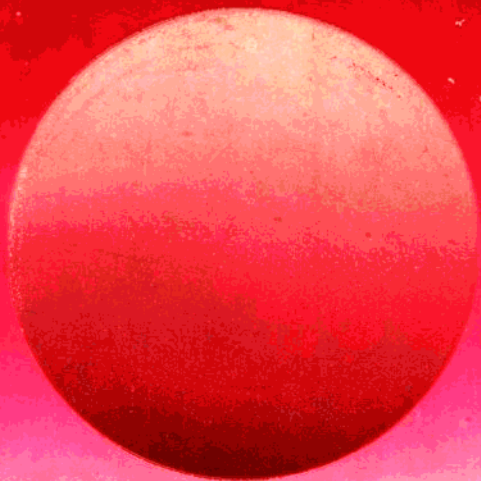


● 研 究 生 用 书 ●

DATABASE DESIGN
AND ANALYSIS

华中理工大学出版社



刘 云 生

数据库设计与分析

内 容 简 介

本书是作者在多年数据库教学与科研的基础上编写的。它包含三部分,第一至七和十二章叙述数据库设计方法学,包括数据库需求分析与规范说明,概念数据库结构设计及其语义数据分析与建模方法,逻辑数据库结构(模式)设计,物理数据库结构设计和分布式数据库设计(第十二章);第八至九章分析数据库实现的各种功能、结构及其特征;第十至十一章讨论数据库系统的运行、维护与管理。

本书可作为高等学校计算机及有关专业的研究生教材,亦可供本科高年级学生和数据库分析员、开发与设计者、数据库管理员及有关应用人员学习与参考。

ABSTRACT

This book is written based on a many-year teaching and science research in database area. The book includes three sections, chapters 1 to 7 and 12 describe database design methodologies, covering database requirements analysis and specification, conceptual database structures design with its semantic data analysis and modelings, logical database structures (schemas) design, physical database structures design, and distributed database design (chapter 12). Chapters 8 to 9 analyze functions, structures and their characteristics of a database implementation. Chapters 10 to 11 discuss issues of database system operation, maintenance and administration.

The purpose of this book is to provide a first-year graduate course, It could be used for senior, database analyst, designer, database administrator, and other related personnel.

“研究生用书”总序

研究生教材建设是提高研究生教学质量的重要环节,是具有战略性的基本建设。各门课程必须有高质量的教材,才能使学生通过学习掌握各门学科的坚实的基础理论和系统的专门知识,为从事科学研究工作或独立担负专门技术工作打下良好的基础。

我校各专业自 1978 年招收研究生以来,组织了一批学术水平较高,教学经验丰富的教师,先后编写了公共课、学位课所需的多种教材和教学用书。有的教材和教学用书已正式出版发行,更多则采用讲义的形式逐年印发。这些讲义经过任课教师多年教学实践,不断修改、补充、完善,已达到出书的要求。因此,我校决定出版“研究生用书”,以满足本校各专业研究生教学需要,并与校外单位交流,征求有关专家学者和读者的意见,以促进我校研究生教材建设工作,提高教学质量。

“研究生用书”以公共课和若干门学位课教材为主,还有教学参考书和学术专著,涉及的面较广,数量较多,准备在今后数年内分批出版。编写“研究生用书”总的要求是从研究生的教学需要出发,根据各门课程在教学过程中的地位和作用,在内容上求新、求深、求精,每本教材均应包括本门课程的基本内容,使学生能掌握必需的基础理论和专门知识;学位课教材还应接触该学科的发展前沿,反映国内外的最新研究成果,以适应目前科学技术知识更新很快的形势;学术专著则应充分反映作者的科

研硕果和学术水平,阐述自己的学术见解。在结构和阐述方法上,应条理清楚,论证严谨,文字简炼,符合人们的认识规律。总之,要力求使“研究生用书”具有科学性、系统性和先进性。

我们的主观愿望虽然希望“研究生用书”的质量尽可能高一些,但由于研究生的培养工作为时尚短,水平和经验都不够,其中缺点、错误在所难免,尚望校内外专家学者及读者不吝指教,我们将非常感谢。

华中理工大学研究生院院长

黄树槐

1989. 11.

