

21 世纪高等学校计算机教育实用规划教材

数据库原理与DB2应用教程

杨鑫华 主编
赵慧敏 丁传华 副主编



清华大学出版社

数据库原理与DB2应用教程

杨鑫华 主编

赵慧敏 丁传华 副主编

清华大学出版社
北 京

内 容 简 介

本书以帮助读者建立数据库基本概念,提高数据库应用、开发能力为目标。在讲述传统的“数据库原理”的同时,结合软件市场对数据库人才的需求,介绍 IBM 公司的旗舰产品——DB2 通用数据库及其典型应用,并以 DB2 通用数据库作为实践环境,阐述数据库相关概念以及常用的数据库应用开发方法。

本书取材适宜、难度适当、实用性强,可使读者在学习数据库基本原理的同时,掌握一种实用的大型数据库应用技术,掌握从事信息处理工作应具备的数据库技术的基本知识、基本技能和操作能力。

本书可作为计算机、软件工程、电子商务、信息管理等专业本科生数据库课程的教材;也可作为参加 IBM DB2 专业认证考试的辅导教材,以及从事数据库应用的开发人员的参考书。

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 13501256678 13801310933

图书在版编目(CIP)数据

数据库原理与 DB2 应用教程/杨鑫华主编. —北京:清华大学出版社,2007.8

(21 世纪高等学校计算机教育实用规划教材)

ISBN 978-7-302-15554-6

I. 数… II. 杨… III. 关系数据库—数据库管理系统, DB2—高等学校—教材
IV. TP311.138

中国版本图书馆 CIP 数据核字(2007)第 096732 号

责任编辑:闫红梅

责任校对:李建庄

责任印制:王秀菊

出版发行:清华大学出版社 地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn> 邮 编:100084

c-service@tup.tsinghua.edu.cn

社 总 机:010-62770175 邮购热线:010-62786544

投稿咨询:010-62772015 客户服务:010-62776969

印 刷 者:北京市世界知识印刷厂

装 订 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

开 本:185×260 印 张:18.25 字 数:438 千字

版 次:2007 年 8 月第 1 版 印 次:2007 年 8 月第 1 次印刷

印 数:1~5000

定 价:24.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题,请与清华大学出版社出版部联系调换。联系电话:(010)62770177 转 3103 产品编号:026254-01

出版说明

随着我国高等教育规模的扩大以及产业结构调整的进一步完善,社会对高层次应用型人才的需求将更加迫切。各地高校紧密结合地方经济建设发展需要,科学运用市场调节机制,合理调整和配置教育资源,在改革和改造传统学科专业的基础上,加强工程型和应用型学科专业建设,积极设置主要面向地方支柱产业、高新技术产业、服务业的工程型和应用型学科专业,积极为地方经济建设输送各类应用型人才。各高校加大了使用信息科学等现代科学技术提升、改造传统学科专业的力度,从而实现传统学科专业向工程型和应用型学科专业的发展与转变。在发挥传统学科专业师资力量强、办学经验丰富、教学资源充裕等优势的同时,不断更新其教学内容、改革课程体系,使工程型和应用型学科专业教育与经济建设相适应。计算机课程教学在从传统学科向工程型和应用型学科转变中起着至关重要的作用,工程型和应用型学科专业中的计算机课程设置、内容体系和教学手段及方法等也具有不同于传统学科的鲜明特点。

为了配合高校工程型和应用型学科专业的建设和发展,急需出版一批内容新、体系新、方法新、手段新的高水平计算机课程教材。目前,工程型和应用型学科专业计算机课程教材的建设工作仍滞后于教学改革的实践,如现有的计算机教材中有不少内容陈旧(依然用传统专业计算机教材代替工程型和应用型学科专业教材),重理论、轻实践,不能满足新的教学计划、课程设置的需要;一些课程的教材可供选择的品种太少;一些基础课的教材虽然品种较多,但低水平重复严重;有些教材内容庞杂,书越编越厚;专业课教材、教学辅助教材及教学参考书短缺,等等,都不利于学生能力的提高和素质的培养。为此,在教育部相关教学指导委员会专家的指导和建议下,清华大学出版社组织出版本系列教材,以满足工程型和应用型学科专业计算机课程教学的需要。本系列教材在规划过程中体现了如下一些基本原则和特点。

