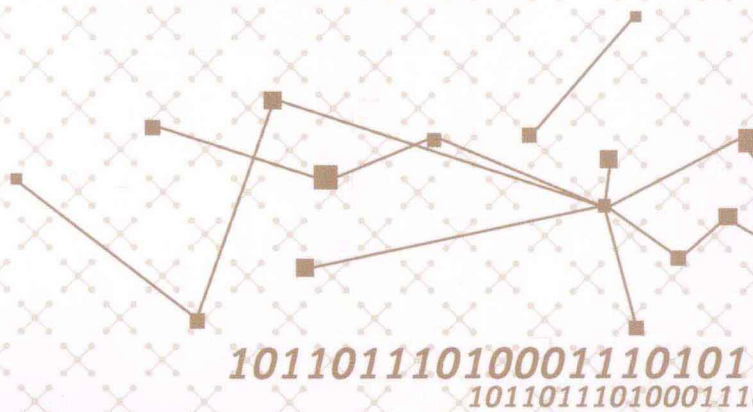




“十二五”国家重点图书出版规划项目

中国科学技术大学精品教材



苏仕华 贾伯琪 顾为兵 / 编著

# 数据库技术 与应用

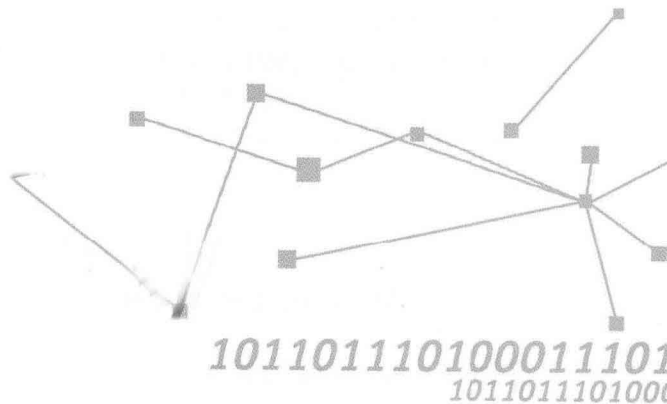
Database Technology and Application

中国科学技术大学出版社



“十二五”国家重点图书出版规划项目

中国科学技术大学 **精品** 教材



苏仕华 贾伯琪 顾为兵 / 编著

Database Technology and Application

# 数据库技术 与应用

中国科学技术大学出版社

## 内 容 简 介

本书较详细地介绍了数据库系统的原理、方法和发展趋势,并重点介绍了数据库系统的设计、实践与应用。内容包括:数据库系统的基本概念、关系模型的运算、关系数据库设计理论、关系数据库设计与应用、关系数据库语言 SQL、数据库保护、数据库技术的新发展以及 ORACLE 数据库设计实践等。

本书可作为高等院校“数据库技术与应用”课程的教材,也可以作为一般科研人员、工程技术人员的参考书。

### 图书在版编目(CIP)数据

数据库技术与应用/苏仕华,贾伯琪,顾为兵编著. —合肥:中国科学技术大学出版社,2013.6

(中国科学技术大学精品教材)

“十二五”国家重点图书出版规划项目

ISBN 978-7-312-03178-6

I. 数… II. ①苏… ②贾… ③顾… III. 数据库系统—高等学校—教材  
IV. TP311.13

中国版本图书馆 CIP 数据核字(2013)第 013679 号

中国科学技术大学出版社出版发行

地址:安徽省合肥市金寨路 96 号,230026

网址:<http://press.ustc.edu.cn>

合肥市宏基印刷有限公司印刷

全国新华书店经销

开本:710 mm×960 mm 1/16 印张:21.25 插页:2 字数:410 千

2013 年 6 月第 1 版 2013 年 6 月第 1 次印刷

定价:38.00 元

# 总 序

2008年,为庆祝中国科学技术大学建校五十周年,反映建校以来的办学理念和特色,集中展示教材建设的成果,学校决定组织编写出版代表中国科学技术大学教学水平的精品教材系列。在各方的共同努力下,共组织选题281种,经过多轮、严格的评审,最后确定50种入选精品教材系列。

