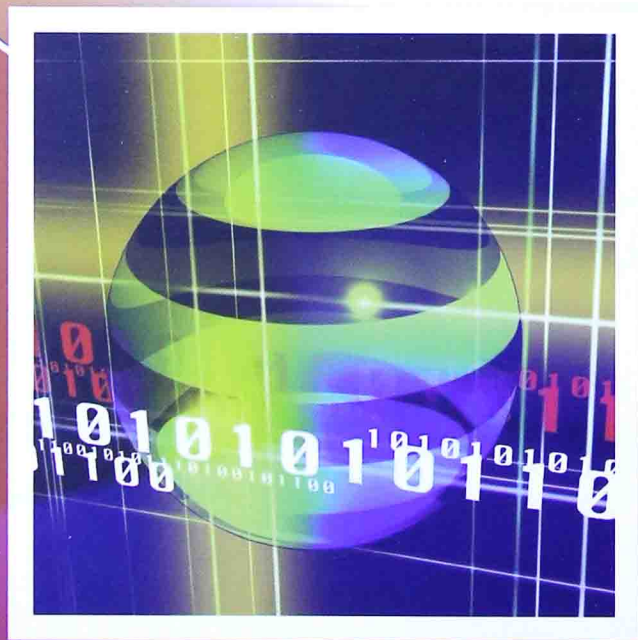


Access数据库

与程序设计研究

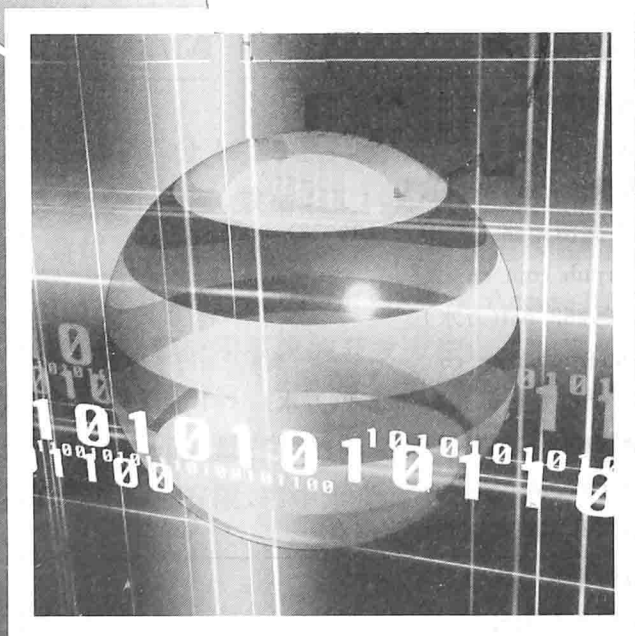
主编 赵娟 司小玲 王飞



中国水利水电出版社
www.waterpub.com.cn

Access数据库 与程序设计研究

主 编 赵 娟 司小玲 王 飞
副主编 史乙力 林烨秋 靳 婷
张 瑛 百 顺 赵 韬



中国水利水电出版社
www.waterpub.com.cn

TP311.138 AC
222

内 容 提 要

全书以 Microsoft Access 2003 中文版为平台,分 12 章,内容包括数据库基础知识,Access 数据库系统概述、关系数据库理论及设计、数据表的基本操作、结构化查询语言 SQL,以及查询、窗体、报表、数据访问页、宏、模块与 VBA 的设计,数据库安全管理,数据库应用系统的开发过程,并在最后给出了一个教学管理系统设计案例。

本书条理清晰,通俗易懂。不仅可以作为各类办公人员、中小型数据库应用开发人员的参考用书,同时对于计算机应用人员和计算机爱好者也是一本实用的参考书。

图书在版编目(CIP)数据

Access 数据库与程序设计研究/赵娟,司小玲,王
飞主编. —北京:中国水利水电出版社,2014.6

ISBN 978-7-5170-1890-2

I. ①A… II. ①赵…②司…③王… III. ①关系数
据库系统—程序设计 IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2014)第 070311 号

策划编辑:杨庆川 责任编辑:杨元泓 封面设计:马静静

书 名	Access 数据库与程序设计研究
作 者	主 编 赵 娟 司小玲 王 飞
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路 1 号 D 座 100038) 网址:www.waterpub.com.cn E-mail:mchannel@263.net(万水) sales@waterpub.com.cn
经 售	电话:(010)68367658(发行部)、82562819(万水) 北京科水图书销售中心(零售) 电话:(010)88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排 版	北京鑫海胜蓝数码科技有限公司
印 刷	三河市天润建兴印务有限公司
规 格	184mm×260mm 16 开本 24.75 印张 633 千字
版 次	2014 年 6 月第 1 版 2014 年 6 月第 1 次印刷
印 数	0001—3000 册
定 价	86.00 元

凡购买我社图书,如有缺页、倒页、脱页的,本社发行部负责调换

版权所有·侵权必究

前 言

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,随着计算机与网络技术的飞速发展,作为计算机应用的一个重要领域,数据库技术得到了广泛的应用与发展。

近些年,有关数据库的研究成果和新产品不断涌现,数据库技术与网络通信技术、面向对象技术、多媒体技术、人工智能等技术相互渗透、结合,逐步成为数据库新技术发展的主要特征,最终形成了新一代数据技术体系。典型的数据库管理系统的种类有很多,如 SQL Server、Oracle 等,相对于这些大型数据库,Access 这一桌面数据库管理系统为数据管理提供了简单实用的操作环境,如可视化操作工具及向导。同时,Access 中还提供了强大的程序设计语言 VBA(Visual Basic for Application),配合 VBA 代码,能够开发出性能更高的数据库应用系统。此外,由于 Access 易于使用和功能强大的特性,无论对于经验丰富的数据库设计人员,还是刚刚接触数据库管理系统的初学人员,都能够获得高效的数据处理能力。

