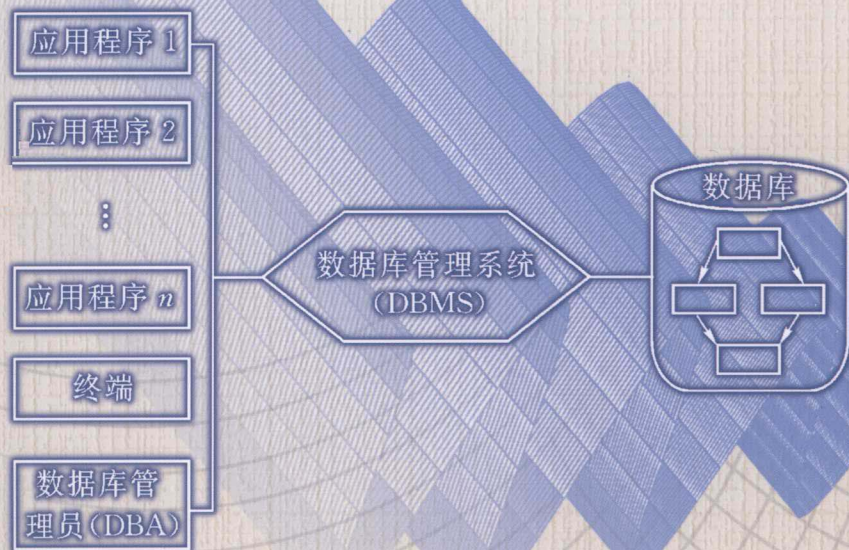




普通高等教育“十二五”规划教材

数据库系统原理

沈钧毅 侯迪 冯中慧 何亮 编著



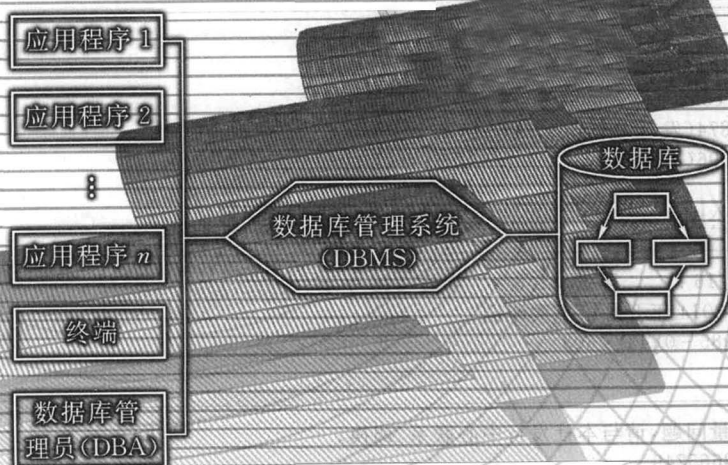
西安交通大学出版社
XI'AN JIAOTONG UNIVERSITY PRESS



普通高等教育“十二五”规划教

数据库系统原理

沈钧毅 侯迪 冯中慧 何亮 编著



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

内容简介

本书融数据库系统基础、数据库设计、数据库管理系统和数据库新技术为一体,以关系数据库为重点,全面系统地介绍数据库系统原理与技术的基本内容。全书共分四大部分十四章。第一部分(第1~3章)为数据库系统基础,分别介绍数据库系统的基本概念、关系数据库和数据库语言。第二部分(第4~5章)数据库设计,主要介绍关系数据库理论、数据库设计。第三部分(第6~9章)介绍数据库管理系统内部机制系统,详细、深入、系统地阐述数据库管理系统内部机制。第四部分(第10~14章)数据库新技术,分别简要地介绍分布式数据库与并行数据库系统、数据库访问技术、面向对象数据库与对象关系数据库、XML数据库技术以及数据仓库与联机分析处理技术。为方便读者学习,各章均附有习题。

本书可作为高等学校计算机专业或其它专业学生的数据库课程基础教材。全书在系统阐述数据库基本概念和原理的基础上,强调理论与应用相结合,力图反映当前数据库技术发展的水平和趋势。本书便于自学,也可供广大科技人员学习、参考。

图书在版编目(CIP)数据

数据库系统原理/沈钧毅,侯迪,冯中慧,何亮编著;

