

21世纪大学计算机专业教材

数据库系统原理与设计

李俊山 孙满囤 韩先锋 李艳玲 编著
沈钧毅 沈绪榜 主审

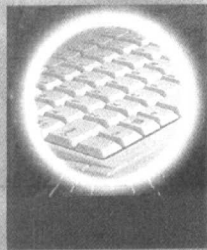
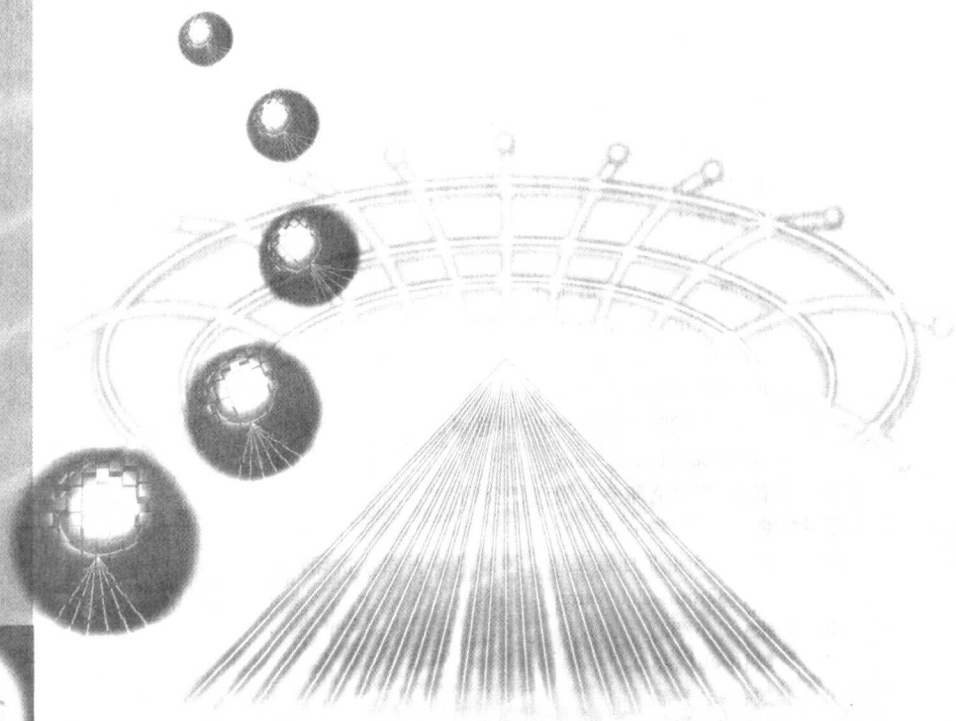


西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

21世纪大学计算机专业教材

数据库系统原理与设计

李俊山 孙满囤 韩先锋 李艳玲 编著
沈钧毅 沈绪榜 主审



西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS

· 西安 ·

内 容 简 介

本书较全面地介绍了数据库系统的基本概念、基本原理和基本设计方法。全书共分为 11 章,内容包括:数据库系统概述,关系运算,关系数据库语言 SQL,改善数据库的查询性能,关系数据库设计理论,数据库管理系统,数据库的保护、控制与实现技术,数据库应用系统设计,高等数据库体系结构与访问技术,分布式数据库系统,数据库技术的新发展。基本上覆盖了关系数据库系统原理与设计的主要内容。

本书内容选材新颖精练,理论与实践并重,系统性强。表述通俗,图文并茂,并配有大量的习题。

本书可作为高等院校大学本科计算机及信息类专业的数据库课程教材,也可作为电类和信息类专业研究生的数据库课程教材,还可供从事计算机软件工作的科技人员和工程技术人员及其他有关人员参考。

图书在版编目(CIP)数据

数据库系统原理与设计/李俊山,孙满囤,韩先锋,李艳玲编著. — 西安:西安交通大学出版社, 2003.9

(21 世纪大学计算机专业教材)

ISBN 7-5605-1720-X

I. 数… II. ①李… ②孙… ③韩… ④李…
III. 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(2002)第 095553 号

书 名:数据库系统原理与设计

编 著:李俊山 孙满囤 韩先锋 李艳玲

策划编辑:贺峰涛 屈晓燕

文字编辑:宗立文

出版发行:西安交通大学出版社

地 址:西安市兴庆南路 25 号(邮编:710049)

网 址:<http://unit.xjtu.edu.cn/unatjtupress>

电 话:(029)2668357,2667874(发行部)

(029)2668315,2669096(总编部)

电子信箱:eibooks@163.com

印 刷:陕西友盛印务有限责任公司

经 销 者:各地新华书店

版 次:2003 年 9 月第 1 版 2003 年 9 月第 1 次印刷

开 本:787mm×1 092mm 1/16

印 张:20.625

印 数:0 001~3 000

字 数:495 千字

书 号:ISBN 7-5605-1720-X/TP·337

定 价:28.00 元

版权所有,翻版必究!