(1) 面向工程型与应用型学科专业,强调计算机在各专业中的应用。教材内容坚持基本理论适度,反映基本理论和原理的综合应用,强调实践和应用环节。

(2) 反映教学需要,促进教学发展。教材规划以新的工程型和应用型专业目录为依据。教材要适应多样化的教学需要,正确把握教学内容和课程体系的改革方向,在选择教材内容和编写体系时注意体现素质教育、创新能力与实践能力的培养,为学生知识、能力、素质协调发展创造条件。

(3) 实施精品战略,突出重点,保证质量。规划教材建设仍然把重点放在公共基础课和

专业基础课的教材建设上;特别注意选择并安排一部分原来基础比较好的优秀教材或讲义修订再版,逐步形成精品教材;提倡并鼓励编写体现工程型和应用型专业教学内容和课程体系改革成果的教材。

(4) 主张一纲多本,合理配套。基础课和专业基础课教材要配套,同一门课程可以有多种具有不同内容特点的教材。处理好教材统一性与多样化,基本教材与辅助教材、教学参考书,文字教材与软件教材的关系,实现教材系列资源配套。

(5) 依靠专家,择优选。在制订教材规划时要依靠各课程专家在调查研究本课程教材建设现状的基础上提出规划选题。在落实主编人选时,要引入竞争机制,通过申报、评审确定主编。书稿完成后要认真实行审稿程序,确保出书质量。

繁荣教材出版事业,提高教材质量的关键是教师。建立一支高水平的以老带新的教材编写队伍才能保证教材的编写质量和建设力度,希望有志于教材建设的教师能够加入到我们的编写队伍中来。

21 世纪高等学校计算机教育实用规划教材编委会

联系人: 丁岭 dingl@tup.tsinghua.edu.cn

前 言

数据库技术是计算机科学技术中发展最快、应用最广泛的技术之一,其在计算机辅助设计、人工智能、电子商务、行政管理、科学计算等诸多领域均得到了广泛应用,已经成为计算机信息系统和应用系统的核心技术和重要基础。

目前在国内外软件人才市场上,对数据库应用和开发人才,特别是对具有大型数据库管理、开发经验和技能人才的需求量非常大,用人单位在强调应聘者的数据库基本理论知识的同时,更加注重其数据库应用和开发的实际能力。

本教材以帮助学生建立数据库基本概念,提高其数据库应用开发能力和分析解决问题的能力为目标。在讲述传统的“数据库原理”的同时,结合社会需求,介绍 IBM 的旗舰产品——DB2 通用数据库及其典型应用,并以 DB2 通用数据库作为实践环境,阐述数据库相关概念;以期使学生在学数据库基本原理的同时,掌握一种实用的大型数据库应用技术,掌握从事信息处理工作应具备的数据库技术的基本知识、基本技能和操作能力。本教材也可作为参加 IBM DB2 专业认证考试的辅导教材。

全书共分为 13 章。第 1~3 章介绍数据库的一些基本概念和 IBM DB2 通用数据库;第 4 章介绍 SQL 语言;第 5 章介绍关系数据库规范理论;第 6 章是数据库设计的内容;第 7 章讲授 DB2 数据库的存储管理和维护;第 8~11 章分别结合 DB2 数据库系统,讲授数据库的完整性、安全性、并发性,以及数据库恢复技术;第 12 章介绍基于嵌入式 SQL、OLEDB、JDBC 以及 SQLJ 技术的数据库的应用程序开发方法;第 13 章介绍数据库及其应用的前沿技术。

本书由大连交通大学杨鑫华主编,赵慧敏、丁传华任副主编,由多年从事数据库课程教学经验的一线教师编写。其中,第 1 章由牛一捷编写;第 2 章、第 5 章由赵慧敏编写;第 3 章、第 7 章由吴博编写;第 4 章由杨鑫华编写;第 6 章由刘丽娟编写;第 8 章由邓武编写;第 9 章由李雪梅编写;第 10 章由李媛媛编写;第 11 章由王斐编写;第 12 章由李荣健、孙爱香编写;第 13 章由丁传华编写。

