

数据库系统基础

冯玉才 编著



华中工学院出版社

TD311.13
FYC/2

数据库系统基础

冯玉才 编著

邹海明 主审



1016582

清华大学出版社

内 容 简 介

本书较全面地介绍了数据库系统的概念、基本原理和基本技术。全书共分九章：第一章数据库系统概论；第二章物理数据库组织；第三至六章分别介绍层次、网络和关系模型，并重点介绍关系数据库的设计理论；第七章数据库保护；第八章数据库设计，较详细地介绍数据库设计的概念、特点、方法和过程；第九章DBMS的设计分析，主要介绍DBMS的设计原则、方法、分层思想、模块划分、模块功能分析、DBMS与用户及操作系统的接口等。本书既注重理论联系实际，又尽量反映国内外的最新研究成果；叙述简练易懂，除八、九两章外，其余各章均配有习题。

本书可作为高等学校计算机、经济管理、信息和自动控制等专业的教材，亦可供从事计算机应用、管理科学、数据处理的广大科技人员学习与参考。

数据库系统基础

冯玉才 编著

责任编辑 唐元瑜

华中工学院出版社出版

新华书店湖北发行所发行

湖北省通城县印刷厂印刷

*

开本：850×1168 1/32 印张：14 插页：2 字数：352,000

1984年4月第一版 1986年3月第二次印刷

印数：15,001—35,000（其中精装200本）

统一书号：15255—025 定价：（平）2.60元（精）2.85元

序 言

计算机始于数值计算，却在非数值计算中得到了广泛的应用，显示了它强大的生命力。在现代计算机应用领域中，数据处理约占70~80%。数据库技术是数据处理的最新研究成果。它的出现，使得计算机应用更加广泛地渗透到工业、农业、商业、文教、卫生及军事等各个领域。

数据库系统是一门综合性的软件技术，它是编译原理、数据结构、操作系统、程序设计等许多软件知识的综合应用，其理论性和实践性都很强，是使用计算机进行各种信息管理的必备知识。通过本书的学习，读者可以基本上掌握数据库系统的基本概念、基本原理和基本技术，以便解决信息管理的实际问题。

本书广泛参考了国内外数据库方面的大量书籍、文献、资料和论文。其中主要吸取了J.D.Ullmán的《数据库系统原理》，C.J.DATE的《数据库系统导论》，James Martin的《计算机数据库组织》，姚诗斌教授在普陀大学讲授“数据库系统”的讲稿和他精选编辑的各种数据库方面的论文，Michael L.Brodie的《数据库设计》，Leo J.Cohen主编的《数据库管理系统》，B.Davis的《数据库软件选择》，A.T.F.Hutt的《关系数据库管理系统》等著作中的好思想好内容。当然书中也包含了笔者几年来在教学和科研中的一些体会和经验总结。本书选材的原则是理论和实践并重，尽量反映这一领域的最新研究成果。为使读者学以致用，特编写了第八章“数据库设计”和第九章“DBMS的设计分析”。在这两章中较详细地介绍了数据库设计的概念、特点、方法和步骤以及DBMS的设计原则、分层思想、模块划分、模块功能分析和DBMS的设计步骤，以供从事计算机系统和应用的同志学习参考。

本书大部分内容，笔者曾对计算机专业的学生、研究生及计算机应用短训班讲授过多次。书中除第八、九两章外，其他各章均配有习题，较难的题目冠以“*”号，在教学中，作为数据库系统基本理论的介绍，建议讲授第一至第七章，可安排72学时左右。第八、九两章，可由使用本书的单位根据具体情况自行安排。

中国计算机学会教育分会主席、华中工学院自控计算机系主任邹海明教授主审全书原稿，在文字和内容上都提出了许多指导性的意见。刘健教授仔细审阅了部分章节，并提出了宝贵的意见。徐则琨副教授，黄文奇副教授对编写本书给予了很多帮助。洪帆，张金隆老师分别对第五章和第一、第四章提出了一些有益的意见。华中工学院自控计算机系和计算机教研室的领导及有关老师对本书的编写给予了热情的支持。特别要指出的是，本书能这样快地同读者见面是与华中工学院出版社的大力支持和责任编辑唐元瑜同志的辛勤工作分不开的。唐元瑜同志不仅认真负责地编辑加工此书，而且对该书的内容也提出了宝贵的意见。在此，对所有为本书的编写、审阅和出版给予支持和帮助的同志一并表示衷心的感谢。

