

● 计算机应用系列丛书

数据库系统导论

● 马玉书 主编 ● 石油大学出版社



数据库系统导论

马玉书 主编

石油大学出版社

新登字 10 号

内 容 提 要

本书系统而又全面地介绍数据库的一般概念、基本理论和应用技术。全书分为两篇,第一篇为数据库的基本理论及数据库系统,主要内容包括数据库的发展,网状、层次和关系三种数据模型及数据库系统、关系数据理论、数据库保护、数据库设计技术和方法、分布式数据库系统,最后介绍了我国目前流行的几种主要的数据库管理系统 ORACLE、dBASE、FoxBase、SQL/DS 等。第二篇数据库技术的新发展和新一代数据库系统,介绍了数据库的最新发展方向和新出现的各种数据库,例如智能数据库、面向对象数据库、工程数据库、多媒体数据库、模糊数据库和数据库机等。

本书内容新颖,重点突出,条理清楚,叙述严谨,覆盖面广。除包括了目前各类数据库教材和专著的主要内容外,对数据技术的最新发展给予足够的篇幅。它可以作为高等院校计算机及有关专业数据库课程的教材及参考书,也可作为从事计算机和数据库工作的科研与应用技术人员的业务参考书,经适当取舍后,也可供各类数据库技术培训班和专科学校教学使用。

数据库系统导论

马玉书 主编

*

石油大学出版社出版

(山东省 东营市)

石油大学出版社照排室排版

山东省东营新华印刷厂印刷

新华书店发行

*

开本 787×1092 1/16 20.375 印张 515 千字

1993 年 11 月第 1 版 1993 年 11 月第 1 次印刷

印数 1—3000 册

ISBN 7-5636-0317-4/TP·13

定价: 19.50 元

前 言

数据库技术是研究用电子计算机对数据进行管理、实现数据共享的一门高技术学科。它是计算机科学及其应用的重要组成部分。数据库技术是 60 年代末期在文件管理的基础上,作为数据管理的一项新技术而登上数据处理(信息管理)这一计算机应用的舞台的。从 60 年代至今,它已有了 30 多年的发展历史,逐步形成了一套较完整的理论、方法,并开发了许多数据库管理系统,生成和建立了大量的数据库及其应用系统,所有这些都大大促进了数据库技术的不断普及和发展,对计算机的广泛应用起着十分重大的作用。

随着电子计算机应用领域的不断扩展及延拓,社会的信息量激增和数据处理在计算机应用中所占比重的不断上升,数据库技术在数据处理(信息管理)及其它新兴领域中已扮演了越来越重要的角色。今天,数据库技术和计算机网络及通信已成为当前计算机应用中两个重要的基础领域,电子计算机的一些重要应用,如管理信息系统(MIS)、决策支持系统(DSS)、办公室自动化(OA)、计算机辅助系统(CAD/CAM/CASE 等)、专家系统(ES)和知识库系统(KDS)等,大都离不开这两个基本手段。与此同时,数据库技术本身也因其不断地引进一些新思想、新技术、新方法,得到了不断的完善和发展,从而成为计算机科学中一个十分活跃而重要的分支。

石油工业是一个高技术、高资金的密集性行业,石油勘探、开发工作,从某种意义上讲就是把巨额的投资变为有用的信息资源。目前,我国石油工业正在建立一个用于生产和经营的大型石油勘探开发数据库及庞大的石油综合信息系统。它将综合应用包括数据库、计算机网络、卫星通讯在内的各种先进技术,最大限度地收集、利用地下和地面的各种有用信息资源,能为快速而有效地勘探、开发油气田和实现油田的自动化生产及管理提供充足的数据,为经营管理作出正确可靠的决策。总之,数据库技术的引入,将对石油工业的生产和管理带来一场深刻的革命,石油工业部门几十万职工,不同层次的工程技术人员和管理干部,石油大专院校的师生学习掌握数据库技术,开展信息管理和综合应用数据库技术,已成为当务之急。

本书是有关石油院校的教师,在多年的数据库课教学和科研的基础上编写的一本教材和技术参考书。它系统地阐述了数据库的基本概念、基本原理、基本理论以及有关的实现技术和方法。本书既保持了数据库技术本身的完整性、系统性,吸取了国内外同类教材的经验,同时又力求做到内容新颖,重点突出,科学性强,条理清楚,叙述严谨,简单易懂,以克服部分同类教材存在内容枯燥乏味、一大堆概念罗列的缺点。书中还十分注重联系实际,突出关系数据库方法、ORACLE 关系数据库系统、关系数据库的优化设计及其在石油工业中的应用。此外,书中还对人工智能与数据库技术的结合——智能数据库、面向对象数据库、工程数据库、多媒质数据库和数据库机等新一代数据库及当代数据库技术的最新发展给予了足够的重视和一定的篇幅。

本书共分两篇,第一篇为数据库的基本理论及数据库系统,其中第一、二章简明扼要地介绍了数据库技术的基本概念与原理,读者通过对这两章的学习,能对数据库技术的特点、数据库的组织 and 结构有一个概貌性的了解,并初步掌握将现实世界中的对象转化为数据库中数据的方法;第三、四章介绍了网状和层次这两个早期的数据库模型及其数据库管理系统,通过对格式化数据模型的学习和分析,使读者能更深入地了解有关数据库系统的概念与原理;第五、六、七章为本书的重点,它突出地叙述了关系数据库方法,分析解剖了具有代表性的 SYS-

