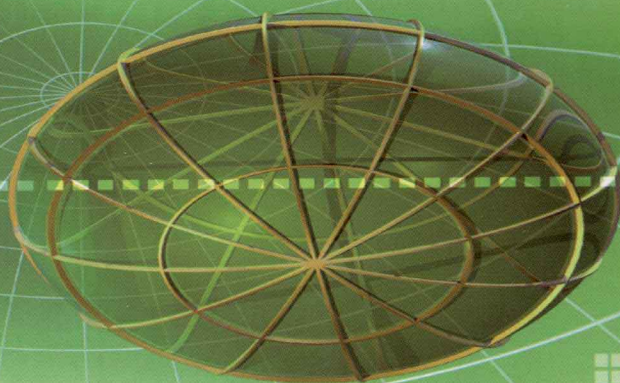




高等教育
电气工程与自动化系列规划教材

数据库原理及应用

高等教育教材编审委员会 组编
主编 王新房 邓亚玲 黄 元



大连理工大学出版社



高等教育
电气工程与自动化系列规划教材

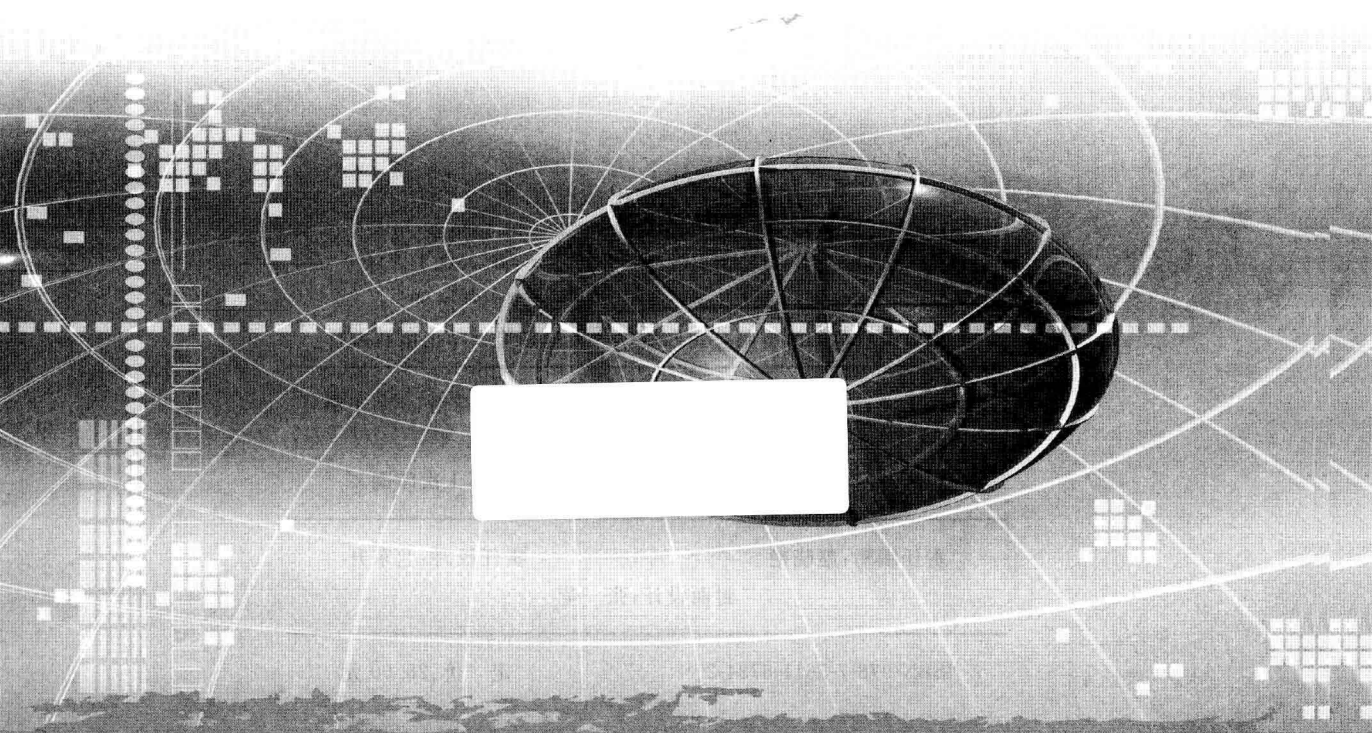
数据库原理及应用

SHUJUKU YUANLI JI YINGYONG

高等教育教材编审委员会 组编

主编 王新房 邓亚玲 黄 元

编



大连理工大学出版社

图书在版编目(CIP)数据

数据库原理及应用 / 王新房, 邓亚玲, 黄元主编

· 一 大连 : 大连理工大学出版社, 2012. 2

高等教育电气工程与自动化系列规划教材

ISBN 978-7-5611-6734-2

I. ①数… II. ①王… ②邓… ③黄… III. ①数据库
系统—高等学校—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2012)第 011444 号

大连理工大学出版社出版

地址:大连市软件园路 80 号 邮政编码:116023

发行:0411-84708842 邮购:0411-84703636 传真:0411-84701466

E-mail:dutp@dutp.cn URL:http://www.dutp.cn

大连美跃彩色印刷有限公司印刷 大连理工大学出版社发行

幅面尺寸:185mm×260mm 印张:13.25 字数:309 千字

印数:1~2000

2012 年 2 月第 1 版

2012 年 2 月第 1 次印刷

责任编辑:赵晓艳

责任校对:兰东升

封面设计:张莹

ISBN 978-7-5611-6734-2

定 价:28.00 元

《数据库原理及应用》是新世纪应用型本科教材编审委员会组编的电气工程及自动化系列规划教材之一。

数据库技术是计算机科学技术中发展最快的领域之一,也是应用最广的技术之一。它已成为计算机信息系统与应用系统的核心技术和重要基础。数据库技术已经成为人们存储数据、管理信息的最先进最常用的技术。普通高校电类专业都开设了数据库课程,但是,不同学校、不同专业对数据库课程的要求是不相同的。本书主要是针对普通高等学校电类专业(电子信息、自动化、通信、信息管理、电气、计算机等专业)而编写的一本数据库实用教材。通过对本书的学习,既能掌握数据库的基本概念、原理、技术和方法,又能够利用实际的数据库和开发工具开发数据库应用系统,达到学用结合、学以致用目的。

在已出版的众多数据库教材中,数据库理论内容涉及的较多,而有很多从事数据库应用的人员并不需要关心理论内容这部分。本书力求理论与实际应用紧密结合,侧重实际应用问题的解决方法。通过大量实例说明在应用中可能遇到的问题的解决方法。本书的定位就是在掌握基本概念的基础上,从实际应用的需求出发,重点介绍 SQL 语言、数据库的设计方法、数据库应用系统的开发和数据库系统的管理这四个方面,而不太关心 DBMS 内部实现所涉及的技术细节问题。要开发一个实际的数据库应用系统,掌握一些基本的 SQL 语句是必需的。数据库设计非常灵活,重要的是对基本概念和方法的掌握。数据库应用的开发涉及数据访问技术和开发工具的选取,但更重要的是对程序设计实现方法的掌握。

