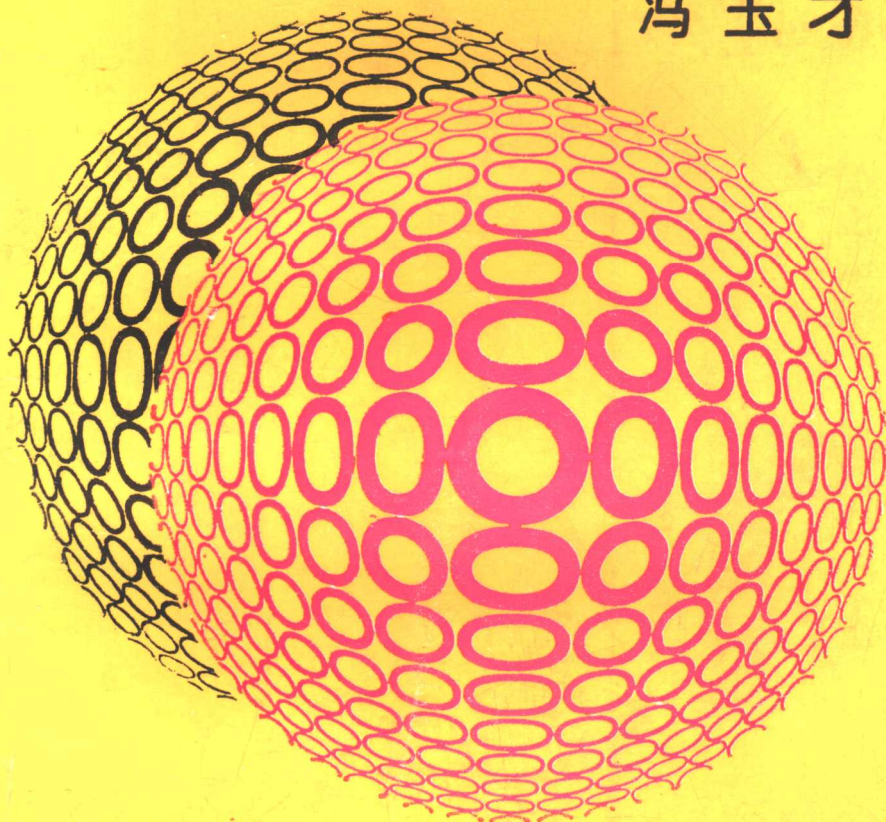


数据库系统 基础

(第二版)

冯玉才



华中理工大学出版社

数据库系统基础

第 二 版

(本书第一版荣获国家教委
高等学校优秀教材一等奖)

冯玉才 编著

华中理工大学出版社

数据库系统基础

冯玉才 编著

郑兆昭

责任编辑

唐元瑜

*

华中理工大学出版社出版发行

(武昌 喻家山 邮编 430074)

新华书店湖北发行所经销

华中理工大学出版社沔阳印刷厂印刷

*

开本:850×1168 1/32 印张:15.75 字数:385 000

1984年4月第1版 1993年9月第2版 1995年2月第7次印刷

印数:52 001—57 000

ISBN7-5609-0873-X/TP·99

定价:9.50元

(鄂)新登字第10号

第二版序言

本书第一版从1984年问世以来,得到了兄弟院校同行和广大读者的鼓励和厚爱。十年来已印刷6次,发行5万余册。并在1987年获国家教委高等学校优秀教材一等奖。

十年来,计算机技术尤其是数据库技术发展很快,取得了令人瞩目的新的成就,传统的数据库技术的理论和系统已日趋成熟。而复杂对象的应用需求将数据库技术推向新的发展领域,数据库技术进入了高层次多学科交叉研究的发展阶段。数据库与分布式技术,数据库与并行处理,数据库与人工智能,数据库与面向对象技术,数据库与多媒体技术,数据库与图象识别和图象处理,数据库与语言识别等都非常自然地需要互相补充和结合。于是出现了分布式数据库,并行数据库,智能数据库,专家数据库,演绎数据库,面向对象数据库,多媒体数据库,图象数据库,语音数据库等。本书第一版已不能全面反映当今数据库技术发展的现实情况。这是促使作者对第一版内容作全面修订更新的主要原因。

