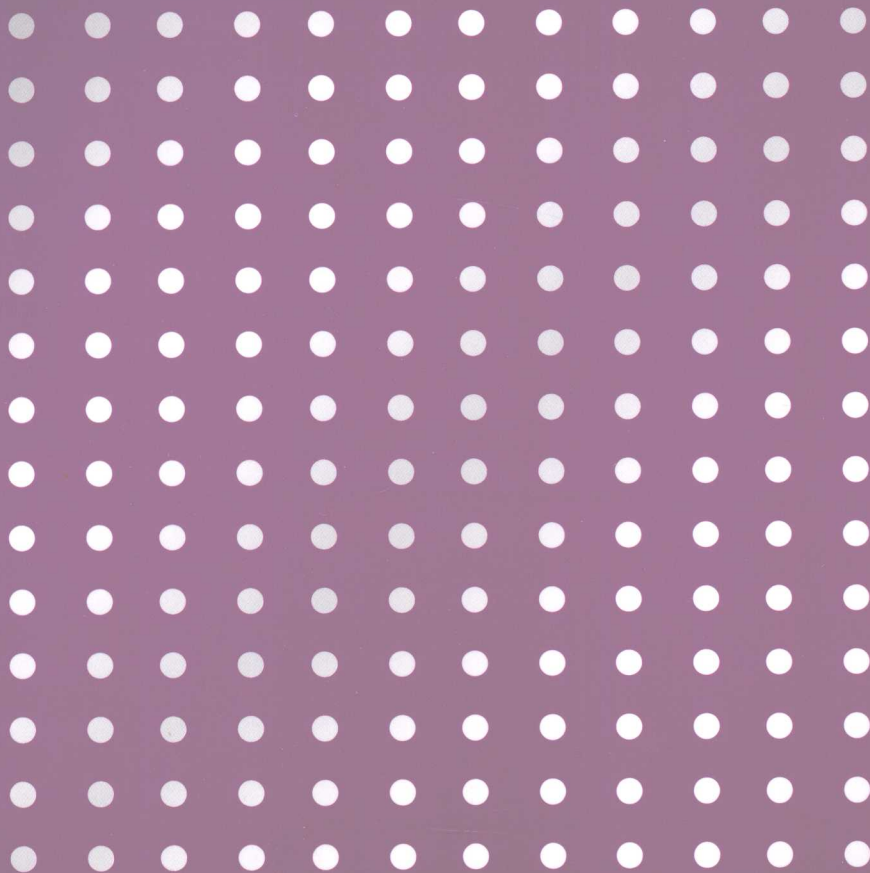


国家级精品课程配套教材

高等院校信息技术规划教材

数据库系统原理与设计

万常选 廖国琼 吴京慧 刘喜平 编著



清华大学出版社

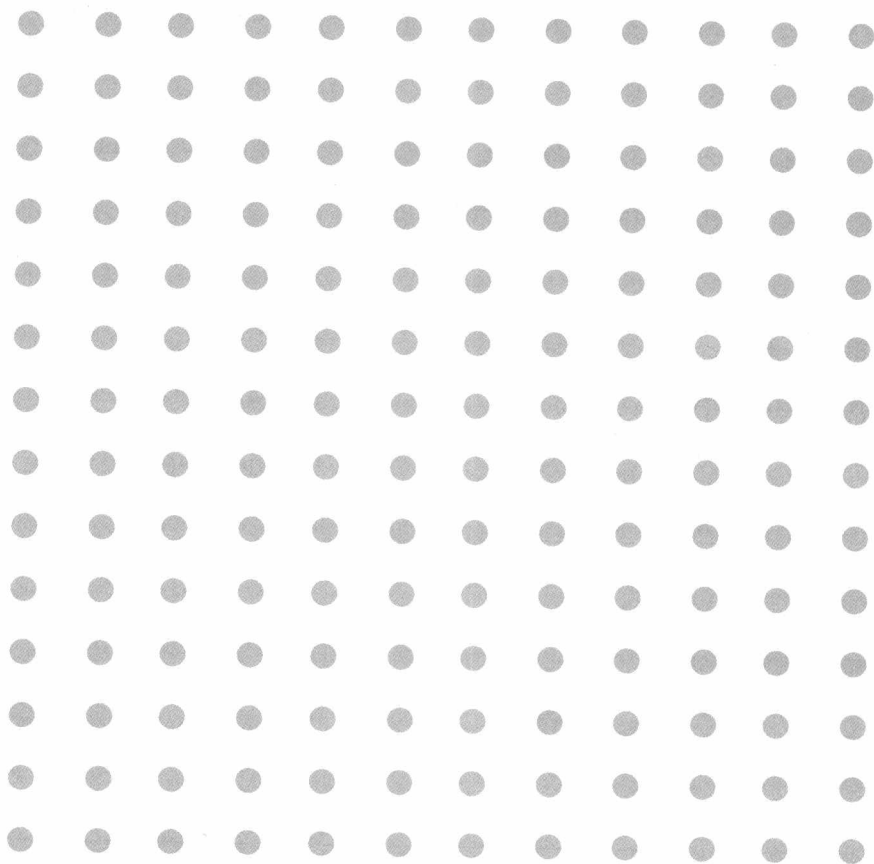


国家级精品课程配套教材

高等院校信息技术规划教材

数据库系统原理与设计

万常选 廖国琼 吴京慧 刘喜平 编著



清华大学出版社
北京

内 容 简 介

全书共分 12 章。第 1 章是数据库系统基本概念;第 2 和第 3 章是关系数据库基础;第 4~第 6 章是关系数据库设计;第 7~第 10 章是关系数据库管理系统;第 11 和第 12 章是数据库应用开发。本书注重数据库应用与设计能力的培养,将数据库设计的内容分散在第 4~第 6 章以及 7.6 节和 9.6 节等章节逐层推进。

本书及配套的实验教程是国家精品课程《数据库系统及应用》的建设教材,有配套的电子教案和教学网站(<http://skynet.jxufe.edu.cn/jpkc/sjk>),可作为计算机及相关专业本科生的数据库系统原理课程教材,也可供数据库爱好者自学和参考。

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

数据库系统原理与设计/万常选等编著. —北京:清华大学出版社,2009.9
(高等院校信息技术规划教材)

ISBN 978-7-302-20590-6

I. 数… II. 万… 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(2009)第 117048 号

