

面向 21 世纪课程教材
Textbook Series for 21st Century

数据库系统教程

(第 2 版)

施伯乐 丁宝康 汪 卫 编著



高等教育出版社

内 容 提 要

本书第 1 版列为教育部“面向 21 世纪课程教材”。这次再版时作了较大的修改和补充,增加了关系逻辑、递归 SQL、对象联系图、UML 的类图、ODMG 标准、ODBC、JDBC 和 XML 等内容。本书详细介绍数据库系统的基本原理、方法和发展趋势。内容包括:数据库系统的基本概念、关系模型的运算理论、SQL 语言、规范化设计理论、数据库设计与 ER 模型、存储结构、系统实现技术、对象数据库、分布式数据库、中间件技术、数据库与 WWW、XML 技术。

本书第 1 版在 2002 年获教育部优秀教材二等奖。本书是作者长期从事数据库教学和科研活动所获得的结晶。全书内容丰富,叙述严谨清楚,每章后均有适量的习题,适于教和学。与本书配套的《数据库系统教程习题解答及上机指导》已经出版。本书的教学课件(Power Point)也已上网,需要者可到计算机教学资源网 <http://cs.hep.com.cn> 处下载。

本书可作为高等院校计算机有关专业的数据库课程教材,也可供从事信息领域工作的科技人员和工程技术人员以及其他有关人员参阅。

第 2 版前言

《数据库系统教程》第 1 版自 1999 年 12 月出版以来,在计算机界同仁和学生那里得到了很大的反响,并于 2002 年获教育部优秀教材二等奖。

进入 21 世纪以后,无论是市场的需求还是技术条件的成熟,对象数据库技术和网络数据库技术的推广和普及已成定局,这些技术将成为新时代数据库技术应用的主流技术。为此,我们对第 1 版的结构作了如下的改动:

- 在第 2 章关系运算理论中增加了关系逻辑内容,在第 3 章的 SQL 语言中增加了递归查询内容,并把 QBE 内容缩写成对 SELECT 语句的图示解释。

- 把数据库设计从第 4 章中抽出来,单独成为新的第 5 章,并着重介绍 ER 模型,并列举了四个较大的 ER 模型实例。原来的第 5、6 章就顺延为新的第 6、7 章。在第 7 章中,对系统实现技术进行了改写,读起来更为通顺。

- 原来的第 8、9 章内容改写为新的第 8 章对象数据库系统,并增加了 ODMG2.0 标准内容,以及使用 UML 类图概念对象建模的内容。同时删去了原来的第 10 章。

- 对原来的第 7 章分布式数据库内容,改写成新的第 9 章。

- 增加了三章新的内容:第 10 章中间件,第 11 章数据库与 WWW,第 12 章 XML 技术。

本书的主要目的是使读者掌握数据库技术的基本原理、方法和应用技术,能有效地使用现有的数据库管理系统,掌握数据库结构的设计和数据库应用系统的开发方式,同时也适时地掌握对象数据库和网络数据库等技术。

全书共分 12 章,具体内容如下:

第 1 章介绍数据库技术的基本概念,包括数据模型、数据库的体系结构和数据库系统的全局结构。

第 2 章介绍关系模型和关系运算理论,包括关系代数、关系演算和关系逻辑。

第 3 章介绍关系数据库标准语言 SQL 的数据定义、操纵、嵌入式等内容。

第 4 章介绍关系数据库的规范化设计理论,包括函数依赖、模式分解特性、范式和模式设计方法等问题。

第 5 章介绍数据库设计的全过程、ER 模型以及 ER 模型到关系模型的转换规则。

第 6 章介绍数据库的存储结构,包括存储介质层次、文件结构、索引技术和散

列技术等内容。

第7章绍数据库系统实现的技术,包括事务概念及恢复、并发控制、完整性和安全性等实现机制。

第8章介绍对象数据库系统的两种形式:ORDB和OODB。介绍其定义语言和操纵语言,还介绍了概念对象建模的两种图示形式:对象联系图和UML的类图。

第9章介绍分布计算的两种形式:客户机/服务器系统和分布式数据库系统。

第10章介绍数据库技术的两种中间件技术:ODBC技术和JDBC技术。

第11章介绍数据库与WWW的关系和结合,以及JSP和PHP等语言。

第12章介绍XML技术的基本概念和XML查询语言。

本书对某些较深的内容,在章节前加了符号“*”;在每一道例题或定理的结尾处,加了符号“□”,以示与正文的区别。

全书内容丰富,具有理论性、实用性和可操作性。每章后均有适量的习题,以配合对知识点的掌握。教师讲授时应根据需要对内容作适当取舍。

本书可采取如下的教学方法:

(1) 讲授的重点应是第1~5、7章。对于第2章和第4章中理论性较强的内容(目录中加*的节),可作适当压缩。第2章的重点是关系代数,第4章的重点是函数依赖和范式。

(2) 对于第6章的存储结构内容,教师可作适当引导让学生自学,以拓宽知识面。

(3) 对于第8章的对象数据库内容,教师可作适当引导让学生自学,以提高学生的工作能力和研究水平。

(4) 对于第9~12章,教师可有针对性地选择某些内容向学生讲授。

本书是我们数据库教学小组20多年教学和科研成果的结晶,凝聚了几代人的心血。在此向给青年教师做出表率的罗文化、楼荣生等老教师致以崇高的敬意。本次修订工作第5章由许建军执笔,第10章由周华奇执笔,第11、12章由汪卫执笔,其余章节由丁宝康执笔。参与本书修订工作的还有陈长洪、张谧、吴爱华、庞引明、谢闽峰、吴清茂和曾宇昆等老师。最后由施伯乐、丁宝康和汪卫一起对全稿进行了修改和统稿。

为了满足大家对使用本书作为主教材的教和学的要求,我们还组织了《数据库系统教程习题解答及上机指导》随之一起出版。

限于水平,书中欠妥之处,敬请广大读者和专家批评指正。对本书的意见请按地址 dn@citiz.net 反馈给我们,谢谢。

施伯乐

2003年5月于复旦大学

第 1 版前言

数据库系统产生于 20 世纪 60 年代末。30 年来,数据库技术得到迅速发展,已形成较为完整的理论体系和一大批实用系统,现已成为计算机软件领域的一个重要分支。

在 20 世纪 50 年代,数据管理还处于人工管理阶段,而计算机应用主要是用于科学计算。60 年代,出现了操作系统,从而使数据管理进入文件系统阶段。随着计算机应用的普及,计算机逐步从科学计算转向企业管理。为了克服文件系统的数据库冗余弱点,方便用户操作和提高程序开发的生产率,在 60 年代末产生了数据库系统。数据库系统产生之后,显示了其强大的生命力。70 年代,层次、网状数据库系统迅速研制成功,并在商业上得到广泛应用。当时,关系数据库的研究还集中在理论和实验系统的开发上,直至 80 年代初才形成产品。由于关系数据库有较好的理论基础,并具有操作方便等优点,因此关系数据库的商用系统迅速占领市场,并逐步取代了层次和网状数据库系统。1987 年 ISO 组织研究并颁布关系数据库语言 SQL 标准。90 年代数据库技术进一步发展,推出许多新型数据库系统,以适应用户提出的新需求,并进而渗透到多媒体、人工智能、网络等领域。

随着数据库系统的推广使用,计算机应用已深入到工农业生产、商业、金融、行政管理、科学研究和工程技术的各个领域;当今的管理信息系统(MIS)、办公自动化(OA)、计算机辅助设计与制造(CAD/CAM)、计算机集成制造系统(CIMS)、地理信息系统(GIS)和知识库系统等,也都以数据库技术为基础。90 年代初,我国已在邮电、银行、电力、铁路、气象、民航、情报、公安、军事、航天、财税等行业装备了以数据库为基础的大型计算机系统。

在人类迈向 21 世纪知识经济时代的时候,信息变为经济发展的战略资源,信息技术已成为社会生产力中重要的组成部分。人们充分认识到,数据库是信息化社会中信息资源管理与开发利用的基础。对于一个国家来说,数据库的建设规模和使用水平已成为衡量该国信息化程度的重要标志。因此,数据库课程是计算机领域中一门重要课程。