前 言

数据库技术产生于 20 世纪 60 年代末。经过 30 多年的迅猛发展,已经形成了较为完整的理论与技术体系,并已成为计算机科学技术中的一个重要分支。

随着数据库技术的迅猛发展,数据库已经成为各类管理信息系统(MIS)、决策支持系统(DSS)、地理信息系统(GIS)、办公信息系统(OIS)、计算机辅助设计(CAD)、计算机辅助制造(CAM)、计算机集成制造系统(CIMS)等各类信息系统的核心技术,并在商业、行政管理、工农业生产、文化教育、科学研究、国防和军事等领域得到了十分广泛的应用,成为信息化社会中的最重要的支撑技术之一。也因此使数据库及其设计技术受到了人们的普遍关注。数据库不但早已是高等院校计算机、信息工程、管理信息系统等专业的骨干专业课之一,而且越来越多的专业及相关技术人员都要求具有数据库知识。

在近 20 年的数据库教学和科研工作中,有一个问题一直萦绕在我的脑海中,这就是采用什么样的内容组织和选材方式,采用什么样的教学思路,才能使数据库这门课程的内容既能体现其坚实的理论基础,又能突出数据库系统的总体概念和设计技术?如何才能使学习者从抽象的规范化与函数依赖理论中理解数据库设计方法与设计技术的内涵?本书将作者近 20 年的数据库教学和科研实践经验,与国内外各种数据库教科书在教学内容组织与选材方面的成功经验相结合,比较好地回答了这些问题。从总体上讲,本书在内容的组织和选材上具有以下特色:

(1)着眼于数据库的基本理论、基本设计技术和基本设计方法。在内容选取上以关系数据库为主,并对分布式数据库、数据库技术的新发展做了恰如其分的介绍。

(2)从加深理解数据库的工作机理和应用数据库设计角度出发,比较全面详细地介绍了数据库查询语言 SQL,特别是比较详细地介绍了应用程序的设计技术,并给出了若干个完整的应用程序实例。对理解和掌握数据库应用系统的设计具有促进作用。

(3)在关系数据库设计理论和应用数据库系统设计两章中较好地说明了如何将规范化理论和函数依赖理论用于数据库设计的问题;并在数据库应用系统设计一章中首次给出了关系数据模型的规范化设计方法和关系数据模型的优化设计方法。对于数据库规范化理论的理解及规范化理论在数据库应用系统设计中的应用具有重要的指导意义。

(4)改变了以往的对数据库存储组织的描述思路,首次从提高数据库查询性能的角度出发介绍了索引技术和查询优化技术,略去了传统数据库存储组织中较低层次的内容。

(5)突出了数据库的设计技术,特别是对数据库概念模式设计中最常用到的 E-R 模型及其设计技术进行了比较详细的介绍,并在数据库应用系统设计一章中增加了 E-R 模型的操

作内容。为读者全面掌握数据库设计技术具有重要的意义。

(6)增加了高等数据库体系结构与访问技术内容,介绍了与数据库应用系统设计有关的C/S结构数据库系统、B/S结构数据库系统及ODBC技术,为全面理解数据库系统的体系结构和掌握网络环境下的数据库系统设计打下了良好的基础。

(7)在读者已经掌握了数据库基础、关系运算、SQL关系数据库语言、索引技术与查询优化、关系数据库设计理论等知识的基础上,首次开辟独立的一章(第6章)专门介绍数据库管理系统(DBMS)的内容,这对于促进读者对数据库系统总体概念的掌握,加深读者对DBMS软件系统的理解,提高读者的数据库应用系统设计能力和水平都具有重要的作用。

(8)书中配有大量例子,理论联系实际,便于学习者学习和理解。

本书共分为11章,第1章介绍数据库的基本概念,第2章介绍关系运算,第3章介绍关系数据库语言SQL,第4章介绍改善数据库查询性能的途径和技术,第5章介绍关系数据库设计理论,第6章介绍数据库管理系统的组成和功能,第7章介绍数据库的保护、控制与实现技术,第8章介绍数据库应用系统的设计过程和设计技术,第9章介绍高等数据库体系结构和访问技术,第10章介绍分布式数据库系统的概念、结构和技术方法,第11章介绍数据库技术的新发展。