在教材编写过程中,得到了 IBM 大中华区大学合作部程毓佳的大力支持和帮助,在此表示诚挚的谢意。此外,在编写过程中,笔者参阅了大量的有关著作、教材,谨对这些著作、教材的编著者表示衷心感谢。

本书由上海交通大学过敏意教授主审,唐飞龙博士协审。审稿人对书稿提出了许多修改意见,在此表示衷心感谢。

作为“教育部-IBM 精品课程”建设项目的内容之一,本书还免费为教师提供电子教学课件、实验指导书及其解决方案等相关课程建设资料,教师可从清华大学出版社网站(www.tup.com.cn)下载。

加入 IBM 学院计划、成为注册会员的教师可以在学校内使用从 IBM 学院计划中获得的教学资料和 DB2 产品 (<http://www-304.ibm.com/jct09002c/university/scholars/academicinitiative/>), 用于教学和非商业性的研究, 同时可以将这些资源提供给自己的学生, 利用这些资源进行软件应用程序的设计、开发和测试。

计算机技术日新月异, 由于编者水平有限, 书中难免存在疏漏、不足之处, 恳请读者批评指正。

编 者

2007 年 5 月

目 录

第 1 章 概述	1
1.1 数据管理技术的发展	1
1.1.1 人工管理阶段	1
1.1.2 文件系统阶段	2
1.1.3 数据库系统阶段	3
1.2 数据库基本概念	4
1.3 数据模型	6
1.3.1 数据模型的组成要素	7
1.3.2 概念模型	7
1.3.3 三种常见的逻辑数据模型	10
1.4 数据库的系统结构	15
1.4.1 数据库的三级模式结构	15
1.4.2 数据库的二级映像功能与数据独立性	17
1.5 数据库管理系统	18
1.5.1 数据库管理系统的主要功能	18
1.5.2 数据库管理系统的程序组成	18
1.5.3 用户访问数据的过程	19
1.6 本章小结	20
习题	21
第 2 章 关系数据库	22
2.1 关系模型概述	22
2.2 关系数据结构及形式化定义	23
2.2.1 关系及相关概念	23
2.2.2 关系模式	25
2.2.3 关系数据库	26
2.3 关系的完整性	26
2.4 关系代数	28
2.4.1 关系代数概述	28
2.4.2 传统的集合运算	29

2.4.3 专门的关系运算	31
2.5 关系演算	37
2.5.1 关系演算概述	37
2.5.2 元组关系演算语言 ALPHA	37
2.5.3 域关系演算语言 QBE	41
2.6 本章小结	41
习题	42
第 3 章 DB2 通用数据库	44
3.1 DB2 通用数据库产品概述	44
3.1.1 DB2 的发展历史	44
3.1.2 DB2 UDB 版本分类	45
3.1.3 DB2 产品组成和架构	46
3.1.4 DB2 客户端	48
3.2 DB2 数据库的对象	49
3.2.1 DB2 实例	49
3.2.2 DB2 数据库对象和层次结构	50
3.3 DB2 数据库目录	51
3.3.1 数据库目录	51
3.3.2 数据库内部目录结构	52
3.3.3 安装路径下数据库内部目录结构	54
3.4 DB2 UDB 图形用户界面	55
3.4.1 工具总览	55
3.4.2 命令行工具	56
3.4.3 开发工具	57
3.4.4 一般管理工具	58
3.4.5 信息工具	62
3.4.6 监控工具	62
3.4.7 安装工具	62
3.4.8 其他工具	64
3.5 本章小结	65
习题	65
第 4 章 关系数据库语言 SQL	67
4.1 SQL 概述	67
4.1.1 SQL 语言概述	67
4.1.2 SQL 语言的特点	68
4.2 SQL 的数据定义功能	69
4.2.1 模式的创建和撤销	69