五十周年校庆精品教材系列于2008年9月纪念建校五十周年之际陆续出版,共出书50种,在学生、教师、校友以及高校同行中引起了很好的反响,并整体进入国家新闻出版总署的“十一五”国家重点图书出版规划。为继续鼓励教师积极开展教学研究与教学建设,结合自己的教学与科研积累编写高水平的教材,学校决定,将精品教材出版作为常规工作,以《中国科学技术大学精品教材》系列的形式长期出版,并设立专项基金给予支持。国家新闻出版总署也将该精品教材系列继续列入“十二五”国家重点图书出版规划。

1958年学校成立之时,教员大部分来自中国科学院的各个研究所。作为各个研究所的科研人员,他们到学校后保持了教学的同时又作研究的传统。同时,根据“全院办校,所系结合”的原则,科学院各个研究所在科研第一线工作的杰出科学家也参与学校的教学,为本科生授课,将最新的科研成果融入到教学中。虽然现在外界环境和内在条件都发生了很大变化,但学校以教学为主、教学与科研相结合的方针没有变。正因为坚持了科学与技术相结合、理论与实践相结合、教学与科研相结合的方针,并形成了优良的传统,才培养出了一批又一批高质量的人才。

学校非常重视基础课和专业基础课教学的传统,也是她特别成功的原因之一。当今社会,科技发展突飞猛进、科技成果日新月异,没有扎实的基础知识,很难在科学技术研究中作出重大贡献。建校之初,华罗庚、吴有训、严济慈等老一辈科学家、教育家就身体力行,亲自为本科生讲授基础课。他们以渊博的学识、精湛的讲课艺术、高尚的师德,带出一批又一批杰出的年轻教员,培养

了一届又一届优秀学生。入选精品教材系列的绝大部分是基础课或专业基础课的教材,其作者大多直接或间接受到过这些老一辈科学家、教育家的教诲和影响,因此在教材中也贯穿着这些先辈的教育教学理念与科学探索精神。

改革开放之初,学校最先选派青年骨干教师赴西方国家交流、学习,他们在带回先进科学技术的同时,也把西方先进的教育理念、教学方法、教学内容等带回到中国科学技术大学,并以极大的热情进行教学实践,使“科学与技术相结合、理论与实践相结合、教学与科研相结合”的方针得到进一步深化,取得了非常好的效果,培养的学生得到全社会的认可。这些教学改革影响深远,直到今天仍然受到学生的欢迎,并辐射到其他高校。在入选的精品教材中,这种理念与尝试也都有充分的体现。

中国科学技术大学自建校以来就形成的又一传统是根据学生的特点,用创新的精神编写教材。进入我校学习的都是基础扎实、学业优秀、求知欲强、勇于探索和追求的学生,针对他们的具体情况编写教材,才能更加有利于培养他们的创新精神。教师们坚持教学与科研的结合,根据自己的科研体会,借鉴目前国外相关专业有关课程的经验,注意理论与实际应用的结合,基础知识与最新发展的结合,课堂教学与课外实践的结合,精心组织材料、认真编写教材,使学生在掌握扎实的理论基础的同时,了解最新的研究方法,掌握实际应用的技術。

入选的这些精品教材,既是教学一线教师长期教学积累的成果,也是学校教学传统的体现,反映了中国科学技术大学的教学理念、教学特色和教学改革成果。希望该精品教材系列的出版,能对我们继续探索科教紧密结合培养拔尖创新人才,进一步提高教育教学质量有所帮助,为高等教育事业作出我们的贡献。

中国科学技术大学校长  
中国科学院院士  
第三世界科学院院士

# 前 言

数据库技术自 20 世纪 60 年代中期产生到现在仅仅五十多年的历史,但其发展速度之快、应用范围之广是其他技术所不及的。短短五十年间,已从第一代的层次、网状数据库系统,第二代的关系数据库系统,发展到第三代以面向对象模型为主要特征的数据库系统。数据库技术的研究和发展已成为现代信息化社会中具有强大生命力的一个重要领域。数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等技术之间互相渗透、互相结合,已发展成为当前数据库技术的主要特征。