TEM R 实验关系数据库管理系统,并深入讨论了关系数据库理论,这不仅可帮助读者加深对关系方法的理解,而且也今后的关系数据库设计奠定坚实的理论基础;第八章数据库保护,讨论了实现数据库管理系统的一些主要技术问题(安全性、完整性、并发性及数据库恢复与维护)及其解决方法;第九章全面介绍了数据库设计的基本技术和方法;第十章给出了目前我国及石油部门广泛采用的 ORACLE、VAX-DBMS、SQL/DS 等数据库管理系统及 MDBS3、dBASE III、FOXBASE 等微机数据库系统,通过了解这些实际系统,读者可进一步了解数据库系统的原理及掌握开发一个实用的数据库系统应具备的知识;第十一章分布式数据库系统,介绍了分布式数据库的产生背景、概念、特点及其主要实现技术。第二篇为数据库技术的新发展和新一代数据库系统,其中第十二章数据库技术的发展和展望指出传统数据库的缺陷与不足和新一代数据库的发展途径;第十三章智能数据库指出将数据库和人工智能这两项技术相结合实现数据库的智能化和智能化的数据库的途径,并以演绎数据库为重点介绍了演绎数据库、专家数据库系统、主动数据库、知识库系统的特点和实现方法;第十四章专门介绍了面向对象数据库系统的原理、实现技术和方法,以及面向对象数据库系统的标准化问题。最后,第十五章介绍了当前数据库技术的其它新的研究课题,以及工程数据库、图形库和图象库、多媒质数据库、统计数据库和科学数据库、历史数据库和时态数据库、模糊数据库和数据库机等其它各种新型数据库系统。第二篇对本科学生不作为讲授内容,但鉴于当前数据库技术正处于更新换代的前夜,通过本篇内容的学习,可使读者对当前数据库技术的最新发展有一全面的了解,以跟踪当代数据库技术的发展趋势。

本书是在石油高校计算机专业指导委员会及石油高校计算机教材编委会组织下,由各石油院校计算机系担任数据库课程教学的有关教师,通力协作,在两年的时间内共同编写完成的。本书第一篇的第一、二、三章由叶飞跃、马玉书编写,第五、六、七章由柴之聪、李莹编写,第四、八章由冷英男编写,第九章由冷英男、柴之聪编写,第十章由叶飞跃、冷英男编写,第十一章和第二篇中的第十二、十三、十四、十五章由马玉书、叶飞跃编写。最后由马玉书、叶飞跃负责全书的统编、修改和校订工作。

中国科学院软件所曹东启研究员认真审阅了全部原稿,并提出了许多宝贵的意见和建议,在此表示衷心的感谢。

由于编者水平有限,加之数据库技术本身尚在不断地完善和发展,在本书内容的选材和安排方面可能存在不少问题和不妥之处,恳请有关专家和读者批评、指正。

编 者

1992 年 5 月

目 录

第一篇 数据库基本理论和数据库系统

第一章 数据库系统概述.....	1
§ 1.1 信息和数据	1
§ 1.2 数据管理技术的发展	2
§ 1.3 数据库系统的基本概念	4
§ 1.4 数据库系统的特点	9
习 题	11
第二章 数据模型	12
§ 2.1 信息结构的三个世界.....	12
§ 2.2 E-R 模型.....	13
§ 2.3 层次模型.....	16
§ 2.4 网状模型.....	16
§ 2.5 关系模型.....	17
§ 2.6 E-R 模型向经典数据模型的转化.....	18
习 题	21
第三章 网状方法	22
§ 3.1 DBTG 系统的产生	22
§ 3.2 网状数据库的数据结构.....	22
§ 3.3 DBTG 的存储策略	27
§ 3.4 数据描述语言 DDL	35
§ 3.5 数据操纵语言 DML	40
习 题	47
第四章 层次方法	48
§ 4.1 IMS 系统的层次结构及数据模型	48
§ 4.2 数据子模型.....	54
§ 4.3 IMS 的存储结构	57
§ 4.4 IMS 的数据操纵及 DL/1	62
§ 4.5 IMS 的逻辑数据库	70
§ 4.6 IMS 的辅助索引	75
习 题	76
第五章 关系方法	79
§ 5.1 关系及基本概念.....	79
§ 5.2 关系数据语言概述.....	83
§ 5.3 关系代数.....	84
§ 5.4 关系演算.....	90

§ 5.5	ALPHA 语言	93
§ 5.6	QBE 语言	96
§ 5.7	SQL 语言	99
§ 5.8	查询优化	104
第六章	实验关系数据库系统 SYSTEM R	111
§ 6.1	SYSTEM R 的总体结构	111
§ 6.2	关系数据系统(RDS)	112
§ 6.3	研究存储系统(RSS)	121
第七章	关系数据库理论	127
§ 7.1	关系模式的一般表示及设计中的问题	127
§ 7.2	函数依赖	128
§ 7.3	函数依赖的公理系统	130
§ 7.4	关系模式的分解	133
§ 7.5	关系模式的规范化	137
§ 7.6	多值依赖和第四范式	139
§ 7.7	连接依赖和第五范式	141
第八章	数据库保护	144
§ 8.1	数据库保护的一般问题	144
§ 8.2	安全性	146
§ 8.3	完整性	154
§ 8.4	数据库的恢复	157
§ 8.5	并发控制	159
习 题		168
第九章	数据库设计	169
§ 9.1	数据库设计概述	169
§ 9.2	数据库设计方法综述	173
§ 9.3	数据库设计的理论与工具	180
§ 9.4	需求分析	183
§ 9.5	概念设计	184
§ 9.6	数据库的逻辑结构设计	190
§ 9.7	数据库物理设计	194
§ 9.8	数据库的实施和维护	196
第十章	几个具体的 DBMS	198
§ 10.1	VAX-DBMS 网状数据库系统	198
§ 10.2	ORACLE 关系数据库系统	201
§ 10.3	SQL/DS	204
§ 10.4	微机或微机网上的 DBMS	208
第十一章	分布式数据库系统	213
§ 11.1	分布式数据库及其产生背景	213
§ 11.2	分布式数据库的特点	215