Microsoft Access 2003 关系型数据库管理系统是 Microsoft Office 系列应用软件的一个重要组成部分,它界面友好,功能全面且操作简单,不仅可以有效的组织和管理、共享和开发应用数据库信息,而且可以把数据库信息与 Web 结合在一起,为局域网络与互联网共享数据库信息奠定了基础。

本书以 Access 2003 为实践环节,分 12 章内容,包括数据库系统基础知识、关系数据库理论及设计、Access 数据库管理系统操作基础、数据表的基本操作、查询的创建与操作、结构化查询语言 SQL、数据访问页设计、窗体与报表设计、宏的设计、模块与 VBA 程序设计、Access 数据库的安全管理、Access 2003 应用系统开发实例。更好地体现了 Access 的基本知识体系,突出数据库的本质概念和应用要求,强调数据库应用系统的开发方法。书中案例中用到的人名、电话、电子邮件等均为虚构,如有雷同,实属巧合。

本书在编写过程中,参考了国内外大量有价值的文献与资料,并得到了众多前辈的支持与帮助,在此一一表示感谢。由于编者水平有限,书中难免存在疏漏或不妥之处,恳请广大专家学者批评指正。

编 者

2014.3

目 录

第 1 章 绪论	1
1.1 数据库系统概述	1
1.2 数据模型	7
1.3 数据库系统结构	16
1.4 数据库管理系统	21
1.5 数据库新技术	27
第 2 章 关系数据库理论及设计	35
2.1 关系数据库	35
2.2 关系模式	38
2.3 关系操作	40
2.4 函数依赖	42
2.5 关系模式的规范化	46
2.6 关系代数与关系演算	53
第 3 章 Access 2003 数据库管理系统操作基础	61
3.1 Access 数据库概述	61
3.2 Access 的开发环境	65
3.3 Access 的数据库对象	70
3.4 Access 的联机帮助	75
3.5 Access 数据库的创建	77
3.6 Access 数据库的基本操作	85
第 4 章 数据表的基本操作	90
4.1 创建数据表	90
4.2 维护数据表	109
4.3 操作表中数据	117
4.4 表间关联关系的建立	124
第 5 章 查询的创建与操作	129
5.1 查询概述	129
5.2 选择查询	134
5.3 参数查询	145
5.4 交叉表查询	148
5.5 操作查询	154
5.6 查询中的计算	161

第 6 章 结构化查询语言 SQL	165
6.1 SQL 概述	165
6.2 SQL 的数据定义	171
6.3 SQL 数据的查询	180
6.4 SQL 的数据操纵	193
第 7 章 窗体与报表设计	197
7.1 窗体与报表概述	197
7.2 窗体设计	204
7.3 报表设计	224
第 8 章 数据访问页设计	252
8.1 数据访问页概述	252
8.2 数据访问页的创建	256
8.3 数据访问页的编辑	266
8.4 数据访问页的发布与访问	271
第 9 章 宏的设计	273
9.1 宏概述	273
9.2 创建与编辑宏	277
9.3 运行与调试宏	288
第 10 章 模块与 VBA 程序设计	304
10.1 模块概述	304
10.2 VBA 编程基础	308
10.3 VBA 程序结构	321
10.4 过程的定义与参数传递	326
10.5 面向对象程序设计	329
10.6 VBA 数据库编程	332
10.7 VBA 程序调试与错误处理	335
第 11 章 Access 数据库的安全管理	339
11.1 数据库安全维护	339
11.2 用户安全机制	345
11.3 数据库的压缩和恢复	355
11.4 数据库的备份	360
11.5 数据的导入与导出	362
第 12 章 Access 2003 应用系统开发实例	364
12.1 数据库应用系统的开发过程	364
12.2 教学管理系统设计	374
参考文献	390

第 1 章 绪 论

1.1 数据库系统概述

1.1.1 数据与信息

在信息技术(Information Technology, IT)领域,数据和信息是两个重要的基本概念,它们之间既有联系又有区别。数据是信息的基础,如何实现数据的存储、操纵、管理和检索,进而从中获取有价值的信息,已成为当今计算机技术研究和应用的重要课题。

数据是数据库系统存储、处理和研究的对象,它是指描述客观事物的数、字符,以及所有能输入计算机并被计算机程序处理的符号的集合。因此,在计算机科学技术中,数据的含义是十分广泛的,它不仅可以是数值,其他诸如字符、图形、图像乃至声音等都可以视为数据。数据集中的每一个个体称为数据元素,它是数据的基本单位。

数据有“型”和“值”之分。数据的型是指数据的结构,而数据的值是指数据的具体取值。数据的结构指数据的内部构成和对外联系。

信息在管理和决策中起着主导作用,是管理和决策的依据,是一种重要的战略资源。在信息技术中,信息通常是指“经过加工而成为有一定的意义和价值且具有特定形式的数据,这种数据对信息的接收者的行为有一定的影响”。

信息具有以下一些基本特征:

①时间性。即信息的价值与时间有关,它有一定的生存期,当信息的价值变为零时,则其生命结束。

②事实性。即信息需是正确的、能够反映现实世界事物的客观事实,而不是虚假的或主观臆造的。

③明了性。即信息中所含的知识能够被接收者所理解。

④完整性。即信息需详细到足够的程度,以便信息的接收者能够得到所需要的完整信息。

⑤多样性。即信息的定量化程度、聚合程度和表示方式等都是多样化的。可以是定量的也可以是定性的,可以是摘要的也可以是详细的,可以是文字的也可以是数字、表格、图形、图像、声音等表示形式。

