

陈晓云 徐玉生 编著

Shujuku 数据库 原理与设计 yuanli yusheji



兰州大学出版社

S hujuku
数据库
原理与设计
yuanliyusheji

陈晓云 徐玉生 编著



兰州大学出版社

图书在版编目(CIP)数据

数据库原理与设计/陈晓云,徐玉生编著. —兰州:兰州大学出版社,2009. 2

ISBN 978-7-311-03188-6

I. 数... II. ①陈... ②徐... 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(2009)第 013528 号

策划编辑 陈红升
责任编辑 郝可伟 陈红升
封面设计 张芳芳

书 名 数据库原理与设计
编 著 陈晓云 徐玉生
出版发行 兰州大学出版社 (地址:兰州市天水南路 222 号 730000)
电 话 0931-8912613(总编办公室) 0931-8617156(营销中心)
0931-8914298(读者服务部)
网 址 <http://www.onbook.com.cn>
电子信箱 press@onbook.com.cn
印 刷 兰州德辉印刷有限责任公司
开 本 787×1092 1/16
印 张 21
字 数 523 千字
版 次 2009 年 2 月第 1 版
印 次 2009 年 2 月第 1 次印刷
书 号 ISBN 978-7-311-03188-6
定 价 32.00 元

(图书若有破损、缺页、掉页可随时与本社联系)

前 言

数据库是计算机科学的一个最重要、最活跃的分支,从20世纪60年代出现的基于文件系统的第一个商用数据库管理系统以来,已经经历了大约50年的发展历程。数十年发展取得的知识和技术的结晶都凝聚在数据库管理系统(DBMS)软件之中。DBMS是实现数据冗余度最低、满足独立于软硬件实现技术、基于开放标准、具备广泛接口的强大的数据管理工具,DBMS是最复杂的软件系统之一。数据库技术根据数据模型的发展经历了三个重要阶段:第一代的网状、层次数据库系统;第二代的关系数据库系统;第三代的以面向对象模型为主要特征的数据库系统。关系代数、范式理论奠定了关系数据库基础,结构化查询语言(SQL)标准进一步推动了数据库技术的发展。随着DBMS和计算机网络技术的不断发展与成熟,数据库已经成为先进信息技术的重要组成部分,成为现代计算机信息系统和计算机应用系统的基础和核心。

鉴于数据库技术的重要性及其应用领域的广阔性,国内外高等学校理、工科的大多数计算机科学相关专业都系统地开设了数据库方面的课程。本书是参照国内外多所大学的本科数据库课程教学大纲编写的一本教材。

全书共分为十二章,内容涵盖了数据库基本理论、基础知识、数据建模、存储结构、事务、结构化查询语言、数据库设计、数据库优化、数据库安全以及数据库技术的最新发展等,较系统地介绍了数据库及相关知识。

第一章数据库系统概述,讲述了数据库相关的基本知识。第二章实体联系模型,涉及了实体联系模型与扩展实体联系模型的基本概念、实体联系模型的设计、以及实体联系模型向关系模式的转化。第三章关系数据库,主要内容为关系模型、关系代数、关系演算三个方面,是关系数据库的理论基础。第四章结构化查询语言,介绍了SQL的数据定义、数据查询、数据操作功能、嵌入式SQL、动态SQL以及特殊的空值处理问题。第五章关系模型规范化,主要内容包括函数依赖、多值依赖、范式和模式分解。第六章数据库设计,探讨了数据库设计、应用系统设计、业务规则处理,以及数据库设计工具简介。第七章存储结构,重点内容为多种索引结构及实现方法:顺序索引、B树索引、哈希索引、位图索引,以及在数据库管理系统中索引的管理。第八章存储过程与触发器,重点介绍DBMS的常见后台编程机制,包括存储过程、触发器、函数等。第九章事务,围绕事务的基本概念、事务的特性、事务的操作、事务的并发执行等问题详细介绍了冲突、事务的可串行化、基于锁的并发控制、基于图的协议、基于时间戳的协议等。第十章查询处理与查询优化,首先分析了查询的处理代价,在此基础上,探讨了查询的优化方法。第十一章安全性,主要内容为计算机系统安全性评测标准、数据库安全访问控制机制及实现技术、数据库备份与恢复技术、数据完整性约束等。第十二章数据库新技术,主要介绍数据自动化管理、分布式数据管理、事务性复制的应用、数据库应用系统的架构、XML技术与数据库、数据抽取转换与加载工具、数据仓库技术、联机分析处理、数据挖掘等。