本书较为详细地介绍了数据库系统的基本概念、原理、方法及应用技术。全书共分 10 章。

第一章介绍数据库系统的基本概念,包括数据模型、数据库的体系结构、数据库系统的全局结构。

第二章介绍关系模型的运算理论:关系代数和关系演算。

第三章介绍关系数据库标准语言 SQL 和另一个对数据库发展有重大作用的 QBE 语言。

第四章介绍关系数据库的规范化设计理论以及数据库设计的全过程。

第五章介绍数据库的存储结构,包括存储介质、文件组织、索引和散列技术。

第六章介绍数据库系统的实现技术,包括事务的概念及并发控制、恢复、完整性和安全性实现机制。

第七章介绍分布式数据库的概念、结构和方法。

第八章介绍基于面向对象程序设计思想的面向对象数据模型的概念以及持久化 C++ 系统。

第九章介绍基于关系模型的对象关系模型的概念和语言。

第十章回顾数据库技术的发展历史、综述现状、介绍新的应用、分析面临的挑战并提出新的研究方向。

全书每章后面均配有适量的习题,以加强对数据库系统原理、方法的理解和掌握。

本书是计算机有关专业的教材。讲授时应根据需要对内容作适当取舍。建议如下:

① 重点讲授第一章至第四章及第六章内容。其中对于第二章中关系演算和第四章中理论性较强的内容可根据情况适当压缩。

② 对于第五章的存储结构内容,教师可作适当引导并让学生自学,以提高学生的工作能力和研究水平,拓宽知识面。

③ 对于第七章以及后面涉及到数据库系统的进一步发展的内容,教师可有针对性地选择某些内容,向学生传授。

本书由施伯乐负责内容的取材、组织和审定。第一章至第六章由丁宝康执笔,第七章由田增平执笔,第八章至第十章由周傲英执笔。

复旦大学楼荣生教授仔细校阅了全稿,在此向他表示感谢。

上海交通大学尤晋元教授审阅了全稿,并提出许多宝贵的意见,在此向他致以衷心的感谢。

限于水平,书中难免有欠妥之处,欢迎广大读者和专家批评指正。

作 者

1999 年 4 月于复旦大学

第 1 章 数据库概论

从 20 世纪 50 年代中期开始 ,计算机的应用由科学研究逐渐扩展到企业、行政等社会各领域 ,数据处理已成为计算机的主要应用。在 60 年代末 ,数据库技术就是作为数据处理中的一门新技术发展起来的。它是计算机软件领域的一个重要分支 ,经过 30 多年的发展 ,形成了较为完整的理论体系和实用技术。本章先回顾数据库管理技术的发展过程 ,然后介绍数据库中的基本概念 ,以使读者对数据库的概貌有所了解。

1.1 引言

当前人类社会正处于信息社会 ,人类知识以惊人的速度增长。如何组织和利用这样庞大的知识成为急需解决的问题。在 20 世纪 60 年代的“软件危机”中 ,数据库技术作为软件学科的一个分支应运而生了。

现在“数据库”这个名词已是家喻户晓。据查 ,数据库(DataBase)这个名词起源于 20 世纪 50 年代初 ,当时美国为了战争的需要 ,把各种情报集中在一起 ,存储在计算机里 ,称为 Information Base 或 DataBase。

1963 年 ,美国 Honeywell 公司的 IDS(Integrated Data Store)系统投入运行 ,揭开了数据库技术的序幕。1965 年 ,美国一家火箭公司利用该系统帮助设计了阿波罗登月火箭 ,推动了数据库技术的产生。当时美国社会上产生了许多形形色色的 DataBase 或 DataBank ,但是 ,它们基本上都是文件系统的扩充或是倒排文件系统。1968 年 ,美国 IBM 公司推出层次模型的 IMS 数据库系统(1969 年形成产品);1969 年 ,美国 CODASYL(Conference On Data System Language 数据库系统语言协会)组织的数据库任务组(DBTG)发表关于网状模型的 DBTG 报告(1971 年正式通过);1970 年 ,IBM 公司的 E. F. Codd 发表论文提出关系模型。这三件事奠定了现代数据库技

术的基础。

20世纪70年代是数据库蓬勃发展的年代,网状系统和层次系统占据了整个数据库商用市场,而关系系统仅处于实验阶段。许多计算机厂商研制了各种数据库管理系统,许多学者发表了大量研究性论文,使数据库技术在实践中和理论上得到飞速发展。数据库技术也日趋成熟。

20世纪80年代,随着计算机硬件性能的改善,关系系统由于使用简便,逐步代替网状系统和层次系统占领了市场。同时关系数据库的理论日趋完善,走向更高级的阶段。数据库的理论以及新系统的研究继续进行,不断深化。此时关系数据库语言 SQL 成了国际标准语言。