责任编辑:焦虹 李玮琪

责任校对:时翠兰

责任印制:杨艳

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

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

印 刷 者:北京市清华园胶印厂

装 订 者:北京市密云县京文制本装订厂

经 销:全国新华书店

开 本:185×260 印 张:25.75

字 数:604 千字

版 次:2009 年 9 月第 1 版

印 次:2009 年 9 月第 1 次印刷

印 数:1~3000

定 价:38.00 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:033895-01

数据库系统是计算机系统的重要组成部分,是企业、机构、互联网乃至整个信息社会赖以运转的基础,在当今社会中扮演着越来越重要的角色。正是由于数据库具有重要的基础地位,数据库理论与技术教育已成为现代计算机科学和相关学科教育中的核心部分,所有计算机及其相关专业的学生都有必要掌握和熟悉数据库理论与技术。

通过多年的数据库课程教学,我们发现学生在学习数据库课程之后,仍然不会“用”数据库——不会设计数据库,不会管理数据库,不会开发数据库应用程序。带着这些问题,我们进行了一系列的数据库课程的教学改革探索与实践,并取得了一定的成绩,如《数据库系统及应用》于2007年获得国家精品课程立项。目前,虽然数据库教材很多,但是很难找到完全适合教学需要的教材。于是,我们决定动手编写一套让学生会“用”数据库的教材,一本“够用”的教材,这便是编写本书的初衷。本书虽然不一定能完全达到目标,但至少开始了有益的尝试。