本书可作为高等院校大学本科计算机及信息类专业的数据库课程教材,也可作为电类和信息类专业研究生的数据库课程教材。在讲授时可根据教学需要和学习者的基础对内容作适当取舍。本书教学参考学时数为60学时至70学时,其中标有“*”的内容可以不讲。

本书的第1章至第5章、第8章、第9章、第11章内容由李俊山编写,第6章内容由李艳玲编写,第7章内容由韩先锋和李俊山共同编写,第10章内容由孙满囤和李俊山共同编写。

西安交通大学沈钧毅教授和西北工业大学沈绪榜院士仔细审阅了全稿,并提出了许多宝贵意见,在此向他们表示真诚的感谢。特别感谢西安交通大学出版社贺峰涛和屈晓燕编辑对本教材出版的关心和精心编辑。

由于水平有限,书中难免有不当之处,敬请广大读者和专家批评指正。

李俊山

2003年元旦于砺剑园

(电子邮件:ep_lijunshan@163.net)

目 录

第 1 章 数据库系统概述	(1)
1.1 引言	(1)
1.1.1 数据管理技术的发展	(1)
1.1.2 数据库技术的基本概念	(4)
1.2 数据描述	(5)
1.2.1 数据与信息	(5)
1.2.2 现实世界的描述	(6)
1.2.3 数据之间的联系	(7)
1.3 实体-联系模型	(8)
1.3.1 实体-联系模型的基本要素	(8)
1.3.2 实体-联系图	(10)
1.3.3 实体和属性的确定准则	(11)
1.3.4 实体-联系模型设计中的一些特殊情况	(12)
1.4 数据模型	(17)
1.4.1 层次模型	(17)
1.4.2 网状模型	(18)
1.5 关系模型	(20)
1.5.1 关系模型的基本概念	(20)
1.5.2 关系的定义	(22)
1.5.3 关系的键	(23)
1.5.4 实体-联系模型向关系模型的转换	(23)
1.6 数据库的结构	(25)
1.6.1 数据库的三级模式结构	(25)
1.6.2 概念模式	(26)
1.6.3 外模式	(27)
1.6.4 内模式	(28)
1.6.5 数据库的体系结构	(28)
1.7 数据语言	(30)
1.7.1 数据描述语言	(30)
1.7.2 数据操作语言	(30)
1.7.3 宿主系统	(30)
1.7.4 应用程序	(31)
习题 1	(31)

第 2 章 关系运算	(33)
2.1 关系代数	(33)
2.1.1 关系代数的五种基本运算	(33)
2.1.2 可以用五种基本运算表示的其它代数运算	(35)
2.1.3 扩充的关系代数运算	(39)
2.1.4 关系代数运算在关系数据库查询操作中的应用	(41)
2.2 关系演算	(43)
2.2.1 元组关系演算	(43)
2.2.2 域关系演算	(48)
* 2.3 三种关系运算表达能力的等价性	(50)
2.3.1 关系代数表达式与元组演算表达式的等价性	(50)
2.3.2 元组演算表达式与域演算表达式的等价性	(53)
2.3.3 域演算表达式与关系代数表达式的等价性	(53)
2.4 基于关系运算的关系数据库语言简介	(57)
2.4.1 基于关系代数的语言 ISBL	(57)
2.4.2 基于域演算的语言 QBE	(58)
2.4.3 SQL 语言	(60)
习题 2	(60)
第 3 章 关系数据库语言 SQL	(63)
3.1 表的基本操作	(63)
3.1.1 表的定义、修改与撤消	(63)
3.1.2 数据的插入、修改、删除、提交与撤消	(68)
3.2 SQL 的数据查询	(71)
3.2.1 简单查询	(71)
3.2.2 SQL 语言中的常用函数及其使用方式	(76)
3.2.3 SQL 高级查询技术	(80)
3.3 SQL 的视图操作	(86)
3.3.1 视图的概念	(86)
3.3.2 视图的定义	(87)
3.3.3 在视图上进行查询操作	(88)
3.3.4 在视图上进行数据更新	(88)
3.3.5 视图的撤消	(88)
3.3.6 用户视图对数据库系统设计和系统性能带来的影响	(89)
3.4 SQL 的数据更新	(89)
3.4.1 数据插入操作	(89)
3.4.2 数据更新操作	(90)
3.5 嵌入式 SQL	(91)
3.5.1 嵌入式 SQL 的概念	(91)
3.5.2 嵌入式 SQL 遇到的问题	(91)