⑥共享性。即信息可以广泛地传播,为人们所共享。

⑦模糊性。即由于客观事物的复杂性、人类掌握知识的有限性和对事物认识的相对性,信息往往具有一定的模糊性或不确定性。

由上可知,数据是信息的素材,是信息的载体;而信息则是对数据进行加工的结果,是对数据的解释,是数据的内涵。数据与信息的关系如图 1-1 所示。这里指出,尽管数据与信息的概念是有区别的,但在某些场合人们通常并不去严格地区分它们。

数据库系统的每项操作,均是对数据进行某种处理。数据输入计算机后,经存储、传送、排

序、计算、转换、检索、制表以及仿真等操作,输出人们需要的结果,即产生信息。

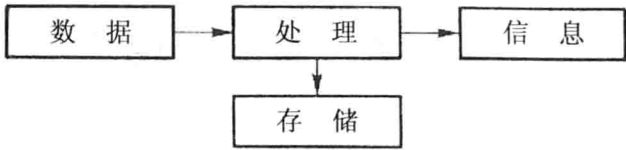


图 1-1 数据与信息的关系

1.1.2 数据库系统的特点

数据库系统(DataBase System,DBS)是引入数据库后的计算机系统。它由数据库、数据库管理系统及其开发工具、应用系统、数据库管理员和用户构成,如图 1-2 所示。

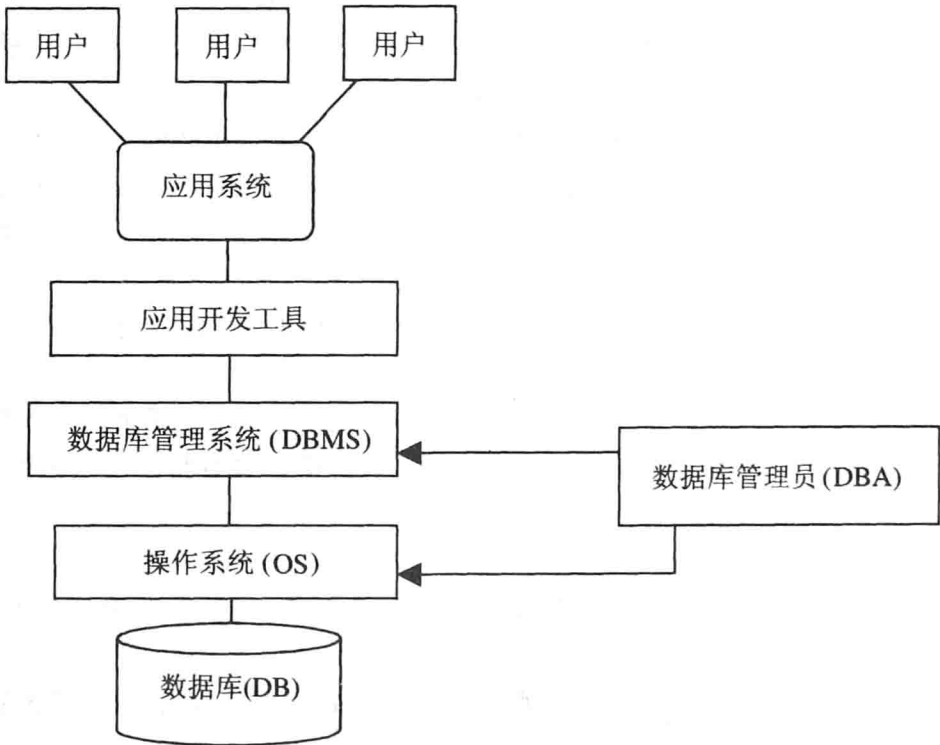


图 1-2 数据库系统的构成

数据的重要价值是使用而非收集,数据库系统就是为了方便使用数据而设计的,它对数据进行集中控制,能有效地维护和利用数据。

1. 数据是结构化的

这是与文件系统的根本区别。目的是节省空间、增强灵活性,可以为多个应用提供服务。在数据库系统中,不仅数据是结构化的,而且存取数据的方式也很灵活,可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录。

数据库系统中不仅要考虑某个应用的数据结构,还要考虑整个组织的数据结构,以便为各部门的管理提供必要的记录。例如,图 1-3 为一个学校的管理信息系统中,不仅要考虑学生的人事管理,还要考虑学籍管理、选课管理等,根据图中所示方式为该校的管理信息系统组织学生数据。