数据库技术发展至今,已经相当成熟,相关知识博大精深。本书定位于面向计算机及其相关专业本科生的第一本数据库入门教材,在内容选择上颇费思量。在构思本书之前,本书作者一直在思考:对于计算机及其相关专业的本科生来说,需要掌握哪些数据库知识?回答这个问题并不容易,这是因为数据库知识非常丰富,而且由于课时有限,学生不可能了解所有数据库知识。另外,不同用户在使用数据库时,他们的视角是不同的。数据库系统的用户大致可分为4类:数据库管理员、数据库系统分析员、数据库设计人员、数据库应用程序员及终端用户。我们认为,计算机及其相关专业的学生既可能做数据库管理员,也可能成为系统分析员和数据库设计员,更可能是数据库应用程序员。因此,作为一本数据库入门教材,必须要提供这些方面的知识,为学生以后的深入学习打下基础。

基于这些考虑,本书内容包括以下几个部分。

- **数据库系统基本概念**(第1章)。该部分介绍了数据、数据管理、数据库、数据库管理系统和数据库系统等基本概念。同时,也介绍了数据模型、数据抽象、数据库模式等概念。
- **关系数据库基础**(第2章和第3章)。第2章介绍了关系模型(关系数据结构、关系操作和关系完整性约束条件)以及关系代数;第3章介绍了关系数据库的标准语言——SQL,包括数据定义 DDL 语言、数据控制 DCL 语言和数据操纵 DML 语言。
- **关系数据库设计**(第4章~第6章)。第4章介绍了数据库建模方法,包括实体-联系模型基本概念、概念模型设计过程以及如何将 E-R 模型转化为关系模型;第5章介绍了关系数据库设计理论,着重讲述了函数依赖及规范化理论;第6章通过一个实例演示了关系数据库设计过程。
- **关系数据库管理系统**(第7章~第10章)。第7章介绍了关系数据库物理存储结构,包括文件组织、记录组织、索引技术以及物理数据库设计;第8章介绍了查询处理技术,包括查询处理过程、各种关系操作算法以及查询优化技术;第9章讲述了数据库完整性和安全技术,包括数据库安全性、完整性的基本概念和措施,游标、存储过程和触发器,以及应用与安全设计;第10章阐述了事务管理和恢复相关技术。
- **数据库应用开发**(第11章和第12章)。第11章介绍了数据库应用系统的体系结构、常用数据库访问技术和数据库应用开发技术;第12章介绍了 Web 数据库基本概念,讨论了 Web 数据库访问技术,并介绍了 XML 数据库基本概念。

与其他教材相比,本书的主要特点如下。

① 强化对关系数据库模式的理解,以模式导航图为工具,使学生深刻理解关系数据库查询原理,灵活掌握关系代数和 SQL 查询语言。

② 以案例为驱动,通过分析应用需求来介绍数据库建模的基本方法,使学生深刻理解关系数据库设计思想,提高数据库设计能力。

③ 增加关系数据库物理存储结构、查询处理和查询优化等内容,一是有利于学生从底层的角度理解 SQL 查询,二是有利于学生更好地理解关系数据库的物理设计,三是有利于学生更好地理解关系数据库的优点和缺点。该部分内容可根据授课对象的不同有选择地教学。

④ 从使用者的角度,以应用需求为驱动介绍数据库的事务、并发、恢复和完整性、安全性等内容。

⑤ 增加数据库应用开发和 Web 数据库技术介绍,使学生能了解数据库常用访问方法和 Web 数据库的原理和实现技术,为后续学习和数据库应用开发实践打下一定的基础。

本书由万常选、廖国琼、吴京慧和刘喜平编写,其中,第1、第2、第7、第8章由万常选执笔,第4、第5、第6、第10章及7.6节、9.6节由廖国琼执笔,第3和第9章及11.1节由吴京慧执笔,第11和第12章由刘喜平执笔。万常选提出本书的编写大纲,并对全书的

初稿进行了修改、补充和总纂。

本书及配套的《数据库系统原理与设计实验教程》是国家精品课程《数据库系统及应用》的建设教材,有配套的教学 PPT 和教学网站(<http://skynet.jxufe.edu.cn/jpkc/sjk>),可作为计算机相关专业本科生的数据库系统原理课程教材,也可供数据库爱好者自学和参考。