前 言

数据库系统越来越普遍地受到人们的关注与重视，其应用领域已不仅仅是传统的商务数据处理，近年来广泛地应用于各种CAX (CAD/CAM, CAT, CAE, CASE 等等)、地理信息系统、办公自动化等。数据库与人工智能、程序设计语言的结合在数据库的理论、系统与应用研究领域形成了极其热闹的局面，取得了丰硕的成果。

本书旨在为数据库的基础理论、系统开发和应用实现研究提供一个统一的框架，并以理论为基础，以系统为手段，以应用为目的，由底向上形成一个层次体系结构。同时，也强调了理论与实践相结合，并以应用为轴心，将基础理论、基本原理，数据库管理系统功能、结构与特性，数据库设计与应用的实现方法、技术与工具、各种特殊考虑的问题等各个方面有机地集成为一个整体。力求使读者通过本书的学习获得有关数据库系统的完整知识，且能设计与实现数据库系统以解决信息管理的实际问题。

本书是笔者在多年从事研究生的数据库教学和科研的基础之上，广泛地参阅了大量的国内外有关书籍、论文、文献和资料编写而成的。书中的内容已对计算机及有关专业的研究生、留学研究生、助教进修班等连续讲授过多次，原稿也曾由他们多次阅读并提出了许多宝贵意见，借此向他们致以谢意。

在内容的编排上，按数据库设计方法学（第一至七章）、数据库系统实现方法与技术（第八至九章）、数据库运行与管理（第十至十一章）的顺序进行组织，最后概要地讨论了有关分布式数据库设计的一些特殊问题（第十二章）。在编写时，始终注重“分析”二字，尤其在数据库系统实现方法与技术部分是以分析、评价与应用的观点来讨论的。读者可以根据自己的需要分别选择其

中的有关部分，不一定要通读全书。

本书是作为计算机科学与工程及有关专业研究生的教材而编写的，也可供数据库系统分析、开发与研究人员，数据库设计者，数据库管理员，数据库应用程序员及其他有关应用的工程技术人员阅读参考。阅读本书的读者被认为已熟悉数据库系统原理、数据结构、操作系统与文件管理等的基本知识。

华中理工大学博士生导师康华光教授，武汉大学计算机系（原系主任）黄俊杰教授对本书给予了热情的支持与帮助，他们在百忙中审阅了全部原稿，提出了宝贵的意见，在此向他们表示衷心感谢！

诚挚感谢对本书的编写与出版给予大力支持与帮助的华中理工大学研究生院，华中理工大学出版社的各位领导与同志们以及其他各方面的同志们。

还要感谢我的妻子，她为本书作出了巨大的牺牲与贡献。

本书中有的内容是笔者的实践经验和研究心得，加之水平有限，难免有谬误与欠妥之处，恳请专家与读者批评指教。

刘 云 生

1992年6月于武昌华工园

目 录

第一章 概论	(1)
1.1 数据库设计问题	(1)
1.2 分步的数据库设计	(3)
1.2.1 设计过程的发展	(3)
1.2.2 设计的过程	(5)
1.2.3 设计的步骤	(7)
1.3 数据库设计技术	(13)
1.4 数据库设计方法学	(14)
1.5 应用与数据特征的影响	(17)
1.5.1 数据特征	(17)
1.5.2 应用环境特征	(23)
1.5.3 应用与数据特征的设计考虑	(27)
第二章 需求分析与系统说明	(29)
2.1 引言	(29)
2.2 应用环境分析	(30)
2.3 元数据收集	(32)
2.4 应用功能分析	(34)
2.5 开发系统说明	(36)
2.5.1 组织模型转换	(36)
2.5.2 任务说明	(37)
2.5.3 静态数据说明	(38)
2.5.4 数据使用说明	(39)
2.5.5 运行需求说明	(42)
第三章 概念设计	(43)
3.1 概述	(43)
3.1.1 信息表示	(43)

