



实用数据库技术

唐常杰 韩仲清 编著

四川大学出版社

实用数据库技术

唐常杰 韩仲清 编著

四川大学出版社

一九八六年八月

内 容 提 要

本书共四篇十四章。第一篇数据库基础知识；第二篇关系数据库dBASEⅢ；第三篇关系数据库技术，包括关系模式规范化的Smith依赖图方法，数据库的设计与实现及关系代数；第四篇数据库模型，包括网状模型和层次模型的基本思想以及第二代数据库模型（语义模型和对象历史）的研究动向、成果和参考文献。书末的附录给出了上机实习指南和一个供实习时参考的DBMS的Pascal源程序。

本书前三篇可作为高等学校财经、情报、图书馆、管理专业的数据库课程教材，配合上机实习dBASEⅢ，全书可作为计算机专业学生的数据库课程教材，配合实习数据库管理系统程序的编写。

本书由浅入深，便于自学，可供对数据库原理及技术感兴趣的广大科技人员及大专院校师生参考。

本书第一、三、四篇及附录由唐常杰编写，张一立校审。第二篇由韩仲清编写，张陞楷校审。

实用数据库技术

唐常杰·韩仲清著

四川大学出版社出版

四川省新华书店发行 成都体育学院印刷厂印刷

开本787×1092毫米 1/16 17.875印张 字数43,736千

1986年8月第一版 1986年8月第一次印刷

印数：1—6500册

统一书号：15404·2 定价：3.40元

序 言

人类进入了信息社会时代，数据库技术渗透进了各行各业。

由于其重大意义和深刻影响，数据库课程已成为财经，管理，图书馆，情报等专业，当然，更是计算机科学各专业学生的必修课。本书正是为上述各专业学生写的一本教科书。

本书在编写特色上力求作到广（读者广），实（强调实习），浅（入门知识浅）、新（内容新）。

为了使读者尽量广泛，对内容作了精心的安排，使得前三篇适合财经、管理，图书馆，情报等专业的学生、研究生或非计算机专业人员使用，而全书满足了计算机科学各专业学生的大纲要求。

本书强调上机实习。在机器上用数据库，写数据库是学习数据库技术的重要环节。财经、管理、图书馆、情报专业的学生应结合第二篇内容的学习，上机实习dBASE III。本书附录一给出了使用IBM—PC和dBASE III的一些基本常识。计算机科学各专业的学生，凡有条件者都应学习编写一个简单的数据库管理系统程序，附录二给出了这个实习的指南，并以采用PASCAL语言为例，给了若干提示。需要这个实习的学生可以尽早地粗读附录二，以便带着问题学，留心有关知识。附录三给出了一个用PASCAL语言编写的DBMS的实例。可供借鉴。

门槛浅，才易入门。本书假定读者开始对数据库是完全陌生的。在内容安排上由浅入深，先看看dBASE III，再用用dBASE III。通过上机实习获得足够感性认识后，再学习理论和技术。了解关系数据库后，再利用触类旁通的优势去探讨网状，层次模型以及第二代数据库模型的基本思想。

新，才有存在价值，才能跟上计算机科学知识更新的步伐。本书收入了若干新的成果。第九章的依赖图方法是H·C·Smith于1985年8月才发表的。它避开了艰深的模式分解理论，使普通用户也能很容易地合成出规范化模式。第十四章介绍了第二代数据库模型的动机，主要思想及研究动态，还列出了若干文献，供学生作毕业论文时参考。

由于编著者水平有限，书中难免有欠妥之处，恳请批评指正。

编 者

1985.10.于四川大学

目 录

第一篇 数据库导论

第一章 绪 论	1
1.1 什么是数据库.....	1
1.2 数据库发展史.....	1
1.3 关于阅读本书的建议.....	3
第二章 实体及其联系	3
2.1 dBASE III 一瞥.....	3
2.2 一览表的缺点.....	8
2.3 实体—联系图.....	10
2.4 把E—R图转换为关系模式.....	12
习 题	12