§ 11.3	分布式数据库的基本概念	216
§ 11.4	分布式数据库管理系统	219
§ 11.5	典型的分布式数据库管理系统	223
§ 11.6	分布式数据库的主要技术问题	225
第二篇 数据库技术的新发展和新一代数据库系统		
第十二章	数据库技术的发展和展望	230
§ 12.1	传统数据库与新一代数据库	230
§ 12.2	传统数据库的缺陷和不足	231
§ 12.3	当前数据库技术的发展途径	233
§ 12.4	两个宣言与新一代数据库系统	234
第十三章	智能数据库	237
§ 13.1	数据库和人工智能(DB&AI)	237
§ 13.2	数据库的智能化和智能化的数据库	238
§ 13.3	演绎数据库	239
§ 13.4	智能数据库	254
§ 13.5	专家数据库系统	255
§ 13.6	主动数据库	260
§ 13.7	智能数据库的特点及主要的实现技术	260
§ 13.8	知识库系统	263
第十四章	面向对象数据库系统	269
§ 14.1	面向对象程序设计方法产生的背景	269
§ 14.2	面向对象程序设计的基本概念和特点	270
§ 14.3	面向对象程序设计的发展	274
§ 14.4	面向对象程序设计语言	275
§ 14.5	面向对象数据库管理系统的实现方法和途径	278
§ 14.6	OODBMS 体系结构	280
§ 14.7	典型的 OODBMS 举例	281
§ 14.8	OODBS 的应用及其相关研究领域	283
§ 14.9	OODBS 有待解决的问题及 OODBS 的标准化	284
第十五章	其它新一代数据库	287
§ 15.1	工程数据库	287
§ 15.2	图形库和图象库	295
§ 15.3	多媒质数据库	302
§ 15.4	统计数据库和科学数据库	304
§ 15.5	历史数据库和时态数据库	305
§ 15.6	空间数据库	306
§ 15.7	模糊数据库	308
§ 15.8	数据库机	312
参考文献		318

第一篇 数据库基本理论和数据库系统

第一章 数据库系统概述

数据库技术是数据管理的最新技术,是计算机科学的一个重要分支。从 60 年代初期,由于人们对数据处理的要求越来越高,单纯的文件系统已渐渐不能满足需要,于是开发了各种专用的数据管理软件包,并由此逐渐发展成为一门通用的系统软件——数据库管理系统。正是因为数据库系统是逐渐形成的,而且目前还在不断发展之中,人们对它的认识也在不断地深化,所以很难用几句话严格地概括出它的全部特征。因此,这一章将从数据管理的进展情况,给出数据库的一个大体上的轮廓性描述。

§ 1.1 信息和数据

信息(Information)是自然环境和人们一切活动所产生的各种状态和消息的总称。我们所赖以生存的世界是一个物质的世界,也是一个信息的世界。过去,人们主要注重于处理那些物质的、有形的财产。例如:存款、家具、房屋等,而忽略了信息的价值。然而信息是极为重要的。例如,市场的需求信息对于一个工厂的生产有着极其重要的参考价值;又如,某一地区的地质情况如何,对于是否决定在该地区开采石油也有着重要的影响。也就是说物质的存在伴随着信息的存在,物质的变化会引起信息的变化。人们在政治、经济、军事、文化、科学、教育、生产和商业等活动中将会产生大量的信息。信息如果能被迅速地处理、加工、传播和利用,又会反过来促进物质世界的发展,推动社会的前进。但是在科学技术不甚发达的时代,信息的作用及利用价值被限制在较低的程度和较小的范围以内。随着计算机技术的迅速发展,计算机具有的高速处理能力和巨大的存储容量,使人们有可能对大量的信息进行保存和加工处理,为了记载信息,人们使用各种物理符号来表示信息,这些物理符号就是数据(Data),即数据是信息的具体表示形式。由此可见,数据和信息是不可分离而又有一定区别的概念。信息是消化了的数据,数据是信息的具体表现。因此,也有人认为数据的含义是指原始的、未加工的数据,信息是指在某种情况下可使用的或经过处理、解释后的数据。不过,在某些场合下,一般并不严格地区分两者,信息和数据可以相互替代使用。

有了数据就产生了数据处理的问题,这是由于随着人们对信息的需求的增加,信息量急剧上升,数据的形式也多样化,有数字、文字、声音、图形、图象等等,结构越来越复杂。数据处理就成为了一个极其重要的问题。

所谓数据处理(Data Processing)指的是对数据进行收集、组织、加工、储存、抽取、传播的过程。数据处理的目的是要从大量的、杂乱无章的、难以理解的数据中,获得对某个特定的应用领域来说是有价值、有意义的信息,借以作出决策的依据。例如,气象预报中的数据处理和石油地震勘探资料的数据处理,都是先采集到原始数据,然后按预先设计的数学模型进行处理,得