3.1.2	设计的观点	(45)
3.1.3	设计中的决策	(46)
3.2	概念设计的步骤	(49)
3.3	概念数据建模	(51)
3.3.1	数据抽象	(51)
3.3.2	实体建模	(53)
3.3.3	实体、联系与属性的选择	(58)
3.4	视图合并	(59)
3.4.1	恒等合并	(60)
3.4.2	聚合并	(61)
3.4.3	概括合并	(63)
3.4.4	小结	(64)
3.5	概念设计方法学	(67)
3.5.1	实体分析	(67)
3.5.2	属性综合	(68)
第四章	语义数据库建模	(71)
4.1	语义模型	(71)
4.1.1	概述	(71)
4.1.2	语义模型与面向对象的程序设计语言	(73)
4.1.3	语义数据模型的优点	(74)
4.1.4	利用语义模型的数据库设计	(76)
4.2	数据分析的关系模型	(78)
4.3	语义建模的 E-R 方法	(81)
4.3.1	E-R 模型	(81)
4.3.2	E-R 建模	(83)
4.3.3	实体、联系到关系	(89)
4.4	实体模型	(92)
4.5	作用模型	(93)
4.5.1	作用模型	(94)
4.5.2	作用的建模结构	(97)
4.5.3	作用模型结构的关系表示	(101)

4.6	类模型	(106)
4.7	聚集与概括	(111)
4.8	二元模型	(113)
4.9	函数模型	(115)
第五章	关系的数据库设计	(120)
5.1	关系模型结构	(120)
5.1.1	基本概念	(120)
5.1.2	关系的一般特征	(122)
5.1.3	关系数据完整性	(124)
5.1.4	关系数据库的操作	(126)
5.2	函数依赖	(129)
5.2.1	基本概念	(129)
5.2.2	关键字的定义	(131)
5.2.3	函数依赖的性质	(131)
5.2.4	最小函数依赖集	(133)
5.3	关系规范化	(137)
5.3.1	范式	(137)
5.3.2	多值依赖与第四范式	(144)
5.3.3	其他范式	(146)
5.3.4	规范化过程	(149)
5.4	关系设计标准	(150)
5.4.1	表示标准	(151)
5.4.2	冗余标准	(153)
5.5	关系的分解	(154)
5.5.1	无损分解	(154)
5.5.2	无损分解的测试	(155)
5.5.3	分解算法	(157)
5.5.4	推广到多值依赖	(158)
5.5.5	分解法的问题	(161)
5.6	关系的综合法	(164)
5.6.1	函数依赖的语义	(165)

5.6.2 综合算法开发	(166)
第六章 实现设计	(170)
6.1 概述	(170)
6.2 实现设计的步骤	(172)
6.3 概念模型到关系模型的转换	(175)
6.4 概念模型到记录类型的逻辑转换	(176)
6.4.1 E-R 模型的逻辑转换	(176)
6.4.2 作用与类模型的逻辑转换	(178)
6.4.3 关系模型的逻辑转换	(183)
6.4.4 逻辑记录结构的最小化	(185)
6.4.5 概念模型到网状模型的逻辑实现	(185)
6.4.6 概念模型到层次模型的逻辑实现	(188)
6.5 应用设计说明	(190)
6.6 数据库模式性能分析	(191)
第七章 物理设计	(195)
7.1 引言	(195)
7.2 物理设计的步骤	(197)
7.3 存贮记录结构	(199)
7.4 记录群集	(203)
7.5 存取方法设计	(205)
7.6 性能估测	(208)
7.6.1 性能模型	(208)
7.6.2 I/O 服务时间的计算	(212)
7.6.3 外存空间	(216)
第八章 数据库实现	(219)
8.1 数据库实现过程	(219)
8.2 用户请求实现的层次结构	(220)
8.3 数据的存取	(224)
8.3.1 逻辑数据存取	(224)
8.3.2 物理数据存取	(226)

8.4	顺序组织	(228)
8.5	直接组织	(229)
8.5.1	直接地址结构	(229)
8.5.2	Hashing 结构	(230)
8.5.3	索引结构	(234)
8.5.4	B-树	(237)
8.5.5	直接组织小结	(240)
8.6	索引顺序与 B ⁺ -树组织	(241)
8.7	链表结构	(245)
8.7.1	单链表结构	(245)
8.7.2	多链表结构	(246)
8.7.3	链表结构的数据维护	(247)
8.8	倒排表结构	(249)
8.9	重复组的物理存贮	(251)
第九章	数据库管理系统结构	(254)
9.1	数据管理软件概述	(254)
9.2	DBMS 的功能与结构	(255)
9.3	数据字典系统	(258)
9.3.1	数据字典结构	(258)
9.3.2	数据字典管理	(259)
9.3.3	数据字典系统的实现	(260)
9.4	数据集成	(262)
9.4.1	数据独立	(262)
9.4.2	数据共享	(268)
9.4.3	数据的集中控制	(268)
9.4.4	数据可用性	(270)
9.5	数据完整性	(271)
9.5.1	数据验证	(272)
9.5.2	数据一致性	(273)
9.5.3	锁机构	(275)
9.5.4	死锁	(277)

