



湖南大学现代远程教育
全国税务系统远程教育 系列教材

数据库

CHINESE FISCAL AND TAXATION LONG-DISTANCE LEARNING

郑若忠 宁洪 主编



湖南电子音像出版社

数 据 库

郑若忠 宁 洪 主编

TP311.13/43

湖南电子音像出版社



湖南大学现代远程教育
全国税务系统远程教育

系列教材

数 据 库

郑若忠 宁 洪 主编

策划:谭慧渊 刘镜波 蒋菊香

责任编辑:杨许国 肖家红

装帧设计:黄 弋 赵 慧

湖南出版集团

湖南电子音像出版社出版发行

长沙市展览馆路66号 邮编:410005

长沙青山印刷厂印刷

2002年2月第1版 2002年2月第1次印刷

开本:850×1168 1/32 印张:15

字数:370千字 印数:6000册

ISBN-7-900352-06-6/G4·98

定价:98.00元/套(光盘配书)

编写说明

现代远程教育是 20 世纪 80 年代以来国际教育发展的共同趋势。1998 年 9 月,教育部批准湖南大学等四所大学首批试办现代远程教育,标志着我国现代远程教育已正式启动。湖南大学的现代远程教育,在探索中不断前进,特别是与国家税务总局合作开办的主要面向行业的财税远程教育,在办学模式、教学手段等方面正在实现跨越式发展。

在全国税务系统远程学教育领导小组的领导下和在全国税务系统远程学历教育教学指导委员会的指导下,我们根据湖南大学本科学历教育教学大纲和新形势下社会对财经类人才素质的要求,组织全国相关专业的著名教授、学者、专家编写了这套系列教材及学习指导书,并配有电子光盘、VCD 光盘、网络课件等教学资源。

本书由郑若忠、宁洪、郑倩冰、王永文、周婧编写。

由于时间原因,错漏之处在所难免,敬请同行专家批评指正。

目 录

1. 数据库系统概论	(1)
1.1. 数据库	(1)
1.2. 现实世界的数据库描述	(10)
1.3. 数据库的结构	(28)
1.4. 数据库管理系统	(34)
2. 数据的物理组织	(48)
2.1. 数据的物理存储	(48)
2.2. 文件的基本组织方法	(53)
2.3. 文件的索引结构	(83)
2.4. * B 树	(93)
2.5. 多关键字检索	(102)
2.6. * SQL Server 的存储结构	(109)
3. 关系数据库	(115)
3.1. 关系模型	(115)
3.2. 关系运算	(131)
3.3. 关系数据库语言	(161)
3.4. 几种常用的关系数据库	(195)
4. 查询优化	(208)
4.1. 查询处理	(208)
4.2. 查询处理的访问例程	(210)
4.3. 查询优化	(219)
4.4. 数据库性能测试	(233)
5. 事务管理	(237)

5.1. 事务管理的引入	(237)
5.2. 事务及其形式描述	(239)
5.3. 事务的特性	(243)
5.4. 事务模型	(246)
5.5. * 并发控制	(251)
5.6. * 恢复机制	(266)
5.7. * SQL Server 中的事务管理	(270)
6. 数据库保护	(277)
6.1. 数据库保护的基本内容	(277)
6.2. 完整约束	(280)
6.3. 安全保护	(284)
7. 关系模型设计理论	(314)
7.1. 函数相关性	(314)
7.2. * 多值相关性	(328)
7.3. * 连接相关性	(332)
7.4. 关系模型的规范理论	(333)
8. 数据库应用系统设计	(358)
8.1. 数据库应用系统设计概述	(358)
8.2. 系统需求分析	(361)
8.3. 概念模型设计	(367)
8.4. 逻辑模型设计	(370)
8.5. * 物理模型设计	(377)
8.6. * 数据库应用程序的开发	(379)
9. 数据库发展前沿	(386)
9.1. 面向对象数据库	(387)
9.2. 多媒体数据库	(403)
9.3. * 基于逻辑的数据模型	(410)
9.4. 分布数据库	(415)

9.5. * 并行数据库	(434)
9.6. 数据仓库	(444)
9.7. 数据库与 Web 的连接	(452)
9.8. * 其它数据库	(460)