出的结果,提供给专业人员作为预报或预测的依据。广义地说,数据处理已经历了漫长的历程。可分为三个阶段:第一阶段,人们利用各种初级的计算工具,如算盘等,这是手工数据处理阶段;第二阶段,从1880年美国进行人口统计,采用Herman Hollerith发明的卡片制表机编制人口普查表,利用穿孔卡片来存储信息,用机械方法进行数据处理,开始了机械数据处理的阶段;第三阶段,本世纪40年代电子计算机的发明,则使数据处理进入了电子数据处理的新时代。由于电子计算机处理速度快,存储容量大,自动化程度高,使数据处理工作得到飞速发展。后面我们所说的数据处理,均系指电子数据处理而言。

电子数据处理中,数据是处理的对象,数据的组织方式、结构、存储形式是影响数据处理效率的根本因素。如何组织、管理好数据是数据管理工作的任务。所谓数据管理(Data Management)是指对数据的组织、存储、检索和维护等工作。它是数据处理的基础。随着计算机技术的发展,磁芯存储器、半导体存储器、磁带、磁鼓和磁盘等存储技术的出现,使数据的保存和管理得到了迅速的发展,它主要围绕着如何改进存储方法、数据组织结构,提高数据独立性,降低数据冗余度、共享数据,提高数据的安全性和方便地运用及管理这些数据资源。数据库就是60年代作为数据管理的最新技术登上数据处理舞台的。下面我们就数据管理技术的发展作一简单介绍。

§ 1.2 数据管理技术的发展

§ 1.2.1 人工管理阶段

这是计算机用于数据处理的初级阶段。50年代及以前,计算机仅提供基本的输入输出程序,程序员除了编制处理程序外,还要考虑数据的组织形式,亲自设计物理组织。数据的逻辑组织和物理组织往往是一致的。程序和数据混为一体,在需要存取数据时,直接按物理地址访问。一旦数据的物理组织或存储设备改变,其应用程序必须重新编制。在这个阶段,数据一般不长期保存,没有专门软件对数据进行管理。一组数据对应一个应用程序,数据无法共享,应用程序负责数据的管理。其对应关系如图1.1所示。

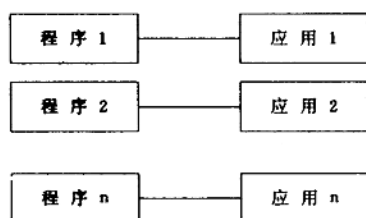


图 1.1 手工管理阶段

§ 1.2.2 文件系统阶段

人工管理阶段,应用程序依赖于数据的存储形式,数据独立性差,不同应用之间的共同数据需要重复存储,造成很高的数据冗余,给数据的维护带来很多困难。到50年代后期,操作系统中包含了文件系统后,有了专门来进行数据管理的软件。程序员只涉及数据的逻辑结构,它与物理结构之间由文件系统实现转换。文件可按不同的组织方式组成顺序文件、索引文件和随机文件。一个应用程序可以建立、维护和处理一个或几个文件,按文件组织方式的不同,顺序地或直接地以记录为单位存取数据,无需关心外部设备和地址,这一阶段数据和程序之间的关系可用图1.2表示。

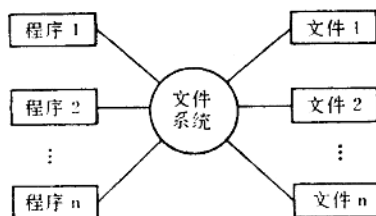


图 1.2 文件系统阶段

在这一阶段,文件系统作为应用程序和数据文件之间的接口,起到了将数据的逻辑结构和物理结构独立分开的作用。当物理结构或外部设备变化时,应用程序不需修改照样可以使用,提高了数据的物理独立性,并实现了应用程序之间以文件为单位的数据共享。但这种方法,仍存在许多弱点。例如,未实现以记录甚至以数据项为单位的数据共享,数据冗余现象还很严重,数据的逻辑结构还是根据具体应用来设计,即一个文件基本上对应于一个应用程序,数据独立性不高,而且文件之间无联系,文件中结构单一,表示复杂数据的能力有限。不过不管怎么说,文件系统在数据处理技术的发展过程中曾起过相当重要的作用。目前,它仍然是一种使用相当广泛的数据管理办法。

§ 1.2.3 数据库管理阶段

随着电子计算机工业的发展,到60年代初,出现了大容量的直接存取设备——磁盘。而文件系统的局限性,数据处理技术对数据管理提出了更高的要求。要求具有更高的数据共享,保证更高的数据独立性,表达更复杂的数据结构,因而导致了数据库管理技术的发展。数据库实际上是一个存储在计算机内的所有相关数据构成的集合。其基本思想是要对所有用户数据实行统一的、集中的管理、操作和维护。数据独立于程序而存在,并可提供给不同应用、广泛地共享使用。如图1.3所示。