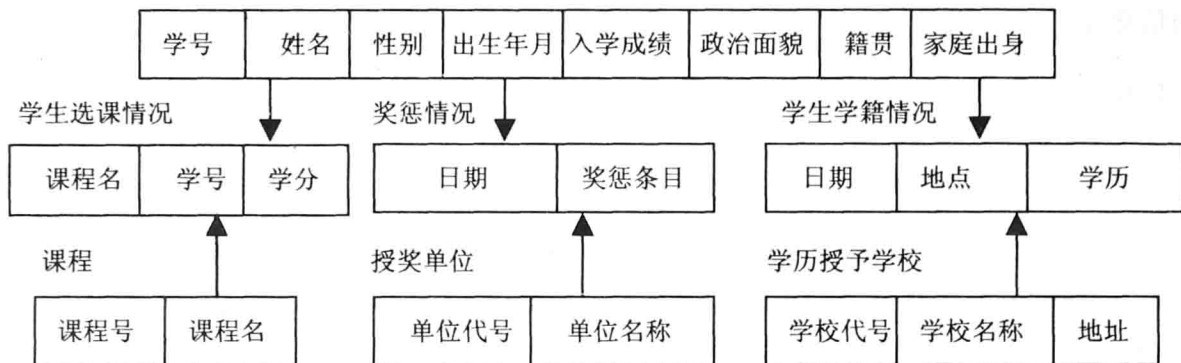


图 1-3 适应多种管理的学生数据记录

采用这种数据组织方式为多种管理提供必要的记录,使学校的学生数据结构化。这就要求在描述数据时不仅要描述数据本身,还要描述数据之间的联系。

2. 实现广泛的数据共享

- ①不同的用户可以并发地访问同一数据。
- ②具有广泛的适应性,有多种语言的接口。
- ③某个应用所使用的只是数据库的一个子集,不同的子集可以任意重叠。

3. 保证数据的安全可靠

①确保数据的安全存取。系统只对有权使用数据的用户授有限使用权,任何一个用户都不能无限制地使用数据库中的数据。系统提供一套有效的安全性检查功能和控制设施,确保数据的合法使用。

②保证数据的完整一致。数据完整性在系统应用中是十分重要的,如果某个应用程序破坏了数据的完整性,使其他的应用程序使用了不正确的数据,可能产生错误的处理结果,甚至会造成重大的经济损失。因此,为了保证用户所使用的数据是正确、合理和相容的,系统提供了数据完整性约束条件和控制机制。

③并发控制。不同的应用程序同时访问数据库,就有可能使数据受到损坏而失去完整性。系统提供的并发控制机制,就是用来排除和避免损坏数据的一致性,保证数据库的完整、准确。

④故障的发现、排除与系统的恢复。数据库系统在运行时,随时都会受到局部或全局性的破坏。系统提供一套完整的中断和后援方案,确保能及时发现并排除故障。

4. 尽可能减少数据的冗余度

数据库系统只能尽量减小,而不能消除数据的冗余。这是因为满足某些要求时,同一数据的多次存储是必须的,系统会对这些冗余进行控制,保证不会因此引起数据不一致性。

5. 保证数据独立性

当数据库系统的物理性质发生变化(如更换存储设备、改变组织方法等)和逻辑性质发生变化(如改变模式和外模式等)时,都不用修改应用程序。

6. 实现标准化

数据库对数据的集中控制管理便于实现数据的标准化。标准化的实施,有利于行业间、国家

间的信息交流与技术协作。

1.1.3 数据库系统的发展史

1. 第一代数据库系统

20 世纪 70 年代,主要的数据库系统都是网状和层次型的。其中,1968 年 IBM 公司研制的 IMS(Information Management System)是层次型数据库系统的典型代表,而 1969 年 CODASYL 系统,为 CODASYL(Conference on Data System Language)下属的数据库任务组提出的一个方案系统,也称 DBTG 系统。它是网状模型的典型代表。

在第一代的数据库系统中,无论是层次型的还是网状型的系统都支持数据库系统的三级模式结构和两级映像功能,可以保证数据与程序间的逻辑独立性和物理独立性;它们都使用记录型及记录型之间的联系来描述现实世界中的事物及其联系,并用存取路径来表示和实现记录型之间的联系;同时,它们都用导航式的 DML 来进行数据的管理。

这一时期,由于硬件价格相对较贵,各 DBMS 的实现方案都关注于能提供对信息的联机访问,注重处理效率的提高,以降低高价硬件的成本。

这两种数据库系统具有以下几个共同特点:

- ①支持三级模式的体系结构。
- ②独立的数据定义语言。
- ③导航的数据操纵语言。
- ④用存取路径来表示数据之间的联系。

2. 第二代数据库系统

20 世纪 70 年代末,IBM 公司 San Jose 研究实验室的高级研究员 Edgar Frank Codd 在《Communication of ACM》上发表了题为“A Relational Model of Data for Large Shared Data Banks”的文章提出了崭新的关系数据模型,这成为数据库发展历史上的一个划时代的标志和伟大的转折点。1981 年 IBM 推出了具有 System R 所有特性的数据库软件产品 SQL/DS。与此同时,美国加州大学伯克利分校也研究出了 INGRES 这一关系数据库系统的实验系统,被 INGRES 公司采用并发展成了 INGRES 数据库产品。自此,关系数据库系统如雨后春笋,相继出现众多商用的关系数据库产品,取代层次和网状关系数据库系统称为主流产品。

关系数据库系统采用的关系模型,是建立在严格的数学基础之上,概念简单清晰,使用关系(二维表)来描述现实世界中的事物及其联系,并用非过程化的 DML 对数据进行管理,易于用户理解和使用。凭借这种简洁的数据模型、完备的理论基础、结构化的查询语言和方便的操作方法,关系数据库系统深受广大用户的欢迎。20 世纪 80 年代,几乎所有新开发的数据库系统都是关系型的。

目前,关系数据库系统在全球信息系统中得到了极为广泛的应用,基本上满足了企业对数据管理的需求,世界上大部分企业的数据都是由这种关系数据库系统来管理的。

但是,随着数据库新的应用领域尤其是 Internet 的出现,传统的关系数据库受到了很大的冲击。其自身所具有的局限性也愈加明显,很难适应建立以网络为中心的企业级快速事务交易处理应用的需求。因为关系数据库是用二维表格来存放数据的,因此不能有效地处理大多数事务

