

数据库原理 与工程应用

陈学俭 陈洪亮 编著

中国科学技术大学出版社

数据库系统与应用

陈学佺 陈洪亮 编著

中国科学技术大学出版社

1996·合肥

图书在版编目(CIP)数据

数据库原理与工程应用 / 陈学俊 等编著. —合肥: 中国科学技术大学出版社, 1996 年 5 月
ISBN 7-312-00789-9

- I. 数据库……
- II. 陈学俊 等编著
- III. ①数据库 ②计算机 ③应用
- IV. TP

凡购买中国科大版图书, 如有白页、缺页、倒页者, 由承印厂负责调换。

中国科学技术大学出版社出版发行

(安徽省合肥市金寨路 96 号, 邮编: 230026)

中国科学技术大学印刷厂印刷

全国新华书店经销

开本: 850X1168/32

印张: 10.5 字数: 280 千

1996 年 5 月第 1 版

1996 年 5 月第 1 次印刷

印数: 1 — 5000 册

ISBN 7-312-00789-9/TP·140

定价: 10.80 元

前 言

众所周知，当今世界正处在由工业化社会向信息化社会转变的进程中。以电子计算机为代表的高新技术的迅速发展和普遍应用，使人类解决信息爆炸、进行信息革命、发展信息产业、走向信息时代已成为现实。80年代以来，信息革命的浪潮冲击着科技界、工业界、商贸界……乃至社会的方方面面。用新的知识武装自己，改变传统的、陈旧的知识结构，已成为时代的潮流。

数据库在计算机信息处理与管理中有着十分重要的地位，是整个信息系统建设中不可缺少的部分。了解数据库技术的基本概念和应用方法，已成为许多人迫切的愿望。尤其是从事信息科学的研究和管理人员，全面了解数据库的原理，掌握其应用技术和一般设计方法，更是时代的需要，自我发展的需要。

作者在多年从事信息科学专业的科研和教学的实践中，深感需要一本适合于电子技术与信息科学有关人员学习和掌握数据库原理和工程应用的书籍，它既不同于计算机专业人员的用书（理论性较强，从基本概念、理论分析、典型系统、设计方法等方面详细讨论，分为概论、设计、分布式系统等若干门课程来学习。）又不同于一般的非计算机专业的人员的用书（浅近地介绍一般原理，仅涉及现在流行的某个关系型数据库，只要求掌握典型数据库管理系统的数据语言和编程方法）。而是把原理与应用有机地结合起来，着眼于应用，同时又注意必要的理论基础，清晰的概念，从历史、现状和发展中选择最主要的内容加以阐述、说明。力求达到使读者能在较短的时间里不仅掌握数据库的原理和方法，而且能实际应用，进行数据库系统的设计，同时还能了解数据库技术的新的进展和新的方法。

有鉴于此,呈献给读者的这本书,从内容安排上可以分为两大部分:第1~9章是数据库的基本原理与应用。其中,1~7章介绍了数据库的基本概念,一般理论和技术,分别讨论了层次方法、网状方法、关系方法及各自典型的数据库管理系统,重点介绍目前应用最多的关系数据库;8~9章介绍了数据库的设计理论和设计方法。为读者进行数据库系统的设计与实现打下基础。第10~14章是进一步的内容,涉及高层次的应用,及新的技术和方法。其中,依次介绍了正在迅速发展的分布式数据库系统;随着图象信息应用的扩大而发展起来的图象数据库;目前在程序设计等方面都很受欢迎的面向对象的方法和面向对象的数据库;当前热门的多媒体数据库;在最后一章中,还介绍了各种新技术、新方法的综合应用,主要涉及的是工程数据库系统。

两大部分实际上是关于数据库系统的初级教程和高级教程的有机结合。前者适用于电子与信息专业的本科生学习,后者适合有关学科研究生阅读、参考。

由于作者水平有限,加之时间仓促,书中不妥之处在所难免,恳请读者、同行指正。

本书的出版,得到了各方面的关心和帮助,凝聚着许多同事的辛劳。在此,谨向中国科技大学电子工程与信息科学系微机实验室和信息处理中心支持和帮助本书编写的老师、参与本书录入的工作人员致以衷心的感谢。

