

钟恢扶 王仲东 编著



DBASE-Ⅱ

关系数据库的 设计与应用

湖北科学技术出版社

关系数据库的设计与应用

钟恢扶 王仲东 编著

湖北科学技术出版社

内 容 提 要

这是一本普及数据库软件设计技术和应用技术的书。

本书以在苹果机(APPLE-Ⅱ)和IBM-PC/XT及其兼容机上配置的最为流行的“大众数据库”—dBASE-Ⅱ为对象,详细的介绍了关系数据库的设计及应用技术。从层次结构上分为两部分,第七章至第十二章为第一部分,介绍关系数据库dBASE-Ⅱ各种命令的功能及使用,命令文件的编制及上机操作方法;第一章至第六章、第十三章为第二部分,以E-R模型(即实体联系模型)和规范理论为指导,详细的论述了关系数据库应用程序设计的基本理论和基本方法,以及从调查现实问题入手到应用程序设计完成这一设计过程。

本书适合于中等以上文化程度的广大科技干部的管理干部作为数据库的培训教材或自学读物,也适合于有苹果机或IBM-PC/XT及其兼容机的大专院校,中专等教学单位,作为微机数据库教材或参考资料。

关系数据库的设计与应用

钟恢扶 王仲东 编著

湖北科学技术出版社出版 新华书店湖北发行所发行

湖北省阳新县印刷厂印刷

787×1092毫米 32开本 18.875印张 416,000字

1986年6月第1版 1986年6月第1次印刷

印数1—7,000

统一书号: 15304·138 定价3.45元

24455/03

前 言

数据库技术是六十年代初开始发展起来的一门数据信息管理自动化的新技术，是计算机科学的一个重要分支。它日益广泛地应用于国民经济、文化教育、军事指挥、科学研究等各个社会领域，为计算机应用开辟了无限广阔的前景。

关系数据库，尤其是中、小型数据库是世界上最受欢迎、销售量最大的数据库，有“大众数据库”的美称。它学习容易（难度只有BASIC语言的三分之一）、操作直观、使用灵活、编程方便；它适应的环境非常广泛（一般八位机及十六机均可运行这种软件）；它又有很强的数据处理能力。在社会生产力高速发展，信息量急剧增加，人们对信息和数据的利用与处理要求越来越高的今天，关系数据库定将成为我国计算机应用领域强有力的工具。现在有越来越多的人希望了解它，应用它。

目前我国微型机算机的应用日益普及，以关系数据模型为基础的关系数据库系统在微型计算机的应用领域中具有重要的地位。使用关系数据库管理系统（如dBASE—II等）来开发应用系统，一是要掌握具体关系数据库管理系统所提供的数据库语言，二是要具有较扎实的关系数据库逻辑设计的基础知识。关系数据库的逻辑设计（对于dBASE—II、III等就是库中主文件的设计）必须以正确地反映现实世界、数据冗余（即重复）最小、共享程度高、确保数据的完整性

(即正确性)、尽量保证应用程序与数据的独立性、访问时间短等为目标。一个“好”的数据库能够收到明显效益,一个“坏”的数据库将会带来不堪设想的后果。因此,关系数据库的逻辑设计是开发应用系统的基础和关键。数据库的设计理论目前虽然还很不完善,但关系数据模型的规范理论为数据库的逻辑设计奠定了理论基础、提供了有力的工具和正确的指南。为了使广大的计算机使用者能够掌握数据库设计的基本理论和方法,多、快、好、省地开发各种应用系统,我们编写了这本书,奉献给广大读者。

在我国所拥有的十几万台微型计算机中,绝大多数都配上dBASE—II小型数据库管理系统,这就为普及数据库知识奠定了物质基础。但目前介绍dBASE—II数据语言即命令系统的书为数不多,因此我们在书中用了较大的篇幅介绍dBASE—II命令系统的使用方法和应用程序的设计方法和设计过程。