4.2.2	定义、删除与修改基本表	70
4.2.3	建立与删除索引	73
4.3	数据操纵——查询	74
4.3.1	SELECT 语句的基本格式	74
4.3.2	简单查询	75
4.3.3	多表连接查询	85
4.3.4	标量函数与运算	89
4.3.5	列函数与分组	93
4.3.6	子查询	97
4.4	视图操作	100
4.5	数据操纵——更新	102
4.5.1	插入数据	102
4.5.2	修改数据	104
4.5.3	删除数据	104
4.6	数据控制	105
4.6.1	授权	105
4.6.2	收回权限	106
4.7	本章小结	107
	习题	108
第5章	关系数据库规范化理论	112
5.1	关系规范化的必要性	112
5.1.1	关系数据库的逻辑设计问题	112
5.1.2	规范化理论研究的内容	114
5.2	函数依赖	114
5.2.1	数据依赖	114
5.2.2	函数依赖	114
5.2.3	键的形式化定义	116
5.2.4	候选键的求解理论和算法	117
5.3	范式	118
5.3.1	范式的定义	118
5.3.2	第一范式(1NF)	118
5.3.3	第二范式(2NF)	119
5.3.4	第三范式(3NF)	119
5.3.5	BCNF	120
5.3.6	多值依赖与第四范式(4NF)	120
5.4	关系模式的规范化	122
5.4.1	关系模式规范化的目的和基本思想	122
5.4.2	关系模式规范化的步骤	122

5.4.3 关系模式规范化的要求	123
5.5 本章小结	126
习题	126
第6章 数据库设计	128
6.1 数据库设计概述	128
6.1.1 数据库设计的特点	128
6.1.2 数据库设计的方法	129
6.1.3 数据库设计的步骤	129
6.2 需求分析	130
6.2.1 需求分析的任务	131
6.2.2 需求分析的调查步骤和常用方法	131
6.2.3 数据字典	135
6.3 概念结构设计	140
6.3.1 概念结构的设计方法与步骤	140
6.3.2 局部视图设计	141
6.3.3 视图的集成	143
6.4 逻辑结构设计	145
6.4.1 E-R 图向关系模型的转换	145
6.4.2 数据模型的优化	146
6.4.3 设计用户子模式	147
6.5 数据库的物理设计	147
6.5.1 数据库物理设计的内容	148
6.5.2 选择关系存取方法	148
6.5.3 确定数据库的存储结构	149
6.5.4 评价物理结构	150
6.6 数据库的实施	150
6.6.1 数据的载入和应用程序的调试	150
6.6.2 数据库的试运行	151
6.7 数据库的运行和维护	151
6.8 本章小结	152
习题	152
第7章 DB2 数据库的存储管理和维护	154
7.1 DB2 数据库的存储管理	154
7.1.1 DB2 存储模型	154
7.1.2 缓冲池	155
7.1.3 表空间	156
7.1.4 容器	157

7.1.5	创建表空间	158
7.1.6	表空间状态信息查询	160
7.1.7	查看表空间的容器信息	162
7.1.8	向表空间增加容器	162
7.2	DB2 数据库的数据维护	162
7.2.1	数据维护实用程序介绍	162
7.2.2	数据移动的文件格式	163
7.2.3	DB2 导出实用程序	165
7.2.4	DB2 导入实用程序	168
7.2.5	DB2 装入实用程序	169
7.2.6	db2move 实用程序	173
7.2.7	DB2 数据库维护实用程序	173
7.3	本章小结	177
	习题	177
第 8 章	数据库完整性	179
8.1	数据库完整性概述	179
8.1.1	完整性约束条件	179
8.1.2	完整性控制	180
8.2	实体完整性	181
8.2.1	基本概念	181
8.2.2	DB2 实体完整性	182
8.3	参照完整性	183
8.3.1	基本概念	183
8.3.2	DB2 参照完整性规则	184
8.4	域完整性约束	186
8.5	触发器	188
8.5.1	基本概念	188
8.5.2	DB2 触发器	188
8.6	本章小结	191
	习题	191
第 9 章	数据库安全性	193
9.1	安全性概述	193
9.1.1	网络系统层次的安全技术	193
9.1.2	宿主操作系统层次的安全技术	194
9.1.3	数据库管理系统层次的安全技术	194
9.2	数据库安全控制	194
9.2.1	计算机系统的安全模型	194
9.2.2	安全性控制的一般性方法	195