由于编者水平有限，加之时间匆促，书中错误和不妥之处，在所难免，诚望读者批评指正。

编 者

1984年1月于武汉

目 录

第一章 数据库系统概论	(1)
1.1 引言	(1)
1.1.1 数据库系统研究的对象	(1)
1.1.2 数据库系统产生的背景	(2)
1.1.3 数据库系统的产生和发展	(3)
1.1.4 数据库系统与其他课程	(4)
1.2 实体·信息·数据	(5)
1.2.1 现实世界	(6)
1.2.2 信息世界	(7)
1.2.3 计算机世界	(8)
1.3 数据模型	(10)
1.3.1 实体间的联系	(10)
1.3.2 数据模型	(12)
1.4 数据库系统结构	(19)
1.4.1 硬件资源	(19)
1.4.2 软件资源	(25)
1.4.3 数据库管理系统 (DBMS)	(27)
1.4.4 数据库管理员	(34)
习题一	(34)
第二章 物理数据组织	(36)
2.1 物理组织综述	(36)
2.1.1 与存贮介质有关的问题	(36)
2.1.2 选择物理组织策略时要考虑的因素	(48)
2.1.3 几个基本概念	(49)
2.2 文件结构综述	(52)
2.2.1 顺序结构	(52)
2.2.2 随机结构	(54)
2.2.3 链表结构	(55)

2.2.4	树状结构	(56)
2.3	HASH文件	(57)
2.3.1	溢出处理技术	(58)
2.3.2	一个实用的Hash函数	(60)
2.4	索引文件	(62)
2.4.1	稀疏索引文件	(63)
2.4.2	稠密索引文件	(67)
2.5	B-树	(67)
2.5.1	多层索引文件	(67)
2.5.2	基本B-树	(70)
2.5.3	B ⁺ 树	(75)
2.6	在非关键字上检索的文件	(78)
2.7	部分匹配检索	
	——一个特殊的Hash函数	(79)
	习题二	(89)
第三章	层次模型系统	(92)
3.1	层次模型的概念	(92)
3.2	IMS的概念	(95)
3.3	IMS的结构层次	(96)
3.4	IMS的存储结构	(99)
3.4.1	层次序列码	(99)
3.4.2	存取方法综述	(100)
3.4.3	HSAM	(101)
3.4.4	HISAM	(103)
3.4.5	HDAM	(105)
3.4.6	HIDAM	(107)
3.4.7	四种存取方法的比较	(109)
3.5	IMS的物理数据库	(111)
3.6	IMS的逻辑数据库	(123)
3.7	程序说明块PSB	(125)
3.8	IMS的数据操纵语言DL/1	(128)
3.8.1	主语言接口	(128)
3.8.2	程序通讯区	(129)

3.8.3	节受控索引条件SCA	(130)
3.8.4	DL/I 命令	(131)
3.8.5	DI/I 命令的执行概况	(133)
	习 题 三	(137)
第四章	网络模型系统	(140)
4.1	网络模型的一般概念	(140)
4.1.1	网络数据结构的类型	(140)
4.1.2	网络数据结构的分解	(145)
4.1.3	网络模型的系统概况	(146)
4.2	DBTG报告的基本概念	(147)
4.2.1	DBTG系统的数据单位	(147)
4.2.2	DBTG系统的层次和语言	(148)
4.2.3	DBTG系统的几个特殊概念	(150)
4.3	DBTG系统的数据模型	(151)
4.3.1	层次结构的系模型	(152)
4.3.2	网状结构的系模型	(153)
4.3.3	DBTG系	(156)
4.4	DBTG系统的存取策略	(160)
4.5	DBTG系统的模式描述语言	(163)
4.5.1	模式条目	(164)
4.5.2	域条目	(164)
4.5.3	记录条目	(164)
4.5.4	系条目	(166)
4.6	DBTG系统子模式描述语言	(169)
4.7	应用程序的运行环境	(173)
4.8	DBTG系统的数据操纵语言DML	(176)
4.8.1	DML和主语言的接口	(176)
4.8.2	DML命令简介	(178)
4.8.3	检索命令	(181)
4.8.4	更新命令	(184)
	习 题 四	(186)
第五章	关系模型系统	(188)
5.1	关系模型概述	(188)

