

935517

TP311.13
1020

高等学校教材

数据库设计与实现

王能斌 董逸生

华中理工大学出版社

高等学校教材

数据库设计与实现

王能斌 董逸生

华中理工大学出版社

内 容 简 介

本书面向数据库的应用,以关系数据库为主,兼顾网状数据库,系统地阐述了数据库的需求分析、概念设计、逻辑设计、物理设计、分布式数据库设计和数据库设计工具。书中还介绍了有代表性的非传统的数据模型以及数据库建立、维护和管理中的若干技术问题。对于数据库应用中所需的 DBMS 知识,书中亦有所涉及。本书比较重视基本概念、基本原理和设计思想的阐明,读者只要具有数据库的基本知识,不难自学。

本书可作为有关专业的本科高年级学生和研究生的教材,也可作为从事数据库应用的科技人员的参考书。

高等学校教材

数据库设计与实现

王能斌 董逸生

责任编辑 陈少华

*

华中理工大学出版社出版发行

(邮码:430074)

新华书店湖北发行所经销

华中理工大学出版社印刷厂印刷

*

开本:850×1168 1/32 印张:12 字数:301 000

1991年9月第1版 1991年9月第1次印刷

ISBN 7-5609-0580-3/TP·55

印数:1·3 000

定价:3.10元

出版说明

根据国务院关于高等学校教材工作分工的规定,我部承担了全国高等学校、中等专业学校工科电子类专业教材的编审、出版的组织工作。由于各有关院校及参与编审工作的广大教师共同努力,有关出版社的紧密配合,从1978年至1985年,已编审、出版了两轮教材,正在陆续供给高等学校和中等专业学校教学使用。

为了使工科电子类专业教材能更好地适应“三个面向”的需要,贯彻“努力提高教材质量,逐步实现教材多样化,增加不同品种、不同层次、不同学术观点、不同风格、不同改革试验的教材”的精神,我部所属的七个高等学校教材编审委员会和两个中等专业学校教材编审委员会,在总结前两轮教材工作的基础上,结合教育形势的发展和教学改革的需要,制订了1986~1990年的“七五”(第三轮)教材编审出版规划。列入规划的教材、实验教材、教学参考书等近400种选题。这批教材的评选推荐和编写工作由各编委会直接组织进行。

这批教材的书稿,是从通过教学实践、师生反映较好的讲义中经院校推荐,由编审委员会(小组)评选择优产生出来的。广大编者、各编审委员会和有关出版社为保证教材的出版和提高教材的质量,作出了不懈的努力。

限于水平和经验,这批教材的编审、出版工作还会有缺点和不足之处,希望使用教材的单位,广大教师和同学积极提出批评建议,共同为不断提高工科电子类专业教材的质量而努力。

电子工业部教材办公室

前 言

本教材由电子工业部计算机教材编审委员会组稿和推荐出版。责任编辑为谢希仁教授。

读者在学习了数据库系统概论课以后,如要进一步深入学习数据库,一般有三方面的课程可供选择:一是面向数据库应用的课程;二是面向数据库管理系统的课程;三是面向数据库理论的课程。在数据库领域中,从事数据库应用的人也许是最多的。本书是一本面向数据库应用的教材和参考书。读者在学习本书前应具有数据库的基本知识。在数据库系统概论课中,也多少涉及到数据库设计和实现的内容,但本书论述的角度不同,对有关内容已有所了解的读者,可以快速浏览这部分内容,但最好不要跳过,以免影响全面理解本书的内容。

数据库设计这个课题虽然早在70年代初就提出了,在这方面也发表了很多文章,但迄今只有一个框架,还没有形成一套为大家所接受的设计方法。目前很多单位都在进行数据库设计,但每个单位的数据库设计方法或多或少有些不同。本书并不企图提供一套完整的数据库设计方法,而是以一般公认的数据库设计步骤为序,阐述每一设计步骤的背景知识、设计原则、基本方法以及存在的问题。我们认为,只要读者掌握这些知识,就不难根据具体的设计环境,形成一个适合本单位的数据库设计的具体方法。对于有志于研究数据库设计方法的读者,也可以通过本书比较快地熟悉这个领域的基本内容及问题,以便进一步阅读有关文献,深化这方面的知识。

本书以关系数据库设计为主,兼顾网状数据库设计,如果读者对网状数据库设计不感兴趣,可以跳过这部分内容,不影响阅读的

连续性.