处理应用中包含的多维数据,结果往往是建立了大量的表,用复杂的方式来处理,却仍然很难模仿出数据的真实关系。同时,由于 RDBMS 是为静态应用(例如报表生成)而设计的,因此在具有图形用户界面和 Web 事务处理的环境中,其性能往往不能令人满意,除非使用价格昂贵的硬件。

3. 第三代数据库系统

第二代数据库系统的数据模型虽然描述了现实世界数据的结构和一些重要的相互联系,但是仍不能捕捉和表达数据对象所具有的丰富而重要的语义,因此还只能属于语法模型。第三代的数据库系统将以更丰富的数据模型和更强大的数据管理功能为特征,从而可以满足更加广泛、复杂的新应用的要求。

新一代数据库技术的研究和发展导致了众多不同于第一、二代数据库的系统诞生,构成了当今数据库系统的大家族。这些新的数据库系统无论是基于扩展关系数据模型的(对象关系数据库),还是面向对象模型的;是分布式、C/S 还是混合式体系结构的;是在并行机上运行的并行数据库系统,还是用于某一领域的工程数据库、统计数据库、空间数据库,都可以广泛地称为新一代数据库系统。

1990 年,高级 DBMS 功能委员会发表了题为《第三代数据库系统宣言》的文章,提出了第三代 DBMS 应具有以下 3 个基本特征:

(1) 第三代数据库系统应支持数据管理、对象管理和知识管理

第三代数据库系统不像第二代关系数据库那样有一个统一的关系模型。但是,有一点应该是统一的,即无论该数据库系统支持何种复杂的、非传统的数据模型,都应该具有面向对象模型的基本特征。数据模型是划分数据库发展阶段的基本依据,因此第三代数据库系统应该是以支持面向对象数据模型为主要特征的数据库系统。但是,只支持面向对象模型的系统不能称为第三代数据库系统。第三代数据库系统除了提供传统的数据管理服务外,将支持更加丰富的对象结构和规则,应该集数据管理、对象管理和知识管理为一体。

(2) 第三代数据库系统必须保持或继承第二代数据库系统的技术

第三代数据库系统必须保持第二代数据库系统的非过程化数据存取方式和数据独立性,应继承第二代数据库系统已有的技术。这不仅能很好地支持对象管理和规则管理,而且能更好地支持原有的数据管理,支持多数用户需要的即时查询等功能。

(3) 第三代数据库系统必须对其它系统开放

数据库系统的开放性表现在:支持数据库语言标准;在网络上支持标准网络协议;系统具有良好的可移植性、可连接性、可扩展性和可互操作性等。

就数据库技术而言,在许多新的数据库应用领域前面,传统的数据库技术和系统已不能满足需求,对传统的数据库技术和研究工作提出了挑战。数据库技术面临的挑战主要表现在以下几个方面:

①环境的变化。数据库系统的应用环境由可控制的环境转变为多变的异构信息集成环境和 Internet 环境。

②数据类型的变化。数据库中的数据类型由结构化扩大至半结构化、非结构化和多媒体数据类型。

③数据来源的变化。大量数据将来源于实时和动态的传感器或监测设备,需要处理的数据量成倍剧增。

④数据管理要求的变化。许多新型应用需要支持协同设计和工作流管理。

为了应付这些挑战,许多数据库技术研究和实践人员认为有两条可行的途径。第一条可行途径是反思原先的研究和开发思路,将原有的思想和技术进行扩充、推广和转移来解决面临的难题。第二条可行路径是拓宽研究思路,研究全新的技术,提出新的数据库管理系统概念。实际上,只有结合这两个方面,才有可能开辟新的数据库技术研究局面。

1.1.4 组成数据库系统的软硬件

数据库系统由支持系统的计算机硬件设备,数据库、相关的计算机软件系统,以及开发管理数据库系统的人员组成。

1. 数据库系统对硬件的要求

数据库系统是建立在计算机硬件基础之上的,必须有相应的硬件资源支持才能运行。支持数据库系统的计算机的硬件资源需要 CPU、内存、外存及其他设备。外部设备主要包括某个具体的数据库系统所需的数据通信设备和数据输入/输出设备。

数据库系统因数据量大、数据结构复杂、软件内容多,因而要求其硬件设备能够快速处理它的数据。所以在配置数据库系统的硬件时,要满足以下 3 个方面的要求。

(1) 尽量大的内存

数据库系统的软件结构包括操作系统、数据库管理系统(DBMS)、应用程序及数据库,构成较为复杂。工作时它们都需要占用一定的内存作为程序工作区或数据缓冲区,因此,与其他计算机相比需要更大的内存。而内存的大小对数据库系统性能的影响非常明显,内存大可以建立较多、较大的程序工作区或数据缓冲区,以管理更多的数据文件和控制更多的程序进程,进行更复杂的数据管理,更快地进行数据操作。每种数据库系统对内存有一定的要求,如果内存达不到要求,便无法正常运行。

(2) 尽量大的外存

计算机外存主要有磁带、光盘和硬盘,其中硬盘为主要的存储设备。数据库系统要求硬盘有尽量大的容量,这样的优势在于:

①能够为数据文件和数据库软件提供足够的空间,满足数据和程序的存储要求。

②能够为系统的临时文件提供存储空间,保证系统的正常运行。

③数据检索时间缩短,加速数据存取速度。

(3) 计算机尽量快的数据传输速度

由于数据库的数据量大而操作复杂度不大,数据库运行时要经常进行内、外存之间的交换操作,要求计算机不仅有较强的运算能力,而且数据存取和数据交换的速度要快。对一般的系统来说,计算机的运行速度是由 CPU 和数据 I/O 的传输速度两者决定的,但对于数据库系统来说,数据 I/O 的传输速度是提高运行速度的关键问题,而提高数据传输速度则是提高数据库系统效率的重要部分。