第二篇 dBASE III

第三章 dBASE III 概述	13
3.1 引言.....	13
3.2 dBASE III 文件.....	15
3.3 常量、变量、函数及数据类型.....	15
3.4 表 达 式.....	19
3.5 dBASE III 的 运 行.....	21
3.6 几个简单命令的用法.....	22
3.7 dBASE III 技术规格和操作总则.....	26
习 题	28
第四章 数据文件及其操作	29
4.1 数据库文件的建立和数据的装入.....	29
4.2 工作区的选取和数据库文件的打开.....	35
4.3 检索命令.....	38
4.4 修改命令.....	43
4.5 定位记录指针.....	49
4.6 添加记录的命令.....	52
4.7 删除记录的命令.....	56

4.8 统计方面的命令	59
4.9 排序和索引的命令	61
4.10 数据库文件的连接	71
4.11 报表格式文件和标记格式文件	73
4.12 数据库备注文件	80
习 题	81
第五章 函数、运行特征、存贮文件、文本文件和系统通讯	83
5.1 dBASE III 函数	83
5.2 设置运行特征	94
5.3 存贮文件	98
5.4 文本输出文件	99
5.5 系统通讯	101
习 题	104
第六章 命令文件和格式文件	105
6.1 命令文件	105
6.2 命令文件中可能用到的新命令	109
6.3 格式文件	122
6.4 命令文件实例	124
6.5 过程文件	129
6.6 命令文件的调试	133
第七章 dBASE III 附录	135
7.1 dBASE III 与 dBASE II 的主要区别	135
7.2 dBASE 转换程序	140
7.3 关于 dBASE III 的完备性问题	146
7.4 光标控制键	149
7.5 命令一览表	150
第三篇 关系数据库技术	
第八章 数据库模式规范化的 Smith 方法	163
8.1 数据依赖和关键字	163
8.2 规范化的类型及目的	164
8.3 依赖图 (Smith 图) 的图素	167
8.4 Smith 依赖图	169
8.5 规范数据库模式的合成	171
8.6 提高检索效率的措施	173
8.7 依赖图方法的规范化程度	174

习 题	174
第九章 一个详例——教务管理数据库的设计, 实现和维护	176
9.1 需求分析	176
9.2 数据库的逻辑设计	176
9.3 数据库的物理设计	178
9.4 数据库的实现	179
9.5 数据库的维护, 安全性控制及系统恢复	181
习 题	182
第十章 关系代数	183
10.1 开发应用程序的工具	183
10.2 三种集合运算—并差交	185
10.3 五种专门的关系运算	190
10.4 关系代数恒等式	192
10.5 复杂查询的表示	194
10.6 关系代数基本运算转换为应用程序	196
10.7 由查询到应用程序	196
习 题	198
第十一章 数据库系统的结构和特点	198
11.1 带有数据库的计算机系统	198
11.2 数据库管理系统的组成	200
11.3 数据库的特点	200
习 题	202
 第四篇 数据库模型	
第十二章 网状模型	203
12.1 从实体—联系图到网状模型	203
12.2 网状模型的基本特点	205
12.3 架次法	205
12.4 数据间的逻辑联系	206
12.5 逻辑联系的实现	207
12.6 DBTG简介	209
习 题	209
第十三章 层次模型	210
13.1 基本特点	210
13.2 从实体—联系图到层次模型	210
13.3 逻辑联系的实现	211

13.4 IMS简介	213
习 题	214
第十四章 第二代数据库模型	215
14.1 语义模型的基本特点	215
14.2 IS—A联系和派生数据	217
14.3 语义模型的优点	219
14.4 语义模型的数据结构	220
14.5 语义模型的研究动态	220
14.6 对象历史模型的背景和目的	222
14.7 对象历史的数学模型	223
14.8 对象历史的研究动态	224
附录一：IBM—PC使用常识	226
1. 硬件配置	226
2. 软件配置	227
3. 启动和停机	227
4. 初学者须知	228
5. 使用dBASE III 注意事项	230
6. 关于使用汉字	232
7. 关于 Turbo—Pascal	233
8. 关于文件操作	233
9. 行编辑程序 EDLIN	235
附录二：编写数据库管理系统程序实习指南	238
1. 引言	238
2. 实习的一般要求	238
3. DBMS的描述	239
4. 实现元组和关系的数据结构	239
5. 关于编写DBMS源程序的一些提示	242
6. 关于小组分工的建议	246
附录三，简单DBMS源程序示范	247
参考文献	274