本书可作为计算机类专业本科生选修课教材,并可作为计算机、管理、电子技术、自动控制等专业的研究生的教材,授课约需40学时左右.本书也可作为从事数据库应用的科技工作者的参考书.

本书的第一、二、五、六、七、九章由王能斌编写,第三、四、八章由董逸生编写.

徐洁磐教授担任本书的主审,详细地审阅了全稿,提出了不少宝贵的意见,谢希仁教授对本书的编写给予了热情的关心和鼓励,作者特此表示衷心的感谢.

本书在编写过程中,得到了东南大学计算机科学与工程系计算机应用技术教研组和研究室的同仁以及系图书室的支持,作者对他们表示诚挚的谢意.

本书的部分内容还在发展中,尚欠成熟.国内在数据库设计,尤其大型数据库设计方面的经验还不多.作者仓促受命,勉作尝试,水平有限,错误难免.尚祈广大读者和海内同行,不吝赐教,以资改进.

王能斌 董逸生

1990年12月

于东南大学六朝松畔

目 录

第一章 概 论

§ 1-1 数据库系统	(1)
§ 1-2 DBMS 结构简介	(2)
§ 1-3 数据模型和数据模式	(5)
§ 1-4 数据库设计概述	(9)
§ 1-5 数据库的生命周期	(13)
§ 1-6 数据库设计与软件工程	(14)
习题	(16)
参考文献	(16)

第二章 数据模型

§ 2-1 数据和数据模型	(17)
§ 2-2 对传统数据模型的评价	(20)
§ 2-3 E-R 数据模型	(25)
§ 2-4 函数式数据模型	(30)
§ 2-5 语义数据模型	(36)
§ 2-6 面向对象数据模型	(43)
习题	(53)
参考文献	(53)

第三章 数据库设计的需求分析

§ 3-1 引言	(54)
§ 3-2 确认设计范围	(56)
§ 3-3 面向数据的方法	(59)

§ 3-4 面向过程的方法	(71)
§ 3-5 计算机辅助需求分析	(84)
习题	(90)
参考文献	(91)

第四章 数据库的概念设计

§ 4-1 引言	(92)
§ 4-2 局部视图设计	(95)
§ 4-3 视图集成概述	(113)
§ 4-4 预集成	(118)
§ 4-5 实体类的集成	(119)
§ 4-6 联系类的集成	(126)
§ 4-7 新老数据模式的集成	(133)
习题	(134)
参考文献	(136)

第五章 数据库的逻辑设计

§ 5-1 引言	(137)
§ 5-2 关系数据模型的简短回顾	(138)
§ 5-3 E-R 数据模型到关系数据模型的转换	(140)
§ 5-4 关系数据库设计中的一些数据语义问题	(146)
§ 5-5 函数依赖	(151)
§ 5-6 多值依赖	(159)
§ 5-7 连接依赖	(163)
§ 5-8 关系模式的分解及其问题	(165)
§ 5-9 关系模式的规范化	(177)
§ 5-10 关系数据库概念模式的调整	(190)
§ 5-11 关系数据库外模式的设计	(193)
§ 5-12 网状数据模型概述	(196)
§ 5-13 网状数据库的概念模式和外模式的设计	(205)
习题	(217)
参考文献	(219)

第六章 数据库的物理设计

§ 6-1 数据库物理设计引论	(220)
§ 6-2 存储记录的设计	(222)
§ 6-3 关系数据库的簇集设计	(225)
§ 6-4 网状数据库的簇集设计	(230)
§ 6-5 文件结构和访问方法概述	(241)
§ 6-6 索引的选择	(257)
习题	(270)
参考文献	(272)

第七章 分布式数据库的设计

§ 7-1 分布式数据库系统	(273)
§ 7-2 数据的分布策略和分布式数据库的设计	(275)
§ 7-3 数据的分割设计	(282)
§ 7-4 数据的分布设计	(284)
习题	(286)
参考文献	(287)

第八章 数据库的计算机辅助设计工具

§ 8-1 引言	(288)
§ 8-2 NITDBDA 概貌	(289)
§ 8-3 NITDBDA 概念设计子系统	(291)
§ 8-4 NITDBDA 关系数据库设计子系统	(295)
§ 8-5 代价估算中处理需求的表示	(306)
§ 8-6 常用的事务执行方式	(312)
§ 8-7 NITDBDA 的代价估算模型	(314)
§ 8-8 多元事务的分解	(318)
§ 8-9 NITDBDA 网状数据库设计子系统	(321)
习题	(324)
参考文献	(324)