3.5.3	嵌入式 SQL 的实现	(92)
3.5.4	SQL 语句的嵌入形式	(92)
3.5.5	宿主程序的组成与结构	(92)
3.5.6	数据更新及其嵌入式 SQL 应用程序设计	(97)
3.5.7	数据查询及其嵌入式 SQL 应用程序设计	(99)
* 3.5.8	动态嵌入式 SQL 技术	(102)
习题 3	(107)
第 4 章	改善数据库的查询性能	(109)
4.1	改善数据库查询性能的途径	(109)
4.2	索引技术	(110)
4.2.1	索引的概念	(110)
4.2.2	线性索引	(111)
4.2.3	B-树	(112)
4.2.4	B ⁺ 树	(114)
4.2.5	SQL 语言中的索引及其操作	(118)
4.3	查询优化	(120)
4.3.1	查询优化问题的提出	(120)
4.3.2	优化的一般策略	(122)
4.3.3	关系代数表达式的等价变换规则	(122)
* 4.3.4	关系表达式的优化算法	(124)
习题 4	(128)
第 5 章	关系数据库设计理论	(129)
5.1	关系模式设计中的问题	(129)
5.1.1	泛关系与数据库模式	(129)
5.1.2	关系模式的操作异常	(130)
5.2	函数依赖	(131)
5.2.1	函数依赖的定义	(131)
5.2.2	函数依赖的逻辑蕴涵	(132)
5.2.3	键	(133)
5.2.4	关系的约束与关系的表示	(134)
5.3	函数依赖的公理体系	(134)
5.3.1	Armstrong 公理	(134)
5.3.2	公理的推论	(135)
5.3.3	公理的完备性	(136)
5.3.4	闭包的计算	(137)
5.3.5	函数依赖集的等价与覆盖	(139)
5.3.6	最小函数依赖集	(139)
5.4	关系模式的分解	(141)
5.4.1	关系模式分解的概念	(141)

5.4.2	无损联接分解	(142)
5.4.3	保持依赖的分解	(148)
5.5	关系模式的规范化	(149)
5.5.1	第一范式(1NF)	(149)
5.5.2	第二范式(2NF)	(150)
5.5.3	第三范式(3NF)	(151)
5.5.4	鲍依斯-柯德范式(BCNF)	(152)
5.5.5	向 3NF 的模式分解算法	(153)
5.5.6	向 BCNF 的模式分解算法	(156)
* 5.6	多值依赖与第四范式	(158)
5.6.1	多值依赖的定义	(158)
5.6.2	多值依赖的公理体系	(160)
5.6.3	多值依赖公理的推论	(160)
5.6.4	依赖基	(161)
5.6.5	无损联接	(162)
5.6.6	第四范式	(162)
* 5.7	联接依赖与第五范式	(163)
	习题 5	(165)
第 6 章	数据库管理系统	(168)
6.1	数据库系统的视图级别与数据库管理员	(168)
6.2	DBMS 的功能	(170)
6.2.1	DBMS 的基本功能	(170)
6.2.2	数据字典与数据字典系统	(170)
6.3	DBMS 的组成	(173)
6.3.1	DBMS 的组成模块	(173)
6.3.2	DBMS 的层次结构	(175)
6.4	DBMS 与 OS 的关系	(177)
6.4.1	DBMS 的存取控制及其与 OS 的关系	(177)
6.4.2	DBMS 的进程结构	(178)
6.5	访问数据库数据的过程	(179)
6.6	关系型 DBMS 的十二条准则	(180)
6.6.1	RDBMS 的十二条准则	(180)
6.6.2	RDBMS 产品的发展	(181)
	习题 6	(182)
第 7 章	数据库的保护、控制与实现技术	(183)
7.1	事务	(183)
7.1.1	事务的概念及其特性	(183)
7.1.2	事务的提交与回退	(184)
7.2	数据库的安全性	(185)