在编写本书的过程中,如何把握理工科专业学习数据库课程的要求,教材内容应该如何选取和组织,确实是一个很值得探讨的课题,但更重要的是要考虑学生毕业后可能从事什么样的工作。学生毕业后可能会从事数据库管理工作、数据库设计工作、应用软件的开发工作。对直接从事 DBMS 产品设计和开发所需的技术细节没有进行深入讨论。

本书需要40学时左右,重点是第1章、第2章、第3章、第5章和第6章,其他章节可作为自学内容来安排。为了方便教学,本书配有电子教案和有关程序源代码,读者可在出版社网站上下载。尽管本书的例子是在SQL Server环境下实现通过的,但对其他数据库产品也具有通用性。

本书共8章,分别是:绪论,关系数据库基本概念,SQL语言,SQL Server编程,关系数据库设计,数据库应用开发,数据库系统管理,数据仓库及应用。

本书由西安理工大学王新房、邓亚玲、黄元担任主编,全书由王新房老师组稿和定稿。大连理工大学吕蕾蕾老师审阅了全书并提出了许多宝贵意见和建议,在此深表谢意!

在写作过程中,我们尽量通过和实际应用密切相关的例子来说明有关的基本概念和方法,力求深入浅出,但由于学识浅陋,错误和不足之处在所难免,希望读者提出宝贵的批评意见,更希望学术界同仁不吝赐教。

所有意见和建议请发往:dutpbk@163.com

欢迎访问我们的网站:<http://www.dutpbook.com>

联系电话:0411-84707424 84706676

编 者
2012年2月



第 1 章 绪 论	1
1.1 数据管理技术的发展	1
1.2 数据库技术发展的现状及存在的问题	2
1.3 数据库技术的基本概念	5
1.4 数据模型	6
1.5 数据库系统的三级模式结构	9
1.6 数据库系统的构成	10
本章小结	11
习 题	11
第 2 章 关系数据库基本概念	12
2.1 关系数据结构	12
2.2 关系的完整性约束	14
2.3 关系模型规范化	15
2.4 关系数据库中的对象	17
2.5 样例数据库	18
本章小结	19
习 题	19
第 3 章 SQL 语言	21
3.1 Transact-SQL 的语法规约规则	21
3.2 数据定义语言	22
3.3 数据更新语句	25
3.4 数据查询语言	26
3.5 视图	35
3.6 数据控制语言	37
本章小结	37
习 题	38
第 4 章 SQL Server 编程	40
4.1 SQL Server	40
4.2 SQL Server 基础知识	43
4.3 流程控制语句	47
4.4 游标(Cursor)	51
4.5 存储过程	57
4.6 函数	60
4.7 触发器	66
4.8 事务	74
4.9 批处理	76