9.3	DB2 数据库的安全性措施	197
9.3.1	DB2 安全性简介	197
9.3.2	DB2 中的身份验证	198
9.3.3	DB2 中的权限与特权	200
9.3.4	DB2 审计	205
9.4	本章小结	206
	习题	206
第 10 章	数据库并发性	207
10.1	并发概述	207
10.1.1	事务的概念	207
10.1.2	事务的串行调度、并发调度及可串行化	208
10.1.3	并发操作带来的四个问题	210
10.2	封锁	211
10.2.1	封锁	211
10.2.2	封锁协议	212
10.2.3	活锁和死锁	215
10.2.4	两阶段锁协议	216
10.3	DB2 隔离级别	217
10.4	本章小结	218
	习题	219
第 11 章	数据库恢复	220
11.1	数据库故障的类型与恢复策略	220
11.1.1	故障的类型	220
11.1.2	不同故障的恢复策略	221
11.2	DB2 数据库的恢复	222
11.2.1	DB2 数据库恢复的类型	222
11.2.2	数据库恢复必须考虑的因素	223
11.3	DB2 数据库日志文件	225
11.3.1	日志的类型	226
11.3.2	日志配置参数	228
11.4	DB2 数据库的恢复方法	229
11.4.1	崩溃恢复	229
11.4.2	版本恢复	229
11.4.3	前滚恢复	233
11.5	本章小结	235
	习题	235
第 12 章	数据库应用开发方法	236
12.1	嵌入式 SQL	236

12.1.1	嵌入式 SQL 概述	236
12.1.2	嵌入式 SQL 的一般形式	236
12.1.3	嵌入式 SQL 语句和主语言之间的通信	237
12.1.4	不用游标的 SQL 语句	239
12.1.5	使用游标的 SQL 语句	241
12.2	基于 OLE DB/ADO 的数据库开发方法	243
12.2.1	OLE DB	243
12.2.2	ADO 与 OLE DB	244
12.2.3	ADO 应用程序开发	246
12.3	基于 JDBC 技术的数据库开发方法	248
12.3.1	JDBC 驱动	248
12.3.2	JDBC 的接口和类	249
12.3.3	JDBC 应用程序开发	251
12.4	SQLJ 及其在数据库开发中的应用	254
12.4.1	SQLJ 与 JDBC 比较	254
12.4.2	SQLJ 应用程序开发	255
12.5	本章小结	257
	习题	257
第 13 章	数据库及其应用前沿技术	258
13.1	并行和分布式数据库	258
13.1.1	三种多 CPU 数据库体系结构	259
13.1.2	客户-服务器体系结构	261
13.2	数据仓库和数据挖掘技术	261
13.2.1	数据仓库	261
13.2.2	数据挖掘技术	264
13.3	多媒体数据库	265
13.3.1	多媒体数据的特点	265
13.3.2	多媒体数据库简介	265
13.4	移动数据库	266
13.4.1	移动数据库的特点	266
13.4.2	移动数据库系统对数据管理的要求	267
13.4.3	移动计算模型	267
13.5	本章小结	267
	习题	268
附录	DB2 样本数据库表	269
	参考文献	273

数据库技术就是数据管理的技术,它所研究的问题是如何科学地组织和存储数据,高效地获取和处理数据,是当代计算机系统的重要组成部分。作为计算机学科中的一个重要分支,它几乎涉及所有的应用领域,从小型事务处理到大型信息系统,从联机事务处理到联机分析处理,从一般企业管理到计算机辅助设计与制造(CAD/CAM)、从电子商务到电子政务,乃至地理信息系统等,都用到了数据库技术。