此外这十年来,作者除了教学工作,还一直在进行数据库的研究,并取得了不少在理论上或实用上有价值的成果。如候选关键字的求解理论与算法, B^+ 树技术,聚集技术等。作者感到将这些内容系统地、集中地加以总结和介绍,也是十分必要的,因此,它们也就成为修订时增加内容的一个重要部分。而这些部分正是本书较有特色的部分。

修订是在第一版的基础上,力求充分反映当今数据库技术的最新成就和研究动态,其中也展示了作者的研究成果。还吸收了广大读者的宝贵意见。对全书的章节层次和内容都作了相当大的修改和更新。具体有:

1. 删除了曾经作为层次模型代表的 IMS 系统的介绍,这一方面因为 IMS 的概念术语很不标准,作为教材内容不太适合;另一方面 IMS 在中国很少应用;而且许多学校在教学中也不讲授这一章。因此修订时只将层次模型的通用概念和技术保留下来安排在第一章中。

2. 删去了第一版第七章数据库保护,而将完整性、一致性、并发控制作为数据库特性安排在第一章中。

3. 第一章的讲法和内容全部重写,宗旨是不以说教方式提出概念,而以解决实际问题的过程引出数据库系统的基本概念。

4. 第二章中加进了作者研制数据库管理系统所提出的新的索引技术——Bⁿ 树。

5. 第五章加进了作者的理论研究成果——关系模式候选关键字的求解理论及算法研究。

6. 第六章数据库设计全部重写,以三个管理信息系统中数据库设计的经验教训为背景讨论数据库设计的方法和技术。

7. 第七章 DBMS 的设计分析全部进行重写,以作者参与实现的几个 DBMS 的工作为基础,介绍其关键技术和体系结构。

8. 第八章数据库最新研究动态,着重介绍当前最活跃的数据库研究领域,反映当今数据库的世界研究前沿课题。

前六章内容可供有关专业的本科教学用,后两章可作为研究生专题讲座内容。

本书第二版的写作得到了许多读者和同仁的关注和建议。特此表示感谢,书中不妥和错误之处,恳请读者批评指正!

冯玉才

1993. 7

第一版序言

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. Ullman 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。刘键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的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 数据库系统引论	(1)
1.1 绪论	(1)
1.1.1 实体、信息、数据	(1)
1.1.2 数据库发展简史及趋势	(5)
1.1.3 研究数据库技术的方法	(13)
1.2 什么是数据库	(14)
1.2.1 一个管理信息系统的实例	(14)
1.2.2 数据库技术的本质是变换	(18)
1.3 数据库系统的功能特性	(23)
1.3.1 数据独立性	(23)
1.3.2 数据安全性	(24)
1.3.3 数据完整性	(25)
1.3.4 数据一致性	(26)
1.3.5 数据共享	(26)
1.3.6 控制冗余	(27)
1.3.7 集中管理	(27)
1.3.8 并发控制	(28)
1.3.9 故障恢复	(29)
1.3.10 数据字典	(31)
1.4 数据模型	(33)
1.4.1 实体间的联系	(33)
1.4.2 实体-联系模型	(35)
1.4.3 层次模型	(37)
1.4.4 网状模型	(40)
1.4.5 关系模型	(40)