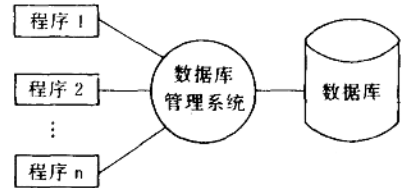


图 1.3 数据库阶段

数据库技术有别于文件系统的最大特点之一就是数据共享。数据不再从属于特定的应用而是面向整体来组织数据。虽然在操作系统下数据仍然是以文件的形式存放于磁盘,但这些文件内部的组织已经与文件系统中的文件的组织有所不同。文件之间也不再是松散无关的,常采用某种数据模型将全部数据按文件组织成一个结构化的整体,数据库管理系统提供它们之间交叉访问的手段,这就大大减少了数据的冗余度。数据库技术还提供较强的数据独立性、安全性和完整性,使得数据受应用程序的影响大大减少,数据的可靠性得到了保证。另外,数据库系统还提供了统一的数据操作手段,为用户提供了方便的接口,从而使数据的应用更加有效。

据目前统计,数据库管理系统中采用的数据模型主要有网状模型、层次模型和关系模型三种。这三种模型以60年代后期和70年代初期的三个事件为标志,这也是数据库发展史上的三个里程碑。

(1) 1968年美国IBM公司推出了IMS系统。这是一个以层次模型为基础的数据库管理系统;

(2) 1969年美国数据系统协商委员会(COADSYL)的下属组织,数据库工作小组(DBTG)发表了“DBTG报告”,提出了以网状模型为基础的数据库系统方案,也可认为是一个数据库规范;

(3) 1971年美国IBM公司的E. F. Codd发表了第一篇关系数据库论文,并随后提出了一系列有关数据库的理论。自此,数据库技术的发展逐渐以关系数据模型为主导地位。

数据库是一项新型的数据管理技术,由于它的出现,数据处理的效率越来越高,成本大大下降,可靠性不断增加,已成为现代管理信息系统不可缺少的强有力的工具。

§ 1.3 数据库系统的基本概念

§ 1.3.1 数据库、数据库管理系统和数据库系统

数据库(Database)广义地讲可以是任何数据的集合,也并不一定是驻留在可由计算机快速存取的存储设备上。一套处方卡片,一个书店的书库,一个地区的地质资料档案,磁带上的数据、程序、歌曲或录像,甚至山洞里的史前壁画都可以认为是数据库。然而这并不是我们所感兴趣的。我们所讲的数据库是计算机中的一种数据管理技术。因此,我们将数据库的概念局限在计算机中按一定的组织方式存储在一起的、相互有关的、为用户共同关心的全部信息的集合。它具有最少的数据重复(非冗余),能为多种应用服务(数据共享),数据的存储独立于应用程序(数据独立性),更新和检索库中数据时,系统提供统一的控制方法(操纵方法)并确保数据的安全性和完整性。

有了数据库,就必须有一套管理数据库的通用软件。它负责数据库的建立,数据结构的定义,数据库中数据的更新和查询,多用户并发访问数据库时的集中控制,并进行安全性和完整性检查以及数据库初始数据的装入,系统性能的监测,数据库的更新和再组织,数据库数据的转储、出错或硬软件故障后的恢复等。完成这些任务的软件就称为数据库管理系统(Database Management System)。简言之,数据库管理系统是一套用于建立、管理和维护数据库的大型系统软件。

数据库管理系统,一般是通过计算机上的操作系统来实现数据库功能的。把这种引入了数据库管理系统后,具有管理数据功能的计算机系统称为数据库系统(Database System)。因此,数据库系统包括:计算机硬件、操作系统、数据库管理系统、数据库和在其上面开发的应用软件的全体。图 1.4 表示了一个数据库系统的组成及其相互之间的关系。



图 1.4 数据库系统

由图 1.4 可以看出,操作系统给计算机硬件(裸机)作了第一次扩充,而数据库管理系统为具有操作系统的计算机系统作了第二次扩充。用户通过数据库管理系统提供的数据库操纵方法来开发各种应用,进行数据处理。数据库管理系统则利用操作系统提供的设备输入输出手段来实现用户和二级存储器之间的数据交换。由于数据库管理系统是借助于操作系统在磁盘上存取数据的,所以,

存放在磁盘上的数据库实际上是以多个有关连的相关文件的形式出现的。因此,有人就从这个角度将数据库也称之为由数据库管理系统建立和维护的相互有关文件的集合。

§ 1.3.2 数据库系统结构

数据库技术中,为了提高数据的独立性,采用将数据库系统结构划分为多个级别的方法。美国数据库管理系统研究小组 ANSI/SPARC (ANSI/SPARC 是 70 年代初由美国国家标准局(ANSI)建立的组织,开展数据库系统的标准化工作。SPARC 为 Standards Planning and Requirements Committee 的缩写)。1975 年提出了将数据库划分为三级的标准化建议。虽然目前运行的数据库系统并不都是按这种结构组织,但是,它为研究数据库系统提供了有用的框架。图 1.5 表示了这种结构的基本思想。这种结构划分为三级:外部级、概念级和内部级。

1. 外部级

外部级是数据库系统的用户级。一般地说大多数用户只对数据库的某部分感兴趣。外部

级就是用于定义与各用户有关的数据库中部分数据的局部逻辑结构。这个局部的逻辑结构称之为数据库的外部视图(External View)。一个数据库可以对应于多个外部视图,各外部视图各不相同,但也允许相互交叉重叠,以起到共享数据的目的。用户使用局部逻辑结构访问数据,隔离了与他无关的数据,简化了程序的编写,也保护了他不该访问的那部分数据。外部视图使用专用的数据描述语言进行定义,这个定义称之为外部模式(External Schema)或简称子模式(Subschema)。