7.2.1	对数据库安全的威胁	(186)
7.2.2	DBMS 的安全性级别与评估标准	(187)
7.2.3	数据库安全控制	(189)
7.2.4	视图机制	(194)
7.2.5	审计	(195)
7.2.6	数据加密	(196)
7.3	数据库的完整性	(196)
7.3.1	完整性约束条件	(197)
7.3.2	完整性控制	(200)
7.3.3	ORACLE 完整性控制	(202)
7.4	数据库恢复	(206)
7.4.1	数据库的故障分类	(206)
7.4.2	数据库的基本恢复方式	(207)
7.4.3	恢复策略	(209)
7.4.4	具有检查点的恢复技术	(210)
7.4.5	数据库镜像	(212)
7.5	并发控制	(213)
7.5.1	数据库并发操作带来的数据不一致性问题	(213)
7.5.2	锁	(215)
7.5.3	锁协议	(216)
7.5.4	活锁和死锁	(217)
7.5.5	并发调度的可串行性	(218)
7.5.6	两段锁协议	(219)
7.5.7	锁的粒度	(220)
	习题 7	(222)
第 8 章	数据库应用系统设计	(224)
8.1	数据库应用系统设计概述	(224)
8.2	数据库设计规划	(226)
8.3	用户需求分析	(227)
8.3.1	用户需求分析过程	(227)
8.3.2	数据流图	(229)
8.3.3	数据字典	(229)
8.4	概念结构设计	(231)
8.4.1	概念结构设计的步骤和方法	(231)
8.4.2	E-R 模型的操作	(233)
8.5	逻辑结构设计	(235)
8.5.1	E-R 图向关系模型的转换	(236)
8.5.2	关系数据模型的规范化设计	(237)
8.5.3	关系数据模型的优化	(238)

8.6	物理结构设计	(238)
8.6.1	ORACLE 数据库的存储结构	(238)
8.6.2	ORACLE 数据库的组织方式	(241)
8.6.3	ORACLE 数据库的存储空间管理	(243)
8.6.4	数据库物理存储结构的设计	(245)
8.7	数据库物理存储模式创建	(247)
8.8	数据库应用行为设计	(247)
8.9	数据库的运行与系统维护	(249)
	习题 8	(249)
第 9 章	高等数据库体系结构与访问技术	(251)
9.1	数据库体系结构的发展	(251)
9.1.1	主机-终端结构的数据库系统	(251)
9.1.2	客户机/服务器结构的数据库系统	(251)
9.1.3	浏览器/服务器结构的数据库系统	(253)
9.2	客户机/服务器结构的数据库系统	(253)
9.2.1	客户机/服务器结构的组成	(253)
9.2.2	客户机/服务器结构中的服务器类型	(253)
9.2.3	客户机/服务器结构数据库系统的组成与实现思想	(254)
9.2.4	客户机与数据库服务器的连接与内部结构	(255)
9.2.5	两层客户机/服务器结构数据库系统的优点及存在的问题	(255)
9.2.6	三层客户机/服务器结构的数据库系统	(257)
9.3	浏览器/服务器结构的数据库系统	(258)
9.3.1	三层浏览器/服务器结构的数据库系统	(258)
9.3.2	多层浏览器/服务器结构的数据库系统	(259)
9.3.3	浏览器/服务器结构数据库系统的优点	(259)
9.4	数据库访问接口	(260)
9.4.1	问题的提出	(260)
9.4.2	ODBC 的组成与结构	(260)
9.4.3	ODBC 的接口函数	(262)
9.4.4	ODBC 的基本工作流程	(263)
	习题 9	(263)
第 10 章	分布式数据库系统	(265)
10.1	分布式数据库系统概述	(265)
10.1.1	分布式数据库系统的定义及其理论基础	(265)
10.1.2	分布式数据库系统的特点	(266)
10.1.3	分布式数据库系统得以发展的技术和应用背景	(267)
10.1.4	分布式数据库系统的分类	(268)
10.1.5	分布式数据库系统的目标	(268)
10.1.6	分布式数据库管理系统的组成	(269)