本章小结	78
习 题	79
第 5 章 关系数据库设计	80
5.1 数据库设计概述	80
5.2 需求分析	83
5.3 概念结构设计	90
5.4 逻辑结构设计	96
5.5 物理结构设计	98
5.6 数据库实施	100
5.7 数据库运行和维护	103
本章小结	104
习 题	105
第 6 章 数据库应用开发	106
6.1 数据库应用体系结构	106
6.2 数据库应用系统开发平台	107
6.3 ADO	111
6.4 基于 ADO 的数据库应用开发	114
6.5 ADO.NET	121
6.6 基于 ADO.NET 的数据库应用开发	124
6.7 基于 ADO 的应用系统开发实例	131
6.8 基于 ADO.NET 的应用系统开发实例	150
本章小结	160
习 题	161
第 7 章 数据库系统管理	162
7.1 安全管理	162
7.2 数据复制	167
7.3 数据导入与导出	172
7.4 备份与还原	175
7.5 管理索引	182
本章小结	185
习 题	186
第 8 章 数据仓库及应用	187
8.1 OLTP 与 OLAP	187
8.2 数据仓库的概念	189
8.3 数据集市	190
8.4 数据仓库系统的体系结构	192
8.5 数据挖掘	195
8.6 数据挖掘和数据仓库的关系	201
本章小结	202
习 题	202
参考文献	203

第 1 章

绪 论

随着计算机技术、通信技术和网络技术的飞速发展,信息系统已渗透到社会的各个领域。各行各业都在推进自己的信息化进程,大量的数据需要借助计算机进行处理,而数据管理是数据处理的一个重要方面,当前用于数据管理的最新技术就是数据库技术。可以说,数据库技术是现代信息科学与技术的重要组成部分,是计算机数据处理与信息管理的核心。数据库的建设规模、信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。数据库技术要解决的主要问题就是如何高效地组织和存储信息处理过程涉及的大量数据的问题。

1.1 数据管理技术的发展

数据管理技术的发展经历了人工管理、文件管理、数据库系统和高级数据库系统四个阶段。

1.1.1 人工管理阶段

20 世纪 50 年代中期以前属于人工管理阶段。当时的计算机主要用于科学计算。外部存储器只有磁带、卡片和纸带,还没有磁盘等直接存取存储设备。这个阶段没有专用的软件对数据进行管理,数据的组织方式由程序员自行设计安排,数据是面向应用的。这个阶段的特点是:数据 and 应用程序密切相关,数据不具有独立性,无法实现数据共享。

1.1.2 文件管理阶段

20 世纪 50 年代后期至 60 年代中期,计算机开始用于信息管理,与其相应的数据结构和数据管理技术也迅速发展起来。这个时期外部存储器已有了磁盘、磁鼓等直接存取存储设备。软件领域也出现了高级语言和操作系统,通过操作系统中的文件系统实现对外部存储的数据的管理。这个阶段的特点是:数据可以长期保存在外部存储设备中。尽管在数据共享和独立性方面有所进步,但数据共享性差,数据独立性差。

1.1.3 数据库系统阶段

20 世纪 60 年代后期,计算机管理的对象规模越来越大,应用范围越来越广泛,数据量急剧增大,多种应用共享数据的需要越来越强烈。在这种背景下,产生了用于统一管理数据

的专用系统软件,这就是数据库系统。数据库系统通过数据高度结构化,很好地满足了数据共享需求,使得数据具有高度的独立性,高效的存储效率,且易于扩展。这个阶段的数据库系统可以说是集中式的。

1.1.4 高级数据库系统阶段

从 20 世纪 80 年代,进入高级数据库系统阶段。这个阶段出现了分布式数据库系统、对象数据库系统和网络数据库系统。

分布式数据库可将物理上分布在不同场合的数据在逻辑上构成一个统一的整体。也就是说数据的物理位置对用户是透明的,各个应用通过网络可以透明地访问各类数据。

对象数据库采用面向对象的概念和技术,可以完整地描述现实世界的各类数据结构,可以方便地表达数据间的嵌套、递归联系。比如,对于多媒体数据、多维表格、CAD 数据等,对象数据库可以更好地对其进行管理。

计算机网络技术的发展,要求对网络资源透明访问。在网络环境中,通过中间件技术可以隐藏各种复杂性,保证不同的操作系统、通信协议和 DBMS 之间进行通信,实现互操作。当前流行的数据库都是通过中间件技术来实现这种通信和互操作的,这就是网络数据库系统。

