

新

世纪高等学校计算机系列教材



数据库原理与技术

刘方鑫 主编
罗昌隆 刘同明 副主编
孙志挥 主审



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

URL: <http://www.phei.com.cn>

新世纪高等学校计算机系列教材

数据库原理与技术

刘方鑫	主 编
罗昌隆 刘同明	副主编
孙志挥	主 审

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书是一本高等院校计算机专业或相关专业的数据库课程教学用书,它以关系数据库系统为核心系统,完整地论述了数据库系统的基本概念、基本原理和应用技术,力图使读者对数据库系统有一个全面、深入、系统的了解,为进一步从事数据库系统的研究、开发和应用奠定坚实的基础。

本书主要内容包括:数据库系统的产生与发展、数据库系统特点、数据库系统的数据模型、数据库系统结构、关系数据库方法、关系数据库标准语言 SQL、数据依赖和关系规范化理论、数据库保护、数据库设计、分布式数据库系统和数据库技术的新发展等。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究。

图书在版编目(CIP)数据

数据库原理与技术/刘方鑫主编. —北京:电子工业出版社,2002.2

(新世纪高等学校计算机系列教材)

ISBN 7-5053-7163-0

I. 数... II. 刘... III. 关系数据库—数据库系统—高等学校—教材—IV. TP311.13

中国版本图书馆 CIP 数据核字(2001)第 090182 号

责任编辑:李 影

印 刷:北京牛山世兴印刷厂

出版发行:电子工业出版社 <http://www.phei.com.cn>

北京市海淀区万寿路 173 信箱 邮编 100036

经 销:各地新华书店

开 本:787×1092 1/16 印张:20 字数:512 千字

版 次:2002 年 2 月第 1 版 2002 年 5 月第 2 次印刷

印 数:6 000 册 定价:25.00 元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。
联系电话:(010)68279077

《新世纪高等学校计算机系列教材》组织结构

顾问 (以汉语拼音为序)

顾冠群 (中国工程院院士, 东南大学校长)

李 未 (中国科学院院士, 国务院学位委员会委员, 北京航空航天大学教授)

陆汝钤 (中国科学院院士, 中国科学院软件所研究员)

潘云鹤 (中国工程院院士, 浙江大学校长)

孙钟秀 (中国科学院院士, 南京大学教授)

王 珉 (教授, 博导, 江苏省人民政府副省长)

王 荣 (研究员, 江苏省教育厅厅长)

杨芙清 (中国科学院院士, 北京大学教授)

周兴铭 (中国科学院院士, 国防科技大学教授)

11/15/05

编辑委员会

主任 徐宝文

副主任 (以汉语拼音为序)

史忠植

陈道蓄

文宏武

委员

蔡家楣

陈 纯

陈道蓄

程自强

胡学龙

付育熙

何炎祥

怀进鹏

金远平

鞠时光

刘方鑫

吕 建

梅 宏

钱培德

秦小麟

瞿裕忠

史忠植

宋方敏

王 茜

王怀民

王汝传

王士同

王绍棣

王志坚

文宏武

夏士雄

徐宝文

徐汀荣

徐福培

须文波

张 宏

张 明

张茂坤

张荣琴

张岳新

张在跃

周傲英

总 序

为了认真贯彻《中国教育改革和发展纲要》和教育部“面向 21 世纪教育振兴行动计划”精神,适应现代社会、经济、科技、文化,特别是教育的发展方向,适应培养新世纪计算机人才的需要,根据计算机科学技术学科发展和新世纪高等学校教学内容和课程体系改革的要求,我们决定编写这套《新世纪高等学校计算机系列教材》。

教学改革和教学质量是高等学校的头等大事。教学改革的关键在于教材的改革,我们组织编写这套系列教材的指导思想是:立足于新世纪培养创造型开拓型人才的培养目标,立足于高速发展的计算机科学技术,科学地预测本学科前沿发展趋势;注重教材内容的思想性、科学性、先进性、系统性和广泛的适应性,继承与改革现有教材内容和体系;注重加强能力的培养,切实提高学生的综合素质。同时,为了适合我国国情,适应绝大多数高等学校计算机科学与技术专业类的教师知识结构和课程改革的实际情况,一方面,我们尽力拓宽教材内容,充实和加强有关理论,引用新的科技成果;另一方面,我们兼顾各门课程的历史、现状与发展,在具体章节的内容编排上力求循序渐进,顺理成章。在保证学科系统的前提下,充分考虑到学生的接受能力,使教材体系尽量符合学生的认知规律,便于学生接受、理解、掌握和巩固所学知识,促使学生学以致用,理论密切联系实际。