本书较详细地介绍了数据库系统的原理、方法和发展趋势,并重点介绍了数据库系统的设计、实践与应用。内容包括:数据库系统的基本概念、关系模型的运算、关系数据库的设计理论、关系数据库设计与应用、关系数据库语言 SQL、数据库保护、数据库技术的新发展以及 ORACLE 数据库设计实践等。

本书在内容表述上,以数据库理论为基础,以数据库应用为主线,始终贯彻理论和实践相结合的思想和方法,特别注重实践与应用。

本书层次结构清晰,各章节内容安排结构合理,语句条理清楚,做到了融会贯通。本书语言流畅、通俗易懂,每章都是从最基本的概念入手,由浅入深、循序渐进地介绍相关知识,列举出大量例题,着重分析重点、难点,并附有适当的习题。作者都是从事“数据库技术”以及“数据库原理”课程的教学工作多年,有丰富的教学经验,而且编写过数据库方面书籍的专家教授。因此,在本书的编写过程中作者紧密结合了教学经验,并处处体现以人为本。所以,本书具有既便于老师教学,又方便学生学习的特点。

本书加入了“数据库课程设计”方面的内容,这将为老师教授数据库课程和学生学习数据库课程带来很大的方便,而且对培养学生分析问题和解决问题的实际动手能力都会有极大的帮助。



# 目 次

总序 .....	( i )
前言 .....	( iii )
<b>第 1 章 数据库导论</b> .....	( 1 )
1.1 数据管理技术的发展 .....	( 1 )
1.2 数据库系统概述 .....	( 5 )
1.3 数据模型 .....	( 12 )
<b>第 2 章 关系数据库基础</b> .....	( 32 )
2.1 关系模型基础 .....	( 32 )
2.2 关系模型的完整性 .....	( 37 )
2.3 关系代数运算 .....	( 39 )
2.4 关系演算 .....	( 47 )
<b>第 3 章 关系数据库设计理论</b> .....	( 53 )
3.1 数据依赖 .....	( 53 )
3.2 函数依赖性 .....	( 55 )
3.3 关系模式的范式 .....	( 57 )
3.4 多值依赖与高级范式 .....	( 61 )
3.5 关系模式的规范化 .....	( 63 )
<b>第 4 章 数据库设计</b> .....	( 67 )
4.1 概述 .....	( 67 )
4.2 需求分析 .....	( 70 )
4.3 概念结构设计 .....	( 80 )
4.4 逻辑结构设计 .....	( 88 )





# 第 1 章 数据库导论

自从 1946 年世界上第一台计算机问世之后,计算机就是快速计算的代名词,它主要的功能是用于科学计算,而从 20 世纪 50 年代中期开始,在计算机的三大应用(科学计算、数据处理和过程控制)中,已逐步转向以数据处理为主,数据处理所占比重在 70%左右。在 20 世纪 60 年代末期,数据库技术就是作为数据处理的一门技术而发展起来的。

20 世纪 70 年代是数据库技术发展速度最快的年代。层次和网状模型数据库系统占据了整个市场,而关系数据库系统才仅仅处于实验阶段。当进入到 20 世纪 80 年代后,关系数据库系统由于其操作方便、使用简单,而逐步替代层次和网状系统占领了市场。目前市场上使用的基本上都是关系型数据库系统。

本章主要介绍数据库的基本概念,包括:什么是数据库、数据库系统、数据库管理系统、数据模型及数据库技术的发展等,是全书学习的基础。

## 1.1 数据管理技术的发展

当今,随着数据库技术的迅猛发展,数据库技术已被广泛应用到现实世界的各行各业中,并已成为计算机应用领域最重要的技术之一,是计算机软件学科中一个重要的独立分支。

### 1.1.1 数据与数据管理

**数据**是数据库系统的管理对象,它与人们一般认为的数据不同,是一种抽象化了的数据,它不仅仅指数字,还包括文字、图像、图形、声音等。实际上,人们为了认

识世界和交流信息,就有描述事物的需要,而数据就是用来描述事物的记录。比如,图书的信息、职工的档案、会计的账目等都是数据。

**数据处理**是指从某些已有的数据出发,加工推导出一些新的数据作为决策者决策的依据。数据处理包括数据收集、数据管理、数据加工利用以及数据输出等。

在数据处理中,通常数值的计算是比较简单的,而数据的管理则是比较复杂的。数据管理是指对数据的收集、整理、组织、存储、维护、检索、传输等操作。它不仅是数据处理的基本环节,而且是任何数据处理业务中必不可少的中心问题。所以说数据处理与数据管理是相互联系的,因此,数据管理技术的好坏,直接影响数据处理的效率。