9.6	数据库恢复	(279)
9.6.1	数据库的后备	(279)
9.6.2	日记	(280)
9.6.3	恢复与重起动过程	(281)
9.7	安全与保密	(283)
9.7.1	用户识别	(283)
9.7.2	权限审定	(284)
9.7.3	密码数据	(286)
9.8	关系数据库管理	(287)
9.8.1	一般特征	(287)
9.8.2	RDBMS 体系结构	(289)
9.8.3	关系数据定义	(290)
9.8.4	关系数据操作	(293)
9.9	网状数据库管理	(296)
9.9.1	网状数据模型概述	(296)
9.9.2	系结构	(297)
9.9.3	DBTG 模型特征	(298)
9.9.4	DBTG 数据库的检索	(301)
9.9.5	数据维护与控制	(302)
9.9.6	网状数据库评议	(304)
9.10	层次数据库管理	(305)
9.10.1	数据结构特征	(305)
9.10.2	层次数据库检索	(307)
9.10.3	层次数据库的维护	(308)
9.10.4	层次数据库的重构	(309)
9.10.5	层次数据库评议	(310)
第十章	数据库运行	(312)
10.1	数据库运行环境	(312)
10.1.1	运行环境的构成	(312)
10.1.2	运行环境的类型	(313)
10.1.3	处理方式	(314)

10.2	数据库的建立	(317)
10.3	数据库应用程序	(321)
10.3.1	应用程序开发	(321)
10.3.2	应用程序运行	(322)
10.4	最终用户接口	(324)
10.4.1	用户类型	(324)
10.4.2	最终用户特性	(327)
10.4.3	接口特征	(329)
10.5	性能监视与分析	(331)
10.6	数据库重组与重构	(333)
第十一章	数据库的管理	(337)
11.1	引言	(337)
11.2	数据库管理	(338)
11.2.1	数据库开发周期	(338)
11.2.2	数据库管理职责	(340)
11.2.3	数据库管理接口	(342)
11.3	数据库管理的组织	(344)
11.3.1	数据库管理员的选择	(344)
11.3.2	数据库管理机构	(345)
11.3.3	数据库管理组织结构	(346)
11.3.4	分散的数据库管理组织结构	(348)
11.4	数据库规划	(350)
11.4.1	数据库的战略目标	(350)
11.4.2	应用环境描述	(352)
11.4.3	规划过程与任务	(354)
11.5	数据库设计的管理	(355)
11.6	代价与效益分析	(358)
11.6.1	效益的标识与定量表示	(358)
11.6.2	代价因子	(360)
11.6.3	代价效益比分析	(361)
第十二章	分布式数据库设计	(364)

12.1	引言.....	(364)
12.2	分布式 DBMS 体系结构.....	(366)
12.3	分布式数据库设计问题.....	(367)
12.3.1	分布式数据库设计模型	(367)
12.3.2	数据分布策略	(369)
12.3.3	分布式数据字典设计	(373)
12.3.4	同型与异型系统	(376)
12.4	数据库分布设计.....	(377)
12.4.1	数据分布设计的目标	(377)
12.4.2	数据分布设计方法	(378)
12.4.3	数据库划分	(379)
12.4.4	数据分配	(380)
参考文献.....		(382)

第一章 概 论

计算机最主要的应用方面在于构成各种信息系统或信息处理系统。信息系统用来收集、组织、处理、存贮、检索与显示信息，故它必然离不开数据库技术，并往往包含一个数据库作为其核心。因此，数据库设计技术在信息系统开发中起着关键作用。

设计一个组织（或部门）的数据库是一项很困难、很费时的的工作，它不但要使用各种数据处理技术与基础理论，以及其他领域里的许多知识，而且要涉及整个组织的各个方面，例如，从组织单位的战略目标、职能结构直到各业务部门的事务处理过程。数据库设计往往不是一个非常结构化的过程，而是一个可以有多种不同设计的方法学。在各种方法学中使用不同的设计技术与工具。数据库设计的质量就依赖于设计方法学和所使用的技术与工具。