愿本书的出版,能为培养和造就高素质、多层次、复合型的信息与计算机应用方面的人才有所帮助。

作 者

1996年春于中国科学技术大学

目 次

前言	i
第一章 绪论	1
1.1 信息科学与数据库	1
1.2 数据管理技术的发展	4
1.2.1 人工管理阶段	4
1.2.2 文件系统阶段	5
1.2.3 数据库系统阶段	6
1.3 数据库系统的特点	8
第二章 数据库系统的组成与结构	13
2.1 数据库系统的体系结构	13
2.2 数据库系统的组成	15
2.3 数据库管理系统 DBMS	17
2.4 用户存取数据库数据的过程	21
第三章 数据模型	24
3.1 数据模型的概念	24
3.2 层次模型	25
3.3 网状模型	27
3.4 关系模型	28
第四章 层次数据库	30
4.1 层次数据模型	30
4.2 物理数据库的描述	33
4.2.1 DBD 控制语句	33
4.2.2 层次序列	35

4.3 逻辑数据库描述	36
4.3.1 逻辑数据库的第一种类型	36
4.3.2 程序通讯块(PCB)和程序说明块(PSB)	38
4.3.3 逻辑数据库的第二种类型	40
第五章 网状数据库	45
5.1 网状数据模型	45
5.1.1 系型	46
5.1.2 系值	48
5.1.3 系的种类	51
5.1.4 系的实现方式	52
5.2 模式数据描述语言 DDL	54
5.2.1 域(AREA)	54
5.2.2 数据库码 DBK(DATA BASE KEY)	55
5.2.3 记录存放方式(LOCATION MODE)	56
5.2.4 从属记录的属籍类别	57
5.2.5 系序(SET ORDER)	58
5.2.6 当前状态	59
5.2.7 系值选择	60
5.2.8 模式数据描述语言 DDL	60
5.2.9 子模式数据描述语言 SUBDDL	63
第六章 关系方法	65
6.1 基本概念	66
6.1.1 关系的数学定义	66
6.1.2 一些概念和术语	69
6.2 关系代数	70
6.3 关系演算	79
6.4 关系数据语言	81
6.5 ORACLE 关系数据库管理系统	91
6.5.1 ORACLE 交互命令式语言 SQL*PLUS	92

6.5.2 ORACLE 应用程序设计	106
第七章 数据库的安全性、完整性与并发控制	113
7.1 安全性	113
7.2 完整性	118
7.3 并发控制	121
7.4 数据库的恢复	124
第八章 关系数据库设计理论	126
8.1 关系模式的表示和设计中的问题	126
8.2 函数依赖	128
8.3 第一范式、第二范式和第三范式	133
8.4 BC 范式	137
8.5 关系模式的分解	140
8.6 多值依赖与第四范式	145
8.7 连接依赖与第五范式	148
第九章 数据库应用系统的设计	151
9.1 数据库设计的一般过程	152
9.2 需求分析	154
9.3 E-R 图	156
9.4 概念模式设计	160
9.5 逻辑模式设计	163
9.6 性能预测和优化	170
9.7 原型建造法	176
第十章 分布式数据库系统	180
10.1 分布式数据库的产生与发展	180
10.2 分布式数据库的定义	182
10.3 分布式数据库系统的特点	186
10.4 分布式数据库的体系结构	189
10.5 分布式数据库管理系统	195
10.6 分布式数据库的分类	198