1. 数据库系统概论

20 世纪 60 年代末, 数据库技术初露头角, 随即得到迅速发展, 成为数据处理的公用支撑技术。时至今日, 数据库系统已遍布政府机关、社会团体和各个行业, 存储着它们赖以正常运转的数据资源, 显著地提高了工作的效率和质量, 产生出极大的社会、经济效益。数据库技术的发展一方面受到应用需求的牵引和驱动, 另一方面又促进和拓广了计算机的应用领域, 同时也改变了人们运用信息的方式, 有力地推动了社会信息化的进程。本章简要介绍数据库技术的发展历史及有关概念。

1.1 数据库

1.1.1. 数据管理技术的发展

20 世纪后 40 年, 是科学技术和人类社会发展突飞猛进的时期, 新技术、新概念层出不穷, 知识、信息、数据“爆炸”, 使人类社会进入了知识经济、信息社会的时代。它的特征集中表现在信息技术对人类社会的冲击上。信息技术正以它神奇的速度, 从根本上改变着人们的社会生活方式。数字计算模式正在全球逐渐取代有形计算; 传统企业正在发现他们最主要的财富不再是他们生产的产品或

提供的服务，而是他们运用信息的方式。

数据库技术正是在满足这种日益增长的数据、信息处理需求的过程中，由文件系统演变而来，与计算技术和新应用需求同步发展的数据管理技术。

60 年代早期的文件系统是数据管理方法的雏形，其文件在外存的物理结构与用户观点的逻辑结构完全一致，用户的数据文件主要存储在磁带上，其组织方式是顺序式的，这时的数据管理软件属于操作系统的一部分，其主要功能是完成 I/O 设备的输入/输出操作。显然，这种组织形式只能应付批处理，不适应于实时访问。由于让用户各自建立文件，用户不但要编制应用程序，还要花费很大精力为数据文件设计物理细节，而且文件不宜共享，数据冗余度高，当文件的物理结构发生变化或更换外存设备时，须修改或重编应用程序，用户负担沉重，人们不得不探索新的数据管理方法。

大约到 60 年代中后期，直接访问设备——磁鼓、磁盘的性能改善使数据组织发生了如下变化：除磁带外，磁鼓、尤其是磁盘成为联机的主要外存设备；文件的物理结构与逻辑结构之间已有所区别，在文件的物理结构中增加了链接和索引等形式，因而对文件中的记录可顺序地和随机地访问；数据管理软件（仍属操作系统的一部分）提供从逻辑文件到物理文件的“访问方法”是这一时期数据管理的主要特征。对系统而言，组成文件的最小单位是记录，记录之间仅存在顺序关系而无其它结构；在管理软件中还增加了安全、保密检查机构；部分系统允许用户之间以文件为单位共享数据，但未能实现以记录和数据项为单位的数据共享；用户仍以文件标识（文件名）与系统交往，也允许以文件中的记录标识访问数据。它不但适应于批处理，也可用于实时联机任务；系统更换外设也无需用户修改应用程序；可以实现以文件为单位的数据共享等。但这种文件系统仍不够理想，未能体现用户观点下的数据逻辑结构较大地独立于数据在外存的物理结构。因此，数据物理存储的改变，仍然

需要修改用户的应用程序。再者，以文件而不以记录或数据项为单位共享数据，必然导致数据的大量冗余，用户也不能以记录或数据项为单位访问数据。同时，文件系统亦难于增删新旧数据以适应新的应用要求。这些亟待解决的问题，促使人们研究一种新的数据管理技术。

60年代后期出现的三件大事，标志着数据管理技术进入数据库时代。

第一件事是美国 IBM 公司 1968 年研制成功、1969 年正式推出的商品化的层次数据库管理系统 IMS (Information Management System)。

第二件事是美国 CODASYL 组织 (Conference On Data Systems Language) 1968 年公布的 DBTG (Database Task Group) 报告，提出网络数据库系统。

第三件事是自 1970 年起，IBM 公司 E.F.Codd 发表了一系列论文，奠定了关系数据库的理论基础。

初期的数据库方法通常具有如下特点：1) 对用户观点的数据进行严格细致的描述，使得文件的记录、数据项等数据单位之间的联系清晰，结构简单；2) 允许用户以记录或数据项作单位进行访问，也允许许多关键字检索和文件之间的交叉访问；3) 数据的物理存储可以很复杂，同样的物理数据可以导出多个不同的逻辑文件，用户以简单的逻辑结构操作数据而无需考虑数据的存储情况，改动数据的物理位置和存储结构、不必修改或重写应用程序，用户逻辑数据与它们的物理存储之间的转换由数据管理软件完成，从而解决了数据的应用独立于数据的存储的问题。