第九章 数据库的建立、管理与维护

§ 9-1 数据库的建立	(326)
§ 9-2 数据库的恢复	(328)
§ 9-3 数据一致性的保持和并发控制	(333)
§ 9-4 数据库的调整	(342)
§ 9-5 数据库的重组	(347)
§ 9-6 数据库的重构	(351)
§ 9-7 数据库的安全	(352)
§ 9-8 DBA 的作用	(368)
§ 9-9 微机数据库的特点	(369)
习题	(373)
参考文献	(374)

第一章 概 论

§ 1-1 数据库系统

在早期的计算机中,数据附属于应用程序,各个应用程序管理和访问自己的数据.文件系统的出现,仅仅把管理和访问数据的具体操作交给操作系统来执行,虽然各应用程序之间也可以共享一些文件,但终究没有改变数据附属于应用程序这一本质.随着社会和经济的发展,人们逐步认识到数据是一个单位的极重要的资源.数据的分散管理带来一系列的弊端,例如不受控制的数据冗余,数据的不一致性,应用程序对数据的物理组织的依赖性以及在安全性、完整性和标准化方面缺少有效措施等.因此,人们就逐步发展了以统一管理和共享数据为主要特征的数据库系统(database system).

数据库系统的简单结构如图 1-1 所示.图中的数据库是数据的汇集,它们以一定的组织形式存于存储介质上,一般是磁盘;数据库管理系统(database management system,简称 DBMS)是管理数据

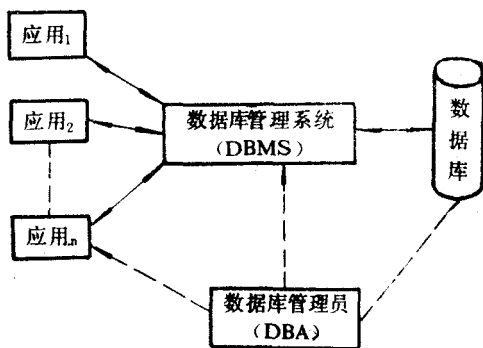


图 1-1

库的系统软件，它实现数据库系统的各种功能；图中的应用是指以数据库为基础的各种应用程序，应用程序必须通过 DBMS 访问数据库。数据库既然是共享的，就需要有人进行数据库的规划、设计、协调、实现、维护和管理等工作，负责这些工作的人员或集体称为数据库管理员 (database administrator, 简称 DBA)。应用程序、DBMS、数据库和数据库管理员构成数据库系统。如果不引起混淆，数据库系统有时也简称为数据库。

DBMS 以及支持它的计算机硬件和操作系统一般由计算机制造厂家提供。数据库以及建立在数据库上的应用程序须根据具体的应用设计，这是开发计算机应用系统的任务。

一个具体的数据库总是应用在一定范围，这可能是一个工厂、一个车间、一个公司、一个部门、一所学校、一个系、一家医院、一个机关……，为方便起见，我们把它们抽象为单位 (enterprise)。一个数据库都要对应一个单位，一个单位的数据库是这个单位的一种抽象或模拟，同时又表示这个单位的状态。

§ 1-2 DBMS 结构简介

DBMS 是数据库系统的核心。对于数据库的设计者和使用者，虽然不要求去设计 DBMS，但为了更好地设计和使用数据库，对 DBMS 的工作原理应有所了解。每个 DBMS 的结构都有其特点，在功能上也有差异，图 1-2 表示一个比较典型的、解释执行的关系型 DBMS 的结构，读者通过这个例子可了解 DBMS 的一般工作原理。

一般的 DBMS 都按照不同用户的需要提供多种用户接口，例如内含式数据语言接口、嵌入式数据语言接口、表格语言接口等。应用程序经过接口软件处理后，抽出需要访问数据库的数据语言语句，形成事务 (transaction)，交词法和语法分析器分析，产生相应的语法树，然后进行授权检查，如果通过授权检查，则继续执