本书在编写过程中,参阅了大量的参考书目和文献资料,在此向参考资料的作者们表示衷心的感谢。

由于作者学识浅陋,书中不足之处在所难免,敬请各位读者批评指正。

编 者

2009 年 3 月

第 1 章 数据库系统概论	1
1.1 数据库系统的作用	1
1.1.1 数据与数据管理	2
1.1.2 数据管理技术的产生与发展	2
1.1.3 数据库应用	8
1.2 数据模型	9
1.2.1 数据模型的分类	10
1.2.2 数据模型的组成要素	11
1.2.3 层次模型	12
1.2.4 网状模型	14
1.2.5 关系模型	14
1.2.6 面向对象模型	16
1.2.7 XML 模型	20
1.3 数据抽象与数据库三级模式	29
1.4 数据库系统	32
1.4.1 数据库系统组成	32
1.4.2 数据库管理系统	33
1.4.3 数据库系统的相关人员	38
本章小结	40
习题 1	42
第 2 章 关系模型与关系代数	44
2.1 关系模型	45
2.1.1 关系数据结构	45
2.1.2 关系完整性约束条件	49
2.1.3 关系操作	55

2.2 关系代数	56
2.2.1 传统的集合运算	57
2.2.2 专门的关系运算	57
2.2.3 关系代数查询综合举例	64
本章小结	67
习题 2	68
第 3 章 SQL 语言	70
3.1 SQL 概述	71
3.1.1 SQL 发展	71
3.1.2 SQL 特点	72
3.1.3 SQL 查询基本概念	73
3.2 简单查询	74
3.2.1 投影运算	76
3.2.2 选择运算	79
3.2.3 排序运算	84
3.2.4 查询表	84
3.3 连接查询	85
3.3.1 等值与非等值连接	85
3.3.2 自表连接	88
3.3.3 外连接	89
3.4 嵌套子查询	91
3.4.1 使用 IN 的子查询	91
3.4.2 使用比较运算符的子查询	95
3.4.3 使用存在量词 EXISTS 的子查询	96
3.5 聚合查询	99
3.5.1 聚合函数	99
3.5.2 分组聚合	100
3.5.3 复杂查询	102
3.6 集合运算	104
3.7 SQL 查询一般格式	106
3.8 SQL 数据定义语言	108
3.8.1 数据库的定义	108
3.8.2 基本表的定义	112
3.8.3 索引的定义	116
3.9 SQL 数据更新语言	118
3.9.1 插入数据	118
3.9.2 删除数据	120

3.9.3	修改数据	120
3.10	视图	121
3.10.1	定义视图	122
3.10.2	查询视图	124
3.10.3	更新视图	125
3.10.4	删除视图	127
本章小结	127
习题 3	128
第 4 章	数据库建模(实体-联系模型)	132
4.1	数据库设计过程	133
4.2	E-R 模型基本概念及表示	134
4.3	约束	137
4.3.1	映射约束	137
4.3.2	码约束	140
4.3.3	参与约束	141
4.3.4	存在依赖与弱实体集	141
4.4	扩展 E-R 特征	143
4.5	E-R 模型设计原则	144
4.6	数据库概念设计实例——大学选课系统	148
4.6.1	概念设计任务	148
4.6.2	系统需求分析	148
4.6.3	数据库概念设计	151
4.6.4	大学选课系统总 E-R 图	152
4.7	逻辑设计——E-R 模型转化为关系模型	153
4.7.1	E-R 模型转化方法	153
4.7.2	大学选课系统 E-R 模型转化实例	156
本章小结	157
习题 4	158
第 5 章	关系数据理论与模式求精	160
5.1	问题提出	161
5.2	函数依赖定义	163
5.3	函数依赖理论	166
5.3.1	函数依赖集闭包	166
5.3.2	属性集闭包	167
5.3.3	正则覆盖	168