但在初期数据库方法中，据的整体逻辑结构仅是用户逻辑文件的简单并集，在用户越来越多，系统为每个用户提供的逻辑文件日渐庞杂的情况下，数据库的组织也就越来越乱。为了提高效率，减少冗余，增加新的数据，常常需要改变数据的整体逻辑结构，这就

必然导致用户逻辑文件的修改,进而导致用户应用程序的修改。特别是对某些系统来说,改变整体逻辑结构已成为系统活动方式,这样就提出了用户的数据逻辑结构尽量不受整体逻辑结构变化的影响问题,促使人们把用户观点的逻辑结构从整体逻辑结构中独立出来,形成今天的数据库方法。

数据库方法的最大特点就是三级结构和两级数据独立性,即在用户数据逻辑结构与数据的物理存储结构之间加入数据的整体逻辑结构,使得数据的物理存储结构的变化尽量不影响数据的整体逻辑结构或用户的应用程序;数据整体逻辑结构的改变也尽量不影响用户应用程序。采用此方法构造的系统的结构如图1.1所示。

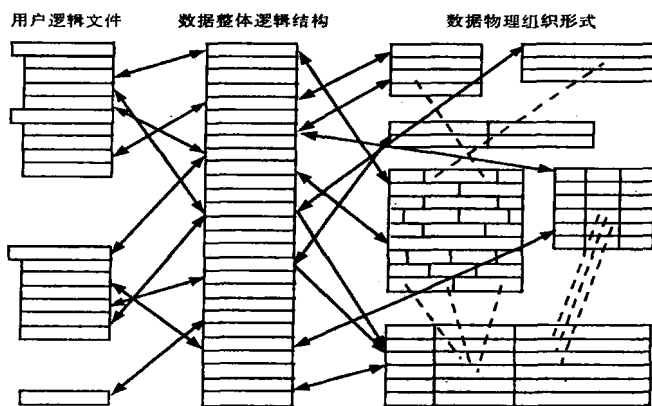


图 1.1 今天的数据库方法

目前全世界已有成千上万个数据库系统在运行,它们几乎触及人类社会生活的所有方面,无论是企事业内部的信息管理,还是各行业(银行、交通、邮电、税务、公安等等)的业务处理,以及一般的信息加工和情报检索无不以数据库技术为基础。近年来随着分布处理、高速网络、多媒体、数据压缩等一批高新技术的发展,数据库技术的应用更加普遍深入,分布式数据库系统也得到了迅速的

发展,目前几乎所有主要的数据库管理系统都提供了分布数据管理的功能。特别是随着 Internet/Intranet 如火如荼地发展,成百上千的站点都以数据库服务器作为低层支持,使人们能更方便有效地共享各类信息。与此同时,数据库技术也遇到了新应用领域的挑战,风靡整个 80 年代的关系数据库技术由于其数据结构平板、数据类型和操作简单、固定,只处理短事务等弊病,不再能有效地支持像计算机辅助设计与辅助加工、计算机辅助软件工程、图像处理、超文本、多媒体等一类新的应用,人们呼唤新一代数据库技术。美国高级 DBMS 功能委员会在 1990 年曾提出一个“第三代数据库系统宣言”。目前较普遍的看法是新一代数据库技术应支持面向对象的数据模型,其体系结构应能适应功能扩展,能处理复杂的数据类型和长寿事务,同时能与 30 多年来陆续开发的各类数据库共存。为此,自 20 世纪 80 年代中后期以来,数据库工作者在如下两方面开展了卓有成效的工作:一是改造和扩充关系数据库,以适应新的应用需求;二是改用新的数据模型(例如面向对象数据模型、基于逻辑数据模型等)构造数据库,人们期待着一个“后关系数据库”(Postrelational Database)时代的到来。此外,新一代数据库技术的另一显著特征是与其它学科的发展高度结合,例如数据库技术与分布处理技术结合导出分布数据库,与并行处理技术结合导出并行数据库,与人工智能技术结合导出演绎数据库、知识库和主动数据库,与多媒体技术结合导出多媒体数据库等等。我们将在第 9 章介绍这些内容。