10.1.7	分布式数据库系统的参考体系结构	(270)
10.2	分布式查询及优化	(272)
10.2.1	数据分片	(272)
10.2.2	数据分配	(273)
10.2.3	全局查询到片段查询的转换	(274)
10.2.4	分布式查询策略的重要性及分布查询处理步骤	(278)
10.2.5	基于半联接的分布式查询优化策略	(279)
10.3	分布式并发控制、恢复及安全性	(280)
10.3.1	分布式事务与两阶段提交协议	(280)
10.3.2	分布式并发控制的主要技术	(283)
10.4	分布式数据库系统应用程序设计	(287)
10.4.1	分布式数据库系统的体系结构	(287)
10.4.2	分布式数据库系统的设计方法	(288)
10.4.3	分布式数据库系统与 Client/Server 体系结构	(289)
10.4.4	客户/服务器结构分布式数据库应用程序开发策略	(290)
10.4.5	多层分布式结构技术	(290)
10.5	分布式数据库系统存在的问题及发展展望	(292)
	习题 10	(293)
第 11 章	数据库技术的新发展	(295)
11.1	数据库新技术的研究与发展	(295)
11.1.1	新应用领域对数据库技术的需求	(295)
11.1.2	关系数据库系统的局限性	(296)
11.1.3	数据库技术的新发展	(297)
11.2	面向对象数据库系统	(298)
11.2.1	对象的逻辑结构	(298)
11.2.2	对象间的限制和联系	(298)
11.2.3	面向对象数据库管理系统的基本要求	(299)
11.2.4	面向对象数据库管理系统的组成	(300)
11.2.5	面向对象数据库系统的实现途径	(301)
11.2.6	面向对象数据库系统的优越性	(302)
11.3	多媒体数据库	(302)
11.3.1	多媒体数据处理和数据管理的新技术要求	(302)
11.3.2	多媒体数据库管理系统的组织结构	(303)
11.3.3	多媒体数据库的研究现状和有待进一步研究的技术问题	(304)
11.4	主动数据库	(305)
11.4.1	主动数据库的知识模型	(305)
11.4.2	主动数据库系统的组成	(306)
11.4.3	主动数据库管理系统的功能	(307)
11.4.4	主动数据库的实现	(307)

11.4.5	事件知识库的实现	(308)
11.4.6	事件监视器的实现	(308)
11.4.7	主动数据库系统的现状和有待进一步研究的问题	(308)
11.5	数据仓库技术	(309)
11.5.1	数据仓库技术的基本思想	(309)
11.5.2	数据仓库的数据组织	(309)
11.5.3	数据仓库的体系结构	(310)
11.5.4	数据仓库的基本特征	(313)
习题 11		(314)
参考文献		(315)

数据库系统概述

数据库技术是计算机领域中最重要技术之一。自从 1969 年第一个商品化的 IMS 数据库系统问世以来,数据库技术得到了迅猛发展,其应用已渗透到工农业生产、商业、金融、行政管理、科学研究、工程技术、国防和军事等领域的每一个部门,成为信息管理、办公自动化、计算机辅助决策、计算机辅助设计、智能信息处理等计算机应用系统的核心部分。

数据库技术是计算机软件学科中的一个十分活跃而重要的独立分支,已形成了一整套的理论与技术体系。本书主要从计算机学科及相近、相关学科的数据库课程教学出发,介绍数据库系统的基本原理和基本设计技术。

在第 1 章中,将从数据管理技术的发展着手,系统地介绍数据库系统及其设计技术所涉及到的基本概念和方法,主要包括实体-联系模型、数据模型、关系模型、数据库的结构、数据语言等,以便使读者对数据库有一个概括的了解。

1.1 引言

1.1.1 数据管理技术的发展

人类社会始终是伴随着对数据的收集、整理、加工、存储和利用等一系列活动而发展的。人类最初仅靠语言、火光和绘画传递信息,并通过刻画、绳结、石块、贝壳等记录信息。当文字、纸张和印刷术发明之后,纸张就变成了数据和文字的主要载体,并成为传递和存储信息的主要手段和途径。电子技术,特别是电子计算机出现以后,人类收集、存储、加工和利用数据的手段和技术发生了革命性的变化。

数据管理是指对数据的收集、整理、组织、存储、维护、检索、传送等操作过程。数据处理是指对数据进行收集、存储、加工和传播的一系列活动的总和,其基本目的是从大量的、杂乱无章的、难以理解的数据中抽取并导出对于那些特定的应用来说是有价值的、有意义的数,借以作为决策的依据。可见,数据管理是任何数据处理业务中必不可少的共有部分,数据管理和数据处理是密不可分的。

随着计算机科学与技术的发展,利用电子计算机进行数据管理大体上经历了从低级到高级的三个阶段,即人工管理阶段、文件系统阶段和数据库系统阶段。

1. 人工管理阶段(20 世纪 50 年代中期以前)

20 世纪 50 年代中期以前,计算机的存储设备只有磁带、卡片、纸带等,没有磁盘等直接存取存储设备,软件尚没有操作系统和数据管理软件。计算机主要用于科学计算。这一时期的