5.3.4	无损连接分解	170
5.3.5	保持依赖分解	170
5.4	范式	171
5.4.1	第一范式(1NF)	171
5.4.2	第二范式(2NF)	171
5.4.3	Boyce-Codd 范式(BCNF)	172
5.4.4	第三范式(3NF)	173
5.5	模式分解算法	174
5.5.1	BCNF 分解算法	174
5.5.2	3NF 分解算法	175
5.6	数据库模式求精	176
	本章小结	178
	习题 5	179
第 6 章	关系数据库设计实例——网上书店	181
6.1	需求描述和系统边界	182
6.2	需求分析	182
6.2.1	业务需求及处理流程	182
6.2.2	功能需求分析	183
6.2.3	业务规则分析	185
6.3	确定实体集及属性	185
6.4	确定联系集及 E-R 图	188
6.5	检查是否满足需求	190
6.6	逻辑数据库设计	191
6.7	模式求精	196
6.8	进一步思考	196
	本章小结	196
	习题 6	197
第 7 章	数据库存储结构	199
7.1	文件组织	199
7.1.1	存储介质	200
7.1.2	存储访问	201
7.1.3	定长记录与变长记录	202
7.2	记录组织	204
7.3	顺序索引	207
7.3.1	索引基本概念	207

7.3.2	索引顺序文件	208
7.3.3	多级索引	209
7.3.4	索引的更新	210
7.3.5	辅助索引	211
7.4	B ⁺ 树索引	212
7.4.1	B ⁺ 树索引的结构	212
7.4.2	B ⁺ 树索引的查询与更新	214
7.4.3	B ⁺ 树文件组织	215
7.5	散列	216
7.5.1	散列文件组织	216
7.5.2	散列索引	217
7.5.3	动态散列	218
7.5.4	散列与顺序索引的比较	218
7.6	物理数据库设计	219
	本章小结	222
	习题 7	225
第 8 章	查询处理	227
8.1	查询处理	227
8.1.1	查询处理过程	228
8.1.2	查询代价度量	229
8.1.3	选择运算	230
8.1.4	连接运算	235
8.1.5	其他运算	242
8.1.6	表达式计算	243
8.2	查询优化	244
8.2.1	查询优化概述	244
8.2.2	关系表达式转换	245
8.2.3	结果集大小估计	247
8.2.4	选择执行计划	249
	本章小结	253
	习题 8	254
第 9 章	数据库完整性与安全	257
9.1	数据库安全性	258
9.1.1	数据库安全的基本概念	258
9.1.2	安全标准	259

9.1.3	SQL 存取控制机制	261
9.1.4	审计机制	266
9.2	数据库完整性	267
9.2.1	完整性约束条件	268
9.2.2	实体完整性	269
9.2.3	参照完整性	271
9.2.4	用户自定义完整性	274
9.2.5	完整性约束的修改	276
9.3	游标	277
9.3.1	游标的使用	277
9.3.2	当前游标集的修改	280
9.4	存储过程	281
9.4.1	创建存储过程	282
9.4.2	执行存储过程	288
9.4.3	修改和删除存储过程	288
9.5	触发器	290
9.5.1	创建触发器	290
9.5.2	修改和删除触发器	294
9.5.3	触发器的作用	295
9.6	应用与安全设计	297
	本章小结	299
	习题 9	300
第 10 章	事务管理与恢复	303
10.1	事务	304
10.1.1	问题背景	304
10.1.2	事务概念	305
10.1.3	事务特性	306
10.1.4	事务并发执行与调度	307
10.1.5	事务调度及正确性准则	309
10.2	并发控制	312
10.2.1	基于封锁的协议	312
10.2.2	两阶段封锁协议	314
10.3	恢复与备份	316
10.3.1	故障分类及恢复策略	316
10.3.2	事务访问数据方式	318