5.2	关系模型的基本概念	(191)
5.3	关系模型及其描述	(191)
5.3.1	关系模型	(195)
5.3.2	关系模式	(196)
5.3.3	关系子模式	(198)
5.4	关系模型的数据操纵语言	(199)
5.5	关系代数	(201)
5.6	关系演算	(205)
5.6.1	元组关系演算	(206)
5.6.2	域关系演算	(210)
5.7	关系运算的安全限制及其相互转换	(212)
5.8	SEQUEL	(216)
5.9	Query By Example	(223)
5.10	查询优化概述	(227)
5.11	优化的一般策略	(229)
5.12	关系代数等价变换规则	(230)
5.13	关系代数表达式的优化算法	(232)
	习题五	(238)
第六章	关系数据库设计理论	(242)
6.1	问题的提出	(243)
6.2	函数依赖	(245)
6.2.1	属性间的关系	(245)
6.2.2	函数依赖	(246)
6.2.3	函数依赖与属性关系	(248)
6.2.4	函数依赖的逻辑蕴涵	(248)
6.2.5	关键字	(249)
6.3	函数依赖公理	(250)
6.3.1	Armstrong公理	(250)
6.3.2	公理的正确性	(250)
6.3.3	公理的推论	(251)
6.3.4	公理的完备性	(252)
6.3.5	闭包的计算	(255)
6.3.6	函数依赖集的等价和覆盖	(260)

6.3.7	函数依赖集的最小集	(261)
6.4	关系模式的分解	(263)
6.4.1	联接不失真性	(263)
6.4.2	依赖保持性	(271)
6.5	关系模式的规范化	(272)
6.5.1	第一范式(1NF)	(272)
6.5.2	第二范式(2NF)	(273)
6.5.3	第三范式(3NF)	(274)
6.5.4	Boyce-Codd范式(BCNF)	(275)
6.5.5	BCNF的分解	(275)
6.5.6	3NF分解	(279)
6.6	多值依赖	(284)
6.6.1	多值依赖的定义	(284)
6.6.2	多值依赖公理	(285)
6.6.3	第四范式(4NF)	(287)
6.7	关系模式规范综述	(288)
	习题六	(289)
第七章	数据库保护	(291)
7.1	数据库保护综述	(291)
7.2	安全性	(293)
7.3	完整性	(296)
7.4	并发控制	(298)
7.4.1	并发控制的基本概念	(298)
7.4.2	可串行化的判断	(300)
7.4.3	两段式调度协议	(304)
7.4.4	读锁与写锁	(305)
7.4.5	树式协议	(307)
7.5	数据库恢复	(311)
	习题七	(313)
第八章	数据库设计	(315)
8.1	数据库设计综述	(315)
8.2	数据库设计的特点及有关概念	(323)
8.3	数据字典	(325)

8.4	数据库设计方法	(322)
8.4.1	基本的设计方法	(323)
8.4.2	关系模型的设计方法	(331)
8.4.3	New Orleans设计方法概要	(333)
8.4.4	一种实用的逻辑设计和物理设计方法	(338)
8.4.5	逻辑记录存取方法LRAM	(340)
8.5	数据库设计步骤	(343)
8.5.1	需求分析	(345)
8.5.2	逻辑设计	(349)
8.5.3	物理设计	(350)
第九章	DBMS的设计分析	(359)
9.1	研制DBMS要考虑的因素	(359)
9.1.1	基本理论和基本技术的准备	(360)
9.1.2	人员配备	(360)
9.1.3	语言工具	(361)
9.1.4	调查应用的目的和要求	(361)
9.1.5	资源条件	(362)
9.1.6	选择数据模型	(363)
9.1.7	用户接口	(364)
9.1.8	与OS的接口	(364)
9.2	研制DBMS的步骤	(368)
9.2.1	确定DBMS的应用目标和范围	(369)
9.2.2	设计DBMS的总体方案	(369)
9.2.3	设计逻辑模型——数据模型部及用户接口	(371)
9.2.4	设计物理模型——数据存取部及存取接口	(371)
9.2.5	研讨并修改系统结构	(371)
9.2.6	设计模块	(372)
9.2.7	研讨整个设计	(372)
9.2.8	编程调试	(372)
9.2.9	模拟、测试及调整	(373)
9.3	系统分层的方法和原则	(373)
9.3.1	什么是层次结构	(373)
9.3.2	层次结构的种类	(374)
9.3.3	分层设计方法	(375)

