是相互独立的,不能反映出不同应用所用的相关数据之间的内在联系。由于不能共享这部分相同的数据,因此数据的冗余度大,浪费存储空间。而相同数据的重复存储、各自管理,容易造成数据的不一致,给数据的维护带来

困难。

例如,学生文件 S 的记录可由学号、姓名、性别、出生年月、所在系等数据项组成;课程文件 C 的记录可由课程编号、课程名称、先修课程号、主讲教师等数据项组成;学生选课

学生文件S的记录结构

学号	姓名	性别	出生日期	所在系
----	----	----	------	-----

课程文件C的记录结构

课程编号	课程名	先修课程号	主讲教师
------	-----	-------	------

学生选课文件SC的记录结构

学号	姓名	课程编号	成绩
----	----	------	----

图 1-3 文件系统中的文件组成结构

文件 SC 的记录可由学号、姓名、课程编号、成绩等数据项组成,如图 1-3 所示。在实际应用中这三个文件的记录之间是有联系的,文件 SC 中的某一记录的“学号”数据项的值应该是文件 S 中某个记录的“学号”数据项的值,文件 SC 中的某一记录的“课程编号”数据项的值应该是文件 C 中某个记录的“课程编号”数据项的值。

但在文件系统中,尽管记录内部是有结构的,但记录之间是没有联系的。学生、课程以及学生选课文件是三个独立的文件,可能是三个不同的用户分别创建的。即使是在同一用户创建的文件 S 和 SC 中,对于学号相同的记录,若其姓名不同,对于系统来说是正常现象;甚至是在同一文件 SC 中,出现同一学生选修同一课程却有不同成绩的记录都是正常现象;若是三个不同用户创建的三个文件,数据间更没有必然的联系,创建文件 SC 的用户无法知道对应某一课程编号的课程的课程名等信息,或对应某一学号的学生性别等信息,如果用户有需要,只能重复存储。而若要使这三个文件中的数据有联系,且相应的数据保持一致,必须再编写相应的应用程序来保证。

2) 数据与程序的独立性差

在文件系统管理阶段中,数据仍然是面向应用的。数据文件中只存储数据,不存储文件记录的结构描述信息。应用程序不仅要确定数据的逻辑结构,还要设计数据的物理结构。应用程序与数据的逻辑结构和物理结构仍然不是相互独立的。一旦数据的逻辑结构或物理结构需要改变,必须修改应用程序,即修改程序中有关数据逻辑结构或物理结构的定义。比如,扩展线性表中记录的数据项,或修改数据项的数据表示法等逻辑结构的改变,或为了提高应用程序的执行效率,而将数据文件由顺序存储改为链式存储等物理结构的改变,必须对应用程序进行修改。因此数据和程序之间仍缺乏数据独立性。

3. 数据库系统管理阶段

20 世纪 60 年代后期以来,计算机管理的数据对象规模越来越大,应用范围也越来越广泛,数据量急剧膨胀,对数据处理的速度和共享性提出了新的要求,对多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。

这个时期外存有了大容量磁盘、光盘,硬件价格大幅度下降;相反,软件的价格不断上升,编制和维护软件及应用程序的成本相对增加,其中维护的成本更高。数据处理上,联机实时处理要求更高,并开始出现分布处理。以文件系统进行数据管理已不能适应数据库管理的需要,为解决多用户、多应用共享数据的需求,使数据为尽可能多的应用服务,数据库系统管理技术应运而生了。

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,美国一家火箭公司利用该系统帮助设计了阿波罗登月宇航器,推动了数据库技术的产生。1968年,美国 IBM 公司研发了基于层次模型的数据库系统 IMS (Information Management System)。1969年,美国数据系统语言研究会 (COncference on DAta SYstem Language, CODASYL) 下属的数据库任务组 (DataBase Task Group, DBTG) 提出基于网状数据模型的一个系统方案。1970年,美国 IBM 的 E. F. Codd 发表论文提出了关系模型。从此,数据库技术进入了蓬勃发展的时期。

1.2 数据库的基本概念

1.2.1 数据库

据有关文献记载,数据库(database)这个名词起源于20世纪50年代初,当时美国把为了战争需要而存储在计算机里的各种情报集合称为 database。数据库,顾名思义,就是存放数据的仓库。只不过这个仓库是在计算机存储设备上,而且数据不是杂乱无章的,是按一定格式存放的。

严格地说,数据库是长期存储在计算机内、有组织的、统一管理的、可共享的相关数据的集合。

数据库在实际使用中具有如下隐含的性质:

(1) 数据指的是用符号记录下来的、可以识别的信息,具有一定的语义。所以数据库应反映现实世界的某些方面,现实世界的变化应在数据库中有所反映。

(2) 数据库是逻辑上一致而且有某种内在含义的数据集合,不是数据的随机归类。

(3) 数据库是为一个特定目标而设计、构建并装入数据的。数据库有目标用户,而且存在这些用户感兴趣的一些预想应用。

由数据库的性质,使得采用数据库技术进行数据管理呈现如下几个方面的特点。

1) 采用数据模型表示数据

在文件系统中,尽管每个文件内部是有结构的(文件由记录构成,每个记录由若干数据项组成),但文件之间是没有联系的。数据文件中只存储数据,不存储文件记录的结构描述信息。例如图 1-3 中的学生、课程以及学生选课文件是三个独立的数据文件。

而在数据库系统中,采用数据模型来描述数据库中的数据,数据模型不仅描述数据本身的特征,还要描述数据之间的联系。例如,在关系数据库中,采用关系数据模型在一个关系数据库中用三个关系表来描述学生、课程以及学生选课文件中的数据,关系表中的元组和属性对应文件系统中数据文件中的记录和数据项;通过定义学生关系表中“学号”属性和课程关系表中“课程编号”属性为主属性(主键),学生选课关系表中“学号”和“课程编号”属性为参照属性(外键),以及参照属性与主属性的对应关系,来描述数据之间的联系,如图 1-4 所示。由数据库管理系统(DBMS)来实现关系数据模型的实体完整性和参照完整性,保证各关系表(文件)中元组的唯一性、相关数据的一致性、相同数据不用重复存储