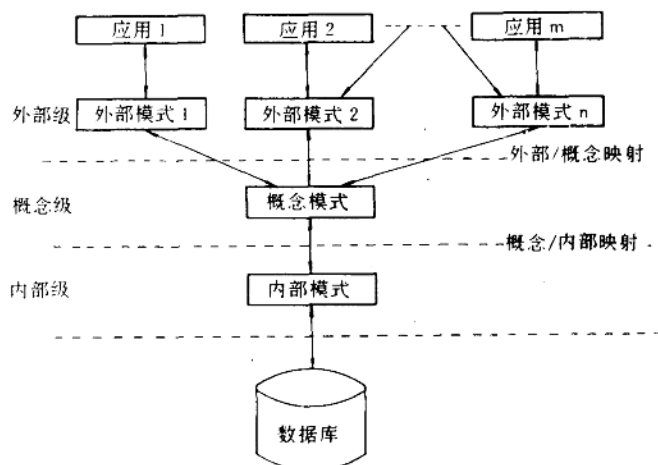


图 1.5 数据库系统结构

2. 概念级

概念级用于描述数据库的全局逻辑结构,这种结构,称之为概念视图(Conceptual View)。它包括所有数据及相互关系的描述。概念视图使用数据描述语言进行定义,此定义称之为概念模式(Conceptual Schema)或简称模式(Schema)。仅从数据结构上看,概念视图是所有外部视图的并集组合。实际上概念模式中除了数据库全部逻辑结构的定义外,还要定义数据的安全性和完整性检查以及其它的数据库特性。即它不仅包括数据的定义,还包括怎样使用这些数据和怎样对这些数据进行检查和控制。目前,实际的数据库系统中,没有一个能够支持这种包罗万象的理想的概念级。大多数系统的概念模式,只是所有外部模式的简单组合,并略加上一点简单的安全性和完整性检查。

3. 内部级

内部级是为了提高数据库的物理独立性而设立的。在概念级中,不涉及数据的任何存储特征,这些有关的存储结构,仅在内部级中予以描述,并称之为数据库的内部视图(Internal View)。内部视图将包括存储字段的说明、存储记录的顺序、索引、散列编址、指针等其它有关存储的细节。内部视图也使用数据库管理系统提供的数据库描述语言进行定义,这个定义称之为内部模式(Internal Schema)或存储模式(Storage Schema)。应当强调,内部模式并非物理数据库,因为它不涉及到物理记录,也不涉及象圆柱面或磁道大小这样的与设备有关的概念。所以说,它距物理层还有一定的距离。假定内部视图是一个无限大的线性的地址空间,数据库在这个空间中进行分配,把这个地址空间如何映射到物理存储设备上则由数据库管理系统和操作系统共同完成。对于数据库,我们不想在二进制字位级别上,而宁愿在更高一级(即认为是简单数据结构记录的集合)上去看待,因而,内部级中省去了有关物理设备的描述。

最后,我们再来看一下图 1.5,会发现结构中有两层映射。一层处在外部级和概念级之间;另一层处在概念级和内部级之间。DBMS 提供这两层之间的映射。外部/概念层映射,定义外部视图和概念视图之间的对应项。例如,字段可以有不同的数据类型,字段和记录名可以改变,多个概念级的数据项可以结合成一个单个外部级数据项等等。概念/内部层映射定义概念视图和内部视图之间的对应项。它确定了概念记录和数据项在内部级怎样表示或如何存储。如果概念视图改变了,那么只要适当修改外部/概念层的映射,外部视图可以不变;同样地,如果存储结构或存储设备改变了,也可以通过修改概念/内部层映射,保证概念视图不用修改。

§ 1.3.3 数据库管理系统的组成

数据库管理系统是数据库系统的核心,其组成和功能随系统的大小而异。在讨论数据库管理系统的组成之前,首先介绍用户存取数据的过程。

1. 存取数据的过程

当数据库建立好之后,程序员根据应用的要求,编制应用程序,在程序中通过数据库管理系统提供的数据操纵语句来实现对数据库中数据的访问。应用程序访问数据库中数据的过程如图 1.6 所示。它表示一个用户“读”数据库中一个记录的过程。

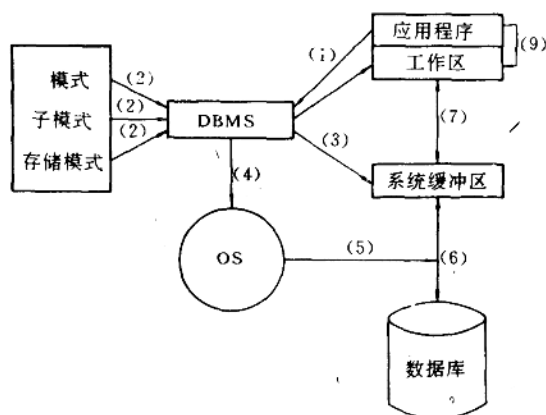


图 1.6 数据库数据存取过程

- (1) 应用程序使用数据操纵语句,向数据库管理系统发出读取一个记录的命令;
- (2) DBMS 查阅数据库中相应的子模式、模式和存储模式,以获得必要的记录定位信息;
- (3) 需要时,DBMS 分配足够的系统缓冲区;
- (4) DBMS 向操作系统(OS)发出物理读/写命令,其中读/写的物理块地址和缓冲区地址是根据(3)给出的;
- (5) 操作系统调用设备 I/O 程序,进行物理块读/写操作;
- (6) 在数据库和系统缓冲区之间传送数据;
- (7) DBMS 根据存储模式、模式和子模式的定义,进行数据映射,将系统缓冲区中的数据转换成子模式中定义的记录,并传送到用户工作区;
- (8) DBMS 将该次访问的状态传送给应用程序;
- (9) 应用程序根据访问的状态,对用户工作区中的记录进行处理。