本书共分十三章,第一章至第三章介绍数据库系统的基本原理。第四章介绍关系数据操纵语言理论基础即关系代数和关系演算。第五和第六章讨论数据理论(即规范理论)和数据库的设计与设计过程。第七至第十一章详细地介绍了dBASE—II各种命令的操作方法、编制应用程序的语句及应用程序的编制过程。第十二章介绍了在IBMPC/XT及其兼容机上汉字dBASE—II的操作,第十三章简单地介绍了关系代数表达式的查询优化方法。

本书用了较多的例子来说明概念和方法,力求通俗易懂。它适合于工矿企业、商业、外贸、银行、统计等诸多领域的广大科技人员和管理干部作为培训教材或自学读物,也可供上述人员及有关大专院校师生作为参考资料使用。随着

计算机在我国的广泛应用，数据库将和BASIC语言一样，被人们认识和广泛应用。

在本书的编写过程中，湖北省暨武汉市计算机学会、中国人民解放军空军雷达学院计算机教研室和三四五六二部队给予了大力的帮助。空军雷达学院计算机教研室郑云峰同志详细地阅读了第一至第六章的初稿，和验证了其中的例子，并提出了许多宝贵意见。中国人民解放军三四五六二部队陈玉荣同志绘制了大部分图。空军雷达学院王士俊副教授、王长胤副教授审阅了全部书稿，在此一并表示诚挚的谢意。

由于编者水平有限，时间又很仓促，书中一定存在不少缺点和错误，敬请读者批评指正。

目 录

绪 论	(1)
第一章 从现实世界到计算机世界	(21)
1·1 现实世界	(21)
1·2 实体联系模型	(25)
1·3 关系数据结构模型	(39)
1·4 数据组织基本概念	(51)
第二章 数据库系统的体系结构	(67)
2·1 数据描述语言和数据模式	(67)
2·2 数据库的分级和体系的分层	(70)
2·3 数据库分级和体系分层的目的	(71)
2·4 数据字典	(74)
第三章 数据库管理系统	(76)
3·1 数据操纵语言	(77)
3·2 数据保护	(83)
3·3 数据库管理系统的组成	(89)
3·4 用户访问数据库的过程	(90)
第四章 关系数据语言	(92)
4·1 概述	(92)
4·2 关系代数	(96)
4·3 关系演算	(125)
4·4 关系数据库的模式和子模式	(132)

第五章 规范理论	(142)
5.1 数据依赖	(142)
5.2 关系模式的范式	(158)
5.3 数据依赖公理系统	(182)
5.4 关系模式的分解	(214)
第六章 关系数据库的设计	(259)
6.1 概述	(259)
6.2 E—R模型设计	(261)
6.3 模式设计	(276)
6.4 数据库物理设计	(290)
第七章 dBASE—II 系统简介	(292)
7.1 系统的特点、功能、性能及基本结构	(293)
7.2 CP/M 操作系统与dBASE—II 系统	(299)
第八章 数据库主文件的建立	(309)
8.1 数据库主文件	(309)
8.2 数据库主文件结构的建立	(313)
8.3 主文件记录的建立	(326)
第九章 内存变量、表达式和函数	(341)
9.1 内存变量	(341)
9.2 表达式	(347)
9.3 函数	(353)
第十章 数据库的操作	(364)
10.1 主文件的排序和索引文件的建立	(364)
10.2 主文件的连接	(371)
10.3 检索操作	(376)
10.4 主文件的合并	(390)
10.5 主文件的间接生成	(394)

10.6	dBASE—II 与外部的通讯	(400)
10.7	统计操作	(405)
10.8	报表生成	(409)
10.9	数据库的修改与维护	(417)
10.10	系统的其它操作	(439)
第十一章	dBASE—II 的应用程序——命令文件	(445)
11.1	命令文件概述	(445)
11.2	命令文件的结构	(450)
11.3	程序交互语句	(471)
11.4	输出格式设计	(477)
11.5	命令文件——应用程序的编制	(499)
第十二章	IBM PC/XT 及其兼容机上汉字 dBASE—II	
	的操作	(516)
12.1	MS—DOS 一般介绍	(517)
12.2	汉字 dBASE—II 的汉字输入操作	(541)
第十三章	查询优化	(552)
13.1	关系代数表达式在优化中的作用	(552)
13.2	关系代数表达式的等价代换规则	(555)
13.3	优化的一般策略	(559)
附录: 附录 1	控制键功能表	(571)
附录 2	dBASE—II 操作中人机对话汉语提	
	示	(575)
附录 3	dBASE—II 命令清单	(582)
附录 4	ASCII(美国标准信息交换码)表	(589)