第一篇 数据库导论

本篇介绍数据库的基本概念。第一章简要地描述了数据库，介绍了它的发展历史，同时给出了阅读本书的建议。第二章从浅识关系数据库管理系统dBASE III出发，引出了数据库逻辑设计问题，介绍了E-R图和数据库模式设计的实体-联系方法。

第一章 绪 论

1.1 什么是数据库

James Martin把数据库比喻为“各种数据用户都可去钓鱼的水库”（强调了数据的共享性）。

J·D·Ullman 写道：数据库是存储在计算机系统的数据”，（强调了计算机系统）。

C·J·Date 写道：“数据库只不过是一个计算机化了的簿记系统”，（强调了簿记，即数据的组织）。

T·J·Teorrey：“数据库是可操作的数据集合”，（强调了可操作性）。

一位读者：“数据库是组织、存储和管理数据的仓库”，（基本正确）。

以上的定义可以一言以蔽之，简单但不准确。下面的定义比较准确但不简单：

数据库是以一定的方式存储在一起的，相互有关的具有“一少三性”特点的数据集合。

“一少”是指冗余数据少，“三性”是指：

数据的共享性：库中数据能为多个用户服务；

数据的独立性：用户的应用程序，数据的逻辑组织和数据的物理存储方式无关。正如普通仓库的用户不必知道库中货物的位置，也不必知道仓库管理员是怎样组织管理库中物资的；

数据的完整性：数据在新陈代谢活动中始终保持正确性。

到第十一章第三节，我们才能把“一少三性”解释清楚，性急的读者也不妨先扫描那一节。

数据库的目的是较快较省地向较多用户提供较好的信息服务。数据库是工具。是帮助信息使用者集团去控制现实世界中信息流的手段。从哲学上讲，数据库是人类观测、通讯、记忆功能的一种扩充。

1.2 数据库发展简史

六十年代前，第一代计算机只能处理数值计算，主要为大量的科学计算服务。到了六十年代，社会大系统中出现了包括人力（Man），物力（Material）和货币（Money）的巨

大物流(又称为三M流)。物流伴随着信息流,即反映物理状态的事实的组合。用以运载信息的数字,文字和符号等称为数据,对这些数据的收集,存贮,加工和传输工作既宏大浩繁又琐碎细小,谁当此任?当然是应运而生的新一代计算机。

六十年代中期,数据银行(Data bank)一词就出现了。当时泛指一个较大的文件管理系统,与现在的数据库含义相差甚大。六十年代末,美国系统发展公司为美国海军基地研制数据库采用了Data Base一词,(Base原意为基地)。信息最大的特点是可复制性(或不守恒性)。能量和物质服从守恒定律,库中物取走一点就少一点,而数据中的信息却可以在原本上无限复制。如果说,早期的数据银行一词还多少意味着一点数据的独占权,则数据库一词则意味着数据的共享,带来了信息社会的曙光。

1969年,IBM公司宣布了它在数据库管理系统方面的成果IMS,其后把它作为产品投向市场。所以许多专家认为数据库诞生于六十年代末。

数据库从一诞生起,就横跨了理论,应用和系统三大领域。

各大计算机厂商和软件公司是早期开发数据库管理系统和推广应用的主力。到1978年,美国信息学会(ASIS)编辑出版的《数据库目录》中收集的数据库就有300多个。1982年出版的《联机数据库目录,工业通讯录》中的数据库就猛增到1350数。据估计,世界上目前有一万多个数据库系统在运行。

系统开发和应用研究的需要,促进了理论研究的发展。1970年E·F·Codd在著名的“大型共享数据库的数据的关系模型”一文中提出了关系数据库模型,并开展了理论奠基工作。经过一大批计算机科学家十多年的努力,关系数据库以简单灵活,数据独立性高,理论严格等优点表现了强大的生命力。为此E·F·Codd获得了1981年的ACM图林奖(计算机科学最高奖)。