20世纪90年代,关系数据库已成为数据库技术的主流。由于受到计算机应用领域以及其他分支学科的影响,数据库技术与面向对象技术、网络技术相互渗透,人们在对象数据库技术和网络数据库技术方面进行了深入的研究。

由于 C. W. Bachman 在网状模型和 DBTG 报告中的贡献,他在 1973 年荣获美国计算机学会(ACM)图灵奖。E. F. Codd 在关系模型上作出了杰出的开拓性贡献,在 1981 年获得了图灵奖。J. Gray 在数据库技术的实用化和事务管理方面发挥了关键的作用,因而在 1998 年也获得了图灵奖。

进入 21 世纪以后,无论是市场的需求还是技术条件的成熟,对象数据库技术、网络数据库技术的推广和普及已成定局。数据库技术使得计算机应用迅速渗透到工农业生产、商业、行政、科学研究、工程技术和国防军事各个部门,渗透到社会的每一角落,并改变着人们的工作方式和生活方式。管理信息系统(MIS)、办公信息系统(OIS)、计算机集成制造系统(CIMS)、地理信息系统(GIS)、因特网(Internet)技术等都是使用了数据库技术的计算机应用系统。

本书第 2~7 章介绍传统的数据库技术,内容包括关系运算理论、SQL 语言、规范化设计理论、ER 模型、存储结构和系统实现技术。第 8 章介绍对象数据库技术的两种形式:对象关系数据库和面向对象数据库。第 9 章介绍分布式数据库系统。第 10 章介绍中间件技术。第 11、12 章介绍数据库与 WWW 及 XML 技术。

1.1.1 数据和信息

数据,是数据库系统研究和处理的对象。数据与信息是分不开的,它们既有联系又有区别。

信息是许多学科广泛使用的概念。在不同的领域中,其含义有所不同。但人们普遍认为信息与能源、材料科学并列,构成现代社会的三大支柱。

在数据处理领域,一般把信息理解为关于现实世界事物存在方式或运动状态

的反映。例如,我们上课用的黑板,颜色是黑的,形状是矩形,尺寸是长 3 m、高 1.2 m,材料是木材,这些都是关于黑板的信息,是黑板存在状态的反映。

信息有许多重要的特征:信息来源于物质和能量;信息是可以感知的;信息是可以存储的;信息是可以加工、传递和再生的。这些特点,构成了信息的最重要的自然属性。作为信息的社会属性,信息已经成为社会上各行各业不可缺少的重要资源之一。人类获取、积累并利用信息是认识和改造客观世界的必要过程。借助信息,人类才能获得知识,才能有效地组织社会各种活动。因此,信息是人类维持正常活动不可缺少的资源。

几乎和信息同样广泛使用的另一个概念是“数据”。所谓数据,通常指用符号记录下来的、可以识别的信息。例如描述黑板的信息,用一组数据“黑色、矩形、3 m、1.2 m”表示。由于这些符号已被人们赋予了特定的语义,因此它们就具有传递信息的功能。

可以看出,信息与数据之间存在着固有的联系:数据是信息的符号表示或称为载体,信息则是数据的内涵,是对数据语义的解释。数据表示了信息,而信息只有通过数据形式表示出来才能被人们理解和接受。尽管两者在概念上不尽相同,但通常人们并不严格去区分它们。

1.1.2 数据处理与数据管理

数据处理也可称为信息处理,因为当把客观事物表示成数据后,这些数据便被人们赋予了特定的含义,从而为人们提供了不必直接观察和度量事物就可以获得有关信息的手段。数据处理是指从某些已知的数据出发,推导加工出一些新的数据,这些新的数据又表示了新的信息。在具体操作中,涉及到数据库收集、管理、加工利用乃至信息输出的演变与推导全过程。

在数据处理中,通常计算比较简单,而数据的管理比较复杂。数据管理是指数据的收集、整理、组织、存储、维护、检索、传送等操作,这部分操作是数据处理业务的基本环节,而且是任何数据处理业务中必不可少的共有部分。对数据管理部分,应研制出一个通用、高效而又使用方便的管理软件,把数据有效地管理起来,以便最大限度地减轻程序员的负担;至于处理业务中的加工计算,因不同业务各不相同,要靠程序员根据业务情况编写应用程序加以解决。所以,数据处理是与数据管理相联系的,数据管理技术的优劣,将直接影响数据处理的效率。数据库技术正是瞄准这一目标研究、发展并完善起来的专门技术。