0.3.4 分层的赋权.....	(376)
9.4 DBMS设计中的分层方法.....	(376)
9.4.1 块式接口分层方法.....	(377)
9.4.2 记录接口分层方法.....	(378)
9.4.3 三级分层方法.....	(381)
9.5 一个DBMS的设计练习.....	(382)
9.6 存储块的格式设计.....	(385)
9.7 元组标识符TID.....	(386)
9.8 DBMS主要功能模块分析.....	(387)
9.8.1 模式编译模块.....	(387)
9.8.2 数据操纵处理模块.....	(389)
9.8.3 空间管理模块.....	(391)
9.8.4 缓冲区管理模块.....	(392)
9.8.5 索引装置模块.....	(393)
9.8.6 存取控制模块.....	(395)
9.8.7 关系DBMS存取路径的选择.....	(397)
9.9 两个典型的关系DBMS.....	(403)
9.9.1 SYSTEM R.....	(403)
9.9.2 INGRES.....	(417)
附录 I	(425)
附录 II	(427)
参考文献	(430)

第一章 数据库系统概论

1.1 引言

数据库技术产生于六十年代末七十年代初，它的出现使得计算机应用渗透到工农业生产、商业、行政管理、科学研究、工程技术和国防军事等每一个部门。在计算机领域中，七十年代是数据库的年代。

数据库是计算机领域中最重要技术之一，它是软件方面的一个独立分支。目前它正处在发展中。

在计算机应用中，数据处理占的比重最大，数据库系统是数据处理的核心机构。它的功能往往决定整个计算机应用的经济效益。要使计算机为我国四个现代化服务，必须把它的应用重点转移到管理方面。为此，我们有必要学习和掌握数据库系统的原理和技术，并用它解决信息管理中的实际问题。

1.1.1 数据库系统研究的对象

数据在人类社会的发展中是一种极为重要的资源。人们的一切社会活动都离不开数据：工厂的自动化系统需要生产线上的实时数据，如电压、电流、温度、压力等；科学工作者需要各种实验数据；军事指挥员需要各种情报，如敌我双方的人员、武器装备等。银行、商店、仓库、图书馆、档案室、公安等部门中的工作人员无时无刻不在和大量的数据打交道。

如何妥善地保存和科学地管理这些数据是人们长期以来十分关注的课题。数据库系统就是人们研究这个课题的最新成果。

数据库系统研究的对象就是如何高效地巧妙地进行数据处理，而又花费最少。

1.1.2 数据库系统产生的背景

与其他技术的发展一样，数据处理也是经历了由低级到高级的发展过程。它的发展大致可分为三个阶段：

- (1) 无管理阶段（五十年代中期以前）；
- (2) 文件系统阶段（五十年代后期至六十年代后期）；
- (3) 数据库系统阶段（七十年代初至现在）。

数据库系统的前身是文件系统，文件系统是数据库系统发展的基础。

文件系统的发展又分为二个阶段：

(1) 六十年代初期，出现了初等的文件系统，其特点是：

- 顺序文件组织；
- 批量处理；
- 物理数据结构等于逻辑数据结构；
- 数据结构的改变导致应用程序的改变；
- 软件处理仅仅涉及I/O；
- 对单个的应用较优；
- 数据高度冗余；
- 数据无集中管理。

(2) 六十年代后期，出现了成熟的文件系统，其特点是：

- 顺序存取和随机存取并用；
- 批量处理和实时处理并用；
- 物理结构和逻辑结构间有了简单变换；
- 软件提供存取方法；
- 对单个应用较优；
- 数据冗余较多；
- 数据无集中管理。

由此我们可以看出，虽然第二阶段的文件系统有了很多改进，但一些根本性的问题仍然没有解决，这主要表现在以下三个方面：

(1) 冗余度大

文件系统下的用户各自建立自己的文件,数据不能共享,造成大量重复,不仅浪费空间,而且增加更新开支,更严重的是容易产生数据不一致性。

(2) 缺乏数据独立性

数据和程序互相依赖,一旦数据结构改变,则与这些数据有关的程序都必须重新编写和调试,这无形中增加了开支。有人统计过这项开支约占总开支的25%。

(3) 数据无集中管理

各个文件没有统一的管理机构,其安全性、完整性等无法得到保证。

所有这些问题,文件系统本身已无法解决,因而造成效率低、成本高,并有很多潜在问题,严重地阻碍了数据处理技术的发展。这就是数据库技术产生的原动力,也就是数据库系统产生的背景。