数据库的设计方法也不断完善。1971年,美国数据系统语言协会(Conference on Data System Language 简称CODASYL)的DBTG报告提出了三级抽象模式,解决了数据独立性问题。1976年美籍华人陈平山(P·P·Chen)提出了数据库逻辑设计的实体一联系方法。1978年新奥尔良(Neworleans)的DBDWD报告把数据库系统的设计全过程分为四步,即需求分析信息分析和定义,逻辑设计,和物理设计四个阶段。

社会的需要使数据库成了大学里的热门课程。1980年J·D·Ullman写出了“数据库系统原理(第一版)”,1984年,David Maier写出了“关系数据库理论”,标志着数据库在理论上的成熟。

数据库是根据库中数据的逻辑组织方式即数据模型来分类的。目前较为普及的有网状模型,层次模型和关系模型。这三类模型的共同特点是面向记录,又称为第一代数据库模型。

七十年代后期,为了使数据库更容易为一般人理解和使用,人们提出了语义模型。到了八十年代,为了能处理历史性的数据记录,人们提出了对象历史模型。这些模型的特点是面向对象,又称为第二代数据库模型。

我国在数据库领域起步较晚。七十年代末,开始引进数据库管理系统,并研制了一些较大的数据库,如森林资源数据库,能源数据库,油田资源数据库等。随着微型计算机的普及,微机关系数据库的应用在国内发展得十分迅速。目前国内流通的微机数据库管理系统,除

dBASE II, dBASE III 外, 还有 Knowledgeman, Lotus 1—2—3 以及我国软件工作者开发的各种系统。它们大都实现了汉字支持功能, 被广泛地用于财务, 人事档案, 物资管理, 生产统计等方面的数据的收集, 处理和报表生成等各种业务中。在四个现代化的建设中, 数据库发挥着越来越大的作用。

1.3 关于阅读本书的建议

下面是一个仿 PASCAL 语言格式写成的有关阅读本书的程序的建议。

《实用数据库技术》阅读程序;

类型说明:

A类读者=(经济、财贸、管理、图书馆情报专业的学生, 研究生或有关人员);

B类读者=(计算机科学各专业的学生或其它需要比较深入了解数据库的人员);

Begin

阅读第一篇“数据库导论”;

如果你不熟悉 IBM-PC (或兼容机器) 则阅读附录一“上机实习使用 IBM-PC”;

阅读第二篇“dBASE III”, 同时上机实习有关内容;

如果你是 B 类读者, 阅读附录二“数据库管理系统程序设计实习”, 着手准备 DBMS 设计实习;

阅读第三篇“关系数据库技术”;

如果你是 B 类读者, 则阅读第四篇“数据库模型”, 且完成程序设计实习

End.

本程序的输入: 数据库初学者。

本程序的输出: 数据库初级工作者。

RUN.....

第二章 实体及其联系

我们的出发点是最朴素的数据库观——数据库是组织、存贮、管理数据的仓库, 本章将从用户的观点描述数据库大致是怎样工作的, 以及如何分析现实世界的实体及其联系, 从而把信息组织得适合数据库系统的管理。

2.1 dBASE III 一瞥

人们都说 dBASE III 易学好用。耳闻不如眼见, 眼见不如亲手干。让我们作为用户, 领略一下关系数据库管理系统 dBASE III 的风味。

数据库不能无中生有地给出数据, 正如仓库管理员不能无中生有地给出货物一样。从数据库中查询出的数据是人们预先直接或间接地输入的。存取信息和在银行存款类似, 先有存, 后有取。

字符型 (Character 简记为C)：字段值为字符串,如“姓名”字段为C型,

日期型 (Date 简记为D)：字段值为日期,如：“84—12—21”。

备注型 (Memo, 简记为M):字段值为不超过4096字节的正文。如“备注”字段为M型。

数值型 (Number, 简记为N)：字段值为数值,可以运算,如“工龄”字段为N型。

逻辑型 (Logical, 简记为L)：字段值在 (真, 假) 或 (是, 否) 中选取。如“是否会员”字段为L型。

把数据分型处理的好处有二,一是计算机易于处理,二是多了一个检查输入数据的手段。比方当你不慎把工资100元中的壹和零错打成字母I和O时, dBASE III 会指出输入 (类型) 错了, 以便能及时纠正。