1.5 数据库管理系统	(43)
1.5.1 数据描述语言	(44)
1.5.2 数据操纵语言	(45)
1.5.3 数据库实用程序	(45)
1.6 数据库系统结构	(46)
1.6.1 硬件资源	(47)
1.6.2 软件资源	(47)
1.6.3 数据库管理员	(49)
1.6.4 数据库系统的工作过程	(51)
小结	(52)
习题一	(53)
第二章 物理数据组织	(54)
2.1 物理组织综述	(54)
2.1.1 与存储介质有关的问题	(54)
2.1.2 选择物理组织要考虑的因素	(64)
2.1.3 几个基本概念	(67)
2.2 文件结构综述	(69)
2.2.1 顺序结构	(69)
2.2.2 随机结构	(71)
2.2.3 链表结构	(72)
2.2.4 树状结构	(73)
2.3 HASH 文件	(74)
2.3.1 溢出处理技术	(74)
2.3.2 一个实用的 Hash 函数	(76)
2.4 索引文件	(79)
2.4.1 稀疏索引文件	(80)
2.4.2 稠密索引文件	(83)
2.5 B 树	(84)
2.5.1 多层索引文件	(84)
2.5.2 基本 B 树	(87)
2.5.3 B ⁺ 树	(91)

2.5.4 B ⁺ 树	(95)
2.6 在非关键字上检索的文件	(100)
2.7 部分匹配检索——一个特殊的 Hash 函数	(102)
2.8 R 树	(111)
2.8.1 基本 R 树	(112)
2.8.2 R ⁺ 树	(116)
2.9 Quad 树	(117)
2.9.1 Quad 树的概念	(117)
2.9.2 Quad 树的操作算法	(119)
2.10 Frame 结构	(120)
2.11 几种树结构的比较	(121)
习题二	(123)
第三章 网状模型系统	(125)
3.1 网状模型的一般概念	(125)
3.1.1 网状数据结构的类型	(125)
3.1.2 网状数据结构的分解	(130)
3.1.3 网状模型的系统概况	(131)
3.2 DBTG 报告的基本概念	(132)
3.2.1 DBTG 系统的数据单位	(132)
3.2.2 DBTG 系统的层次和语言	(133)
3.2.3 DBTG 系统的几个特殊概念	(135)
3.3 DBTG 系统的数据模型	(136)
3.3.1 层次结构的系模型	(136)
3.3.2 网状结构的系模型	(138)
3.3.3 DBTG 系	(141)
3.4 DBTG 系统的存取策略	(144)
3.5 DBTG 系统的模式描述语言	(148)
3.5.1 模式条目	(149)
3.5.2 域条目	(149)
3.5.3 记录条目	(149)
3.5.4 系条目	(151)

3.6	DBTG 系统子模式描述语言	(154)
3.7	应用程序的运行环境	(157)
3.8	DBTG 系统的数据操纵语言 DML	(161)
3.8.1	DML 和主语言的接口	(161)
3.8.2	DML 命令简介	(161)
3.8.3	检索命令	(164)
3.8.4	更新命令	(168)
习题三		(171)

第四章 关系模型系统..... (173)

4.1	关系模型概述	(173)
4.2	关系模型的基本概念	(176)
4.3	关系模型及其描述	(179)
4.3.1	关系模型	(179)
4.3.2	关系模式	(180)
4.3.3	关系子模式	(182)
4.4	关系模型的数据操纵语言	(183)
4.5	关系代数	(185)
4.6	关系演算	(190)
4.6.1	元组关系演算	(190)
4.6.2	域关系演算	(194)
4.7	关系运算的安全限制及其相互转换	(196)
4.8	关系数据库标准语言——SQL	(199)
4.8.1	SQL 标准的基本概念和特点	(200)
4.8.2	SQL 数据定义功能	(204)
4.8.3	SQL 数据操纵功能	(206)
4.8.4	SQL 数据的安全与授权	(229)
4.8.5	交互式 SQL	(231)
4.8.6	嵌入式 SQL	(232)
4.9	Query By Example	(233)
4.10	查询优化	(237)