绪 论

一、数据库技术的产生

随着计算机技术的蓬勃发展，计算机不仅应用于数值计算，而且日益广泛地应用于非数值数据的处理，如图书资料管理、企业管理等。由于非数值数据量庞大、离散、易变和数据间联系密切等特点。早期的文件管理系统就不敷使用了。其主要问题在于：

①数据冗余。数据文件是用户各自建立的，几个用户即使有许多相同的数据也得放在各自的文件中。存贮的数据必然造成大量的重复（称为数据冗余），浪费存贮空间。

②数据不能共享。由①可知，用户“各自为政”，用户间“井水不犯河水”，某用户数据文件不能供其他用户使用。

③数据不独立。即应用程序与数据之间依赖性很强。因为文件完全是根据具体的应用程序的要求而建立的，如果数据文件结构一旦需要修改，那么应用程序也必须作相应地修改。

④用户负担重。文件系统仅为用户提供一种简便的、统一的存取文件的方法，而文件的处理、数据的保护等均交给用户程序去管。要实现一个企业管理系统，用户需要编制大量的程序。

文件系统存在的这些问题，正是数据库技术的目标，数据库技术就是为了解决这些问题提出来的。

这些问题的实质有两个方面：

①文件系统缺乏对数据的统一组织。文件系统仅处理用户数据，对于组成记录各个数据项之间，以及各个文件的记录之间的联系未提供任何信息、不作任何解释，而数据之间的联系正是非数值数据管理至关重要的问题。换句话说，文件系统中的数据是非结构化的数据，是一种分散的、孤立的数据。

②文件系统没有对数据进行集中管理和控制的软件。由于文件系统中的数据是非结构化的，所以无法做到按照一种统一的方法对数据进行管理、控制和维护。

针对文件系统存在的问题，六十年代末人们就开始研究数据管理的新技术——数据库技术。1968年 IBM 公司研究成功IMS系统(层次模型数据库系统)。1969年美国数据系统语言协会提出了 DBTG 系统(网状模型数据库系统)。1970年 IBM 公司的 E·F·Codd 提出了关系数据库模型，随后出现了 SYSTEM R 系统等。

二、数据库系统的组成

数据库系统主要由五部分组成：硬件系统、数据库集合、数据库管理系统、数据库管理员和用户。

运行数据库系统的计算机系统必须有足够大的内存，足够大的外存和具有较高的通道能力。

按照 J·Martin〔1〕的定义，“可以把数据库定义为存贮在一起的相关数据的集合，这些数据无有害的或不必要的冗余，为多种应用服务；数据的存贮独立于使用它的程序；对数据库插入新数据，修改和检索原有数据均能按一种

公用的和可控制的方法进行；数据结构化，为今后的应用研究提供基础。当某个系统中存在结构上完全分开的若干数据库时，则说该系统包含一个数据库集合”。

数据库管理系统(Data Base Management System 简记为 DBMS) 是数据库系统的核心，它是管理和维护数据库的软件系统。其功能及组成将在第三章中详细讨论。

对于大型数据库系统为了对系统中的数据集合实行集中控制，除了在软件上考虑外，还必须要有人来进行总体控制。这种人称为数据库管理员(Database Administrator 简称 DBA)。其任务是：参加数据库的设计和调试，对数据库系统的运行进行监督和控制，维护和修改数据库等。