1.1.3 数据库技术的基本术语

在数据库应用中,常用到 DB、DBMS、DBS 等术语,形式定义如下。

定义 1.1 数据库(DataBase,简记为 DB)

DB 是长期存储在计算机内、有组织的、统一管理的相关数据的集合。DB 能为各种用户共享,具有较小冗余度、数据间联系紧密而又有较高的数据独立性等特点。

定义 1.2 数据库管理系统(DataBase Management System,简记为 DBMS)

DBMS 是位于用户与操作系统(OS)之间的一层数据管理软件,如图 1.1 所示,它为用户或应用程序提供访问 DB 的方法,包括 DB 的建立、查询、更新及各种数据控制。

DBMS 总是基于某种数据模型,可以分为层次型、网状型、关系型和面向对象型等。

定义 1.3 数据库系统(DataBase System,简记为 DBS)

DBS 是实现有组织地、动态地存储大量关联数据、方便多用户访问的计算机硬件、软件和数据资源组成的系统,即它是采用数据库技术的计算机系统。

定义 1.4 数据库技术

数据库技术是研究数据库的结构、存储、设计、管理和使用的一门软件课程。

数据库技术是在操作系统的文件系统基础上发展起来的,而且 DBMS 本身要在操作系统支持下才能工作。数据库与数据结构之间的联系也很密切,数据库技术不仅要用到数据结构中链表、树、图等知识,而且还丰富了数据结构的内容。应用程序是使用数据库系统最基本的方式,在系统中大量的应用程序都是用高级语言(例如 COBOL、C 等)加上数据库的操纵语言联合编制的。集合论、数理逻辑是关系数据库的理论基础,很多概念、术语、思想都直接用到关系数据库中。因此,数据库技术是一门综合性较强的课程。

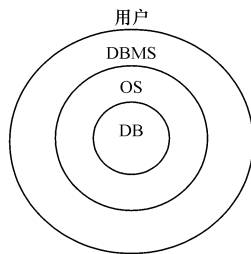


图 1.1 系统层次图

1.2 数据库的由来和发展

计算机的数据处理应用,首先要把大量的信息以数据形式存放在存储器中。存储器的容量、存储速率直接影响到数据管理技术的发展。1956 年生产的第一台

磁盘,其容量仅为 5MB,而现在已达 180 000MB,如图 1.2 所示。目前光盘已经被广泛应用,容量已达数千兆字节。存储器的发展,为数据库技术提供了良好的物质基础。

时间(年)	1956	1965	1971	1978	1981	1985	1995	2003
容量(MB/轴)	5	30	100	600	1 200	5 000	10 000	180 000

图 1.2 磁盘容量的发展

使用计算机以后,数据处理的速度和规模无论相对于手工方式还是机械方式都是无可比拟的,随着数据处理量的增长,产生了数据管理技术。数据管理技术的发展,与计算机硬件(主要是外部存储器)、系统软件及计算机应用的范围有着密切的联系。数据管理技术的发展经历了以下四个阶段:人工管理阶段、文件系统阶段、数据库阶段和高级数据库阶段。

1.2.1 人工管理阶段

在这一阶段(20世纪50年代中期以前),计算机主要用于科学计算,其他工作还没有展开。外部存储器只有磁带、卡片和纸带等,还没有磁盘等直接存取存储设备。软件只有汇编语言,尚无数据管理方面的软件。数据处理的方式基本上是批处理。这个时期的数据管理有以下特点:

(1)数据不保存在计算机内。计算机主要用于计算,一般不需要长期保存数据。在进行某一课题计算时,将原始数据随程序一起输入内存,运算处理后将结果数据输出。随着计算任务的完成,用户作业退出计算机系统,数据空间随着程序空间一起被释放。

(2)没有专用的软件对数据进行管理。每个应用程序都要包括存储结构、存取方法、输入/输出方式等内容。程序中的存取子程序随着存储结构的改变而改变,因而数据与程序不具有独立性。存储结构改变时,应用程序必须改变。此时,由于程序直接面向存储结构,因此数据的逻辑结构与物理结构没有区别。