1.2 数据库技术发展的现状及存在的问题

1.2.1 数据库技术发展的现状

自从 1969 年美国的 IBM 公司开发出第一个 DBMS 系统 IMS 以来,数据库的研究和开发已经走过了四十多年的历程,经历了三代的演变(从层次型数据库系统到网络型数据库系统,再到现在成为数据库主流的关系型数据库系统),取得了辉煌的成就。数据库技术和系统已经成为世界各国信息基础设施的核心技术和重要基础。

美国是世界上数据库业起步最早的国家。目前,在世界范围内,无论是数据库的数量、质量、品种、类型,还是数据库生产者、数据库提供商的数量,还是联机数据库的使用频率和产值方面都处于领先地位,别的国家还都无法与之抗衡。

我国的数据库(主要是中文数据库)建设起步于 20 世纪 70 年代中后期,当时主要引进和学习国外理论和成果。随后,全国许多单位纷纷开始建设数据库。经过多年的努力,我国自建的数据库已经有一定的规模。这为经济建设和社会文明进步起到了积极的推动作用。

目前,我国的数据库数量及其容量都有了很大发展,能发挥效用的数据库大大增加,大部分已在一定范围内提供不同程度的服务,但国产数据库产品占有的比例还比较小。由于使用国外的数据库产品存在着安全、超前消费、售后服务、价格等一系列问题,使得国产数据库有着自己的发展和生存空间。国产数据库产业化的目标是至少占领我国数据库市场的 50%,继而进入国际市场。

多年来,国家一直特别重视开发具有自主知识产权的国产数据库软件产品,并一度列入国家

863 计划项目,但总的来说目前我国数据库软件的发展水平仍令人忧虑,尤其在市场上叫得响的产品更是凤毛麟角。面对数据库“网络化”的发展趋势,现在国内拥有的各类数据库中,还有相当比例的数据库没有上网,资源被束之高阁,这对国家、企业来说无疑是极大的浪费。

电子信息产业已成为全世界发展经济竞争实力的最重要手段之一,党和政府于 1992 年就明确把电子信息产业作为发展国民经济的支柱产业,并于 1997 年 4 月 18 日至 21 日由国务院信息化工作领导小组主持召开全国首次信息化工作会议,全面明确地提出了中国信息化建设的指导方针和原则。虽然取得了很大的成绩,但现在仍是我国数据库产业的形成阶段,国家机关、图书馆、信息中心等部门在信息化的过程中都迫切地需要数据库软件,而且在网络时代数据库也有着更大的发展。近几年来我国数据库新产品的推出,大大增强了国产数据库产品的市场竞争力。但综观我国数据库产品的类型、数量及市场规模等与发达国家的产品相比还有一定的距离,因此为了满足信息资源建设的需要,缩小与国际同类产品的差距,必须加速发展我国的数据库产业。

1.2.2 数据库技术发展存在的问题

(1) 信息特性和来源变化

传统的 DBMS 只能理解、存储和处理简单的数据类型。如整数、浮点数、字符串、日期、货币等。在信息化过程中,数据来源多种多样,涉及的数据类型越来越多。处理这些数据的方法复杂化以及数据量越来越大是当前数据库技术面临的重要挑战。

解决的办法是在原来的 DBMS 系统中增加对复杂数据类型的存储和处理功能,将新的结构移植到传统的架构上,还是应该重新思考 DBMS 基本构架,是当前学术界要研究的问题。

(2) 应用领域的变化

应用领域变化中,Internet 是当前最主要的驱动力。在 Internet 应用环境下所有应用已经从企业内部扩展为跨企业间的应用,需要 DBMS 对信息安全和信息集成提供更有力的保障和支持。这也对 DBMS 研究团体提出了许多新的待解决的问题。

(3) 相关技术的发展

新的信息管理的挑战使传统的 DBMS 技术方面面临着巨大变化,这些变化要求我们对传统数据库存储管理和查询处理算法重新加以评估。另外,处理器高速缓存有了爆炸性增长,并且增加了层次,这就要求 DBMS 能够利用高速缓存来改进实现算法。

Lowell 报告认为,许多影响信息管理的新应用、技术趋势以及相关领域的协作等问题,需要一个全新的信息管理架构并需重新考虑信息存储、组织、管理和访问等方面的问题。

(4) 当前若干研究热点

- 信息集成
- 移动数据管理
- 网格数据管理
- 传感器数据库技术
- DBMS 的自适应管理

1.2.3 数据库技术发展的趋势

技术和应用的发展总是相互作用的。经过多年的积累,企业和部门积累的数据越来越

1.2.3 数据库技术发展的趋势

技术和应用的发展总是相互作用的。经过多年的积累,企业和部门积累的数据越来越多,许多企业面临着“数据爆炸”。如何解决海量数据的存储管理、如何挖掘大量数据中包含的信息和知识,已成为目前亟待解决的问题。因此数据挖掘成为目前发展极为迅速的一个研究领域,数据挖掘又称数据开采,就是从大量的、不全的、有噪声的、模糊的、随机的数据中提取隐含在其中的人们事先不知道的、但又是潜在有用的信息和知识的过程,提取的知识表现为概念、规则、规律模式约束等形式。它综合了机器学习、统计分析和数据库技术,是为数据库中数据的决策型使用服务的。