### 1.1.2 数据库技术的产生

数据库技术是由数据管理任务的需要而产生的。人们利用计算机进行数据处理已经走过了五十多年的发展历程,随着数据处理量的快速增长以及数据处理速度要求的提高,数据管理技术也在不断地发展,这个发展过程大致分为:人工管理、文件管理和数据库系统三个阶段。

#### 1. 人工管理阶段

在 20 世纪 50 年代中期以前,计算机主要用于科学计算,在其他方面的应用还没有发展起来,当时的外部存储设备只有磁带、卡片和纸带等,没有磁盘等直接存取的存储设备。软件只有汇编语言,尚无操作系统及数据管理方面的软件。因此,数据处理方式基本上是人工管理的批处理方式。这个时期的数据管理有以下特点:

##### (1) 数据不具有独立性

由于没有专用的软件对数据进行管理,每个应用程序都要包括设定数据的存储结构、存取方法、输入输出方式等。当数据的逻辑结构或存储结构改变时,所对应的应用程序也必须作相应的修改,因此,数据不具有独立性。

##### (2) 数据不具有共享性

因为数据是面向应用程序的,一组数据只对应一个程序。当多个应用程序需要使用相同的一组数据时,必须在各自的程序中自行定义,无法相互利用实现数据共享。因此,程序与程序之间可能会产生大量的数据冗余。

##### (3) 没有文件概念

由于当时计算机主要用于科学计算,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,运算处理后即输出结果数据,随着计算任务的完成,用户退出计算机系统,数据和程序所占存储空间也一起被释放。

所以,在那个阶段只有程序的概念,而没有文件的概念。

## 2. 文件管理阶段

在 20 世纪 50 年代后期到 60 年代中期,计算机不仅用于科学计算,还用于数据管理。随着数据量的不断增加,数据的存储、检索和维护等问题成为迫切需要解决的问题。因此,数据结构和数据管理技术也就随之得到了迅速的发展。此时,外部存储器也有了磁盘、磁鼓等能直接存取的存储设备;软件方面,出现了高级语言和操作系统,在操作系统中有了专门用来管理外存数据的文件管理系统;在数据处理方式上不仅有批处理方式,而且还有联机实时处理方式。这一阶段的数据管理有以下特点:

### (1) 数据的独立性

由文件系统进行数据管理,即文件系统将数据组织成相互独立的数据文件,程序可按文件名直接对数据记录进行访问,可以对文件进行修改、插入和删除等操作,而不必关心数据的物理位置。程序与数据之间由文件系统提供的存取方法进行转换,从而使应用程序与数据之间有了一定的独立性。但是,由于数据文件是为某一特定应用服务的,文件的逻辑结构对应于相应的应用程序,要想对现有的数据再增加一些新的应用是比较困难的,系统不易扩充。一旦数据的逻辑结构改变,则必须修改应用程序、文件结构。而应用程序的改变,也必将引起文件数据结构的改变,因此,数据与程序之间仍然不具备真正的独立性。

### (2) 数据的共享性

当有了文件的概念之后,数据就可以长期保存在外存设备上,供应用程序反复进行查询、修改、插入和删除等操作;另外,数据不再属于某个特定的程序,而是可以重复使用,但在文件系统中,一个数据文件基本上对应于一个应用程序,即文件仍然是面向应用的。当不同的应用程序需要使用部分相同的数据时,则必须建立各自的文件,而不能实现数据的共享,因此数据冗余度大,浪费存储空间。同时由于相同数据的重复存储和各自管理,很容易造成数据的不一致性。