(3)只有程序(Program)的概念,没有文件(File)的概念。数据的组织方式必须由程序员自行设计与安排。

(4)数据面向程序。即一组数据对应一个程序。

1.2.2 文件系统阶段

在这一阶段(20世纪50年代后期至60年代中期),计算机不仅用于科学计算,

还用于信息管理。随着数据量的增加,数据的存储、检索和维护问题成为紧迫的需要,数据结构和数据管理技术迅速发展起来。此时,外部存储器已有磁盘、磁鼓等直接存取存储设备。软件领域出现了高级语言和操作系统。操作系统中的文件系统是专门管理外存的数据管理软件。数据处理的方式有批处理,也有联机实时处理。

这一阶段的数据管理有以下特点:

(1) 数据以“文件”形式可长期保存在外部存储器的磁盘上。由于计算机的应用转向信息管理,因此对文件要进行大量的查询、修改和插入等操作。

(2) 数据的逻辑结构与物理结构有了区别,但比较简单。程序与数据之间具有“设备独立性”,即程序只需用文件名就可与数据打交道,不必关心数据的物理位置。由操作系统的文件系统提供存取方法(读/写)。

(3) 文件组织已多样化。有索引文件、链接文件和直接存取文件等。但文件之间相互独立、缺乏联系。数据之间的联系要通过程序去构造。

(4) 数据不再属于某个特定的程序,可以重复使用。数据虽然面向应用,但是文件结构的设计仍然基于特定的用途,程序基于特定的物理结构和存取方法,因此程序与数据结构之间的依赖关系并未根本改变。

在文件系统阶段,由于具有设备独立性,因此当改变存储设备时,不必改变应用程序。但这只是初级的数据管理,还未能彻底体现用户观点下的数据逻辑结构独立于数据在外存的物理结构要求。在修改数据的物理结构时,仍然需要修改用户的应用程序。即应用程序具有“程序-数据依赖”性。有关物理表示的知识和访问技术将直接体现在应用程序的代码中。

(5) 对数据的操作以记录为单位。这是由于文件中只存储数据,不存储文件记录的结构描述信息。文件的建立、存取、查询、插入、删除、修改等所有操作,都要用程序来实现。

文件系统阶段是数据管理技术发展中的一个重要阶段。在这一阶段中,得到充分发展的数据结构和算法丰富了计算机科学,为数据管理技术的进一步发展打下了基础。

随着数据管理规模的扩大,数据量急剧增加,文件系统显露出三个缺陷:

(1) 数据冗余(Redundancy)。由于文件之间缺乏联系,造成每个应用程序都有对应的文件,有可能同样的数据在多个文件中重复存储;

(2) 数据不一致(Inconsistency)。这往往是由数据冗余造成的,在进行更新操作时,稍不谨慎,就可能使同样的数据在不同的文件中不一样;

(3) 数据联系弱(Poor Data Relationship)。这是由于文件之间相互独立,缺乏联系造成的。

由于这些原因,促使人们研究新的数据管理技术,因而在60年代末产生了数据库技术。

1.2.3 数据库阶段

随着数据管理规模一再扩大,数据量急剧增长。为了提高效率,人们开始时只是对文件系统加以扩充,研制出倒排文件系统,但这并不能解决问题。恰好在20世纪60年代末,磁盘技术取得重要进展,具有数百兆字节容量和快速存取的磁盘陆续进入市场,成本也不高,这就为数据库技术的产生提供了良好的物质条件。

20世纪60年代中期出现的系统(database或databank)还不能真正地称为数据库系统。数据管理技术进入数据库阶段的标志是20世纪60年代末的三件大事:

- (1) 1968年美国IBM公司推出层次模型的IMS系统;
- (2) 1969年美国CODASYL组织发布了DBTG报告。总结了当时各式各样的数据库,提出网状模型;
- (3) 1970年美国IBM公司的E.F.Codd连续发表论文,提出关系模型,奠定了关系数据库的理论基础。

20世纪70年代以来,数据库技术得到迅速发展,开发出了许多产品,并投入运行。数据库系统克服了文件系统的缺陷,提供了对数据更高级、更有效的管理。概括起来,数据库阶段的数据管理具有以下特点:

(1) 采用数据模型表示复杂的数据结构。数据模型不仅描述数据本身的特征,还要描述数据之间的联系。这种联系通过存取路径实现。通过所有存取路径表示自然的数据联系是数据库与传统文件的根本区别。这样,数据不再面向特定的某个或多个应用,而是面向整个应用系统。数据冗余明显减少,实现了数据共享。