2. 数据库管理系统的组成

数据库管理系统不论大小,一般都具有定义数据库、管理数据库、维护数据库和数据通讯等功能。为实现以上这些功能,数据库管理系统一般由下列几部分组成。

1) 数据描述语言及其翻译程序

数据描述语言 DDL (Data Description Language), 用于定义数据库的各级模式。分为模式 DDL、子模式 DDL 和存储模式 DDL 三种。它们分别用于描述数据库的概念视图、外部视图和内部视图。用户建立数据库, 首先必须使用数据描述语言, 定义数据库的全局、局部的逻辑结构和存储结构, 然后利用数据库管理系统提供的模式语言翻译程序, 将其编译成各级的目标模式, 存储于数据字典之中, 供数据库管理系统在存取数据时使用。

2) 数据操纵语言及其翻译程序

数据操纵语言 DML (Data Manipulation Language) 是 DBMS 提供给用户, 用以查询和更新数据库中数据的工具。

DML 一般分为自含式和嵌入式两种。

自含式语言 (Self Contained Language) 可以独立使用, 进行简单的数据操纵, 它可独立编写程序, 用于非程序员用户调用数据库中的数据, 或数据库设计阶段调试数据库性能。自含式 DML 多为交互式的语言, 使用解释型语言翻译程序, 实现用户指定的任务。

嵌入式语言 (Embedded Language) 不能独立使用, 它必须与某种程序设计语言编写在一起才能运行, 这种程序设计语言叫做宿主语言 (Host Language)。比如, COBOL、PL/I、FORTRAN 或汇编语言等。嵌入式数据操纵语言将有关数据库操纵的语句嵌入到各种程序设计语言编写的程序中去, 组合成一个完整的程序。DML 语句负责数据库的访问, 宿主语言负责数据的处理。根据 DML 语句与宿主接口的不同, 嵌入式语言又可分为三种实现方法: ①显式过程调用接口 (Explicit Procedure Calls Interface) 即宿主语言编写的程序中用过程调用的方式, 调用 DBMS 提供的各种标准例程访问数据库, 这种接口, DBMS 不需要提供语言翻译程序。②隐式过程调用接口 (Implicit Procedure Calls Interface) 是指以某种约定 (比如加前缀), 在编程中直接写出数据操纵语句。因此, DBMS 必须提供相应的翻译程序。通常的办法是预编译方式。即对应用程序进行编译之前, 先利用 DBMS 提供的预编译程序将所有带前缀的语句翻译成宿主语言对数据库管理系统标准例程的过程调用, 然后再进行正常的编译和连接。③原语法接口 (Native Syntax Interface), 即将数据库操作的功能合并到主语言中去, 成为该语言的一部分, 这样 DBMS 就必须提供专门的宿主语言编译程序, 以完成宿主语言和 DML 语言的共同编译任务。

3) 数据库管理例程序

数据库管理例程序用于完成日常的数据库操作和管理, 它由如下几部分组成:

(1) 系统控制程序 (也称为 DBMS 核心程序) 是一个当数据库系统一启动就常驻内存的系统总控程序, 它要完成控制和协调其它各个程序活动的任务。例如, 用户输入一条 DML 命令, 系统控制程序会调用语言翻译程序对用户的命令进行检查, 并转换为相应的执行代码, 然后调用并执行, 如果有对数据库的访问, 再由总控程序调用相应的例程完成对数据库的访问。总之, 系统控制程序也就是一个监控程序, 是数据库管理系统的核心;

(2) 存取控制程序。用于检查用户是否合法, 防止非法用户对数据库的访问, 保证数据库的安全性;

(3) 并发控制程序。当多个用户同时访问数据库中同一数据时, 协调各用户程序对数据库的操作, 完成并发事务的调度, 保证数据一致性;

(4) 完整性控制程序。对数据库进行更新操作时, 检查其是否满足完整性约束条件;

(5) 数据的存取和更新程序。实现对数据库的实际操作。如读、写、删除和修改等;

(6) 通信控制程序。实现用户与 DBMS 之间的通信；

4) 公用程序

完成数据库日常管理中所需的各个功能。

(1) 装入程序。实现大批量数据的加载,常用于数据库建好后的初始数据装入或异种数据库系统中数据的转换；

(2) 恢复程序。当硬、软件出现故障时,利用转储的数据或工作日志,恢复数据库的正确状态；

(3) 转储程序。将数据库中的全部或部分数据转储到其它的存储介质上,作为数据库的备份,供恢复时使用；

(4) 工作日志程序。记录数据库的每一次更新操作及更新前后的数据,用于恢复时使用；

(5) 监测程序。监测系统的执行时间和空间使用情况,为数据库系统性能的调整或重构提供依据；也可用于追踪用户访问数据库的情况作为数据库有力的保安措施；

(6) 数据库重构程序。当需要时可以对数据重新进行组织,修改其结构等,以满足用户的更高要求或改善其性能。

§ 1.3.4 数据字典

正如一个图书馆需要将图书分类编目,以便于读者借阅和图书的存放一样,数据库中也有类似的一部分数据目录,称之为数据字典(Data Dictionary)。数据字典主要存放有关数据库中数据的各种定义。例如,模式、子模式及存储模式均存放在数据字典之中。数据字典可以看成是数据库的一部分,但通常只有数据库管理员才能对它进行访问。数据库管理系统利用数据字典,保持对数据库中数据的解释。数据字典,对于数据库管理系统来说是很重要的,它包含了数据库中所有数据的描述和定位信息,因而,也有人把数据字典称之为数据库的数据库或元数据库(Meta-database)。