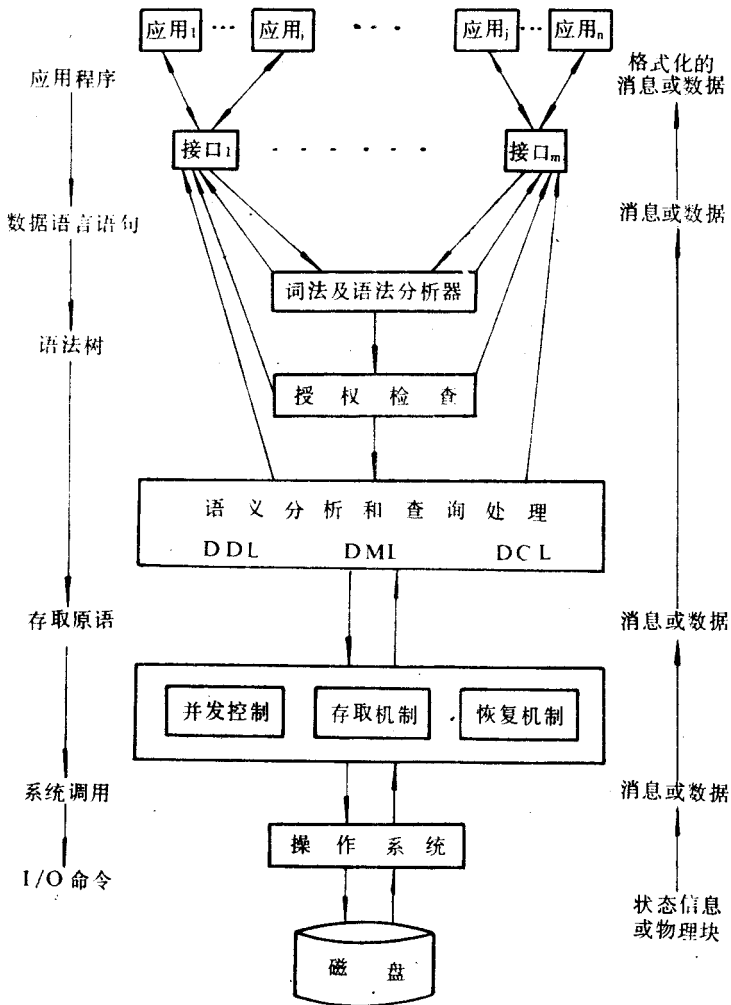


图 1-2

行;否则返回适当消息,拒绝执行。

通过授权检查的语法树即可进行语义分析和查询处理。对于数据定义语言 (data definition language, 简称 DDL)、数据操纵语言 (data manipulation language, 简称 DML) 和数据控制语言 (data control language, 简称 DCL); DBMS 分别作不同的处理,其中 DML 的处理比较复杂,还存在各种可能存取路径的选择问题,这就是所谓查询优化。经过语义分析和查询处理,就形成了事务的执行计划,并以存取原语 (access primitive) 表示。存取原语是 DBMS 内部的命令,例如打开文件、关闭文件、取一纪录、产生索引等。存取原语由存取机制执行。在执行过程中,还须考虑并发控制问题。在并发控制下,即使多个事务并发访问数据库的同一数据,其效果也会和这些事务的某一串行执行结果一样。一般用加锁的办法来避免可能发生的访问冲突。数据是重要的资源,任何破坏都会导致严重的后果。但是再好的系统总会发生故障。在发生故障时,恢复机制能够使数据库恢复到最近的一致状态或先前的某一一致状态。

DBMS 是建立在操作系统上的系统软件,系统的资源是由操作系统统管的,因而 DBMS 如有分配内存、创建或撤销进程、访问磁盘或其他外部设备等要求,必须通过操作系统来执行。DBMS 通过系统调用 (supervisor call, 简称 SVC) 请求操作系统服务。例如,请求访问磁盘上的数据库数据等。操作系统从磁盘取来的是物理块。有些操作系统还将数据的物理形式解释为相应的逻辑形式,例如纪录、属性等。在 UNIX 操作系统中,所有文件都被看成字符流,对物理块的解释工作则由 DBMS 进行。由操作系统传给 DBMS 的数据还须经过适当的处理或变换,例如聚集函数 (平均值、最大值、最小值、计数值以及和等) 的计算、数据类型转换等。由接口提供给应用程序或用户的结果,还须表示成一定的格式,例如表格形式。这些工作由接口来完成。

由上可见, DBMS 是一个比较复杂的软件系统。它对数据库

的功能、性能、用户友好性以及数据的安全、保密等影响极大，DBMS 的选择是建立数据库的重要一步，必须慎重考虑，仔细评价。

§ 1-3 数据模型和数据模式

众所周知，数据是用来描述现实世界的。而数据模型(data model)又是数据特征的抽象，它描述数据的基本结构及其相互间的关系以及定义在数据上的操作。目前，商用数据库常用的数据模型主要有层次、网状和关系三种。这三种数据模型已经成熟并广为应用，因此称为传统数据模型。由于计算机应用的需要，近十几年来，又发展了不少新的数据模型，有关这方面的问题将在第二章中详细讨论。