值得一提的是,存储技术的发展为数据库提供了良好的硬件基础,极大促进了数据库技术的进步。1956 年第一台磁盘机(RAMAC 磁盘)问世,单片磁盘容量仅为 5M 字节,到 1978 年扩大到 6×10^8 字节,1997 年达到 2G 字节,目前则已经达几十 G 字节。近年迅速发展起来的光盘技术,因其容量大、不易磨损、便于保存等诸多特点,很快成为数据库的存储介质,目前许多文献资料数据

库（例如大百科全书等）已存储在光盘上，在市面流通。

1.1.2. 数据库系统

为了更快地理解数据库方法，不妨将其与图书馆做一对比。

图书馆是负责存储和借阅图书的部门；而数据库系统则是存储数据并负责用户访问数据的机构。正像图书馆不能简单地与书库等同起来一样，我们也不能把数据库系统仅仅理解为存储数据的仓库。就图书馆而言，如果把书籍胡乱地堆放在书库中，几乎无法从数以百万册计的浩瀚书海中查找出读者要借的一本书来。因此，没有一套完整的书卡作为图书馆藏书的模型，查借书籍就很困难，管理员也很难掌握藏书全貌。如果不知道书卡与书架的对应关系，管理员也难以按借书单找到该书的存放位置。一个图书馆要想很好地为读者服务，必须完成以下工作：

（1）建立完善的书卡。书卡的内容和格式常包括：书号、分类号、书名、作者名、出版社名、出版时间、内容摘要和其他细节。有时为了方便读者，也按不同分类编排书卡，如以书名、作者名或其它为索引进行编排。

（2）图书应有组织地存放在书库中。图书馆藏书数量很大，书库中房间、书架很多，需要按照一定的顺序和规则（物理结构）分放图书，并列出各类书籍存放的对应关系表，使管理人员能按此表快速查找。

（3）规定借阅权限，并建立周密的借阅管理制度。不同类型读者的借阅权限不同。如善本书只供特定的人借阅；机密图书只供有特权的人借阅；某些书只供读者在馆内翻阅等。

（4）提供图书及各类文献的借阅服务。

读者借书要先出示借书证，图书管理员验明读者身份和借阅权限后，根据读者填写的借书单（访问请求），按照书籍与书架的对应关系表，到书库中查取所借图书交与读者（响应），并作某些登

记（日志）；还书时管理员要按书号把交还的图书送回原来存放的书架位置上；如果管理员又要立即“响应”其他读者借书，可把还书放到“当日暂存书架”上，待有空闲时，再清理还书并送回到书库中原来位置上。

对数据库系统来说，也要完成类似于上述图书馆的工作。

（1）要建立数据模型（见 1.2.3 节），使用户可以根据数据模型访问数据库中的数据（如检索、插入、删除和修改），而不必关心数据的物理存储位置，就像读者可以按书卡填写借书单，而不用顾及书籍存放在书库的什么位置一样。当然数据模型也要像书卡那样能反映各种数据之间的内在联系。

（2）数据也应有组织地存放在存储设备上，并建立数据模型到物理存储位置的对应表（这种对应称为映射），它使系统能够按照用户的访问请求，找到被访问数据的存储位置。建立数据模型和设计数据的物理存储（组织）方法，其目的是使用户对数据的应用与数据的存放位置和存储结构无关，后者的变动不影响前者（正像改变图书的存放位置不影响读者按书卡借书一样），这点也称为数据独立性（见 1.3.3 节）。

（3）数据库系统要为不同用户确定不同的访问权限并进行访问控制。

（4）提供以数据库为基础的各种应用服务。

上述各种功能都是在数据库管理系统（DBMS）的统一管理和控制下实现的。正像一个大型公共图书馆需要有专门的工作人员负责规划、设计、协调、维护和管理一样。数据库系统一般设有数据库管理员（DBA），负责协调和监视对数据库的使用，一旦发现违反安全保密或性能变坏的现象，立即采取相应对策。因此，一个数据库系统，由数据库、数据库管理系统、应用程序和数据库管理员（DBA）4 个部分组成，如图 1.2 所示：