—西安:西安交通大学出版社,2014.2

ISBN 978-7-5605-3985-0

I. ①数… II. ①沈… ②侯… ③冯… ④何… III. 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 011569 号

书 名	数据库系统原理
编 著	沈钧毅 侯迪 冯中慧 何亮

出版发行 西安交通大学出版社
(西安市兴庆南路10号 邮政编码 710049)

网 址 <http://www.xjtpress.com>
电 话 (029)82668357 82667874(发行中心)
(029)82668315 82669096(总编办)

传 真 (029)82669097
印 刷 陕西奇彩印务有限责任公司

开 本	787mm×1092mm	1/16	印张	21.5	字数	516千字
版次印次	2014年2月第1版		2014年2月第1次印刷			
书 号	ISBN 978-7-5605-3985-0/TP·603					
定 价	38.00元					

读者购书、书店添货、如发现印装质量问题,请与本社发行中心联系、调换。

订购热线:(029)82665248 (029)82665249

投稿热线:(029)82665380

读者信箱:banquan1809@126.com

版权所有 侵权必究

前 言

“数据管理”是计算机技术领域中最发展最迅速、应用最广泛的分支之一,数据库系统是继文件系统进行数据管理以后的一种崭新技术,在信息化社会建设中有着重要的作用。当前,数据库系统课程在计算机科学与技术专业及其相关专业的教学计划中占有重要的地位,本书是一部面向本科生教学的数据库系统教程。

全书共分四大部分,涵盖了数据库系统基础、应用(设计)、数据库管理系统内部实现机制及新技术四个部分。我们期望:读者通过这四个部分的学习,全面掌握数据库系统基本概念及原理,为在各个领域信息化建设中熟练应用数据库技术奠定必要的基础。

第一部分介绍数据库系统基础,分三章加以阐述。即第1章 数据库系统概述,从数据库技术的产生和发展、数据库系统、数据库管理系统三个方面概括地阐述了数据库系统的基本概念;第2章 关系模型,阐述了目前流行使用的数据模型——关系模型,从关系代数、关系演算两种数学概念出发,介绍了关系数据库的理论基础;第3章 关系数据库语言 SQL,从应用角度介绍了关系数据库语言,分别阐述了数据定义语言、操纵语言、查询语言、控制语言主要语句的功能和使用。

第二部分介绍数据库设计,从如何应用数据库技术建立计算机信息系统出发,阐述数据库设计的原理与方法,分两章阐述。即第4章 数据依赖与关系模式规范化,概要阐述关系数据库设计的理论依据,第5章 数据库设计,全面阐述数据库设计的过程与方法。

第三部分介绍数据库管理系统内部机制系统,阐述数据库系统的核心——数据库管理系统的内部实现机制,分成四章阐述。即第6章 存储结构,介绍了外存储器及相关的数据库物理组织;第7章 查询处理与查询优化,介绍了数据库查询操作的处理过程,并阐述了提高查询性能的策略;第8章 事务管理,从事务这一重要概念出发,阐述故障恢复与并发控制功能的实现机制;第9章 数据库完整性和安全性,介绍完整性、安全性的基本概念及其实现机制。

第四部分介绍数据库新技术,分成五章阐述。即第10章 分布式数据库与并行数据库系统,简要阐述分布式数据库与并行数据库系统的组成,体系结构与实现机制;第11章 数据库访问技术,分别阐述 C/S、B/S 两种体系结构以及 JDBC 访问技术;第12章 面向对象数据库系统与对象关系数据库系统,从数据模型角度阐述了关系数据库系统之后的新技术;第13章 XML 数据管理技术。第14章 数据仓库与联机分析处理技术,概要地阐述了数据库技术应用于决策分析的新技术。

本书是由我们学校4位从事数据库系统教学工作的教师积多年教学经验撰写而成的。第1章及第14章由沈钧毅执笔,第2章至第5章由何亮执笔,第7章至第9章由冯中慧执笔,第6章及第10章至第13章由侯迪执笔。在本书编写过程中,特别感谢南京大学徐洁磐教授、北京大学唐世渭教授、中国科学院研究生院邵佩英教授、山东大学董继润教授、山东财政学院聂培尧教授以及其他老师提出的修改建议与热情指导。