本套系列教材的组织编写得到了中国计算机学会、江苏省计算机学会的大力支持和帮助,凝聚了参加编审出版工作的许多教师和编辑的心血和汗水,在此一并对他们表示衷心的感谢。

限于我们的水平、能力和经验,本套系列教材的编审与出版工作一定还存在着很多错漏和其他不如意之处,我们真诚地希望使用这套教材的高校教师、学生和广大读者积极提出宝贵意见,以鞭策我们不断提高教材的编写、修订和出版质量,为我国的计算机科学技术专业教材建设服务。

《新世纪高等学校计算机系列教材》编辑委员会

前 言

数据库技术是 20 世纪 60 年代后期产生和发展起来的一项计算机数据管理技术,它的出现和发展使计算机应用渗透到人类社会的广阔领域。目前数据库的建设规模和性能、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。

从 20 世纪 70 年代后期开始,国外许多大学把数据库原理与技术列为计算机专业的学习内容,我国高等院校从 80 年代开始也把数据库原理与技术作为计算机专业的主要课程之一。目前数据库技术已成为信息科学技术领域的重要基础。

本书是一本高等院校计算机专业或相关专业的数据库课程教学用书,它以关系数据库系统为核心系统,完整地论述了数据库系统的基本概念、基本原理和应用技术,力图使读者对数据库系统有一个全面、深入、系统的了解,为进一步从事数据库系统的研究、开发和应用奠定坚实的基础。

本书主要内容包括:数据库系统的产生与发展、数据库系统特点、数据库系统的数据模型、数据库系统结构、关系数据库方法、关系数据库标准语言 SQL、数据依赖和关系规范化理论、数据库保护、数据库设计、分布式数据库系统和数据库技术的新发展等。

本书主要特点:

(1)以关系数据库系统为核心。在系统论述数据库基本知识的基础上,着重讨论了关系数据库的原理与实现,其中对关系数据模型、关系数据库体系结构、关系规范化理论、数据库设计方法、数据库并发控制、查询优化等都有较详细、系统的说明,对层次数据库和网状数据库等传统数据库的内容进行了精简,仅对其数据模型作了简要介绍。

(2)力求反映当前数据库领域的新水平。根据数据库领域的新发展,加强了数据模型和数据表示、分布式数据库系统、面向对象数据库、主动数据库、并行数据库、工程数据库、数据仓库以及基于 Web 的数据库系统等内容,力求反映当前数据库发展的新技术、新水平和新趋势。

(3)注重理论联系实际,加强数据库应用技术。为适应信息技术发展的需要,对数据库设计方法、开发工具、数据库语言(SQL)等数据库应用技术进行了较全面的论述,为读者开发数据库应用系统或维护管理大中型数据库系统打下基础。

(4)在内容选取、章节安排、难易程度等方面充分考虑教学的需要,力求使教材概念准确、清晰、重点明确,内容广泛,便于取舍,每章均配有习题便于教学。

本书由刘方鑫、罗昌隆、刘同明、徐明、潘江波、李毅、吴士军编写。刘方鑫任主编,罗昌隆、刘同明任副主编。其中,刘方鑫编写第 1 章、第 2 章、第 3 章、第 10 章、第 12 章的 12.1~12.4 节;罗昌隆编写第 8 章、第 9 章、第 12 章的 12.5~12.10 节;刘同明编写第 4 章、第 5 章、第 6 章;徐明编写第 7 章;潘江波编写第 4、10 章的部分章节;李毅编写第 1 章、第 2 章、第 3 章的部分章节及负责全书的修改工作,最后由刘方鑫对全书进行了统稿。