随着计算机网络技术及 Internet 应用的日趋普及,势必要求数据库系统的应用平台向 Web 平台迁移,最终达到全球数据信息共享的目的,如何实现数据库平台与 web 平台的无缝对接,即对 Web 数据库技术的研究也成为近期研究的热点。

(1)深度(智能化)

计算机科学的主要目标是使计算机与人的界面尽量靠近人这边,因此,要尽量提高计算机的智能水平。智能化是计算机科学各个分支的研究前沿。在数据库方面,智能化的工作是将人工智能技术与数据库技术相结合,即演绎数据库知识库研究。目前的主要困难在于递归查询处理无法取得满意的性能,硬件技术的革命(大内存、并行机、高速存取的外存储器)将是提高知识库查询效率的重要因素。

(2)广度(多媒体)

多媒体数据处理的困难很多,即使是一般的复杂对象目前也还不能很好地处理。多媒体数据的建模、存储和多媒体数据库的查询及查询处理等都是需要我们研究解决的内容。

(3)分布(网络)

信息分布方面主要是分布式数据库系统的研究。分布式数据库从 20 世纪 70 年代开始研究,当前流行的数据库产品(如:SQL Server, ORACLE7, DB2 等)已具有数据分布特征。但新的计算机应用又对它提出了新的要求,智能化、新型事务模型、多媒体数据的处理、高速信息通信、数据源的高度透明性等将是新型的分布式数据库系统的重要研究内容。

(4)各种数据库间的数据转换

这种转换有两方面的需要。数据库管理系统软件的更新和计算机网络上不同数据库间的数据共享。当数据库管理系统软件更新时,希望将原先的数据库直接转到新系统中来,以保护以前的投资。在计算机网络上,常常运行着多种类型的数据库系统,它们是在不同时间、不同地点建立的,联网后这些投资要保护,所以要进行数据库转换。

(5)数据库性能评价

一个好的数据库应该是数据冗余尽可能少、数据库访问效率尽可能高并且数据库易于维护。这涉及数据语义和许多人为因素,因此定量的评价非常困难。良好的评价方法或衡量模型将产生一个好的数据库设计方法,因此意义重大。

(6)与软件工程的结合

到目前为止,数据库设计与应用程序设计是分离的,并且在具体的应用系统开发中,往往由两个小组各行其是。其后果是优良设计不得不放弃,导致系统性能低下或用户要求无法满足。面向对象技术是二者结合的有效手段。但是,良好的结合还需要很长时间的努力。

■ 4 数据库原理及应用

处理的数据越来越庞大、计算机网络越来越复杂、系统的智能水平越来越高是计算机系统发展的总趋势。因此,未来的信息管理系统的特征将是处理复杂对象、分布、智能。在复杂对象处理方面,面向对象数据库、多媒体数据库将会由于广泛的应用背景而迅速发展。在数据分布方面,客户/服务器数据库系统将快速发展,并在应用上取得良好效果。在智能化方面,数据库和人工智能将在各自的领域不断发展、不断取得新的成果。二者结合方面的研究将不断地利用二者的成果研制出新型的系统。任何时候两方面的结合都是必要的。

数据库是数据管理的最新技术,是计算机科学的重要分支。由于数据库具有数据结构合理、冗余度低、较高的程序独立性、易于扩充、易于编制应用程序等优点,较大信息系统都是建立在数据库设计之上的。因此,作为数据库复杂用户的我们,应该比较熟悉数据库管理系统的各种功能,学会直接使用数据库语言访问数据库,甚至能够基于数据库管理系统的API编写自己的应用程序。

1.2.4 数据库技术的应用举例

数据库技术在各行各业都得到了广泛的应用。下面列举一些应用实例:

- (1)银行业:用于存储客户的信息、账户、贷款以及银行的交易记录。
- (2)航空业:用于存储订票和航班的信息。航空业是最先以地理上分布的方式使用数据库的行业之一,分散于世界各地的终端通过电话线或其他数据网络来访问中央数据库系统。
- (3)大学:用于存储学生的信息、课程注册和成绩的信息。
- (4)信用卡交易:用于记录信用卡消费的情况和产生每月清单。
- (5)电信业:用于存储通话记录,产生每月账单,维护预付电话卡的余额和存储通信网络的信息。
- (6)金融业:用于存储股票、债券等金融票据的持有、出售和买入的信息。
- (7)销售业:用于存储客户信息,产品和购买的信息。
- (8)制造业:用于管理供应链,跟踪工厂中产品的产量、仓库(或商店)中的产品的详细清单以及产品的订单。
- (9)人力资源:用于存储员工、工资、所得税和津贴的信息,以及产生工资单。