10.7 分布式查询处理	199
10.8 分布事务管理和并发控制	202
第十一章 图象数据库系统	206
11.1 概述	206
11.2 图象数据的表达方式	208
11.3 图象数据库的查询方式	212
11.4 图象数据库管理系统	214
第十二章 面向对象的数据数据库系统	218
12.1 概述	218
12.2 面向对象方法的发展与特点	220
12.3 面向对象的基本概念	223
12.4 面向对象的数据库(OODB)	226
12.4.1 OODB 的数据模型	226
12.4.2 类结构的描述	228
12.4.3 OODB 的设计方法	231
12.4.4 OODB 的用户接口	237
第十三章 多媒体数据库	241
13.1 多媒体数据	242
13.1.1 多媒体信息	242
13.1.2 多媒体数据的特点	243
13.2 多媒体数据库系统功能	245
13.2.1 多媒体数据库系统	245
13.2.2 多媒体数据库系统功能	246
13.2.3 对多媒体数据的管理	247
13.3 多媒体数据库主要技术	248
13.3.1 数据模型	249
13.3.2 超媒体技术	250
13.3.3 数据压缩和还原	258
13.3.4 数据采集和存贮	260

13.3.5 分布式传输(多媒体数据通信)	262
第十四章 工程数据库概述	264
14.1 工程数据的特点	264
14.2 工程数据库的设计	266
14.2.1 几个典型工程数据库系统	267
14.2.2 工程数据库技术的发展	268
14.3 工程数据库管理系统的功能	270
14.3.1 工程数据库管理系统的功能要求	270
14.3.2 工程数据库管理系统的结构特点	272
附录 A FOXBASE 数据库管理系统	273
A.1 概述	273
A.2 建立数据库	274
A.3 数据的修改、检索和统计	276
A.4 报表输出	281
A.5 多工作区操作	282
A.6 建立和编写程序文件	283
附录 B OREACLE 数据语言 SQL*PLUS	290
B.1 命名规则	290
B.2 数据类型	290
B.3 运算符	291
B.4 函数	292
B.5 数据字典中的一些表	293
B.6 命令概要	294
B.7 主要命令说明	298
B.8 样例表	307
附录 C SYBASE 数据语言 T_SQL	309
C.1 命名规则	309
C.2 数据类型	309
C.3 运算符	310

C.4 函数	311
C.5 系统表	312
C.6 T_SQL 命令概要	313
C.7 主要命令说明	315
参考文献	323

第一章 绪 论

1.1 信息科学与数据库

随着人类社会不断发展,科学技术迅速提高,信息在当今社会的各类活动中起着越来越重要的作用。采用现代科学技术方法进行信息处理已受到人们的普遍重视。全面地、迅速地、及时地获得所需要的信息,进行处理、分析。从而对面临的问题进行判断决策,无论对经济活动、社会活动、科学研究都是十分重要的。当今社会被称为信息化的社会,信息科学被视为现代科学的三大支柱之一。说明信息对人类活动起到十分重要的作用。

在现代社会中,各方面对信息的需求不断增加,信息获取的手段不断提高,信息的数量急剧增加,信息的形式越来越多,结构也越来越复杂。不仅有一般的数字、文字信息,而且有图形、图象信息,不仅有一维、二维的信息,而且有三维、甚至是 n 维的信息,它们彼此之间时间相关、空间相关,有复杂的结构。多种形式的信息给存贮、加工、处理带来了不少困难,带来了不少值得研究的、有应用价值的新课题。具有直接存贮设备的电子计算机的出现使信息进入了全新的历史时期。电子计算机技术的飞速发展,推动了信息化社会的发展。早年人们把使用计算机的信息处理系统称为电子数据处理系统(EDPS),以区别不使用计算机的信息处理,而现在,人们提到信息处理,几乎都是指EDPS。计算机信息处理已深入到社会活动的许多方面。现在的计算机信息处理不仅数据量大、结构复杂,而且便于进行数据的查询、插入、删除和更新等。在这种情况下,一般的计算机软件已不能满

足日益增长的信息处理的要求，于是数据库技术应运而生。总之，信息的数量不断增加，信息在社会各项活动中的地位不断提高，促使信息管理、信息处理技术不断发展。计算机软硬件的日趋完善，计算机技术的提高，使得数据管理、数据处理成为可能，并得到很快的发展。数据库技术就是在这样的背景下得以产生并迅速发展起来的。今天数据库技术已受到普遍重视，在各行各业中得到了广泛的应用，起到了越来越大的作用。在我国，十几年前“数据库”一词对于大众来说还是一个新鲜的名词，而现在无论是企事业单位管理人员，还是办公室普通工作人员，对此都不再陌生，更不必说许多科技人员已经设计并实现了许多有效的信息管理系统。人们清楚地认识到，在现代化的社会，现代化的管理离不开信息与信息处理，有效的信息处理离不开计算机数据库技术。当前，学习、掌握和使用数据库技术，已成为广大管理人员和科技工作者的共同需要，数据库系统的知识已成为各种信息管理的基础知识。