现在, 着手把图2.1 (c) 的信息存入数据库。手工统计数据步骤是画表, 填入数据, 需要时查表检索数据。使用dBASE III 的步骤是:

- 一、设计关系模式, 相当于把表格框架画出来, 拟好关系名和字段名。这一步与计算机和dBASE III 无关, 称为数据库的逻辑设计。
- 二、把关系输入计算机。
- 三、载入数据和检索数据。

图2.1中已给出了关系模式。现在只需要把关系模式输入计算机。

首先 (按附录一中介绍的步骤) 在IBM-PC上启动磁盘操作系统和dBASE III。等出现园点提示符后, 键入

· CREATE 通讯录↵

dBASE III 知道用户想建立一个名称为“通讯录”的关系, 并等待着用户输入字段信息。如果因故不能用汉字操作系统和汉字dBASE III, 在命令中出现的汉字, 如上面的“通讯录”, 可以用英文或汉语拼音代替。

我们拟建立的关系有七个字段, 每个字段有字段名, 字段类型, 字段长度等三个信息。下面是从键盘上输入这些信息的过程, 其中右边括号中的文字是注释, 不要输入, 其中↵表示IBM-PC上的回车键。

现在, 从键盘上输入如下的字段信息:

姓名↵	(输入字段“姓名”)
C	(字段“姓名”的类型为字符型)
20↵	(约定“姓名”字段最多占20个字符, 姓名字段输入完毕)
性别↵	(输入字段“性别”)
C	(类型为字符型)
2↵	(字段长为2, 性别字段输入完毕)
出生年月日 (CR)	(输入字段“出生年月日”)
D	(dBASE III 规定日期型字段长度为8, 不需用户输入长度)
住址↵	

C
 150↵ (“信址”字段输入完毕)
 工龄↵ (“工龄”字段为数值型)
 N (“工龄”字段输入完毕)
 2↵ (“工龄”字段输入完毕)
 是否会员↵ (类型为逻辑型)
 L (“是否会员”字段输入完毕)
 2↵ (“是否会员”字段输入完毕)
 备注↵ (类型为备注型)
 M (“备注”字段输入完毕)
 512↵ (“备注”字段输入完毕)
 ^ End (同时按下Ctrl键和End键, 表示关系模式输入完毕)

dBASE III把上述关于关系结构的信息记录在一个名叫“通讯录.dbf”的文件中。dbf是database file (数据库文件)的缩写,称为文件后缀,表明文件的范畴,以便人和计算机识别。

“通讯录”关系建立完毕,但目前还没填入内容,相当于图2.1 (b)中的供求关系的框架,即一个空的表格。

接着dBASE III问用户是否想输入信息,正合我们的要求。因此,键入Y (Yes的第一个字母),显示器中立即出现一个空记录,标号为1,供用户填入数据。

RECORD NO1

姓名 _____

性别 _____

出生年月日 _____

信址 _____

工龄 _____

是否会员 _____

备注 _____

显示器上还会出现一个称为“菜单”的提示信息,提示此时可以用哪些命令或功能键。每个字段后的明亮区指出了该字段最大长度。当一个字段输入完毕,击↵键便可进入下一字段。

对键盘上的箭头,退格和Del键,可以顾名思义而知其功能。可用退、删、移等技术来改正输入时的手误。于是,我们输入

RECORD NO1

姓名 王大年↵

性别 男↵

出生年月日 40-3-2↵

信址 川大计算机科学系↵

工龄 20↵

是否会员 是↵

备注 ↵

dBASE III 自动移向下一记录, 即RECORD NO2. 仿上可依次输入关于赵虎, 张凡和林波的数据, 然后按ESC键, 则可回到园点提示符。如果再键入

· Quit↵

则退出dBASE III, 回到操作系统, 装入数据的工作就完成了。

dBASE III 是一个数据库管理系统 (Database Management System 简称DBMS)。如果把DBMS比喻为现成的酒瓶子, 则设计关系模式好比是把原材料酿成酒, 装入数据好比把酒装到瓶子里去。装酒的目的是必要时倒出来品尝。这就是下面要讲的数据检索。