2. 数据库系统的软件组成

(1) 操作系统

操作系统是所有计算机软件运行的基础,在数据库系统中它支持数据库管理系统(DBMS)

及主语言系统的工作。

(2) 数据库管理系统(DBMS)和主语言系统

数据库管理系统(DBMS)是为定义、建立、维护、使用及管理数据库而提供的有关数据管理的系统软件。

主语言系统是为应用程序提供诸如程序控制、数据输入/输出、功能函数、图形处理、计算方法等数据处理功能的系统软件。应用系统的设计与实现,需要将数据库管理系统(DBMS)和主语言系统的结合才能完成。

(3) 应用开发工具软件

应用开发工具是数据库管理系统(DBMS)为应用开发人员和最终用户提供的高效率、多功能的应用生成器等各种软件工具,如报表生成器、表单生成器、查询和视图设计器等,它们为数据库系统的开发提供了良好的环境和有效的支持。

(4) 应用系统及数据库

数据库应用系统包括:为特定应用环境建立的数据库、开发的各类应用程序及编写的文档资料。它们是一个有机整体。数据库应用系统涉及各个方面,如人工智能和计算机控制等。通过运行数据库应用系统,可以实现对数据库中数据的维护、查询、管理和处理操作。

数据库系统人员则由软件开发人员、软件使用人员和软件管理人员组成。其中软件开发人员包括系统分析员、系统设计员及程序设计员,他们的主要职责便是对数据库系统的开发;软件使用人员是数据库的最终用户,他们利用功能菜单、表格及图形用户界面等实现数据的查询与管理工作;软件管理人员是数据库管理员(Data Base Administrator, DBA),他们负责对数据库系统的全面管理与控制。数据库系统与计算机软、硬件的关系如图 1-4 所示。

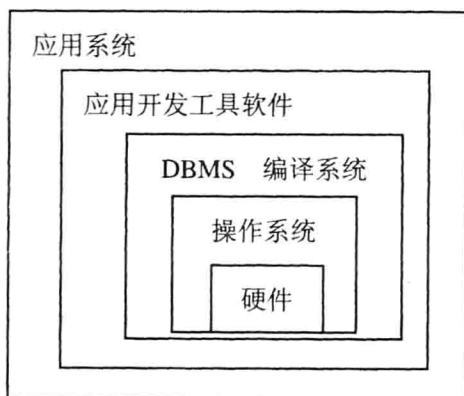


图 1-4 数据库系统与软、硬件的关系

1.2 数据模型

1.2.1 数据模型组成要素

数据模型是数据库系统的核心和基础,通常人们要将现实世界中的具体事物进行抽象、组织成 DBMS 所支持的数据模型,其一般过程是将现实世界抽象为信息世界,再将信息世界转换成数据世界,如图 1-5 所示。

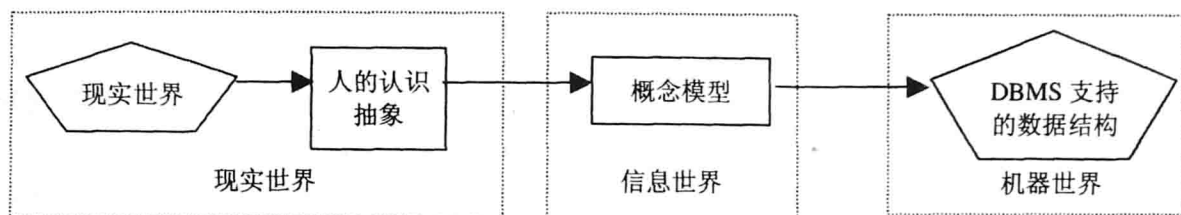


图 1-5 从现实世界到机器世界的过程

数据库的发展集中表现在数据模型的发展。从最初的层次数据模型、网状数据模型发展到关系数据模型,数据库技术产生了巨大的飞跃。数据库系统均是基于某种数据模型的。

模型是现实世界的模拟,在数据库技术中,模型是一组严格定义的概念的集合。数据库管理系统的一个主要功能就是将数据组织成一个逻辑集合,为系统定义该集合的数据及其联系的过程称为数据建模,其使用技术与工具则称为数据模型。数据模型就是关于数据的数学表示,包括数据的静态结构和动态行为或操作,结构又包括数据元素和元素间关系的表示。这些概念精确地描述了系统的静态特性、动态特性和完整性约束条件(Integrity Constraints)。

数据模型的三要素就是数据结构、数据操作和完整性约束三部分。

1. 数据结构

数据结构是数据模型最基本的组成部分,规定了如何把基本的数据项组织成较大的数据单位,以描述数据的类型、内容、性质和数据之间的相互关系。在数据库系统中,通常按照数据结构的类型来命名数据模型。例如,采用层次型数据结构、网状型数据结构、关系型数据结构的数据模型分别称为层次模型、网状模型和关系模型。数据结构是刻画一个数据模型性质最重要的方面。

数据结构描述了数据库的组成对象以及对象之间的联系。一般由两部分组成:①与对象的类型、内容、性质有关的,如网状模型中的数据项、记录,关系模型中的域、属性、关系等。②与数据之间联系有关的对象,如网状模型中的系型(Set Type)。