全书的编写得到了西安交通大学出版社赵丽萍编辑的热情支持与具体帮助。在此,我们表示衷心感谢。

由于编者经验水平有限,缺点与错误在所难免,殷切期望广大读者批评指正。对本书的各种意见与建议,请与编写者联系,谢谢各位。

沈钧毅 侯迪 冯中慧 何亮
2013年11月于西安交通大学

目 录

前 言

第 1 章 绪论	(1)
1.1 数据库技术的产生和发展	(1)
1.1.1 数据管理的发展	(1)
1.1.2 数据库系统的特征	(4)
1.1.3 数据库技术的应用与发展	(5)
1.2 数据库系统	(5)
1.2.1 数据库系统的组成结构	(5)
1.2.2 数据库、数据模型与数据模式	(6)
1.3 数据库管理系统(DBMS)	(9)
1.3.1 DBMS 概述	(9)
1.3.2 DBMS 结构	(9)
1.3.3 DBMS 进程结构	(11)
1.3.4 数据目录	(14)
习题	(14)
第 2 章 关系模型	(16)
2.1 基本概念	(16)
2.1.1 概述	(16)
2.1.2 关系的定义	(17)
2.1.3 关系的完整性	(20)
2.2 关系代数	(21)
2.2.1 传统的集合运算	(21)
2.2.2 专门的关系运算	(23)
2.2.3 关系代数查询实例	(26)
2.3 关系演算	(27)
2.3.1 元组关系演算	(27)
2.3.2 域关系演算	(30)
习题	(30)

第 3 章 关系数据库语言 SQL	(32)
3.1 SQL 语言概述	(32)
3.2 SQL 语言的数据定义功能	(34)
3.2.1 基本表的建立	(34)
3.2.2 基本表的修改与删除	(37)
3.2.3 索引的建立和删除	(38)
3.2.4 模式的建立和删除	(39)
3.3 SQL 语言数据操纵功能	(40)
3.3.1 SQL 数据查询	(40)
3.3.2 SQL 数据插入	(50)
3.3.3 SQL 数据删除	(51)
3.3.4 SQL 数据修改	(52)
3.4 视图	(52)
3.4.1 视图的定义和删除	(53)
3.4.2 视图的查询	(54)
3.4.3 视图的更新	(55)
3.4.4 视图的应用	(56)
3.5 SQL 语言的数据控制功能	(57)
3.5.1 权限与角色	(57)
3.5.2 权限的授予和收回	(58)
3.6 嵌入式 SQL	(58)
3.6.1 嵌入式 SQL 概述	(59)
3.6.2 嵌入式 SQL 数据库访问过程	(60)
3.6.3 不使用游标的嵌入式 SQL 语句	(61)
3.6.4 使用游标的嵌入式 SQL 语句	(62)
3.6.5 动态 SQL 语句	(64)
3.7 SQL/CLI	(65)
3.7.1 SQL/CLI 基本原理	(65)
3.7.2 SQL/CLI 应用编程	(68)
习题	(70)
第 4 章 数据依赖与关系模式规范化	(72)
4.1 问题的提出	(72)
4.2 数据依赖	(74)
4.2.1 函数依赖	(74)
4.2.2 多值依赖	(79)
4.3 关系模式分解	(80)

4.3.1 无损连接分解	(80)
4.3.2 保持函数依赖分解	(82)
4.4 关系模式规范化	(83)
4.4.1 范式	(83)
4.4.2 规范化算法	(85)
习题	(87)
第5章 数据库设计	(89)
5.1 数据库设计概述	(89)
5.1.1 数据库系统生命周期	(89)
5.1.2 数据库设计方法和步骤	(90)
5.2 需求分析	(92)
5.3 概念设计	(92)
5.3.1 E-R 模型简介	(93)
5.3.2 E-R 设计方法	(98)
5.4 逻辑设计	(105)
5.4.1 E-R 模型转换	(105)
5.4.2 优化与调整	(108)
5.5 物理设计	(109)
5.5.1 存储结构概述	(109)
5.5.2 存储结构与存取方式设计	(111)
5.6 IDEF 设计方法简介	(114)
5.7 数据库运行与维护	(116)
习题	(118)
第6章 存储结构	(119)
6.1 物理存储设备	(119)
6.1.1 磁盘存储器	(119)
6.1.2 磁带存储器	(121)
6.1.3 磁盘缓冲处理	(121)
6.1.4 第三级存储	(122)
6.2 文件和文件记录	(122)
6.2.1 定长记录	(123)
6.2.2 变长记录	(125)
6.3 无序文件	(127)
6.4 顺序文件	(128)
6.5 索引文件	(130)