1.1 数据管理技术的发展

数据管理是指如何对数据进行分类、组织、编码、存储、检索和维护,它是数据处理的中心问题。这里所说的数据,不仅是指数字,还包括文字、图形、图像、声音等。凡是计算机中用来描述事物的记录,统称为数据(数据的详细概念,将在下节解释)。

随着计算机硬件和软件的发展,数据管理技术得到了不断的完善和发展,经历了人工管理、文件管理和数据库阶段。目前,该技术仍处在日新月异的发展当中。

1.1.1 人工管理阶段

人工管理数据阶段主要是指 20 世纪 50 年代中期以前的这段时间。在此期间,计算机还比较简陋,主要用于科学计算。外部存储器只有纸带、卡片、磁带,没有磁盘等直接存取的存储设备。软件只有汇编语言,没有操作系统和数据管理的软件。数据处理的方式是批处理。人工管理数据具有以下特点:

(1) 数据不保存。由于计算机主要用于科学计算,一般不需要长期将数据保存,只是在计算某一课题时,将原始数据与程序一起输入内存,运算处理后将结果输出。计算机任务完成之后,数据和程序所占的空间一起被释放。

(2) 数据由应用程序自己管理,没有专门的软件管理数据。应用程序既要规定数据的逻辑结构,又要设计物理结构,包括存储结构、存取方法、输入方式等。因此,程序员负担很重。

(3) 数据不能共享。由于数据是面向应用的,一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时,由于必须各自定义,程序间存在大量的冗余数据。

(4) 数据和程序之间不具有独立性。数据的逻辑结构或物理结构发生变化后,必须对程序做出相应的修改,加大了程序设计和维护的负担。

在人工管理阶段,程序与数据之间的关系,是一种一一对应的关系,如图 1.1 所示。

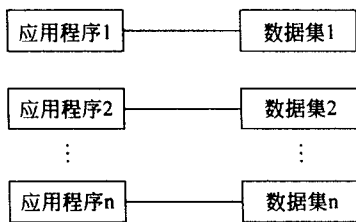


图 1.1 人工管理数据阶段程序与数据间的对应关系

1.1.2 文件系统阶段

文件系统阶段是指 20 世纪 50 年代后期到 60 年代中期以前的这段时间。在此期间,已有磁盘、磁鼓等直接存取的存储设备,在操作系统中也有了专门的数据管理软件,即文件系统。通过操作系统,可以实现文件的打开、读写、关闭等操作。操作方式既可以是批处理,也可以是联机实时处理。此时的计算机,除了用于科学计算,也大量应用于企事业单位的管理。文件系统阶段具有以下特点:

(1) 管理的数据以文件的形式长期保存在计算机外存上。由于计算机大量用于数据处理,采用临时或一次性地输入数据根本无法满足使用要求,数据需要长期保留在外存上,以实现反复进行查询、修改、插入和删除等常见的数据操作。

(2) 数据管理由专门的软件(即文件系统)进行负责。在文件系统中,有专门的计算机软件提供数据的存取、查询、修改和管理功能,能够为程序和数据之间提供存取方法,为数据文件的逻辑结构和存储结构提供转换方法,使应用程序与数据之间有了一定的独立性。程序员可以不必过多地考虑物理细节,将精力集中于算法。而且数据在存储上的改变,不一定反映在程序上,减少了程序维护的工作量。

(3) 文件系统的数据以记录为单位进行存取。文件系统是以文件、记录和数据项的结构组织数据的。文件系统中数据的基本存取单位是记录,即文件系统按记录进行读写操作。只有通过整条记录的读取操作,才能获取其中数据项的信息,而不能直接对记录中的数据项进行数据存取操作。

(4) 文件系统的数据共享性差,冗余度大。文件仍然是面向应用的,一个文件基本上对应一个应用程序,即当不同的应用程序具有大部分的相同数据时,也必须建立各自的文件,而不能共享相同的数据,因此,数据的冗余度大,浪费存储空间。同时,由于文件系统中相同的数据需要重复存储、各自管理,给数据的修改和维护带来困难,容易造成数据的不一致。