数据管理和数据处理主要呈现出以下特点：

(1) 数据和程序一一对应,各个应用程序按计算要求组织各自所需的数据。由于有些应用程序,特别是计算性质相近的应用程序所处理的数据可能会有一定关系,所以程序与程序之间可能会有大量重复数据。程序与数据的对应关系如图 1.1 所示。

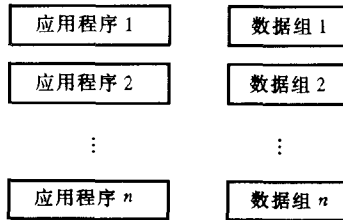


图 1.1 人工管理阶段数据与程序的对应关系

(2) 数据采用顺序存储方式存储,数据的逻辑结构与物理结构相同,即数据在外存的物理结构与用户观点的逻辑结构完全一致,用户程序和数据存储在磁带上,数据的组织方式是顺序存储方式,系统对数据的管理操作仅仅是完成简单的 I/O 操作。数据的逻辑组织与数据的物理组织的对应关系如图 1.2 所示。

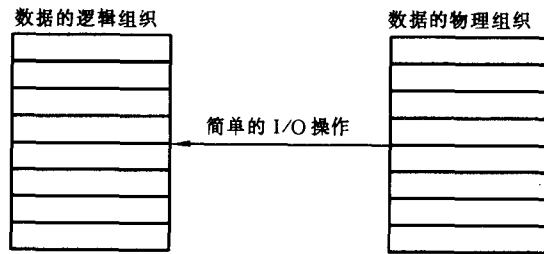


图 1.2 人工管理阶段数据的逻辑组织与数据的物理组织对应关系

(3) 数据对程序不具有独立性。数据的管理完全由应用程序员在程序中进行安排,也就是说应用程序员不仅要规定数据的逻辑结构,而且还要在程序中设计数据的物理结构,包括存储结构、存取方式、输入输出方式等。由于数据的物理组织是由应用程序员根据应用的要求设计的,一旦数据在存储器上改变其物理位置,就需要相应地修改或重编应用程序,所以数据对程序不具有独立性。

(4) 数据不需要保存。因为当时计算机的用途基本上是科学计算,只有在算题时才需要将计算所需的原始数据输入到计算机中,所以数据不需要长期保存。

2. 文件系统阶段(20 世纪 50 年代后期至 60 年代中后期)

大约到了 20 世纪 50 年代中后期,磁盘和磁鼓等直接存取设备开始应用,尤其是磁盘已成为联机应用的主要存储设备。与此同时,计算机的应用也逐步从科学计算扩大到非数值计算领域。随着数据处理的逐渐增多,出现了专门用于数据管理的软件——文件管理系统。这一时期的数据管理和数据处理主要呈现出如下特征:

(1) 组成文件的最小单位和数据存取的基本单位是记录。记录之间仅存在顺序关系而无其它结构。

(2) 文件中的数据(记录)可顺序或随机访问,文件的逻辑结构与物理结构已有一定区别。在文件的物理结构中增加了索引和链接等形式,而且可利用倒排文件进行多码检索。数据管理软件提供了从逻辑文件到物理文件的“访问方法”转换。逻辑文件到物理文件的“访问方法”转换关系如图 1.3 所示。

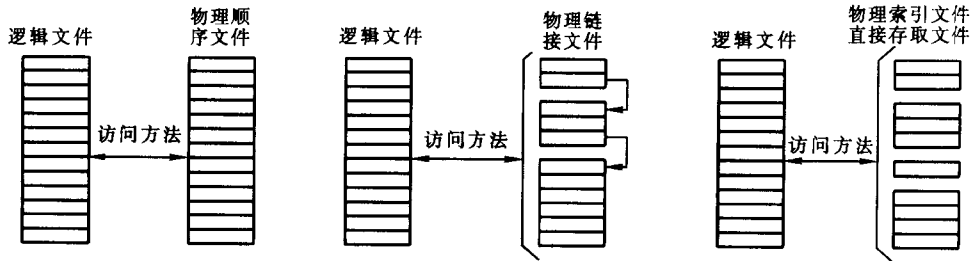


图 1.3 文件系统阶段逻辑文件到物理文件的访问方法转换

(3) 部分系统实现了以文件为单位的数据共享,但用户程序仍以文件标识(文件名)与系统交互,还未能实现以记录和数据项为单位的数据共享,但允许以文件中的记录标识访问数据。所以一个数据文件还基本上是对应一个或几个应用程序,或者说数据还是面向应用的。应用程序与数据文件的对应关系如图 1.4 所示。