一、数据模型

数据模型既要比较自然地模拟现实世界并易于为人们所理解，又要便于在计算机上实现。但这两方面的要求往往是矛盾的。一种数据模型同时满足这两方面的要求至少在目前是困难的。数据库中解决这个矛盾的办法有点类似于程序设计语言，即对于不同的对象采用不同的界面。在计算机系统中，用户使用的一般是高级程序设计语言，而计算机内部执行的却是机器语言。因此，需要由编译系统把高级程序设计语言编译成机器语言。有时在编译过程中还介入一种或多种中间语言。同样的道理，在数据库中也是针对不同使用对象和目的，采用多级数据模型。一般可分为下面三级。

1. 概念数据模型(conceptual data model)

目前，一般商用 DBMS 所提供的用户界面还限于层次、网状和关系这一类数据模型。这类数据模型有诸多规定和限制，不便于非计算机专业人员的理解和应用。而数据库是一个单位的某种

模拟, 必须要有应用人员参加与合作才能正确地设计出来. 显然, 一开始就直接用 DBMS 所提供的数据模型来进行数据库设计是不合适的. 因为这既不便于非计算机专业人员的参与和理解, 同时也会使数据库设计人员一开始就纠缠于一些实现的细节, 不符合自顶向下、逐步求精的软件设计原则. 概念数据模型是一种面向用户、面向现实世界的数据库模型, 它是与 DBMS 无关的. 它主要用来描述一个单位的概念化结构. 采用概念数据模型后, 数据库设计人员在开始阶段可以把主要精力用于正确了解和描述现实世界上, 而把涉及 DBMS 的一些问题放到以后的设计阶段中去考虑. 可供选择的数据库模型将在第二章介绍.

2. 逻辑数据模型(logical data model)

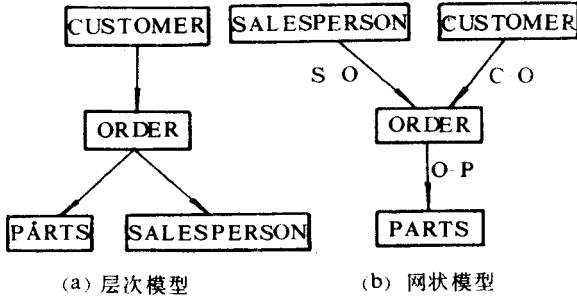
逻辑数据模型是用户从数据库所能看到的数据库模型, 它与 DBMS 有关. 层次、网状和关系数据库模型是常见逻辑数据库模型, 也有些 DBMS 能够支持多种逻辑数据库模型, 例如既支持网状又支持关系数据库模型. 概念数据库模型必须转化为逻辑数据库模型才能在数据库中实现. 逻辑数据库模型既要考虑面向用户, 也要考虑面向实现. 目前流行的几种逻辑数据库模型是在现阶段技术水平上的折衷. 这些逻辑数据库模型在一般数据库教材中已有介绍, 本书不再重复.

3. 物理数据库模型(physical data model)

数据库的数据终须存储到存储介质上, 例如磁盘上. 逻辑数据库模型只反映数据的逻辑结构, 例如关系、记录、属性、主键等, 而不反映数据的存储形式, 例如块(block)、指针、索引结构、溢出区、簇集(clustering)等. 反映数据的物理存储方式的数据库模型称为物理数据库模型. 物理数据库模型不但与 DBMS 有关, 而且与操作系统和计算机硬件有密切关系. 有关这方面的内容, 本书将在讨论数据库的物理设计时介绍.

二、数据模式

在所有的数据模型中，都有“型”（type）与“值”（value）之



CUSTOMER	
CUST-NAME	SALESPERSON

ORDER	
ORDER NO	SALESPERSON

PARTS	
PART NO	ORDER NO

(c) 关系模型

图 1-3

分。型是某一类数据的结构和特性的说明，而值是型的一个实例（instance 或 occurrence）。例如，“整数”是型，而“25”是其一个值。以一定的数据模型对一个单位的数据的类型、结构及其相互间的关系所进行的描述称为数据模式（data schema）。数据模式仅仅是型的描述，而不涉及到具体的值。数据模式实质上是用水数据模型对现实世界某一部分（a slice of the real world）的模拟。例如在图 1-