2. 数据操作

数据操作是指一组用于指定数据结构的任何有效的操作或推导规则,包括操作及有关的操作规则。常见的数据操作主要有两大类:检索和更新(包括插入、删除、修改)、数据模型定义的操作。此外,还有数据模型定义的操作。

数据模型要给出这些操作确切的含义、操作规则和实现操作的语言。因此,数据操作规定了数据模型的动态特性。

3. 完整性约束

完整性约束条件是一系列完整性规则的集合。完整性规则是给定的数据模型中数据及其联系所具有的制约和依存规则,用于限定符合数据模型的数据库状态以及其变化,以确保数据正确、有效地相容。

完整性约束的定义对数据模型的动态特性作了进一步的描述与限定。因为在某些情况下,若只限定使用的数据结构及可在该结构上执行的操作,仍然不能确保数据的正确性、有效性和相

容性。为此,每种数据模型都规定了通用和特殊的完整性约束条件:

①通用的完整性约束条件。通常把具有普遍性的问题归纳成一组通用的约束规则,只有在满足给定约束规则的条件下才允许对数据库进行更新操作。例如,关系模型中通用的约束规则是实体完整性和参照完整性。

②特殊的完整性约束条件。把能够反映某一应用所涉及的数据所必须遵守的特定的语义约束条件定义成特殊的完整性约束条件。例如,关系模型中特殊的约束规则是用户定义的完整性。数据结构、数据操作和数据的约束条件又称为数据模型的三要素。

1.2.2 基本数据模型

模型是对现实世界的抽象。在数据库技术中,用模型的概念描述数据库的结构与语义,对现实世界进行抽象。即数据模型是现实世界数据特征的抽象,是用来描述数据的一组概念和定义。换言之,数据模型是能表示实体类型及实体间联系的模型。

数据模型的种类很多,按照不同的应用层次可将其划分为概念数据模型和逻辑数据模型,如图 1-6 所示。

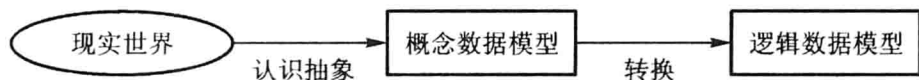


图 1-6 数据模型

1. 概念数据模型

概念数据模型简称为概念模型,也称为信息模型。它是一种独立于计算机系统的数据模型,完全不涉及信息在计算机中的表示,只是用来描述某个特定组织所关心的信息结构,是对现实世界的第一层抽象。概念模型是按用户的观点对数据进行建模,强调其语义表达能力,概念应该简单、清晰、易于用户理解,它是对现实世界的第一层抽象,是用户和数据库设计人员之间进行交流。

概念模型是按用户的观点来对数据进行建模,强调的是语义表达能力。概念模型的设计方法很多,其中最早出现的、最著名的、最常用的方法便是实体—联系方法(Entity-Relationship Approach, E-R 方法),即用 E-R 图来描述现实世界的概念模型。

E-R 数据模型的基本思想是:首先设计一个概念模型,它是现实世界中实体及其联系的一种信息结构,并不依赖于具体的计算机系统,与存储组织、存取方法、效率等无关,然后再将概念模型转换为计算机上某个数据库管理系统所支持的逻辑数据模型。因此,概念模型是现实世界到计算机世界的一个中间层。在 E-R 模型中,只有实体、联系和属性三种基本成分,所以简单易懂、便于交流。

E-R 模型是各种数据模型的共同基础,也是现实世界的纯粹表示,它比数据模型更一般、更抽象、更接近现实世界。

E-R 模型包含 3 个基本成分:实体、属性和联系。

实体(entity)是可区别且可被识别的客观存在的事、物或概念,它是一个数据对象。例如,一把椅子、一个学生、一个产品、一个部门等都是一个实体。具有共性的实体可划分为实体集(entity set)。实体的内涵用实体类型(entity type)表示。在 E-R 图中,实体以矩形框表示,实体名写

在框内。

属性(attribute)是实体所具有的特性或特征。一个实体可以有多个属性,例如,一个大学生有学生的姓名、学号、性别、出生年月、所属学校、院、系、班级、健康情况等属性。在 E-R 图中,属性以椭圆形框表示,属性名写在其中,并用线与相关的实体或联系相连接,表示属性的归属。对于多值属性可以用双椭圆形框表示,而派生属性则可以用虚椭圆形框表示。值得一提的是,不仅实体有属性,联系也可以有属性。

唯一标识实体集中的一个实体,又不包含多余属性的属性集称为标识属性,如实体“学生”的标识属性为“学号”。实体的一个重要特性是能唯一标识。

联系(relationship)表示一个实体集中的实体与另一个实体集中的实体之间的关系,例如,隶属关系、亲属关系、上下级关系、成员关系等。联系以菱形框表示,联系名写在菱形框内,并用连线分别将相连的两个实体连接起来,可以在连线旁写上联系的方式。通常,根据联系的特点和相关程度,联系可分为以下四种基本类型:

①一对一联系。一对一联系(记为 1:1)是指实体集 A 中的一个实体至多对应实体集 B 中的一个实体。例如,学生与教室座位,每位学生都具有一个座位。

②一对多联系。一对多联系(记为 1:N)是指实体集 A 中至少有一个实体对应于实体集 B 中的一个以上的实体。例如,班级与学生,每个班级有多名学生等。

③多对多联系。多对多联系(记为 M:N)是指实体集 A 中至少有一个实体对应于实体集 B 中的一个以上的实体,且实体集 B 中至少有一个实体对应于实体集 A 中的一个以上的实体。例如,学生与课程,每个学生选修多门课程,一门课程可供多名学生选读。