10.3.3 基于日志的故障恢复策略	319
10.3.4 检查点	322
10.3.5 备份与介质故障恢复	324
本章小结	325
习题 10	327
第 11 章 数据库应用开发	329
11.1 数据库系统的体系结构	330
11.1.1 软件开发体系结构概述	330
11.1.2 C/S 体系结构	331
11.1.3 B/S 体系结构	334
11.1.4 C/S 与 B/S 结构的结合	336
11.1.5 常用开发体系结构	338
11.2 数据库访问技术	340
11.2.1 ODBC API 和 ODBC 库	341
11.2.2 MFC DAO	343
11.2.3 RDO	344
11.2.4 OLE DB	344
11.2.5 ADO	346
11.2.6 ADO .NET	346
11.2.7 JDBC	347
11.2.8 数据库访问技术小结	348
11.3 数据库应用开发实践	351
11.3.1 ADO 编程对象模型	351
11.3.2 直接使用 ADO Data 控件访问数据库	356
11.3.3 使用智能指针访问数据库	360
11.3.4 使用 VC++ Extensions for ADO 访问数据库	366
本章小结	371
习题 11	372
第 12 章 Web 数据库	374
12.1 什么是 Web 数据库	375
12.2 Web 基础	375
12.3 脚本语言	376
12.4 Web 数据库访问模式	377
12.5 Web 服务器接口	377

12.6	Web 数据库访问技术	378
12.7	XML 数据库	380
12.7.1	XML 应用场景	380
12.7.2	XML 与数据库	382
12.7.3	什么是 XML 数据库	384
	本章小结	387
	习题 12	389
	参考文献	391

数据库系统概论

学习目标

本章从数据库和数据库管理系统这两个最基本概念入手,引出数据库管理系统所涉及的主要问题并做概括性讨论。因此,本章的教学目标主要有两个,一是要求读者对数据库管理系统有一个初步的认识,并了解数据库管理系统的基本功能;二是要求掌握数据抽象、数据模型、数据库模式等核心概念,并理解这些内容在数据库管理系统中的地位和作用。

学习方法

由于本章主要是一些基本概念的介绍,因此要求牢记这些概念,并把这些概念和已经学过的有关概念进行类比,以便加深理解,达到学习目标。

学习指南

本章的重点是 1.2 节和 1.3 节,难点是 1.2 节。

本章导读

本章主要介绍数据库系统最基本、最重要的概念,例如什么是数据、数据管理、数据库、数据模型、数据独立性、数据库的模式、数据库管理系统和数据库系统。数据模型是数据库的组织基础,根据数据抽象的不同级别,可以将数据模型划分为 3 类:概念模型、逻辑模型和物理模型。数据库是最基本的概念,在理解数据抽象的基础上掌握什么是数据库的三级模式和两级映像。数据库管理系统是数据库系统的核心,数据库管理系统有哪些组成与主要功能;数据库系统是数据库技术的应用系统,要求掌握数据库系统中各个部分有什么作用,特别是 DBA 的职责。

1.1 数据库系统的作用

在系统地介绍数据库系统的基本概念之前,本节首先介绍有关数据与数据管理的基本概念,然后介绍数据管理技术的 3 个发展阶段,包括人工管理、文件系统和数据库管理系统,以及数据库和数据库管理系统等概念。

1.1.1 数据与数据管理

1. 数据

描述事物的符号记录称为**数据**,如数值数据、文本数据和多媒体数据(如图形、图像、音频和视频)等。数据是数据库中存储的对象,也是数据库管理系统处理的对象。

在日常生活中,事物通常采用无结构的文本串形式来描述。例如,一个教师的基本情况可描述为:李天乐,男,1968年9月出生,江西南昌人,现工作于江西财经大学信息学院,教授,主要研究方向包括XML数据库、Web数据管理等。

在日常数据管理中,教师的基本情况通常如表1-1所示。

表 1-1 教师基本情况表

姓名	性别	出生年月	籍贯	工作单位/部门	职称	研究方向
李天乐	男	1968年9月	江西南昌	江西财经大学 信息学院	教授	XML数据库、 Web数据管理
...