数据字典中也可包含一些附加的信息,以帮助 DBMS 实施对数据库的管理和控制。例如,数据字典中可含有①用户授权的描述,记载数据库的合法用户及可以对数据库进行的访问权限；②应用程序信息,记载每个应用程序的名称以及它们的状态(如:测试的、生产的或已废弃不用的等)、程序与程序、程序与用户、程序与数据之间的调用和访问关系；③帮助数据库管理系统维护数据库的附加信息等等。

§ 1.3.5 数据库管理员

数据库管理员 DBA(Database Administrator)是一个特殊的数据库用户,是设计、建立、管理和维护数据库的一个或一组软件人员。一个数据库系统可以有多个用户,但一般只有一个数据库管理员用户。他要参加数据库的设计,负责整个数据库的管理,协调各用户之间的矛盾,进行数据库的日常维护。因而对数据库管理员的要求相当高,他既要对程序设计语言、操作系统和数据库管理系统相当精通,又要熟悉本部门的业务工作。他的职责是：

1. 定义和存储数据库数据

在数据库设计阶段,数据库管理员要参与系统分析和系统设计,和用户一起决定数据库的概念视图和外部视图,编写相应的模式和子模式,根据用户的要求设计数据库的内部视图和存取策略,编写存储模式,建立数据库并装入初始数据。

2. 监控数据库的运行

当用户需使用数据库时,由 DBA 统一进行安排调度和管理(即开帐号、授权等)。为此数据库管理员要定义数据库的保密方案,同时要监督数据库的运行情况。如要掌握用户访问数据

库的情况, DBMS 执行时间及空间的分配情况、数据变化情况、存取概率等各方面资料。并要在进行统计分析基础上适当对用户或某种操作加以控制, 对系统参数进行调整。

3. 数据库的重构造

当数据库运行一段时间后, 数据库中大量发生数据的插入、删除和修改, 时间长了会影响效率, 这就需要 DBA 分析系统性能, 决定是否进行数据库的再组织。DBA 要负责模式和存储模式的修改, 重新定义模式与子模式之间的映射关系, 由此引起数据库的修改, 去掉数据库中无用的数据, 收集碎片, 整理数据库空间等等工作。数据库的重构造工作有时也可能是由于用户对数据库提出新的要求而引起。

4. 日常维护

数据库在运行过程中, 有可能会遇上硬、软件出现故障。因此, DBA 要负责数据库的后援和恢复。即平时要定期进行数据库的后援转储。它可以是全部数据的转储, 也可以是增量数据转储。总之, DBA 要制定一个后援和恢复的策略, 一旦数据库遭到破坏, 就可恢复数据库中的数据。这对于一些实时系统来说尤为重要。

§ 1.4 数据库系统的特点

数据库技术是由文件系统发展起来的一种新型的数据管理技术, 它为用户提供了更广泛的数据共享, 为应用程序提供了更高的独立性, 进一步减少了数据的冗余, 提供了方便的用户接口, 下面我们从数据库与文件系统的区别的角度, 简单介绍一下数据库系统的主要特点。

1. 复杂的数据结构

文件系统中, 把文件通常看成是等长相同格式记录的集合, 且记录与记录之间无任何联系。这样的一种数据结构形式在实际应用中存在着问题。例如, 一个教师的人事记录按图 1.7 的格式组织:

部 门	姓 名	性 别	年 龄	工 资	职 称	学 历				政治历史			社会关系		
						时 间	学 校	时 间	学 校	时 间	事 件	...	姓 名	性 别	关 系

图 1.7 人事记录格式

由于各个教师的情况各不相同, 学历、政治历史和社会关系这三栏的信息量变化很大, 文件系统中实现这个记录时, 必须按信息量最大的教师来定义记录的长度。这样, 对于那些信息量较少的教师来说, 就浪费了大量的存储空间。而在数据库系统中, 由于可以存储复杂的数据结构, 因此上述教师的人事记录可以再细分为相互间联系的几种不同格式的记录, 如图 1.8 所示。这

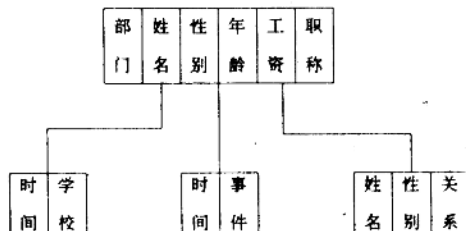


图 1.8 数据库中人事记录格式

样, 虽然各教师的信息量不同, 按照图 1.8 的结构来存储数据时, 有多少数据, 存储多少相应的记录, 减少了空间的浪费。数据库中允许多种不同的记录格式存在, 且记录之间可以有联系, 增加了人们组织数据的灵活性, 提高了描述现实世界的能力, 这是数据库与文件系统的根本区别。