4.11	优化的一般策略	(239)
4.12	关系代数等价变换规则	(240)
4.13	关系代数表达式的优化算法	(242)
	习题四	(247)
第五章	关系数据库设计理论	(251)
5.1	问题的提出	(252)
5.2	函数依赖	(254)
5.2.1	属性间的关系	(254)
5.2.2	函数依赖	(255)
5.2.3	函数依赖与属性关系	(257)
5.2.4	函数依赖的逻辑蕴涵	(257)
5.2.5	关键字	(258)
5.3	函数依赖公理	(259)
5.3.1	Armstrong 公理	(259)
5.3.2	公理的正确性	(259)
5.3.3	公理的推论	(260)
5.3.4	公理的完备性	(261)
5.3.5	闭包的计算	(264)
5.3.6	函数依赖集的等价和覆盖	(268)
5.3.7	函数依赖集的最小集	(269)
5.4	关系模式的分解	(271)
5.4.1	无损联接性	(272)
5.4.2	依赖保持	(279)
5.5	关系模式的规范化	(279)
5.5.1	第一范式(1NF)	(280)
5.5.2	第二范式(2NF)	(281)
5.5.3	第三范式(3NF)	(282)
5.5.4	Boyce-Codd 范式(BCNF)	(282)
5.5.5	BCNF 的分解	(283)
5.5.6	3NF 的分解	(287)

5.6	候选关键字的求解理论和算法	(291)
5.6.1	快速求解候选关键字的一个充分条件	(292)
5.6.2	左边为单属性的函数依赖集的候选关键字成员的 图论判定方法	(294)
5.6.3	算法及其复杂度	(301)
5.6.4	候选关键字求解小结	(303)
5.7	多值依赖	(303)
5.7.1	多值依赖的定义	(303)
5.7.2	多值依赖公理	(305)
5.7.3	第四范式(4NF)	(307)
5.8	关系模式规范综述	(307)
	习题五	(309)
第六章	数据库设计	(311)
6.1	数据库设计综述	(311)
6.1.1	数据库设计的基本概念	(312)
6.1.2	数据库设计与管理信息系统	(313)
6.1.3	数据库设计的特点	(313)
6.1.4	数据库设计方法	(316)
6.1.5	数据库设计步骤	(319)
6.2	需求分析	(321)
6.2.1	需求调查	(321)
6.2.2	需求分析	(324)
6.3	概念设计	(328)
6.3.1	局部视图设计	(329)
6.3.2	全局视图设计	(331)
6.4	逻辑设计	(333)
6.4.1	E-R 图向实际的数据模型的转换	(333)
6.4.2	关系理论的应用	(334)
6.5	物理设计	(335)
6.5.1	选择文件组织策略	(336)
6.5.2	选择存取路径	(337)

6.5.3	选择索引	(339)
6.5.4	记录聚集	(341)
6.6	应用设计	(342)
6.6.1	功能需求分析	(343)
6.6.2	功能设计	(344)
6.6.3	事务设计	(344)
6.7	数据库建立	(355)
6.7.1	数据库的载入	(356)
6.7.2	数据库测试	(357)
6.8	数据库的运行和维护	(358)
6.9	数据库设计的经验教训	(359)
6.9.1	实例之一——物资管理数据库系统的开发	(359)
6.9.2	实例之二——某电力管理信息系统的开发	(360)
6.9.3	实例之三——养路费征收微机管理系统	(361)
6.9.4	经验和教训	(362)
第七章	DBMS 的设计与实现	(365)
7.1	研制 DBMS 的准备工作	(366)
7.1.1	理论和技术准备	(366)
7.1.2	人员配备	(366)
7.1.3	语言工具	(367)
7.1.4	调查应用的目的和需求	(367)
7.1.5	资源条件	(368)
7.1.6	选择数据模型	(368)
7.1.7	用户界面	(369)
7.1.8	与 OS 的接口	(369)
7.2	DBMS 的总体结构	(373)
7.2.1	DBMS 的总体模型	(374)
7.2.2	DBMS 的分层设计法	(377)
7.3	DBMS 的用户接口设计	(383)
7.4	数据定义语言的设计与实现	(384)
7.4.1	模式描述语言	(385)