(5) 数据和程序间的独立性低。文件系统上的文件是为某一特定应用服务的。文件的逻辑结构对该应用程序来说是优化的,因此,要想对现有的数据再增加一些新的应用会很困难,系统不容易扩充。一旦数据结构改变,必须修改应用程序,修改文件结构的定义。而应用程序的改变(例如,应用程序改用不同的高级语言),也将引起文件数据结构的改变。因此,数据与程序之间仍缺乏独立性。

(6) 数据无结构。文件系统是一个不具有弹性的无结构的数据集合。在文件系统中,虽然记录内容是有结构的,但记录之间和数据文件之间是相互孤立的,使得数据间对外联系无法表达,不能反映现实世界中事物之间的相互联系。

在文件系统阶段,应用程序与数据之间的关系如图 1.2 所示。

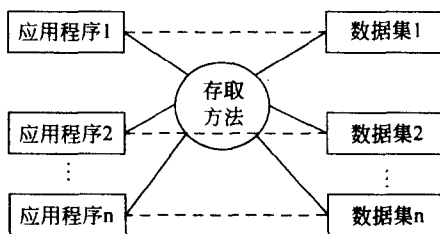


图 1.2 文件系统阶段程序与数据间的对应关系

1.1.3 数据库系统阶段

这一阶段是从20世纪60年代后期开始的。计算机磁盘存储技术取得重大进展,大容量和快速存取的磁盘相继投入市场,为新的数据管理技术的开发提供了良好的物质基础。此外,计算机应用于管理的规模更加庞大,应用越来越广泛,数据量也随之急剧增加,联机实时处理和数据共享的要求也越来越高,文件系统已经不能满足用户在数据管理上的要求。在这种背景下,数据库作为一种新的数据管理技术应运而生。目前,已经开发出了许多商品化的数据库管理系统,数据库技术成为实现和优化信息系统的基本技术。

数据库系统有效地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,提高了数据的一致性、完整性,减少了数据的冗余。在数据库系统管理数据阶段,应用程序与数据之间的关系如图1.3所示。

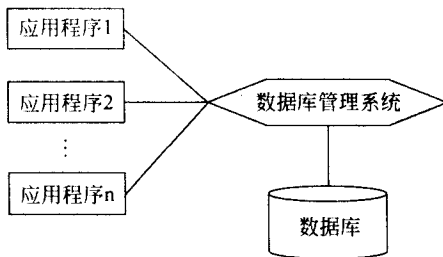


图 1.3 数据库系统阶段程序与数据间的关系

数据库系统管理数据具有以下主要特点:

(1) 数据库系统的数据共享性高、冗余度小。

这是数据库系统阶段的重大改进。由于数据库系统从整体角度来看待和描述数据,数据不再面向某个应用程序,而是面向整个系统。因此,数据库中同样的数据不会多次重复出现,这样便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和不一致性,便于数据维护。

(2) 采用数据模型实现数据的结构化。

数据结构化是数据库和文件系统的根本区别。数据模型不仅描述数据本身的特征,还描述数据间的联系。文件系统的内部虽然有了某些结构,但记录之间没有联系,而数据库系统实现整体数据的结构化。

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,数据的最小存取粒度是记录,粒度不能细到数据项。

(3) 数据独立性高。

数据的独立性是指逻辑独立性和物理独立性。

- 数据的逻辑独立性是指当数据的总体逻辑结构改变时,数据的局部逻辑结构不变,由于应用程序是依据数据的局部逻辑结构编写的,所以应用程序不必修改,从而保证了数据与程序间的逻辑独立性。

例如,在原有的记录类型之间增加新的联系,或在某些记录类型中增加新的数据项,均可确保数据的逻辑独立性。

- 数据的物理独立性是指当数据的存储结构改变时,数据的逻辑结构不变,从而应用程序也不必改变。

例如,改变存储设备和增加新的存储设备,或改变数据的存储组织方式,均可确保数据的物理独立性。

(4) 具有完整的数据管理与控制功能。

数据库为多个用户和应用程序所共享,对数据的存取往往是并发的,即多个用户可以同