(2) 有较高的数据独立性。数据的逻辑结构与物理结构之间的差别可以很大。用户以简单的逻辑结构操作数据而无需考虑数据的物理结构。数据库的结构分成用户的局部逻辑结构、数据库的整体逻辑结构和物理结构三级,如图1.3所示。用户(应用程序或终端用户)的数据和外部存储器中的数据之间的转换由数据库管理系统实现。

数据独立性是指应用程序与数据库的数据结构之间相互独立。在物理结构改变时,尽量不影响整体逻辑结构、用户的逻辑结构以及应用程序,这样我们就认为数据库达到了物理数据独立性。在整体逻辑结构改变时,尽量不影响用户的逻辑结构以及应用程序,这样就认为数据库达到了逻辑数据独立性。

(3) 数据库系统为用户提供了方便的用户接口。用户可以使用查询语言或终端命令操作数据库,也可以用程序方式(如用COBOL、C等高级语言和数据库语言

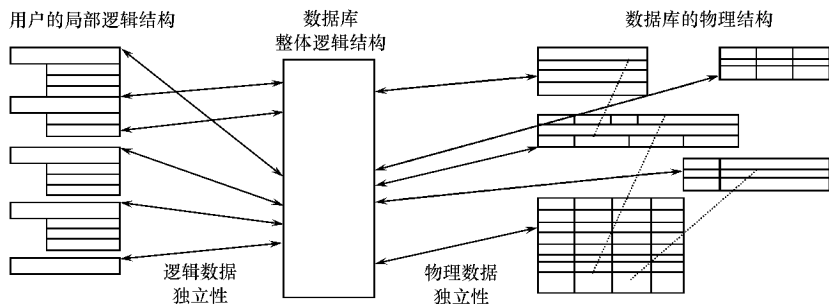


图 1.3 数据库系统的结构

联合编制的程序)操作数据库。

(4) 数据库系统提供以下四方面的数据控制功能：

① 数据库的并发控制：对程序的并发操作加以控制，防止数据库被破坏，杜绝提供给用户不正确的数据。

② 数据库的恢复：在数据库被破坏或数据不可靠时，系统有能力把数据库恢复到最近某时刻的正确状态。

③ 数据的完整性：保证数据库中数据始终是正确的。

④ 数据安全性：保证数据的安全，防止数据丢失或被窃取、破坏。

(5) 增加了系统的灵活性：对数据的操作不一定以记录为单位，可以以数据项为单位。

上述五个方面构成了数据库系统的主要特征。这个阶段的程序和数据的联系可用图 1.4 表示。

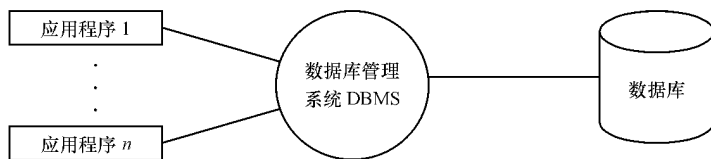


图 1.4 程序和数据间的联系

从文件系统发展到数据库系统是信息处理领域的一个重大变化。在文件系统阶段，信息处理的传统方式如图 1.5(a)所示，人们关注的中心问题是系统功能的设计，因而程序设计处于主导地位，数据只起着服从程序设计需要的作用；而在数据库方式下，信息处理的传统方式已为图 1.5(b)所示的新体系所取代，数据占据