孙志挥教授审阅了全书,并提出许多宝贵意见,编者在此表示诚挚的谢意。

由于编者水平有限,书中疏漏谬误之处在所难免,殷切希望广大读者批评指正。

编 者

2001 年 9 月

目 录

第 1 章 数据库系统引论	(1)
1.1 数据库系统的产生与发展	(1)
1.1.1 数据和数据管理技术	(1)
1.1.2 数据库系统发展的三个里程碑	(3)
1.1.3 数据库系统的三个发展阶段	(4)
1.2 数据库与数据库系统	(5)
1.2.1 数据库	(5)
1.2.2 数据库系统	(6)
1.3 数据库系统的特点	(6)
1.3.1 实现数据的集中化控制	(6)
1.3.2 数据的冗余度小	(7)
1.3.3 采用一定的数据模型实现数据结构化	(7)
1.3.4 避免了数据的不一致性	(7)
1.3.5 实现数据共享	(8)
1.3.6 提供数据库保护	(8)
1.3.7 数据独立性	(8)
1.3.8 数据由 DBMS 统一管理和控制	(8)
习题 1	(9)
第 2 章 数据库系统的数据模型	(10)
2.1 数据模型概述	(10)
2.1.1 数据模型的基本组成	(10)
2.1.2 数据模型的发展	(11)
2.2 概念模型	(11)
2.2.1 E-R 数据模型的基本概念	(12)
2.2.2 E-R 图(E-R Diagram)	(14)
2.2.3 扩充的 E-R 数据模型	(15)
2.3 层次数据模型	(17)
2.3.1 层次数据模型的数据结构	(18)
2.3.2 层次数据模型的数据操作	(18)
2.3.3 层次数据模型的数据约束	(20)
2.3.4 层次数据模型的优缺点	(21)
2.4 网状数据模型	(21)
2.4.1 网状数据模型的数据结构	(21)
2.4.2 网状数据模型的数据操作	(23)
2.4.3 网状数据模型的约束	(25)

2.4.4	网状数据模型的优缺点	(25)
2.5	关系数据模型	(26)
2.5.1	关系数据模型的数据结构	(26)
2.5.2	关系数据模型的数据操作	(27)
2.5.3	关系数据模型的数据约束	(28)
2.5.4	关系数据模型的优缺点	(29)
2.6	面向对象数据模型	(29)
2.6.1	对象和对象标识符	(30)
2.6.2	属性和方法	(30)
2.6.3	封装和消息传递	(31)
2.6.4	类和实例	(31)
2.6.5	类层次结构和继承	(32)
2.6.6	持久性和版本	(33)
2.6.7	多态、重载、重定义与动态联编	(33)
2.6.8	面向对象数据模型与关系数据模型的简单比较	(34)
	习题 2	(34)
第 3 章	数据库系统结构	(36)
3.1	数据库系统的分级结构	(36)
3.1.1	数据库系统的三级模式结构	(36)
3.1.2	数据库模式间的映像与数据独立性	(38)
3.1.3	数据库的抽象层次	(39)
3.1.4	数据库的数据模式与数据模型的关系	(40)
3.2	数据库系统的组成	(40)
3.2.1	硬件支撑环境	(40)
3.2.2	软件系统	(41)
3.2.3	数据库	(41)
3.2.4	数据库管理员、系统分析员、程序设计员及用户	(42)
3.3	数据库管理系统的功能	(43)
3.4	DBMS 的系统结构	(44)
3.4.1	DBMS 的进程结构和多线索结构	(45)
3.4.2	DBMS 的组成	(45)
3.4.3	DBMS 的层次结构	(47)
3.5	DBMS 的语言翻译处理层	(48)
3.5.1	DDL、DCL、DML	(48)
3.5.2	语言翻译处理层工作过程	(50)
3.6	DBMS 的数据存取层	(51)
3.6.1	数据存取层的主要任务	(51)
3.6.2	数据库存取层的功能子系统	(51)
3.7	Client/Server 结构的 DBMS	(52)
3.7.1	Client/Server 结构的 DBMS 的功能划分	(52)