数据库的用户有两类。一类是应用程序员，负责编制使用数据库的应用程序(一般用COBOL或PL/I等语言)，这些程序可以是批处理程序，也可以是“联机”程序(设计成支持终端用户，从联机终端上与系统“交互会话”)。这类用户可以通过批处理程序对数据库进行检索、插入、删除和修改等操作。另一类是联机终端用户。这类用户只能使用查询语言，对数据库进行检索、插入、删除和修改等操作，但以检索为主。这类用户是数据库的直接用户，一般是工程技术人员、科研人员和管理人员。

三、数据模型

数据模型至少是下列三个成分的组合：

- ①若干数据结构类型(数据库装配块)；
- ②若干操作或推理规则，它们能应用到①中列出的任何有效的数据类型，以检索、推导、或修改任意组合起来的结

构的任意部分的数据；

③若干具有完整性（正确性）的通用规则。这些规则隐含地或明显地定义一组相容的数据库状态、或状态变化、或两者兼有；它们有时可以表示成插入、删除、修改规则。这些规则适应于任何使用这种模型的数据库，故称之为完整性通用规则。

下面着重介绍数据结构模型。

数据库中的数据是按照数据结构模型组织的、结构化的数据。为了减少数据冗余和实现数据共享，不仅需要考虑到用户数据（即用户使用的数据库），而且还必须考虑用户数据之间的联系（即同一文件或不同文件的记录与记录之间的联系及记录内部各数据项之间的联系）。反映和实现数据联系的方法称为数据（结构）模型。目前流行的数据模型有三种，按照图论建立的有层次模型和网状模型，按照关系理论建立的有关系模型。

（一） 层次模型

现实世界中许多状态可以用图形来描述。一个图是一些结点和连接两个结点之间的连线所组成，至于连线的长度及结点的位置是无关紧要的。

一个图可以定义为： $G = \langle V, E \rangle$ ，其中 $V(G) = \{V_1, V_2, \dots, V_n\}$ 是结点 V_i 的集合， $E(G) = \{e_1, e_2, \dots, e_m\}$ 是连接结点的边集。

若边 e_i 与结点有序偶 $\langle V_i, V_k \rangle$ 相关联，则称该边为有向边。每一个边都是有向边的图称为有向图。

如果我们把记录看作为图的结点，记录间的联系看作为图的有向边（记录间的联系是有方向的），那么就可以用图的

基本概念来刻画数据间的联系。

记录 R_i 与记录 R_j 之间用 L_{ij} 联结的有向联系称为基本层次联系。始点的记录 R_i 称为双亲，终点的记录 R_j 称为 R_i 的子女。

用图论中树的概念来描述数据间联系即为层次模型。其定义如下：

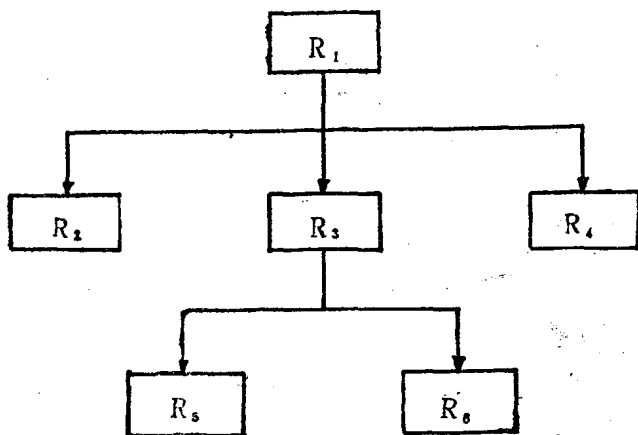
如果一个“基本层次联系”的集合，满足条件：

- ①有而且仅有一个结点无双亲；
- ②其它结点有而且仅有一个双亲。

则称为层次模型。

此定义与图论中树的定义是一致的。

在层次模型中，无双亲的（只有一个）结点称为树根；同一个双亲的结点称为兄弟；无子女的结点称为叶子。如图（1）中， R_1 是根； R_2 、 R_3 和 R_4 是兄弟， R_5 和 R_6 也是兄弟； R_2 、 R_4 、 R_5 和 R_6 是叶子。



图（1）

因为层次模型中的每个记录（根除外）只有一个双亲，所以从任何一个叶到根的映象是唯一的。因此，对于除根以外的记录只要指出它的双亲记录，就可以实现记录间的联系，这样就表示了层次模型的整体结构。