1.1 数据库设计问题

数据库系统是数据库、数据库管理系统软件、应用程序及系统环境（包括操作系统、硬件等）的一个为用户提供信息服务的组合体。对应用程序设计、操作系统及数据库管理系统（DBMS）过去已有不少的研究，开发出了许多有用的技术。在数据库设计领域中也有大量的论著，但如何就不同的应用环境及数据特征有效地使用现有的工具与技术来构造出相应的数据库则仍然是数据库管理员或设计者所面临的主要问题。为此，首先应该明确在整个数据库开发周期中所要解决的主要问题，它们是：

- (1) 用户的需求是什么？如何表示它们？

(2) 这些需求为何转换成有效的数据库结构?

(3) 如何在计算机上有效地实现这种数据库结构?

(4) 怎样使这种数据库实现满足用户当前和将来新的或变化的需求?

数据库的服务是面向团体组织的。组织中有多种不同类型的用户,设计者必须明确他们对系统的处理功能、数据类型、数据量、数据的使用与性能要求,以及系统的各种限制。如何表示或描述用户需求与系统限制,目前在软件需求工程领域内尚是研究课题。尽管已研制出一些形式需求说明语言,但离实用还有一段相当距离。所以,当前通常还是用自由式的自然语言来说明,其中往往有很多含混的术语(如“灵活的”、“合适的”、“充分的”等)和听起来严谨但并无一定定义的术语(如“任选的”、“完全可靠的”等),这给后面的设计与实现带来许多困难与争议。

用户需求与系统限制是整个数据库设计与开发过程的基础与出发点。数据库设计者应以提供一个有效的数据库结构与实现来满足所有的用户需求为目标。然而,这是一个极其困难的任务,除了需要的知识面广(不仅仅是计算机和数据库知识)、使用的技术工具多、与组织的繁多因素关系密切(数据库系统不仅仅是一个技术系统,还是一个社会系统)外,还始终联系到系统的有时甚至是矛盾的各种限制(包括用户需求)。此外,还要考虑到发展变化。当设计者在决策潜在的节省空间与扩展数据可用性时,则必然伴随以某些用户服务功能的潜在降低。设计者虽要尽力避免这种降低,但终究可能只满足所有用户需求的公共部分。

用户通过访问数据库来获取所需信息,并记录他们的决策信息到数据库中,故数据库定位和存取数据库的设施对系统的有效性担负着极其重大的责任。合理地构造与定位数据库,不仅使得容易存取各种数据,而且能使系统很快地响应用户请求。反之,则必然妨碍系统的快速响应。若数据不能及时可用,则用户势必要推迟其任务的完成,或者以不完整的数据进行工作,这样必然最

终导致系统失败。

数据库的结构及其实现不仅要能满足当前的要求，而且还必须具有适应组织变化需要的灵活性。太紧地局限于当前应用的数据库会使许多合作性组织在范围上受到很大限制。在一个组织中，新职能的增加必然跟随着新的工作数据。数据库要能存贮这些新的数据和容纳新的数据关系，系统必须具有容易作出这种变化的性能。一个有效的完整的数据库设计过程应该同时考虑数据库的集成和灵活性两者。

1.2 分步的数据库设计

1.2.1 设计过程的发展

计算机数据库应用的不断增长导致了数据库设计过程的分步倾向，随着新的数据库应用的不断出现，设计步骤也在不断地发生变化。早期的数据库往往支持单个的组织功能，如有的是关于会计的，有的是关于库存的。它们的设计留下了单记录型文件设计法的痕迹，这种系统大多数是提供常规报告（表）的批处理系统。许多当代的数据库系统在很多方面都不同于这些早期系统，它们支持多种组织功能，集中了较多的用户应用，而且数据库的集成与应用环境紧密相关。它们不仅提供常规的标准报告，而且还允许同时以标准和非标准报告两种形式进行联机存取。

由于早期的系统多是用来满足单个用户功能需要，故设计者只需考虑一种或最多几种数据结构。而且，在那种环境下，往往只局限于给定现有的文件结构，而其设计处理则侧重于物理实现，只要选择适当的物理参数（为物理尺寸）和存贮结构等就能满足用户的需要。所以，它能有效地用一步设计法来实现目标，并且过去已有许多实例表明了这种设计法的成功。即便在当今，在局限环境下，这种方法仍不失有效性。但面对当今数据库的复杂要