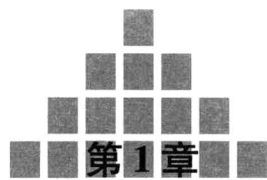
6.5.1	索引的基本结构	(130)
6.5.2	顺序索引	(130)
6.5.3	主索引	(132)
6.5.4	聚簇索引	(133)
6.5.5	辅助索引	(134)
6.5.6	多级索引	(136)
6.5.7	索引的更新	(137)
6.5.8	索引的评价标准	(139)
6.6	B 树与 B ⁺ 树索引	(139)
6.6.1	索引树	(139)
6.6.2	B 树索引	(140)
6.6.3	B ⁺ 树索引	(143)
6.7	散列文件	(147)
6.7.1	简单散列方法	(148)
6.7.2	散列函数	(149)
6.7.3	散列文件的操作	(150)
6.7.4	溢出处理	(151)
6.7.5	动态散列方法	(151)
	习题	(152)
第 7 章	查询处理与查询优化	(154)
7.1	查询处理与查询优化概述	(154)
7.1.1	查询处理	(154)
7.1.2	查询优化分类	(155)
7.1.3	一个启发性的例子	(156)
7.2	代数优化	(158)
7.2.1	代数优化的基本原则	(158)
7.2.2	代数优化的等价变换规则	(158)
7.2.3	代数优化策略	(160)
7.2.4	代数优化算法	(161)
7.3	物理优化	(165)
7.3.1	选择	(165)
7.3.2	连接	(167)
7.3.3	投影	(171)
7.3.4	集合运算	(172)
	习题	(174)

第 8 章 事务管理	(176)
8.1 事务管理概述	(176)
8.1.1 事务的概念	(176)
8.1.2 事务的性质	(178)
8.1.3 事务的操作	(179)
8.1.4 事务的状态	(179)
8.1.5 SQL 的事务管理	(180)
8.2 数据库故障分类	(181)
8.3 数据库恢复技术	(182)
8.3.1 数据转储	(182)
8.3.2 日志文件	(183)
8.3.3 恢复策略	(185)
8.4 事务的并发控制	(187)
8.4.1 并发引起的问题	(188)
8.4.2 事务调度及可串行性	(189)
8.5 基于锁的并发控制	(194)
8.5.1 封锁协议	(194)
8.5.2 两阶段封锁协议	(196)
8.5.3 活锁与死锁	(198)
8.5.4 封锁粒度	(200)
8.5.5 插入和删除操作	(202)
8.6 其他并发控制技术	(203)
8.6.1 基于时间标记的并发控制	(203)
8.6.2 乐观并发控制	(204)
习题	(205)
第 9 章 数据库完整性与安全性	(208)
9.1 数据库完整性与安全性概述	(208)
9.2 数据库的完整性	(208)
9.2.1 数据库完整性约束分类	(209)
9.2.2 完整性约束条件	(209)
9.2.3 DBMS 的完整性控制	(211)
9.2.4 完整性约束的说明	(214)
9.3 数据库的安全性	(216)
9.3.1 访问控制	(217)
9.3.2 视图机制	(222)
9.3.3 数据加密	(222)

9.3.4 数据库审计	(223)
习题	(224)
第 10 章 分布式数据库与并行数据库系统	(225)
10.1 分布式数据库与并行数据库系统概述	(225)
10.2 分布式数据库系统的体系结构	(226)
10.2.1 分布式数据库系统的组成	(226)
10.2.2 分布式数据库系统的特点	(227)
10.2.3 分布式数据库系统的模式结构	(229)
10.3 分布式查询处理和优化	(230)
10.3.1 分布式查询优化的目标和准则	(230)
10.3.2 分布式查询的分类	(231)
10.3.3 分布式查询处理的层次结构	(232)
10.4 分布式事务	(233)
10.5 联邦数据库系统	(234)
10.6 多数据库系统	(236)
10.7 并行数据库系统	(237)
10.7.1 并行数据库系统的目标	(238)
10.7.2 并行数据库系统的体系结构	(238)
10.7.3 并行粒度	(240)
10.7.4 并行方式	(241)
10.7.5 数据划分	(242)
习题	(243)
第 11 章 数据库访问技术	(244)
11.1 数据库访问方式	(244)
11.1.1 数据库应用程序的组成部分	(244)
11.1.2 主机终端系统	(245)
11.1.3 基于网络的 C/S 系统结构	(246)
11.1.4 基于 Web 的 B/S 结构	(249)
11.2 WWW 与数据库访问接口	(250)
11.2.1 公共网关接口 CGI	(250)
11.2.2 ASP 技术	(252)
11.2.3 JSP 技术	(254)
11.2.4 PHP 技术	(256)
11.3 ODBC	(257)
11.3.1 ODBC 的由来和发展	(257)