按照层次模型建立的数据库系统称为层次模型数据库系统，简称层次数据库系统。IMS (Information Management System) 是层次数据库系统的典型代表。

(二) 网状模型

如果一个“基本层次联系”的集合满足下面两个条件，则称其为网状模型。

这两个条件是：

① “基本层次联系”集合中，可以有一个以上结点无双亲；

② 该集合中至少有一个结点有多于一个的双亲。

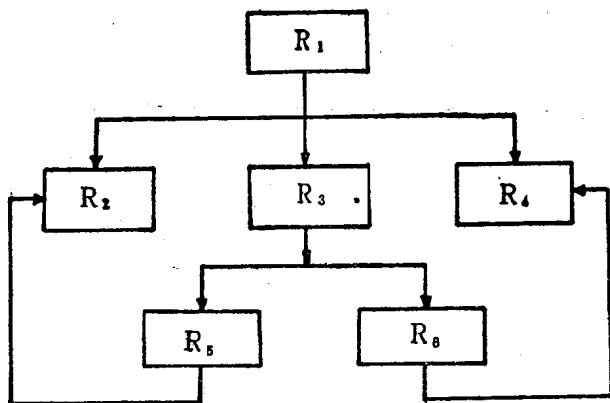


图 (2) (a)

图(2)中的(a)、(b)、(c), 给出了网状模型的几个简单示例。

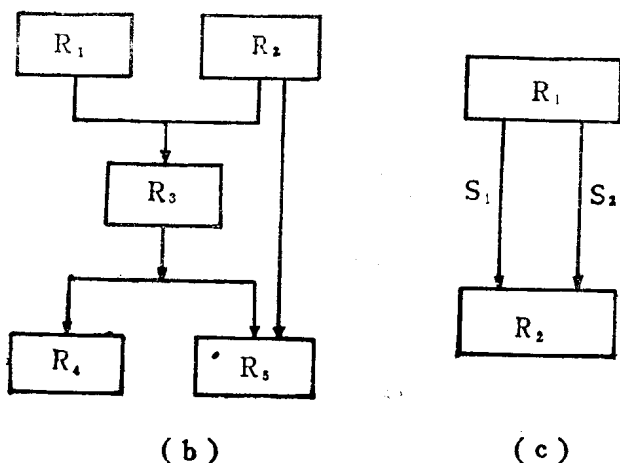


圖 (2)

由于网状模型中的某些子女有一个以上的双亲，其联系不是唯一的，因此对每种联系必须加以命名以示区别。如图(2)(c)中有两种联系，用联系名 S_1 及 S_2 来区分。

按照网状模型建立的数据库系统称为网状模型数据库系统，简称网状数据库系统。其典型代表是 DBTG 系统（DBTG——Data Base Task Group 是美国语言协会下属的一个组织，DBTG 系统是数据任务组提出来的），

无论是层次模型还是网状模型都必须事先定义存取路径。用图论基本概念来描述数据间的联系,其目的之一在于建立存取路径。

(三) 关系模型

在关系模型中，不仅数据用二维表来表示，而且数据间的联系也用二维表来表示。例如表（1）是医生二维表，表（2）是病人二维表，表（3）是医生与病人间的诊断联系二维表。

表（1）医生二维表

R_1

医生号	姓 名	职 称
3401	陈世康	科主任
3402	王 英	主治医生
3403	李卫华	医 生
3404	张小平	见习医生

表（3）医生与病人诊断二维表

R_3

医生号	病床号
3401	2—01
3401	2—02
3402	2—03
3402	2—04
3405	2—05
3406	2—06

表（2）病人二维表

R_2

病床号	姓 名	性 别
2—01	孙玉国	男
2—02	王国华	男
2—03	黄永平	男
2—04	李 英	女
2—05	赵玉珍	女
2—06	周 萍	女

一个二维表的内容称为一个关系。必须指出，算法语言中定义的二维表与这里所说的关系有本质的区别。在关系