了中心位置。数据结构的设计成为信息系统首先关心的问题,而利用这些数据的应用程序设计则退居到以既定的数据结构为基础的外围地位。

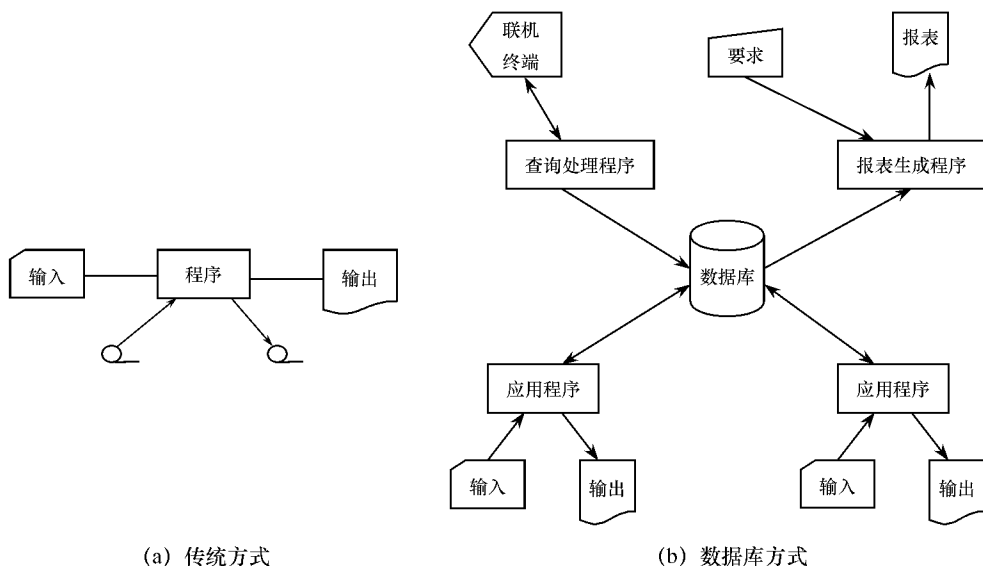


图 1.5 信息处理方式的演变

目前世界上已有大量的大型数据库系统在运行,其应用已深入到人类社会生活的各个领域,从企业管理、银行业务、资源分配、经济预测一直到信息检索、档案管理、普查、统计等,并在通信网络基础上,建立了许多国际性的联机检索系统。20世纪90年代初,我国在全国范围内装备了12个以数据库技术为基础的大型计算机系统,这些系统分布在邮电、银行、电力、铁路、气象、民航、情报、公安、军事、航天和财税等行业,对我国的信息化进程正发挥着巨大的作用。现在几乎各行各业都普遍建立了以数据库为核心的信息系统。

1.2.4 高级数据库技术阶段

这一阶段的主要标志是20世纪80年代的分布式数据库系统、90年代的对象数据库系统和21世纪初的网络数据库系统的出现。

1. 分布式数据库系统

在这一阶段以前的数据库系统是集中式的。在文件系统阶段,数据分散在各

个文件中,文件之间缺乏联系。集中式数据库把数据集中在一个数据库中进行管理,减少了数据冗余和不一致性,而且数据联系比文件系统强得多。但集中式系统也有弱点:一是随着数据量增加,系统相当庞大,操作复杂,开销大;二是数据集中存储,大量的通信都要通过主机,造成拥挤。随着小型计算机和微型计算机的普及,计算机网络软件和远程通信的发展,分布式数据库系统崛起了。

分布式数据库系统主要有以下三个特点:

(1) 数据库的数据物理上分布在各个场地,但逻辑上是一个整体。

(2) 各个场地既可以执行局部应用(访问本地DB),也可以执行全局应用(访问异地DB)。

(3) 各地的计算机由数据通信网络相联系。本地计算机单独不能胜任的处理任务,可以通过通信网络取得其他DB和计算机的支持。

分布式数据库系统兼顾了集中管理和分布处理两个方面,因而有良好的性能,具体结构如图1.6所示。

本书第9章将介绍分布式数据库系统的基本概念。

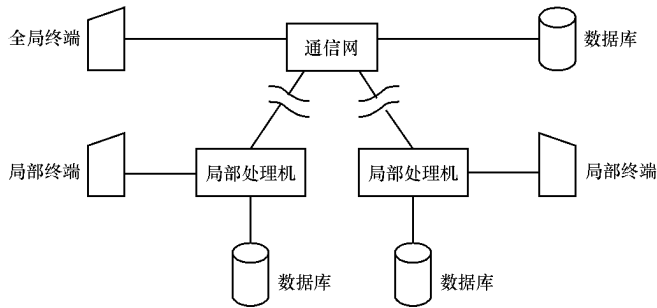


图 1.6 分布式数据库系统

2. 对象数据库系统

在数据处理领域,关系数据库的使用已相当普遍、相当出色。但是现实世界存在着许多具有更复杂数据结构的实际应用领域,已有的层次、网状、关系三种数据模型对这些应用领域都显得力不从心。例如多媒体数据、多维表格数据、CAD数据等应用问题,需要更高级的数据库技术来表达,以便于管理、构造与维护大容量的持久数据,并使它们能与大型复杂程序紧密结合。对象数据库正是适应这种形势发展起来的,它是面向对象的程序设计技术与数据库技术结合的产物。

对象数据库系统主要有以下两个特点:

(1) 对象数据模型能完整地描述现实世界的数据结构,能表达数据间嵌套、递