11.3.2	ODBC 体系结构	(258)
11.3.3	ODBC 的特性	(263)
11.4	JDBC	(263)
11.4.1	JDBC 的由来和发展	(263)
11.4.2	JDBC 的基本功能	(264)
11.4.3	JDBC 数据库应用模型	(264)
11.4.4	JDBC 驱动程序	(265)
11.4.6	JDBC API	(266)
11.4.7	JDBC 应用程序流程	(268)
	习题	(270)
第 12 章 面向对象数据库与对象关系数据库		(272)
12.1	面向对象与对象关系数据库概述	(272)
12.2	面向对象的基本概念	(274)
12.2.1	面向对象方法的基本特征	(274)
12.2.2	对象	(274)
12.2.3	对象结构	(275)
12.2.4	消息	(277)
12.2.5	类	(277)
12.2.6	继承	(277)
12.2.7	动态联编	(279)
12.2.8	对象组合	(279)
12.3	面向对象的数据模型	(280)
12.3.1	基本特征	(280)
12.3.2	基本模型	(281)
12.4	持久化程序设计语言	(283)
12.4.1	对象的持久性	(284)
12.4.2	对象标识和指针	(285)
12.4.3	持久对象的存储和访问	(286)
12.5	对象-关系数据模型	(287)
12.5.1	关系数据模型扩充	(288)
12.5.2	嵌套关系	(288)
12.5.3	组合类型	(290)
12.5.4	聚集类型	(290)
12.5.5	引用类型	(292)
12.5.6	对象类型	(292)
12.6	ORDB 定义语言	(293)

12.6.1 数据类型定义	(294)
12.6.2 继承的定义	(296)
12.6.3 引用类型的定义	(297)
12.7 ORDB 查询语言	(298)
12.7.1 以关系为值的属性	(298)
12.7.2 路径表达式	(299)
12.7.3 嵌套与解除嵌套	(300)
12.7.4 函数的定义和使用	(301)
12.7.5 复合值的创建和查询	(302)
习题	(302)
第 13 章 XML 数据管理技术	(304)
13.1 XML 概述	(304)
13.2 XML 查询语言——XQuery	(306)
13.3 基于关系数据库的 XML 处理	(311)
习题	(316)
第 14 章 数据仓库与联机分析处理	(318)
14.1 数据仓库与联机分析处理概述	(318)
14.1.1 数据管理的层次结构	(318)
14.1.2 数据仓库的产生	(319)
14.2 数据仓库	(320)
14.2.1 概述	(320)
14.2.2 数据仓库的数据模型	(323)
14.2.3 数据仓库的数据模式	(324)
14.2.4 构建数据仓库的步骤	(326)
14.3 OLAP 技术	(326)
14.3.1 概述	(326)
14.3.2 多维分析技术	(327)
14.3.3 OLAP 操作语言	(329)
习题	(330)
参考文献	(331)



第1章

Chapter 1

绪 论

数据库系统的目标是管理大量的、持久的、共享的数据,并支持用户查找和更新所需要的信息,是计算机科学的重要分支。本章主要讲述数据库技术的产生背景、发展历史以及未来的发展方向;数据库系统的组成结构、基本概念和技术;数据库管理系统的基本功能、进程结构和工作原理。

1.1 数据库技术的产生和发展

1.1.1 数据管理的发展

众所周知,电子计算机从 1946 年诞生以来,已成为现代社会各种应用领域中的重要工具。以前,人们经常讲到计算机应用领域时,都会列举哪些领域能使用电子计算机,可能会列举几千种应用方面,将其分类可以概括为:科学计算、过程控制、数据管理等几大领域。现在,人们往往会这样提出问题:哪些部门没有使用电子计算机。这一问题表明“现代社会的方方面面几乎无处不用电子计算机。几十年的应用表明:在这些应用领域中,电子计算机用于数据管理几乎占到了绝大多数。