但是，究竟什么是数据库？它在信息处理中起到什么样的作用？并不是每一个人都能作出正确回答的。在这里我们先对数据库系统作一扼要说明，然后再在后续的章节中逐一对数据库系统的结构与组成，数据库管理系统的功能，数据模型、数据语言、关系数据库的规范化理论及数据库系统设计等方面进行阐述。

数据库技术是数据管理的最新技术，数据库管理系统是现代计算机系统的一个重要组成部分。但是要用几句话严格地、准确地给出数据库的定义，的确不太容易。由于人们从不同的角度用不同的观点讨论数据库，因此现在有关数据库的定义也不完全相同。总体来说，数据库系统是一个以计算机软硬件为基础的数据记录保持系统。它的总体目的是为了很好地保持和记录信息。数据库是存贮数据的仓库，数据库存贮数据是根据特定应用，按综合共享的要求存贮的。数据库中的数据可以为系统中不同的用户共享。这些数据往往分成一个个彼此关联的数据“块”存放，各

个数据“块”中的数据彼此几乎不重复，极大地减少了重复数据所造成的存贮空间浪费，也可以在一定程度上避免数据的不一致性。这有利于数据共享和实施标准化。数据库系统还可以采用一定的措施保证数据的安全、保密及出现意外时的恢复。

数据库方法是信息处理的有效手段，现在信息处理往往也被称为数据处理，似乎信息与数据，信息处理与数据处理是一回事。然而严格来说，它们是有区别的，是不同范围中相应的概念。

通常所涉及的三个范围是：现实世界、信息世界和计算机世界这三个范围。

人们通过对信息的管理来实现对人类社会，也就是现实世界的管理。现实世界中被管理的对象是实际存在的各类事物。在现实世界中，把某类事物的全体称为实体集，把该类事物中能相互区分的每个对象称为实体。实体集是由相同特征的实体组成的，实体集的特征就是某些事物的特征，如：颜色、重量等。实体集中不同的实体有各自的特征值，将一个实体与其它实体区分开来的特征或特征组称为实体的标识特征。

在进行现实世界管理时，这些客观事物必然在人们头脑中产生反映，这被称为信息。人们对这些信息进行记录、整理、归类，成为格式化的信息。所以信息是现实世界状态的反映，而信息管理则是现实世界管理的反映。要管理好现实世界，必须借助信息管理。在信息世界这个范围内，用实体的记录表示实体，用实体记录集表示实体集，用属性表示实体集的特征，具体实体的特征值用属性值来表示。实体记录是由实体的属性值组成。

当数据管理进入计算机管理时，这一范围称为机器世界（或计算机世界）。由于计算机只能处理由数字、字母或符号表示的数字化的信息，所以对信息世界中的信息必须进行数字化，数字化后的信息称之为数据(Data)，数据即为信息的符号表示。在机器世界这一范围内用数据集来表示信息世界中的实体记录集，用

数据或记录值表示实体记录。信息世界中的属性在这里用数据项表示，实体记录的属性值用数据项值来表示。

上面简单介绍了三种范围及术语。但在应用中在某些不需要严格区分“信息”与“数据”这两个术语的地方，将不加区分地使用这两种术语。

1.2 数据管理技术的发展

自60年代开始，随着计算机性能的提高，从事数据处理的科技工作者们进行了卓有成效的工作。研究工作主要围绕着提高数据的独立性，降低数据的冗余度，数据共享，提高数据的安全性和完整性等方面进行，使得用户能方便地存贮、管理、使用数据资源。因而数据管理技术的发展也根据其数据独立性、冗余度、数据相互间的联系、安全性和完整性，分为人工管理、文件系统、数据库系统三个阶段。