1.3 数据库技术的基本概念

数据库技术涉及许多基本概念,包括信息、数据、数据库、数据库管理系统以及数据库系统等。

数据是信息的载体。信息是对数据的解释,是数据中的内容。数据的种类很多,常见的有文本、图形、图像、音频、视频等。

数据库是面向数据的,如何高效地存储数据、实现数据共享、保障数据安全以及高效地检索数据和处理数据是数据库要解决的基本问题。一般认为数据库是长期存储在计算机内有组织的、可以共享的数据集合。数据库中的数据按一定的数据模型组织、存储,具有较小的冗余度、较高的数据独立性和可扩展性,可以为各种用户共享。

数据库中的数据由数据库管理系统(Database Management System, DBMS)进行统一

管理,它屏蔽了数据的存储细节,用户通过简单易学的 SQL 语言实现对数据库中存储的数据进行操作。数据库管理系统属于系统软件,目前流行的有:Oracle、DB2、SQL Server 等。DBMS 由一组程序构成,主要功能是完成对数据库中数据的定义和数据操纵,为用户提供 一个简明的应用接口,实现数据共享及事务处理。

1.4 数据模型

模型是对现实客观世界中某种对象特征的模拟和抽象。数据模型就是对数据特征的抽象,它是用来描述数据、组织数据和对数据进行操作的。因此,数据模型通常由数据结构、数据操作和完整性约束三部分构成。数据结构描述数据库的组成对象及对象之间的联系。数据操作规定对各种对象允许执行的操作,一般包括查询、插入、删除和修改。完整性约束可以保证数据的正确、有效和相容。

数据模型一般应满足:真实模拟各类数据;容易被人们所理解;便于在计算机上实现。当前,一个模型要很好地满足这三个要求尚很困难。因此,根据不同的适用对象和应用目的,需要采用不同的数据模型。数据模型一般分为概念模型、逻辑模型和物理模型。如图 1-1 所示。现有的数据库系统都是基于某种模型实现的。数据模型是数据库系统的核心和基础。

概念模型独立于计算机系统,它不关心数据在计算机中是如何表示的,它关心的是数据结构,按照用户的观点对数据建模,它是对企业主要数据对象的基本表示和概括性描述,主要用于数据库设计。概念模型强调其语义的表达能力,概念应该简单、清晰,易于用户理解,是数据库设计人员和用户之间进行交流的工具。实体联系模型就是概念模型的代表。

逻辑模型直接面向数据库的逻辑结构,通常有一组严格定义 的无二义性的语法和语义的数据库语言,人们可以用这种语言来定义、操纵数据库中的数据。应用程序员根据逻辑模型编程,每种数据库管理系统都支持一种逻辑模型。目前,数据库领域中最常用的逻辑模型有层次模型、网状模型、关系模型和面向对象模型等。

物理模型是对数据最底层的抽象,它描述数据在磁盘或磁带等物理介质上的存储方式和存取方法,是面向计算机系统的。物理模型的具体实现是数据库管理系统的任务,在使用支持关系模型的数据库管理系统时,用户不必考虑物理级的细节。

从概念模型到逻辑模型的转换由数据库设计人员完成,从逻辑模型到物理模型的转换则由数据库管理系统完成。一般人员掌握了逻辑模型就可以很方便地使用数据库。

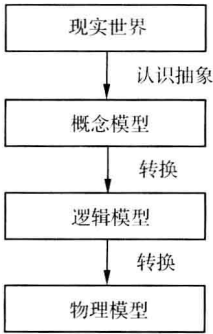


图 1-1 数据模型抽象的层次

1.4.1 实体联系模型

实体联系模型(Entity Relationship Model, 简记为 ER 模型)属于概念模型,它由 P. P. Chen 于 1976 年提出。它从用户的观点对数据建模,强调其语义的表达能力,概念尽量简单、清晰、易于用户理解。实体联系模型中会涉及实体、属性、码、域、实体型、实体集、联系等基本概念。

(1)实体(Entity)是客观存在并可相互区别的事物,它可以是具体的人、事、物,也可以是抽象的概念或联系。比如,一个学生、一个部门、一门课、一次选课,学生与专业的关系等都是实体。

(2)属性(Attribute)是实体具有的某一个特性。一个实体可以有多个属性。例如,学生实体可以有学号、姓名、年龄、性别等属性。

(3)码(Key)是可以唯一表示一个实体的属性集。比如,学号是学生实体的码。

(4)域(Domain)是一组具有相同数据类型值的集合。属性在某个域内取值。比如,学生的性别域为(男,女),学生的年龄为正整数域。

(5)具有相同属性的实体具有共同的特性。用实体名及其属性名集合来抽象和刻画同类实体,成为实体型(Entity Type)。例如,学生(学号,姓名,年龄,性别)就是一个实体型。