电子计算机如何进行数据管理,在历经将近 50 年的过程中,大体可划分为以下三个阶段。

1. 人工管理阶段

在 20 世纪 60 年代计算机操作系统问世以前,人们使用计算机进行数据管理,必须用人工方式组织数据,把用户的数据直接存放在外存设备的有关地址里,然后编写应用程序对这些数据进行管理。在这个阶段,应用程序与数据之间的关系如图 1-1 所示。

显然,这样一种数据管理方式无法解决数据管理的如下需求:

- 1) 数据量很大。如果管理几百、几千个数据,还可以对付,如果数据量是几十 KB、MB,甚至 GB,将是无法解决的;
- 2) 保存数据。每次使用,都得将数据输入到外存,无法保存,下次使用必须重新保存;

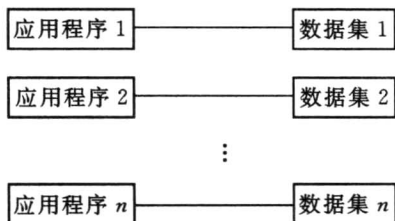


图 1-1 人工管理阶段应用程序与数据之间的相互关系

3)数据共享。数据只能对应于某一个应用程序,其它应用程序使用这些数据时必须自己重新定义,造成多个应用程序使用同一数据时的冗余现象。

正因为这样,操作系统中的文件系统,作为管理外存的一种手段,应运而生,为数据的存放和应用带来了若干方便。

2. 文件系统阶段

20 世纪 60 年代,出现了操作系统,其中的文件系统作为管理外存的系统软件,解决了外存上数据的按“名”存取,为数据管理发展奠定了基础。

文件系统存取数据,用户不必自己指定外存地址来存放,而且只需给定数据集的“名字”,由文件系统加以统一管理。例如,我们将对学生数据进行管理,只要建立学生文件即可。至于这些数据放在磁盘的什么位置,以后如何取出数据,都不需要用户来管理,而由文件系统来管理。这个阶段的应用程序与数据之间对应关系,可用图 1-2(a)来表示。

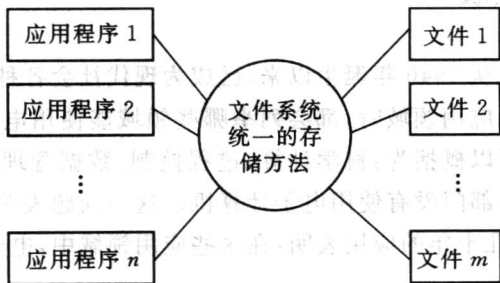


图 1-2(a) 文件系统阶段应用程序与数据之间的对应关系

注意,图 1-2(a)中应用程序中的 n 与文件 m 不一定相同。

可以看出,这一阶段的数据管理,比人工管理阶段显然要好得多:

1)解决数据量大的问题。用户的数据通过文件来存放,其数据量大小与用户自己分配空间的难度完全不同。按“名”存取文件,对数据量大小没有什么影响;

2)数据的持久保存问题。文件系统解决了按“名”持久保存数据的难题,用户一旦建立起了所需的文件,文件系统将统一管理这些数据;

3)部分解决了数据共享问题。用户所建立的文件,可由不同的应用程序加以调用,当然用户必须了解其文件的结构。图 1-2(a)表明:程序与文件之间由文件系统的统一存取方法加

以联系，部分地解决了相互依赖问题，但没有完全解决数据共享问题。

对于文件系统而言，程序与数据有了一定的物理独立性，数据在物理存储上的改变没有反映在程序上。但是，文件为特定的应用程序服务，文件的逻辑结构改变会直接影响应用程序，数据与程序之间缺乏逻辑独立性。而且，前面的 3) 说明，文件系统在数据共享上存在问题，现举一个例子加以解释，如图 1-2(b) 所示。