如果我们想检索通讯录中信息, 启动dBASE III 之后, 在园点提示符后键入

· USE 通讯录↵

则dBASE III 把通讯录作为当前关系。通俗地说, 就是对通讯录作了战斗动员, 把有关的资料从磁盘调入计算机内存, 随时准备接受对通讯录的查询命令。现在键入

· DISPLAY 信址, 是否会员 FOR 姓名='王大年' ↵

则显示器上将显示

信 址	是否会员
川大计算机科学系	是

如果键入

· DISPLAY ALL↵

则显示器上显示出表2.1 (C) 上全部内容。如果键入

· DISPLY 姓名, 信址 FOR 是否会员=是↵

则显示器上显示出三个会员的信址信息。

姓 名	信 址
王 大 年	川大计算机科学系
赵 虎	成都116信箱
林 波	人大信息系

图2.2

一个数据库管理系统的功能远不止这些。dBASE III 还能把一个文件中的记录按某个字段上的内容排序，可以把几个文件中的内容联系起来检索，可以打印出漂亮的报表，可以回答如“列出那些1949年10月1日以后出生的女协会会员的名单”这一类较复杂的问题，可以把某一段上的内容求平均值，求总和，还可以把经常成套使用的命令编成应用程序，存贮起来，使用时键入一个文件名即可，……，这些将在第二篇中系统地学习。

2.2 一览表

上节中看到，使用关系数据库的第一步是设计一个合适的关系模式，或者通俗地说，造出一个表格框架。这一似简实繁工作

牵涉到许多复杂而深刻的理论。

例 2.1. 日常生活中人们爱用“一目了然”，“一览无余”来形容某种信息集的清晰，简明。在造报表时，常常造出一览表。图 2.3 是某厂设备科关于订购设备的一览表。□

设备 分类号	供货厂信息			订货信息			设备信息			注		
	厂名	厂址	联系人	电话	电报 挂号	设备 单价	数量	到货 日期	订货 单号		设备 型号	功率
1	东仪一厂	北京	王永	12345	33333	2万	2	84—1	001	DSA1	100KW	
1	前进二厂	上海	张健	67890	55555	2.1万	1	84—3	002	DSA1	100KW	
2	东仪一厂	北京	王永	12345	33333	3万	2	84—5	003	Q2	70KW	
2	前进二厂	上海	张健	67890	55555	2.8万	5	84—7	002	Q2	70KW	
3	东仪一厂	北京	王永	12345	33333	5万	1	84—9	004	SF6	50KW	

图 2.3 设备订购情况一览表

从数据库的观点看，这个关系设计得不好。

第一，冗余信息多。供货厂信息部分的五行共二十五条数据，属于冗余的占十五个，不

但浪费存贮空间和输入数据时的人工，而且增加数据装入时发生错误的机会。当记录数较大时，这种浪费是相当可观的。

第二，动态特性不好。比方某厂的某个产品的单价改变后，则要把凡是关于该厂该产品的记录都调出来修改，如只改了一些记录，疏忽了另一些记录，则同一厂同一产品会有不同的单价在数据库中，破坏了数据的完整性。

第三，更新异常 (Update anomalies)。例如取消向东仪一厂的订货，即删掉第一、三、五记录，把东仪一厂的地址，电话号码等信息都删掉了，下一次再想联系时，从库中查不到该厂的地址和电话。这是由删除引起的，称为删除异常。又例如广州的南方仪表厂来广告函推销产品，由于尚不打算订货，如果表中不允许空值，则不能把该厂的地址、电话，联系人、信息等装入表格中。这是由插入数据引起的，称为插入异常。删除异常和插入异常合称更新异常。

(a) 设备		设备分类	设备型号	功率	备注
1		DSA1	100K W		
2		Q2	70K W		
3		SF6	50K W		

(b) 订货		设备分类号	供货厂	单价	订货数量	到货日期	订货单号
1		东仪一厂	2万	2	84-1	001	
1		前进二厂	2.1万	1	84-3	002	
2		东仪一厂	3万	2	84-5	003	
2		前进二厂	2.8万	5	84-7	002	
3		东仪一厂	5万	1	84-9	004	

(c) 供货厂		厂名	厂址	联系人	电话	电报挂号
东仪一厂		北京	王永	12345	33333	
前进二厂		上海	张健	67890	55555	

图2.4