3.7.2	在 Client/Server 结构中数据存取的工作过程	(53)
3.7.3	Client/Server 结构 DBMS 中数据库服务器的优点	(54)
3.7.4	三层 Client/Server 结构	(54)
3.8	基于 Web 的数据库系统	(55)
3.8.1	基于 Web 的数据库系统的体系结构	(55)
3.8.2	CGI 和 API 技术	(56)
	习题 3	(58)
第 4 章	关系数据库方法	(59)
4.1	关系数据库的基本概念	(59)
4.1.1	关系的形式化定义	(59)
4.1.2	关系模式	(60)
4.1.3	关系数据库模式	(60)
4.2	关系数据库的用户接口和关系数据语言	(61)
4.3	关系代数	(62)
4.3.1	传统的集合运算	(63)
4.3.2	专门的关系运算	(64)
4.3.3	关系代数应用举例	(68)
4.4	关系演算	(68)
4.4.1	元组关系演算	(69)
* 4.4.2	域关系演算	(72)
4.5	关系系统	(75)
4.5.1	关系系统	(75)
* 4.5.2	全关系系统的十二条基本准则	(76)
4.6	关系系统的查询优化	(78)
4.6.1	查询优化的一般策略	(79)
4.6.2	代数优化	(79)
4.6.3	基于存取路径的规则优化	(85)
4.7	关系数据操纵语言的比较	(92)
	习题 4	(93)
第 5 章	关系数据库的标准语言 SQL	(99)
5.1	SQL 语言概述	(99)
5.2	关系模式的定义	(100)
5.2.1	基本表的定义	(100)
5.2.2	基本表的修改和删除	(104)
5.2.3	关于默认值	(104)
5.2.4	域	(105)
5.2.5	索引的建立和删除	(105)
5.3	数据查询	(106)
5.3.1	简单查询	(106)
5.3.2	复杂查询	(108)

5.4	聚集函数(Aggregation)	(113)
5.4.1	聚集函数的运算符	(113)
5.4.2	元组的分组	(113)
5.4.3	HAVING 子句	(114)
5.5	数据库更新	(114)
5.5.1	元组插入	(114)
5.5.2	元组删除	(115)
5.5.3	元组修改	(115)
5.6	视图	(116)
5.6.1	视图定义	(116)
5.6.2	视图查询	(117)
5.6.3	视图更新	(118)
5.6.4	视图删除	(119)
5.6.5	视图的作用	(119)
5.7	嵌入式 SQL	(120)
5.7.1	程序设计环境下的 SQL	(120)
5.7.2	单行选择语句	(122)
5.7.3	游标	(123)
* 5.8	动态 SQL	(126)
5.8.1	非查询类动态 SQL	(126)
5.8.2	查询类动态 SQL	(128)
	习题 5	(129)
第 6 章	数据依赖和关系规范化理论	(133)
6.1	为什么要规范化	(133)
6.2	函数依赖	(135)
6.2.1	函数依赖的定义	(135)
6.2.2	关系的关键字和超关键字	(137)
6.3	函数依赖理论	(138)
6.3.1	函数依赖的逻辑蕴涵	(138)
6.3.2	Armstrong 公理系统	(139)
6.3.3	属性集闭包	(141)
* 6.3.4	Armstrong 公理的正确性和完备性	(142)
6.4	函数依赖集的等价和覆盖	(144)
6.5	关系模式的分解	(146)
6.5.1	什么叫模式分解	(146)
6.5.2	分解的连接不失真性	(148)
6.5.3	依赖保持性	(151)
6.5.4	模式分解的级别——范式	(152)
6.5.5	模式分解算法	(154)
6.6	多值函数依赖与第四范式	(157)