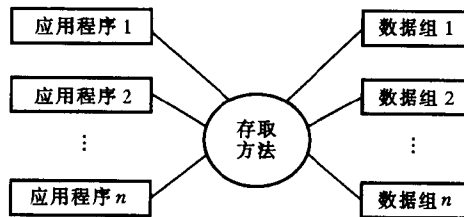


图 1.4 文件系统阶段应用程序与数据文件的对应关系

(4) 数据需要长期保存。由于计算机开始大量用于非数值计算的数据处理等,所以数据需要长期保存在外存储器上。为了实现对这些数据文件的管理,数据管理系统提供了对文件及其中数据(记录)的查询、修改、插入、删除等操作。

3. 数据库系统阶段(20 世纪 60 年代后期开始)

20 世纪 60 年代后期,计算机硬件和软件技术有了较大发展,出现了大容量直接存取设备。计算机在企业管理中的广泛应用,对数据管理提出了更高的要求,要求具有更高的数据共享,要求数据具有更高的独立性。文件系统由于未能完全实现数据的逻辑结构独立于数据的物理结构,未能实现以记录或数据项为单位的数据共享等,所以不能满足上述这些要求。这些亟待解决的问题,促使人们研究一种新的数据管理技术,因而导致了数据库技术的产生和发展。初期的数据库系统具有以下特征:

(1) 可以从用户的要求出发,对数据进行严格细致的描述,使得文件、记录、数据项等数据单位之间的联系清晰,结构简单。

(2) 允许用户以记录或数据项为单位进行访问,也允许多关键字检索和文件之间的交叉访问。

(3) 数据的物理存储可以很复杂,同样的物理数据可以导出多个不同的逻辑文件,用户以简单的逻辑结构操作他的数据而无需考虑数据的存储情况,改动数据的物理位置和存储结构不必修改或重写应用程序。从而实现了数据的应用独立于数据的存储。

数据库系统阶段数据与应用程序的关系如图 1.5 所示。

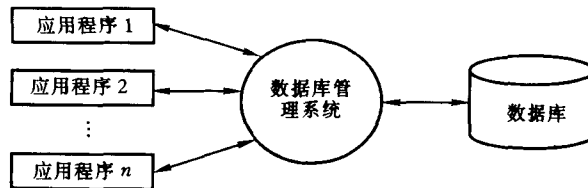


图 1.5 数据库系统阶段应用程序与数据文件的对应关系

4. 高级数据库技术阶段(20 世纪 80 年代初期开始)

初级的数据库系统主要是以 IMS 为代表的层次数据库系统(详见 1.4.1 节)和以 DBTG 报告为标准实现的网状数据库系统(详见 1.4.2 节)。由于层次数据库系统和网状数据库系统的应用涉及到许多与系统实现有关的低层细节问题,使应用有时变得十分复杂,因而限制了它们的进一步应用和发展。

1970 年, E. F. Codd 提出了关系数据库的基本概念,奠定了关系数据库发展的基础。随着关系代数、关系演算理论、依赖理论和规范化理论等的引入,标志着关系数据库理论的日趋完善和数据库技术走向成熟。

随着计算机科学技术的进一步发展,数据库技术和计算机网络、软件工程、人工智能、面向对象技术等相互结合,使数据库技术进入了高级发展阶段,其重要的标志就是 20 世纪 80 年代初开始出现的分布式数据库系统和面向对象数据库系统。此后,数据库技术如雨后春笋,蓬勃发展,并在并行数据库技术、模糊数据库技术等新一代数据库技术与理论方面得到了进一步的发展。

1.1.2 数据库技术的基本概念

本部分介绍数据库技术中的最基本的概念,其目的是为后续内容的介绍和学习奠定一些基础。

(1) 数据库

数据库(Data Base, 简称 DB) 是在计算机存储设备上,按一定的组织方式存储在一起的、相关的数据的集合。数据库中的数据以结构化形式存储,冗余度小,独立于应用程序,为多个用户共享,是数据库系统的核心和管理对象。

(2) 数据库管理系统

数据库管理系统(Data Base Management System, 简称 DBMS) 是建立、管理和维护数据库的软件系统。DBMS 接受并处理用户和应用程序对数据库的操作请求,负责数据库的完整性、一致性和安全性检查,实现数据库系统的并发控制和故障恢复等,是一种位于应用软件和操作系统之间,实现数据库管理功能的系统支持软件。

(3) 数据库应用系统

数据库应用系统(Data Base Application System)是采用数据库技术,以计算机为硬件和