文件管理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,各种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。但随着数据管理规模的扩大,数据量的急剧增加,文件系统仍然存在一些缺陷:数据冗余性、一致性和独立性都比较差。

## 3. 数据库系统阶段

在 20 世纪 60 年代末期,随着计算机用于管理数据的规模越来越大,应用也越来越广泛,多个应用、多种语言同时共享数据集合的要求也越来越强烈。为了提高效率,克服和改进文件系统的缺陷,人们开始研究新的数据管理技术,而此时的磁



可以设计完整性规则以确保数据值的正确性。例如,把数据值限制在某个范围内,或规定数据之间满足一定的关系。

③ 数据库的并发控制。当多个用户程序并发访问数据库时,为避免相互干扰,防止数据库的完整性遭到破坏,杜绝给用户不正确的数据,因此必须对多用户的并发操作进行协调和控制。

④ 数据库的恢复。当因计算机系统故障、人为的操作失误或故意的破坏而造成数据库中数据丢失或不可靠时,数据库系统要能够把数据库恢复到最近某一时刻的正确状态。

## 1.2 数据库系统概述

### 1.2.1 数据库系统

数据库系统可以用简单的一句话来描述,即引进数据库技术的整个计算机系统可称为**数据库系统**(DataBase System,简称 DBS)。它一般由数据库、硬件、软件 and 用户四部分组成。图 1.1 可形象地描述一个最简化的数据库系统。

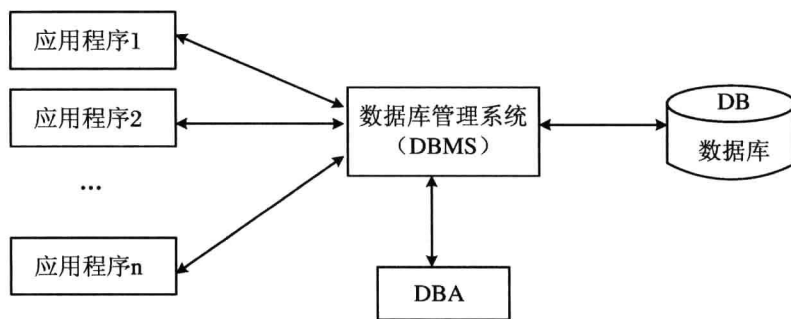


图 1.1 数据库系统示意图

### 1. 数据库

什么是数据库? 简单地说,数据库(DataBase,简称 DB)是以一定的组织方式存储在一起的、能为多个用户所共享的、与应用程序彼此独立的、具有较小冗余的、相互关联的数据集合,其作用在于存储与产生用户所需要的信息,为用户提供快



构可分为单用户结构、主从式结构、分布式结构和客户机/服务器结构。

### 1. 数据库的逻辑结构

数据库系统在逻辑结构上可分成三个层次,即外层、概念层和内层,通常称之为数据库的三级结构(外部级、概念级和内部级),如图 1.2 所示。从某个角度看到的数据特性即称为数据视图,外部级最接近用户,是单个用户所能看到的数据特性,因此单个用户使用的数据视图称之为用户视图;概念级则涉及所有用户数据的定义,也就是全局的数据视图,称为概念视图;而内部级最接近物理存储设备,涉及实际数据的存储方式,物理存储的数据视图称为存储视图或物理视图。

### 2. 数据库的三级模式

数据库系统需要对其抽象的逻辑层次进行描述,描述的结果称之为模式,因此,对应上述的三个数据库逻辑层次分别为外模式(或用户模式)、模式(概念模式)、和内模式(物理模式),如图 1.2 所示。