6.6.1	BCNF 关系模式存在的问题	(157)
6.6.2	多值函数依赖的概念	(158)
6.6.3	第四范式	(158)
* 6.6.4	多值依赖公理系统和多值依赖中的模式分解	(159)
习题 6	(163)
第 7 章	数据库设计	(169)
7.1	数据库设计概述	(169)
7.1.1	数据库设计目标和设计方法	(169)
7.1.2	数据库设计的基本步骤	(171)
7.2	需求分析	(172)
7.2.1	需求分析的任务	(172)
7.2.2	需求分析的实现方法	(173)
7.2.3	需求分析实例	(174)
7.3	数据库概念结构设计	(177)
7.3.1	概念结构设计的引出	(177)
7.3.2	概念结构设计的要求及方法	(177)
7.3.3	局部概念结构设计	(179)
7.3.4	全局概念结构设计	(181)
7.3.5	概念结构设计实例	(184)
7.4	数据库逻辑结构设计及优化	(186)
7.4.1	逻辑结构设计任务与设计准则	(186)
7.4.2	概念模型向关系模型的转换规则	(188)
7.4.3	用关系规范化理论对关系数据模型进行优化	(190)
7.5	数据库的物理设计	(192)
7.5.1	数据库设计人员需掌握的物理设计知识	(192)
7.5.2	数据库物理设计的主要内容	(192)
7.6	数据库设计评价	(195)
7.6.1	数据库评价	(195)
7.6.2	数据库逻辑设计评价和物理设计评价	(196)
7.7	数据库的运行和维护	(197)
7.7.1	数据的载入和应用程序的设计与调试	(197)
7.7.2	数据库的试运行	(198)
7.7.3	数据库的运行和维护	(199)
习题 7	(199)
第 8 章	数据库安全性和完整性	(201)
8.1	数据库安全性概念	(201)
8.2	数据库安全性控制	(202)
8.2.1	用户标识与鉴定	(202)
8.2.2	存取控制	(202)
8.2.3	视图机制	(205)

8.2.4	数据加密	(205)
8.2.5	审计	(205)
8.3	数据库安全性实例——Oracle 系统的安全性	(206)
8.4	数据库完整性	(208)
8.4.1	完整性约束条件的类型	(208)
8.4.2	完整性控制机制的功能	(209)
8.4.3	完整性约束的表达方式	(209)
8.4.4	Oracle 系统的完整性约束表示	(217)
习题 8	(219)
第 9 章	数据库系统的恢复和并发控制技术	(221)
9.1	事务	(221)
9.2	数据库恢复	(222)
9.2.1	故障的类型	(222)
9.2.2	数据库恢复技术	(223)
9.2.3	数据库恢复策略	(225)
9.3	并发控制	(226)
9.3.1	并发操作引起的问题	(226)
9.3.2	调度的可串行性	(227)
9.4	基于封锁的并发控制技术	(229)
9.4.1	X 锁	(229)
9.4.2	(S,X)锁	(230)
9.4.3	(S,U,X)锁	(230)
9.4.4	两段锁协议	(231)
9.4.5	严格的 2PL 实施	(233)
9.5	死锁的检测、处理和防止	(233)
9.5.1	死锁预防	(233)
9.5.2	死锁的检测和处理	(234)
* 9.6	封锁的粒度和多粒度的封锁	(234)
9.6.1	封锁的粒度	(234)
9.6.2	多粒度封锁	(235)
* 9.7	基于时间标记的并发控制技术	(236)
习题 9	(238)
第 10 章	分布式数据库系统	(239)
10.1	分布式数据库系统概述	(239)
10.1.1	什么是分布式数据库系统	(239)
10.1.2	分布式数据库系统的产生与发展	(240)
10.2	分布式数据库系统的特点	(243)
10.2.1	“全能”分布式数据库系统应符合的准则	(243)
10.2.2	分布式数据库系统的特点	(244)
10.3	分布式数据库管理系统 (DDBMS)	(245)