显然,数据的表现形式不能完全表达其内容,其含义即语义需要经过解释才能被正确理解,因此数据和关于数据的解释是不可分的。例如,“1968年9月”可能是指某人的出生年月,也可能指毕业年月,还可能指参加工作年月,等等。但在表1-1中,其语义已由其所列的表头栏目名解释,即为出生年月。对于以表格形式描述的对象,表头栏目名就是对表中数据的语义解释。

将一个教师的姓名、性别、出生年月、籍贯、工作单位/部门、职称等数据组织在一起便构成一条记录,用于描述一个教师的情况。记录是计算机中表示和存储数据的一种格式或一种方法,这样的数据是有结构的。因此,表格描述的数据称为结构化数据。

2. 数据管理

数据处理是指对各种数据进行采集、存储、检索、加工、传播和应用等一系列活动的总和。数据处理的基本目的是从大量的、可能是杂乱无章的、难以理解的数据中抽取并推导出对于某些特定的人们来说有价值、有意义的数据。数据处理贯穿于社会生产和社会生活的各个领域。数据可由人工或自动化装置进行处理。

数据管理是对数据进行有效的分类、组织、编码、存储、检索、维护和应用,它是数据处理的中心问题。数据管理技术的发展及其应用的广度和深度,极大地影响着人类社会发展的进程。对于基于计算机的数据管理离不开数据管理软件的支持,包括用以书写处理程序的各种程序设计语言及其编译程序、管理数据的文件系统、数据库管理系统以及各种数据处理方法的应用软件包等。

1.1.2 数据管理技术的产生与发展

数据管理技术是应数据管理任务的需要而产生的。在应用需求的推动下,在计算机

硬件和软件发展的基础上,数据管理技术经历了人工管理、文件系统和数据库管理系统 3 个阶段。

1. 人工管理阶段

人工管理阶段主要是指 20 世纪 50 年代中期以前的这段时间。此时的计算机还很简陋,尚没有完整的操作系统,主要应用于科学计算,数据处理的方式是批处理。

人工管理阶段的数据是面向应用程序的,一个数据集只能对应于一个程序,程序与数据之间的关系如图 1-1 所示。数据需要由应用程序自己定义和管理,没有相应的软件系统专门负责数据的管理工作。当多个应用程序涉及某些相同的数据时,必须由各自的应用程序分别定义和管理这些数据,无法共享利用,因此存在大量冗余数据。

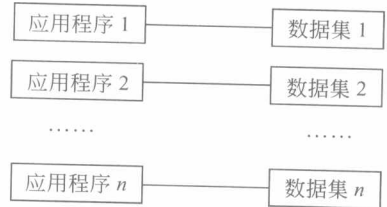


图 1-1 人工管理阶段应用程序与数据之间的对应关系

2. 文件系统阶段

文件系统阶段主要是指 20 世纪 50 年代后期到 20 世纪 60 年代中期的这段时间。此时的计算机已经有了操作系统,在操作系统基础之上建立的文件系统已经成熟并广泛应用。计算机除了应用于科学计算外,已开始应用于数据管理。数据处理方式不仅有批处理,还有联机实时处理。

利用文件系统管理数据,就是由专门的软件对数据进行统一管理。对于一个特定的应用,数据被集中组织存放在多个数据文件(以后简称为文件)或文件组中,并针对该文件组来开发特定的应用程序。文件系统把数据组织成相互独立的文件,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。文件系统阶段程序与数据之间的关系如图 1-2 所示,它的主要特点如下:

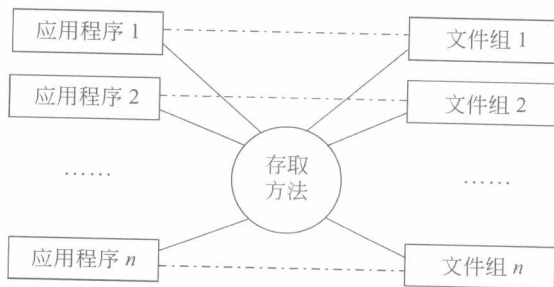


图 1-2 文件系统阶段应用程序与数据之间的对应关系

(1) 文件系统实现了文件内的结构性,即一个文件内的数据是按记录进行组织的,这样的数据是有结构的。但整体上还是无结构的,即多个文件之间是相互独立的,无法建立全局的结构化数据管理模式。

(2) 程序和数据之间由文件系统提供的存取方法进行转换,程序员可以不必过多地