④条件联系。条件联系是指仅在某种条件成立时,实体集 A 中有一个实体对应于实体集 B 中的一个实体,当条件不成立时没有这种对应关系。例如,职工姓名与子女姓名,仅当该职工有子女这个条件成立时,才有确定的子女姓名,对于没有子女的职工,其子女姓名为空。

属性又可分为原子属性和可分属性,前者是指不可再分的属性,后者则是还可以细分的属性。例如,在学生的属性中,学生的姓名、性别、出生年月、所属学校、院、系、班级都是原子属性,而健康情况则是可再细分为身高、体重、视力、听力等属性的可分属性。

属性的可能取值范围称为属性的值域,简称为属性域。属性将实体集中每个实体和该属性的值域中的一个值联系起来。一个实体诸属性的一组特定的属性值,就确定了一个特定的实体,实体的属性值是数据库中存储的主要数据。

2. 层次模型

层次模型(Hierarchical Model)是按照层次结构的形式组织数据库数据的数据模型,是三种传统的逻辑数据模型(层次模型、网状模型和关系模型)之一,是出现最早的一种数据库管理系统的数据模型。

层次模型采用树形结构来表示实体及实体之间的联系,数据被组织成由“根”开始的“树”,每个实体由根开始沿着不同的分支放在不同的层次上。树中的每一个结点代表实体型,连线则表示它们之间的关系。根据树形结构的特点,要建立数据的层次模型需要满足如下两个条件:

①有且只有一个结点没有双亲结点,这个结点就是根结点。

②根结点以外的其他结点有且仅有一个双亲结点,这些结点称为从属结点。

层次模型是按照层次结构(即树型结构)来组织数据的,树中的每一个结点表示一个记录类

型,箭头表示双亲—子女关系。因此,层次模型实际上是以记录类型为结点的有向树,每一个结点除了具有①②性质外,还具有:

③由“双亲—子女关系”确定记录间的联系,上一层记录类型和下一层记录类型的联系是一对多联系。

层次模型的这种结构方式反映了现实世界中数据的层次关系,如机关、企业、学校等机构中的行政隶属关系及商品的分类等,比较简单、直观。

在层次模型中,每个结点表示一个记录类型,它对应实体联系模型中的实体,每个记录类型包含若干字段,它表示实体中的属性。记录类型及其字段都必须命名,各个记录类型不能同名,同一记录类型的各个字段也不能同名。父子结点之间用直线(有向边)相连,它们是一对多的联系,同一双亲的子结点称为兄弟结点(twin),没有子结点的结点称为叶子结点,对于层次模型还需做以下说明。

①结点所表示的记录类型的任何属性都是不可再分的简单数据类型,即具有原子性。

②层次模型中的树为有序树,规定树中任一结点的所有子树的顺序都是从左到右的,这一限制隐含了对层次模型数据库的存取路径的一种控制。

③树中实体间的联系是单向的,即由父结点指向子结点,而且一对父子结点不存在多于一种的联系。这一规定限制了两个实体间可能存在的多种联系的建模。

④层次模型中的联系只能是双亲结点对子结点的一对多的联系,这一规定限制了层次模型对多对多联系的直接表示。

层次模型不能直接表示多对多的联系,通过层次模型表示多对多的联系时,必须首先将其分解成多个一对多的联系。通常的分解方法有两种:冗余结点法和虚结点法。

①冗余结点法,采用冗余结点法就是增加两个结点,将多对多的联系转换成两个一对多的联系,具体如图 1-7 所示。

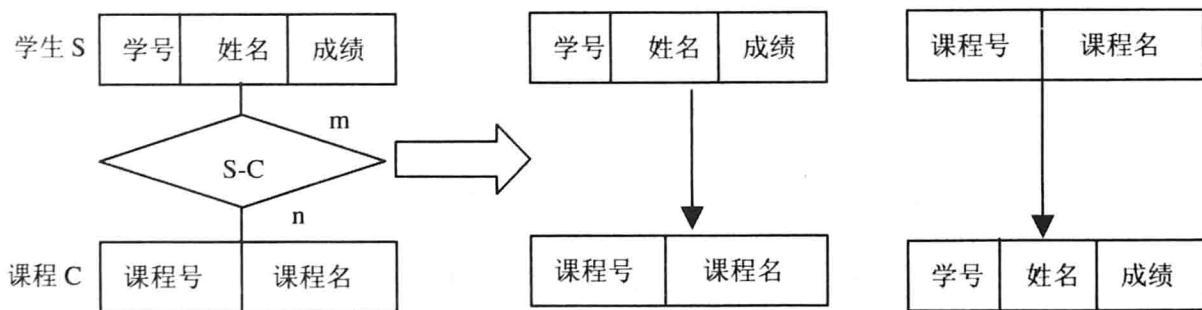


图 1-7 冗余结点法

②虚结点法,采用虚结点法就是将冗余结点法中的冗余结点换为虚拟结点,这个结点存放指向该虚结点所代表的结点的指针,具体如图 1-8 所示。

一般来说,冗余结点法结构清晰,允许结点改变存储位置,但是占用存储空间大,有潜在的不一致性;虚拟结点法占用存储空间小,能够避免潜在的不一致问题,但是改变存储位置时可能会引起虚拟结点指针的改变。

在层次模型中,通过指针来实现记录之间的联系,查询效率较高。但是,由于层次数据模型中的从属结点有且仅有一个双亲结点,所以它只能描述 1:M 联系,且复杂的层次使得数据的查询和更新操作比较复杂。因此,需要使用其他的数据模型来描述实体间更复杂的联系。