10.3.1	DDBMS 的定义和功能	(245)
10.3.2	DDBMS 的组成	(246)
10.3.3	DDBMS 的分类	(247)
10.4	分布式数据库系统的体系结构	(249)
10.4.1	基于 ANSI/SPARC 数据模式的 DDBS 的体系结构	(249)
10.4.2	分布透明性	(252)
10.5	数据分段的规则和类型	(254)
10.5.1	数据分段的目的和规则	(254)
10.5.2	数据分段的类型	(254)
10.5.3	数据分布对数据管理的影响	(256)
* 10.6	分布式查询和优化	(257)
10.6.1	分布式查询处理	(257)
10.6.2	全局查询优化的目标	(258)
* 10.7	分布式事务管理	(258)
10.7.1	分布式事务的基本特征和管理目标	(259)
10.7.2	分布式事务的恢复技术	(259)
10.7.3	分布式数据库中的可串行性	(261)
* 10.8	分布式数据库系统中的并发控制	(262)
10.8.1	基于封锁的并发控制算法	(262)
10.8.2	基于时间戳的并发控制算法	(263)
10.9	异构数据库互连	(265)
10.9.1	异构型 DDBMS	(266)
10.9.2	SAG 规范和 IBM 的 DRDA	(266)
10.9.3	开放型数据库互连 ODBC 的结构	(267)
10.9.4	利用信关技术实现异构数据库互连	(269)
习题 10	(270)
第 11 章	数据库技术的新发展	(271)
11.1	传统数据库系统的局限性	(271)
11.2	新一代数据库系统	(272)
11.2.1	面向对象数据库系统宣言	(272)
11.2.2	第三代数据库系统宣言	(273)
11.2.3	数据库技术与相关技术结合形成新型的数据库系统	(274)
* 11.3	面向对象数据库系统	(275)
11.3.1	面向对象数据库系统概述	(275)
11.3.2	面向对象数据库管理系统 OODBMS	(275)
11.3.3	面向对象数据库系统的基本功能和实现途径	(276)
* 11.4	对象-关系数据库系统	(277)
11.4.1	对象-关系数据库系统的特点	(277)
11.4.2	对象-关系数据库系统实现的方法和途径	(278)
* 11.5	工程数据库系统	(278)

11.5.1	工程数据库系统的特点	(278)
11.5.2	工程数据库的数据模型	(279)
11.5.3	工程数据库的体系结构	(281)
* 11.6	并行数据库系统	(282)
11.6.1	并行数据库系统概述	(282)
11.6.2	并行数据库系统结构	(283)
11.6.3	数据划分	(285)
11.6.4	并行算法	(285)
11.6.5	查询优化方法	(287)
11.6.6	并行数据库系统与分布式数据库系统比较	(288)
* 11.7	主动数据库系统	(289)
11.7.1	主动数据库概述	(289)
11.7.2	主动数据库的模型	(289)
11.7.3	SQL3 中规则表示和执行	(290)
11.7.4	主动数据库管理系统	(291)
* 11.8	数据仓库(Data Warehouse)	(293)
11.8.1	数据仓库的产生	(293)
11.8.2	数据仓库的概念和结构	(294)
11.8.3	联机分析处理(OLAP)	(295)
11.8.4	OLAP 实现技术	(298)
	习题 11	(301)
	参考文献	(302)

第 1 章 数据库系统引论

数据库是数据文件以及用于处理这些数据文件的程序集合,是数据管理的最新技术。数据库技术已成为现代信息技术的重要组成部分,是现代计算机信息系统和计算机应用系统的基础和核心。数据库技术是 20 世纪 60 年代后期产生和发展起来的一项计算机数据管理技术,它的出现使计算机应用渗透到人类社会的广阔领域。目前数据库的建设规模和性能、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志,数据库技术也成为计算机科学技术学科的一个重要分支。

本章主要讨论数据库系统的产生与发展,以及数据库系统的基本概念和特点。

1.1 数据库系统的产生与发展

1.1.1 数据和数据管理技术

1. 数据

数据是载荷信息的媒体,它包括数值型数据和非数值型数据。数值型数据是以数字表示信息,而非数值型数据是以符号及其组合来表示信息。例如字符、文字、图表、图形、图像、声音等均属于非数值型数据。

数据是人类社会发展的一种重要的信息资源,如何有效地保存和科学地管理这些数据是人们长期以来十分关注的课题,从而促进了数据管理技术的发展。

2. 数据管理技术

数据管理是指对数据的分类、组织、编码、存储、查询和维护等活动,是数据处理的中心环节。数据管理技术可根据提供的数据独立性、数据冗余度、数据共享性、数据间相互联系、数据安全性、数据完整性和数据存取方式等水平的高低划分为三个不同的发展阶段。

(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,数据管理处于人工管理阶段,其特点是:

- 数据管理缺乏软件系统支持,依靠应用程序管理数据。在应用程序中程序员不仅要规定数据的逻辑结构,而且还要设计数据的物理结构,包括存储结构、存取方法、输入方式等。
- 数据是面向应用的,一组数据只能对应一个应用程序,数据不能共享。
- 应用程序完全依赖于数据,不具有数据独立性,一旦数据的逻辑结构或物理结构发生变化,应用程序就要做相应的修改。
- 数据一般不需要长期保存。

(2) 文件系统阶段

20 世纪 50 年代后期至 60 年代中期,数据管理进入文件系统阶段。在文件系统中,把数据按其内容、结构和用途组织成若干个相互独立的文件。用户通过操作系统对文件进行打开、读写、关闭等操作。文件系统管理数据具有如下特点:

- 文件系统利用“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、插入和删除操作。数据存取直接由操作系统(OS)提供支持,即采用“应用程序—OS—数据文件”的存取方式。

- 数据可以长期保存在存储设备上供用户使用。

尽管用文件系统管理数据比人工管理阶段有了长足的进步,但面对数据量大且结构复杂的数据管理任务,文件系统仍不能适应,主要表现在如下几点:

- 数据独立性差。文件系统中文件逻辑结构的改变必须修改应用程序;由于语言环境的变化要求修改应用程序时,也将引起文件数据结构的改变,因此数据与程序间仍缺乏数据独立性。

- 数据冗余度大且易产生数据不一致性。在文件系统中,文件一般为某一用户或用户组所有,文件仍是面向应用的,因此数据共享性差,冗余度大。同时由于相同数据重复存储、各自管理,易产生数据的不一致性。

- 在文件系统中,尽管其记录内部有结构,但记录之间没有联系,故其整体是无结构的。

- 数据无集中管理,其安全性、完整性得不到可靠保证,并且在数据的结构、编码、输出格式等方面难以做到规范化和标准化。

- 文件系统一般不支持多个应用程序对同一文件的并发访问,故数据处理的效率较低。

- 使用方式不够灵活。每个已经建立的数据文件只限于一定的应用,且难于对它进行修改和扩充。

这些缺点在规模大、结构复杂的数据管理系统中尤为突出。例如,美国于 20 世纪 60 年代初制定了阿波罗登月计划。阿波罗飞船由二百万个部件组成,且这些部件是由分散在世界各地的若干厂家生产的。为了掌握工程进度和协调工程进展,阿波罗计划的主要合作者 Rockwell 公司,曾研制了一个基于磁带的零部件生产计算机文件管理系统。该文件管理系统的数据库冗余高达 60% 以上,且只能以批处理方式进行工作,系统维护也困难。这些问题一度成为实现阿波罗登月计划的障碍。

上述问题是文件系统本身难以解决的,造成数据处理效率低、成本高,并有许多潜在问题需要解决。这正是数据库系统产生的背景。

(3) 数据库系统阶段

20 世纪 60 年代后期以来,为了克服文件系统的弊病,适应日益迅速增长的数据处理的需求,人们开始探索新的数据管理方法与工具。这一时期,磁盘存储技术取得重要进展,大容量和快速存取的磁盘相继投入市场,给新型数据管理技术的研制提供了良好的物质基础。为了解决多用户、多应用共享数据的需求,使数据为尽可能多的应用服务,数据库技术由此应运而生,出现了统一管理数据的专门软件系统——数据库管理系统(Data Base Management System—DBMS)。世界上最早推出的数据库系统当属美国通用电气公司(GE)C. W. Bachman 等人于 1963 年研制的 IDS(Integrated Data Store)系统。IDS 奠定了网状数据库的基础,它可使多个 COBOL 程序共享数据库中的数据。IDS 是数据库系统的先驱,为此,Bachman 于 1973 年获得了美国计算机协会(ACM)颁发的图灵(Turing)奖。

数据库(Data Base—DB)一词的提出可以追溯到 20 世纪 50 年代后期。据日本的铃木道夫说:“那时,美国军队为了作战和装备,想把各种数据集中在一个地方作为数据基地。于是就把各个部队分散的情报集中起来管理,根据统一取得的最新数据考虑战略部署。据说,这就是数据库术语的起源”。后来美国系统发展公司为海军基地研制数据库借用了 Base(基地)一词,