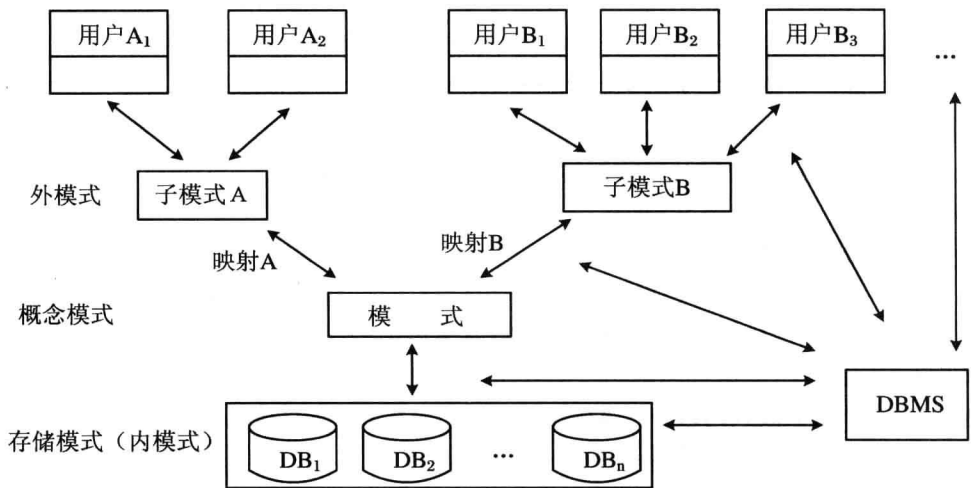


图 1.2 数据库系统结构示意图

#### (1) 内模式

内模式也称为存储模式,是对内层数据的物理结构和存储方式的描述。例如,记录的存储方式是顺序存储,还是按散列(HASH)方法存储;是否需要建立索引,数据是否需要压缩存储等。

#### (2) 模式

模式又称为逻辑模式或概念模式,是对数据库中全部数据概念层的一个逻辑表示或描述,是所有用户的公共数据视图,它是数据库系统模式结构的中间层,不



涉及数据的物理存储细节和硬件环境,与具体的应用程序以及应用程序所使用的开发工具及高级语言无关。实际上,模式是数据库数据在概念级上的视图,一个数据库只有一个模式。

数据库模式通常以某一种数据模型(下一节介绍)为基础,综合考虑所有用户的要求,并将这些需求有机地结合起来,定义数据的整体结构。定义模式不仅要定义数据的逻辑结构,例如,定义数据记录由哪些数据项组成,每个数据项的名称、类型、长度以及取值范围等,同时还需要定义这些记录数据之间的联系以及与数据有关的安全性、完整性约束条件等。

### (3) 外模式

外模式又称为子模式,是单个数据库用户的数据视图,它是外层用户所看到和使用的数据库。外模式通常是模式的一个子集,也是用户与数据库系统的接口。每个用户需要通过一个外模式来使用数据库,但不同的用户可以使用同一个外模式。

因为外模式是单个用户的视图,而不同的用户有不同的需求,因此,不同的用户可以对同一个模式定义出不同的外模式,使它们的外模式在结构、类型、长度、保密程度上均与原来的模式不同。外模式是保证数据安全的一个重要措施,因为每个用户只能看到和访问对应的外模式中的数据,数据库中的其余数据对他们来说是不可见的。

## 3. 数据库模式间的映像

文件管理系统的主要缺点之一就是用户程序过分地依赖存储的数据和数据的逻辑结构,当数据存储方式和数据文件整体逻辑结构发生改变时,用户则不得不重新编写应用程序,以保证原来用户的要求,这种程序对数据的依赖性的缺点在数据库系统中得到了极大的改善。最根本的原因在于数据库系统提供了模式之间的两级映像,即外模式/模式映像和模式/内模式映像。正是这两级映像保证了数据库系统中的数据具有较高的逻辑独立性和物理独立性。

### (1) 外模式/模式映像

这个映像存在于外部级和概念级之间,它用于定义外模式与模式之间的对应关系。模式描述的是数据的整体逻辑结构,而外模式描述的是数据的局部逻辑结构。对应于同一个模式可以有任意多个外模式,对于每一个外模式,数据库系统都有一个外模式/模式映像。当模式改变时(比如,增加新的数据类型、新的数据项或新的关系等),只要相应地改变外模式/模式的对应关系,即可以保证外模式的不变,又使得建立在这一外模式上的应用程序不需要修改。我们把这种通过调整映像,使得当数据的整体逻辑结构改变时,不需改变局部逻辑结构,从而不改变相应