1.1.3 数据库系统的产生和发展

数据库系统的目标首先是克服文件系统的弊病,解决冗余和数据独立性问题,并且用一个软件系统来集中管理所有的文件,从而实现了数据共享。对于完整性、安全性等问题也都得到了相应的解决。

文件系统下的用户需要对他的文件的物理存贮细节有相当的了解,这增加了用户使用数据的困难。而数据库系统下的用户不必了解文件存贮的细节,他可以逻辑地、抽象地使用数据。一切繁琐的存贮细节都由一个软件系统来完成,这个软件系统就是数据库系统的核心——数据库管理系统DBMS(Database Management Systems)。在DBMS的集中管理下,具有较高的数据独立性、较少的冗余并且相互间有联系的文件集合叫数据库。

数据库的出现以六十年代后期的三个事件为标志:

(1) 一九六八年美国IBM公司研制了世界上第一个数据

库管理系统IMS(Information Management System);

(2) 一九六九年美国CODASYL(Conference on Data Systems Languages)组织的DBTG(Data Base Task Group)小组发表了DBTG报告;

(3) 一九七〇年美国IBM公司的研究员E.F.Codd开始发表一系列的关系数据库论文。

此后,美国有很多大学开始了数据库技术的研究;美国的很多公司也先后研制了各种各样大大小小的数据库管理系统。越来越多的部门开始使用数据库技术来管理他们的业务。由于数据库技术的出现,数据处理的成本大为下降,效率越来越高,可靠性不断增加,从而推动了计算机应用的普及。

日本、西德等国家也先后引进并发展了数据库技术,他们都研制了自己的DBMS。

到七十年代后期,数据库系统的技术已趋于成熟,其主要的研究课题已转向:

- 关系数据库;
- 分布式数据库;
- 数据库机器。

1.1.4 数据库系统与其他课程

数据库系统作为软件的一个分支,与其他基础软件和系统软件有密切的关系。它几乎涉及到软件的所有知识,是很多重要软件技术的综合应用。

首先数据库系统是在操作系统OS(Operating System)支持下工作的。它和OS关系十分密切,如同两个齿轮,边界并不清楚,有些工作可以由OS做,也可以由DBMS做,还可以由双方各做一部分,但合起来应是一个完整的整体。所以设计DBMS时应充分熟悉支持它的OS。另一方面OS中用到的许多技术同样可以用到DBMS中,例如缓冲区的管理、并发控制等技术,两个系统中的

处理思想是完全一样的。所以不熟悉OS,要想搞清数据库系统是很困难的。

其次是数据结构,它是软件的一门专业基础课。编译系统和操作系统中都要用到它,数据库系统中用得更多。但在使用上有所不同,一些在其他方面仅有理论价值的数据结构,在大规模的数据库中成为实际可行的组织形式。

再次是编译技术,它在数据库系统中也用得很多。数据库系统中有许多语言,例如数据定义语言,数据操纵语言,查询语言等,这些语言的编译都是数据库系统的任务。

程序设计,它是具体实现数据库系统的最基本的技术,因为数据库系统中有大量的应用程序都是用高级语言加上数据操纵语言来编制的。没有熟练的编程技巧,这些任务也很难完成。

另外离散数学、数理逻辑是关系数据库的理论基础。它们的很多概念、思想甚至名词术语都直接用到关系数据库中。

还有算法分析在数据库中也是经常用到的。

最后软件工程在设计DBMS时,是不可缺少的知识和技术。

所以数据库系统是一门综合性的软件技术,是一门很有意义、很有趣味的学问。要研究和掌握它,必须了解软件的各个方面,以便更加理解和认识到这些知识的内在联系,并在一种观念上将它们统一起来。

1.2 实体·信息·数据

数据是表示信息的,而信息又是反映事物的物理状态的。例如一个国家的钢产量是3000万吨,这是一个数据,它表示了这个国家生产钢的能力这一信息,而这一信息又反映了这个国家的实力这一状态。这种表示信息的数据是自然的,另一种情况是表示信息的数据是人为的,例如一匹马的颜色是枣红色的,为了让计算机能方便地处理它,可用数字来代替它。比如1 = 枣红色,2 = 银白色,