7.4.2	子模式描述语言	(386)
7.5	数据操纵语言的设计与实现	(387)
7.5.1	DML 的编译方法	(388)
7.5.2	全解释与半解释	(389)
7.5.3	预编译技术	(390)
7.5.4	描述型语言语句的翻译	(391)
7.6	存取系统的设计	(394)
7.6.1	存取接口的设计	(394)
7.6.2	逻辑存取路径	(395)
7.6.3	存取机制	(398)
7.7	存贮结构的设计	(401)
7.7.1	记录的存贮	(401)
7.7.2	记录编址	(402)
7.7.3	元组标识符 TID	(402)
7.8	存贮系统的设计	(403)
7.8.1	物理段	(404)
7.8.2	物理块的格式设计	(405)
7.8.3	系统缓冲区	(406)
7.9	并发控制	(408)
7.9.1	三个并发问题	(408)
7.9.2	事务的封锁协议	(410)
7.9.3	封锁的单元与封锁开销	(112)
7.9.4	封锁的种类	(412)
7.9.5	封锁的实现	(419)
7.9.6	封锁作为资源	(421)
7.9.7	死锁及其解决途径	(423)
7.9.8	事务的一致性分析	(427)
第八章	数据库技术的最新研究动态	(432)
8.1	分布式数据库	(433)
8.1.1	分布式数据库的特色	(434)
8.1.2	为什么要有分布式数据库	(436)

8.1.3	分布式数据库的体系结构	(437)
8.1.4	分布式数据库的透明层次	(438)
8.1.5	分布式数据库设计	(441)
8.1.6	分布式数据库查询	(442)
8.1.7	分布式数据库的事务管理	(442)
8.1.8	分布式数据库的并发控制	(443)
8.1.9	分布式数据库的目录管理	(443)
8.2	演绎数据库	(444)
8.2.1	演绎数据库的基本概念	(445)
8.2.2	演绎数据库的基本结构	(445)
8.2.3	演绎数据库的研究现状	(446)
8.2.4	已经实现的演绎数据库系统	(449)
8.2.5	演绎数据库、知识库与智能数据库	(450)
8.3	工程数据库	(451)
8.3.1	工程数据库的应用环境	(451)
8.3.2	工程数据库的特点要求	(452)
8.4	多介质数据库	(455)
8.4.1	多介质数据库管理系统的功能要求	(456)
8.4.2	多介质 DBMS 的体系结构	(457)
8.5	专家数据库	(459)
8.5.1	人工智能与数据库技术的发展	(460)
8.5.2	专家数据库的研究目标和动态	(463)
8.5.3	专家数据库知识数据模型	(466)
8.5.4	逻辑与数据库	(469)
8.5.5	逻辑程序 PROLOG 和关系数据库	(471)
8.5.6	专家数据库 EDS 的系统结构	(473)
8.5.7	几个典型系统	(475)
	参考文献	(478)

第一章 数据库系统引论

1.1 绪 论

数据库技术产生于本世纪 60 年代末 70 年代初。它的出现使得计算机应用进入了新的时期,社会的每一领域都与计算机发生了联系。数据库技术聚集了数据处理最精华的思想,是管理信息最先进的工具。信息社会的紧迫需求使数据库技术成为计算机园地中一支最有生命力的新秀。而与人工智能技术的结合又使它获得了新的血液。80 年代中期数据库技术进入了一个新的层次,智能数据库、演绎数据库、专家数据库、面向对象的数据库、工程数据库、多介质数据库、并行数据库、实时数据库等就是当代数据库技术的研究前沿。

1.1.1 实体、信息、数据

人类生存的世界是一个物质的世界,同时也是一个信息的世界。信息伴随物质而存在,并随物质的变化而变化。物质是信息的基础,信息是物质的反映。表示信息最准确的工具是数据。例如一个国家的存在,同时就有这个国家的人口、军队、工农业产值等信息。如果用一些不确定的形容词来描述一个国家的状态,例如人口很多,军队强大,工农业产值很高,这只能给人一个模糊印象。但如果用数字来描述上述信息:人口 5 亿,军队 500 万,人均产值 2000 美元,则能给人十分清晰的印象。用数字表示信息有时是自然的,有时是人为的。上述描述国家状态的数字是自然的,而在计算机中为了处理的方便,经常用数字来代替一些信息,如令 1 表示红,2 表示绿,……。这就是人为的。但经这样变换后,世界上几乎所有