1.2.1 人工管理阶段

这一阶段大约始于50年代末。当时计算机刚问世不久，主要用于科学计算，计算机的软硬件均不完善，外存贮器没有可以直接存取数据的磁盘，通常只有磁带、卡片及纸带；软件方面还没有操作系统和高级语言，更没有对数据进行管理的软件。数据主要靠人工管理。许多工作均需要程序员来完成，程序员不仅要规定数据的逻辑结构，而且要规定其存贮结构，需自己分配存贮空间，决定存贮方式。因此数据和相应的程序是不可分的，彼此相互依赖，当数据的物理组织或存贮设备改变时，应用程序就要随之改变，即数据对于程序来说不具有独立性。

由于每个程序员都有自己的数据，即使不同的程序员使用同一组数据，也只能各自存贮，不能共享，这样就带来了许多重复的存贮。

综上所述,人工管理阶段的特点是:

①没有专用的数据管理软件。程序员对数据存在严重的依赖。数据的独立性很差。

②基本上不使用文件,偶尔使用,也只是顺序文件。

③数据面向程序,每个程序都使用自己的数据,如果不同程序用到相同的一部分数据,也只能重复存贮,所以数据的冗余度很大。

这一阶段数据的逻辑组织与物理组织的关系如图1.1所示。

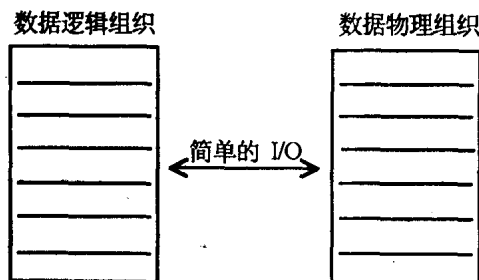


图 1.1 人工管理的特征

1.2.2 文件系统阶段

从50年代末到60年代中期,随着科学技术的发展,计算机技术得到了很大的提高。硬件方面,有了大存贮量的磁盘、磁鼓等直接存取数据的设备;软件方面,有了操作系统、高级语言,还出现了专门用于管理数据的软件,被称为文件系统或称信息管理模块,它往往包含在操作系统之中。在这一阶段,人工数据管理中存在的许多问题得到了改进。系统设置的专门软件“文件管理系统”,负责对数据进行管理,使得物理数据有较高的独立性。数据的逻辑组织与物理组织之间由存取方法实现转换。如图1.2所示。

在数据处理中,常常需要把数据留在外存上反复处理,这促

进了文件处理技术的发展,除顺序文件外,还出现了索引文件,直接存取文件、倒排文件等结构。实现了以文件为单位的数据共享。这时可以以记录为单位存取,但是还不能以数据项为存取单位。文件本身对应着一个或几个应用程序,也就是说数据还是面向应用的。

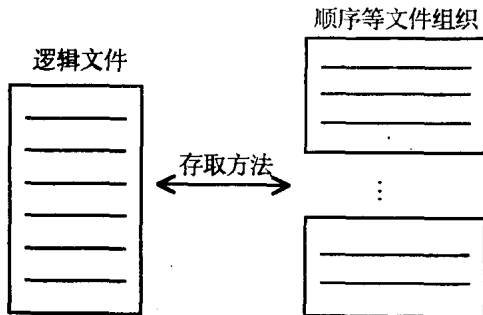


图 1.2 文件系统的数据组织

总之,文件系统较之人工管理有了很大的改进,但是还存在许多不足之处,其特点可归结为:

①程序与数据有一定的独立性,数据的独立性较人工管理有所提高。

②提供了多种文件组织形式。

③实现了以文件为单位的数据共享,数据的冗余度减少了。但还不能实现以数据项为单位的共享。

1.2.3 数据库系统阶段

60年代末,可进行直接存取的磁盘技术的提高和存储容量的增大,为数据管理技术的发展提供了有利的条件。同时大量的数据存在需要大规模的管理,促进了数据管理技术的发展。数据库技术则是在这样的背景下产生并迅速发展、日趋完善的。