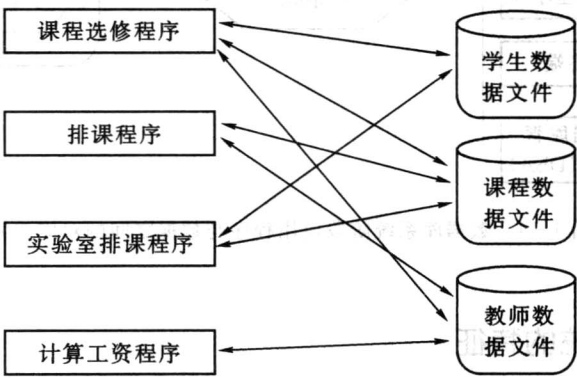


图 1-2(b) 应用程序与文件之间的对应例子

针对文件系统在数据管理上存在的缺陷，在数据统一管理中以共享数据为主要特征的数据库系统应运而生。

3. 数据库系统阶段

20 世纪 60 年代末期，计算机用于数据管理，规模急剧增长，数据共享要求越来越强烈。在文件系统基础上，出现了统一管理数据的专门性系统软件——数据库管理系统 (DataBase Management System, 简称 DBMS)。从数据模型角度看，可以认为：数据库系统是在文件系统基础上增加了文件之间的联系而出现的。(本章 1.3 节将阐述数据模型概念)。图 1-3 表明了 DBMS 是由文件系统发展而来的。

数据库系统阶段，应用程序与数据之间的对应关系，可用图 1-4 来表示。

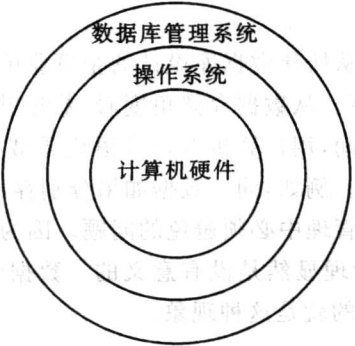


图 1-3 数据库管理系统是文件系统的发展

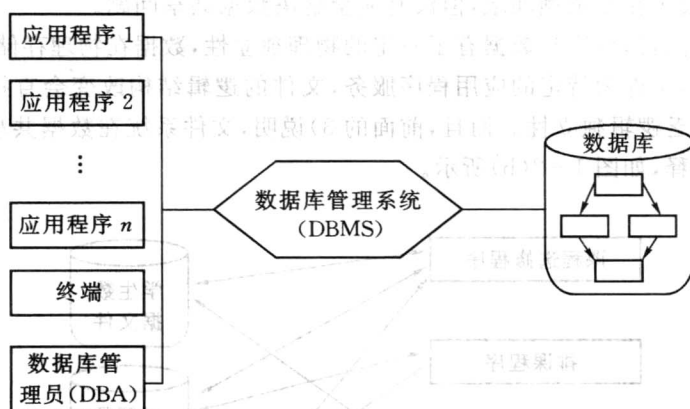


图 1-4 数据库系统阶段应用程序与数据之间的对应关系

1.1.2 数据库系统的特征

数据库系统是现阶段计算机数据管理的主要形式,是数据管理的最新技术。与数据管理前期阶段——人工管理与文件系统——相比较,具有以下几个方面的重要特征。

1. 良好的数据模型

回顾数据库系统阶段以前的数据管理手段,在数据模型上存在严重的缺陷:人工管理阶段,数据都是由用户自行组织的,没有组织结构;文件系统阶段,对于文件内部有了一定的组织结构(如:顺序文件与索引文件),记录之间都有一定的联系,但是不同文件之间一般是相互独立的,使用时由应用程序来完成不同文件记录之间的调用。

数据库系统阶段,数据库模型是数据库存储数据的基础,从开始出现的层次、网状数据库模型,到目前流行的关系数据库模型,直到新一代数据模型的提出,无不反映出构成数据库模型基础的文件之间的联系,数据模型所能描述数据的特性更为良好。

2. 数据共享性好

数据库系统中,数据模型是从所建数据库设计的企事业单位整体角度来描述数据的,所建的数据将被多个用户与应用共享。从数据存储角度看,数据共享大大减少了有关数据分别存放时的重复现象,节省了存储空间,用计算机专业术语说是减少冗余。减少冗余,带来了重要的特性:避免数据之间的不一致。例如,同一数据如果分别存放,势必有可能出现不同的值,造成同一数据的不一致,这是数据管理中必须避免的问题。因为数据的不一致性,即这些数据是垃圾数据,对这样的数据加以管理显然是没有意义的。数据管理中流行的话,垃圾进垃圾出(Garbage In, Garbage Out),指的就是这种现象。

3. 数据由 DBMS 统一管理

数据共享,要求数据由 DBMS 加以统一管理。为此,DBMS 提供如下重要功能,以确保数