(6)实体集(Entity SET)是同一类型实体的集合。例如,全体学生就是一个实体集。

(7)联系(Relationship)是用来描述实体型(或实体)之间的联系。实体型之间的联系有一对一、一对多和多对多。一个班级只能有一个班长,而一个班长只能在一个班中任职,这时,在学生班长和班级之间的联系就是一对一。一个学生只能属于一个专业,但一个专业可以有多个学生,这样,学生和专业之间的联系是一对多。一个学生可以选多门课,一门课可以被多个学生选修,所以学生和选课之间的联系是多对多。同样,实体集内的各实体之间也可以有一对一、一对多和多对多联系。

1.4.2 层次模型

层次模型(Hierarchical Model)用树型(层次)结构表示实体类型及实体间联系。树的节点是记录类型,每个非根节点有且只有一个父节点。上一层记录类型和下一层记录类型之间是一对多联系。

现实世界中许多实体之间存在着自然的层次关系,如组织机构、家庭、物品分类等。如图 1-2 所示是一个层次模型,该模型描述了某一大学的组织机构及其联系。

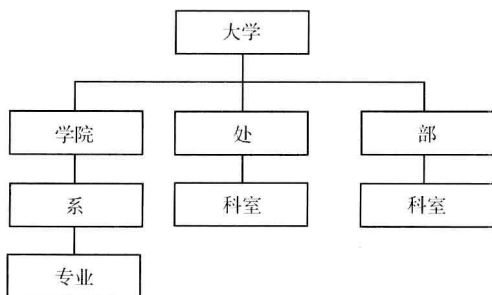


图 1-2 某大学组织机构的层次模型

层次模型中的主要概念有:片断、字段、层次序列和层次路径等。

层次模型的特点是记录之间的联系通过指针来实现,查询效率高,但层次模型只能表示一对多的联系。另外,由于层次顺序的严格和复杂,引起数据的查询和更新的操作很复杂,因此应用程序的编写也比较复杂。1968 年美国 IBM 公司推出的 IMS (Information Management System)系统是典型的层次模型系统,20 世纪 70 年代在商业上得到了广泛的应用。

1.4.3 网状模型

网状模型(Network Model)用有向图结构表示实体类型及实体之间的联系。它的数据结构是一个网络结构。在该模型中,允许一个节点可以有多个双亲节点;多个节点无双亲节点。

在网状模型中每个节点表示一个实体类型,节点间的连线表示实体之间的联系。与层次模型不同,网状模型中的任意节点之间都可以有联系,而且可以表示多对多的联系。网状模型中的主要概念有:纪录、数据项、联系和域等。

网状模型虽然可以表示实体间的复杂关系,但它与层次模型间没有本质的区别,它们都用连线表示是实体之间的联系,在物理实现上也有许多相同之处。例如,都用指针表示实体之间的联系。可以说,层次模型是网状模型的特例,网状模型同样具有数据结构复杂和编程复杂的缺点,因此从 20 世纪 80 年代中期起其市场已被关系数据库产品取代。

1.4.4 关系模型

关系模型(Relational Model)是用二维表表示的实体集,其基本数据结构是二维表。二维表由行和列组成,一张二维表称为一个关系。在关系模型中,实体和实体间的联系都是通过关系表示的。在二维表中既存放实体本身的数据,也存放实体间的联系。如图 1-3 所示是一个表示学生和专业关系模型。图 1-3(a)和(b)分别表示专业关系和学生关系。这两个关系也表示了专业和学生之间的一对多联系。

专业编号	专业名称	学号	姓名	专业编号
0401	自动化	001	张三	0401
0402	电子信息	002	李四	0402
0501	计算机	003	王五	0401

(a)

(b)

图 1-3 学生和专业关系模型

关系模型中的主要概念有:关系、属性、元组、域和关键字等。与层次模型和网状模型相比,关系模型具有以下优点:

- (1)数据结构单一,不管是实体还是实体间的联系都用关系来表示。
- (2)建立在严格的数学概念基础上,具有坚实的理论基础。
- (3)将数据定义和数据操纵统一在一种语言中,使用方便,易学易用。

由于关系模型具有许多优点,在 20 世纪 80 年代后的商品数据库系统几乎都是关系型的。当然,关系模型也有不足之处,主要是其结构单一,缺乏语义信息,不能更好地模拟现实世界的复杂对象。

1.4.5 面向对象模型

面向对象模型(Object Model)中的基本数据结构是对象,一个对象由一组属性和方法组成,属性用来描述对象的特征,方法是用来描述对象的操作。一个对象的属性可以是另外一个对象,另一个对象的属性还可以用其他对象描述,以此来模拟现实世界的复杂实体。

面向对象模型中的对象是封装的,对对象的操作通过调用方法来实现。面向对象模型

中的主要概念有:对象、类、方法、消息、封装、继承和多态性等。

面向对象模型有许多优点,主要有:

- (1)可以表示复杂对象,精确模拟现实世界中的实体。
- (2)模块化的结构,便于管理和维护。
- (3)具有定义抽象数据类型的能力。

面向对象模型是新一代数据库系统的基础,数据库技术发展的方向。表 1-1 对比了四种逻辑模型。

表 1-1 四种逻辑模型的比较

数据模型	层次模型	网状模型	关系模型	面向对象模型
创始	1968 年 IBM 公司的 IMS 系统	1969 年 CODASYL 的 DBTG 报告	1970 年 E. F. Codd 提出关系模型	20 世纪 80 年代
数据结构	复杂(树结构)	复杂(有向图结构)	简单(二维表)	复杂(嵌套、递归)
数据联系	通过指针	通过指针	通过表间的公共属性	通过对象标识
查询语言	过程性语言	过程性语言	非过程性语言	面向对象语言
典型产品	IMS	IDS/II, IMAGE/300	Oracle, SQL Server, DB2, Sybase	ONTOS DB
盛行期	20 世纪 70 年代	20 世纪 70 年代至 80 年代	20 世纪 80 年代至现在	20 世纪 90 年代至现在

1.5 数据库系统的三级模式结构

数据库的体系结构分为三级:外部级(External)、概念级(Conceptual)和内部级(Internal),一般称其为数据库的三级模式结构,如图 1-4 所示。这个结构又称为数据库的体系结构(三级模式结构)。当前大多数的 DBMS 产品都具有这个三级结构特征。

从某个角度看到的数据特性称为数据视图(Data View)。外部级最接近用户,是单个用户所能看到的数据特性。单个用户使用的数据视图的描述称为外模式。概念级涉及所有用户的数据特性,是全局性的数据视图,全局数据视图的描述称为概念模式。内部级最接近于物理存储设备,涉及物理数据存储结构,是存储视图,存储视图的描述称为内模式。

数据库的三级模式结构是对数据的三个抽象级别,它把数据的具体组织留给 DBMS 去做,用户只要抽象地处理数据,而不必关心数据在计算机中的表示和存储细节,这样就大大方便了用户对数据库系统的使用。

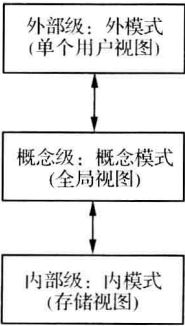


图 1-4 数据库三级模式结构

1.6 数据库系统的构成

数据库系统是基于数据库的计算机应用系统,由四部分构成,即数据库、数据库管理系统、数据库应用程序和用户。如图 1-6 所示。

1.6.1 数据库

数据库(DB)主要是用来存储数据的,数据库中的数据存储在操作系统文件中。这些文件由 DBMS 来统一管理,对用户是透明的,用户不能直接访问文件,必须通过 DBMS 访问数据库中的数据。有些数据库是建立在磁盘的一个或多个分区中,不通过操作系统提供的功能,而使用 DBMS 针对数据库特点实现的功能直接访问数据,从而大大提高了访问速度。

数据库中的数据按一定的数据模型组织、描述和存储,具有较小的冗余度、较高的数据独立性和扩展性,为各种用户所共享。

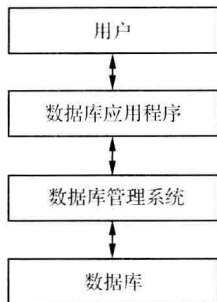


图 1-6 数据库系统的构成

1.6.2 数据库管理系统

数据库管理系统(DBMS)是位于用户和操作系统之间的系统软件,用于管理数据库,完成对数据库的一切操作,包括数据定义、查询、录入、更新、删除以及各种控制。

DBMS 提供数据定义语句,用户通过它们可以方便地定义数据库中的对象。

通过 DBMS 提供的数据操纵语句,用户可以方便地实现对数据的基本操作,比如查询、插入、修改、删除。

DBMS 提供相关的程序供数据库管理人员实施对数据库的管理和维护,包括安全管理、数据库的转储、故障恢复、性能分析等。

1.6.3 数据库应用

用户通过数据库应用实现对数据库中数据的访问。数据库应用是针对特定的企业或单位开发的应用程序,针对性很强。比如,人事管理系统、学生管理系统、财务管理系统等。

数据库应用的开发、管理和使用过程中涉及的人员主要有开发人员(系统分析员、数据库设计人员和应用程序员)、数据库管理员和用户。

1. 开发人员

系统分析员使用软件工程的方法对业务流程进行分析,提出应用系统的需求分析和规范说明,和用户及数据库管理员一起确定系统的硬件和软件配置,并参与数据库系统的概念结构设计。

数据库设计人员根据分析阶段产生的数据流图确定数据库中数据的组织,进行应用系统的设计(包括功能设计和界面设计)。一般情况下数据库设计人员应参加用户需求分析。

应用程序员根据系统设计说明书负责编写程序模块,并进行调试和安装。