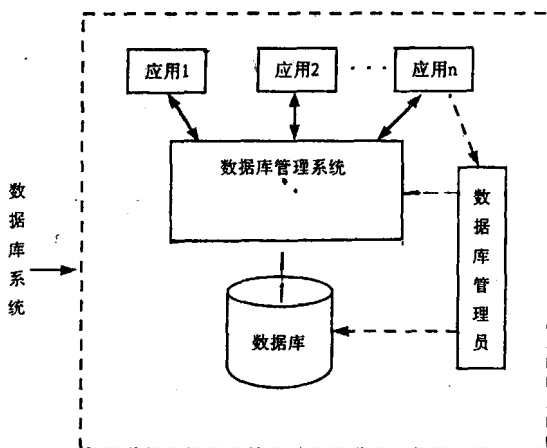


图 1.2 数据库系统环境

(1) 数据库：它是大量的、相互关联的数据的集合。这些数据一般存储在磁盘、光盘等外存设备上。有的也存储在内存、快闪存储器等设备上。如企业的人事、工资、营销数据、学校的教职员工数据、学生信息、成绩数据等。

(2) 数据库管理系统 (DBMS)：它是一个大型的软件系统，统一负责各用户对数据库的访问，如查询、插入、删除、修改、统计等，并保障这些访问是高效、正确、安全的。如市场上能买到的商用数据库管理系统，Oracle、DB2、Sybase、Informix、SQL Server 等。

(3) 应用程序：它是完成特定应用功能的程序。这些程序要通过 DBMS 访问数据库中数据。如打印工资报表、工资条的程序，银行转账程序，电话费计算程序等。

(4) 数据库管理员 (DBA)：对于一个大型的数据库来说，它的用户往往很多，数据库管理员的职责是参与数据库的设计与开发，协调用户对数据库的访问。

在不引起混淆的情况下，数据库系统有时也简称数据库。

至此，我们不难理解：数据库系统是实现有组织地、动态地存储大量关联数据，支持多用户访问的计算机软硬件资源及 DBA 组成的系统。

数据库系统与文件系统相比，显示出下列优越性：

(1) 自包含性。一般使用文件系统时，文件的结构（即数据的组织方式）需在程序中说明，而在数据库中总是有数据模型，它负责描述和说明数据，也就是数据以及描述数据的数据共同构成数据库。

(2) 更强的数据独立性。当数据作某些修改时，应用程序可以不变动。

(3) 更强的数据抽象能力。用户只需关心概念表示，即库中表示的对象及其对象之间的关系，而有关存储访问的细节隐藏在数据字典（见 1.4.4 节）中，由 DBMS 负责从逻辑描述到物理实现的映射。

(4) 支持多视图。不同的用户根据各自的需要认识和使用数据库的一部分，而所有用户视图的综合构成数据库的整体逻辑描述。

(5) 提供对完整性、安全性、并发和恢复四方面的数据控制能力。

(6) 操作既可以以记录也可以以数据项为单位。

(7) 为用户提供应用程序的开发环境和开发工具。如各种 MIS 系统的开发工具。

1.2. 现实世界的的数据描述

1.2.1. 信息的 3 个领域

在图 1.3 所示一般信息控制系统中，信息从客观事物出发，流经数据库，通过控制决策机构，最后又回到客观事物。信息的这一循环经历了 3 个领域：现实世界、观念世界和数据世界。

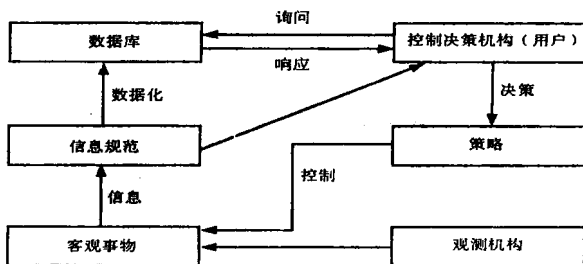


图 1.3 信息控制系统

(1) 现实世界。是存在于人们头脑之外的客观世界，事物及其相互联系就处在这个世界之中。

(2) 观念世界。是现实世界在人们头脑中的反映。客观事物在观念世界中称为实体，反映事物联系的是实体模型。

(3) 数据世界。数据是观念世界中信息的数据化，现实世界中的事物及联系在这里用数据模型描述。

3 个领域的内容及其联系可用图 1.4 所示。

因此，客观事物是信息之源，是设计数据库的出发点，也是使用数据库的最终归宿。实体模型与数据模型是对客观事物及其联系的两级抽象描述。数据库的核心问题是数据模型，数据模型由实体模型导出。为了得到正确的模型，首先要充分了解客观事物。