本书有配套的电子版实习指导书,分为十二个相对独立的实践环节,要求学生独立完

成,其内容既强调了学生对数据库对象的管理与访问知识的掌握,又强调引导学生自学。为了加强对学生实践能力的培养,仅对实习目标、实习内容、实习项目规模做出详细要求,而实习题目由学生自选,实践项目、数据库对象、样例数据由学生自行设计、收集,尽量减少验证性实验内容。

本书内容丰富,教学过程中可以根据学生和专业情况适当调整。例如,如果强调理论教学,则第二、三、四、五、九、十等章节作为重点,反之,则可以适当减少理论教学课时,增加其他章节的教学时间。建议课堂讲授至少安排 54 学时、上机实习 36 学时。建议的教学与实习内容及课时分配如下表所示,仅供参考。

课堂讲授		教学实习	
讲授内容	课时分布	实习内容	课时分布
数据库系统概述	3	DBMS 安装、配置与初步使用	6
实体联系模型	3	创建与管理数据库对象	9
关系数据库	6	数据库对象访问	6
结构化查询语言	6	数据库性能测试	3
关系模型规范化	6	数据库访问控制与数据备份	6
数据库设计	3	数据库复制与自动管理	3
存储结构	6	数据库编程接口	3
存储过程与触发器	3		
事务	6		
查询处理与查询优化	6		
安全性	3		
数据库新技术	3		
总课时	54		36

本书可以作为高等学校计算机科学及相关专业的教学用书或教学参考书,也可以作为从事计算机软件开发的工程技术人员、数据库管理员的参考书、培训教材、自学用书等。

在学习本课程之前,建议先修“数据结构”、“程序设计语言(C、C++、Java等)”、“操作系统”等课程,以便更好地掌握本书原理部分的内容。

本书由陈晓云教授主编并编写大纲。参加编写工作的主要人员有:陈晓云、徐玉生、王雪、李龙杰等。参加本书编写、校对、案例代码验证、实验指导书编写的教师和研究生还有:苏伟、燕昊、程一帆、牛国鹏、穆进超、刘慧玲、吴本昌、祁小丽、兰聪花、马强、陈毅、岳敏、何艳珊、陈鹏飞、张鑫、姚毓凯、刘国华、苏有丽等。本书的编写得到了兰州大学教务处和兰州大学信息科学与工程学院的支持,在此一并表示感谢。由于编者学识浅陋,时间仓促,尽管我们做了最大努力,书中难免有不妥或错误之处,恳请读者批评指正。

编者
2009年1月

目 录

第 1 章 数据库系统概述	(1)
1.1 数据库技术的基本概念	(1)
1.2 数据库管理系统的特点	(4)
1.3 数据库的三级模式结构	(6)
1.4 数据模型	(8)
1.5 数据库系统的组成	(10)
习题	(13)
第 2 章 实体联系模型	(14)
2.1 数据库设计的步骤	(14)
2.2 实体、属性和实体集	(15)
2.3 联系和联系集	(17)
2.4 实体联系图	(20)
2.5 扩展的实体联系图	(23)
2.6 数据库 E-R 模型的设计	(29)
2.7 E-R 模型向关系模式的转化	(37)
习题	(38)
第 3 章 关系数据库	(40)
3.1 关系模型	(40)
3.2 关系代数	(45)
3.3 关系演算	(53)
习题	(56)
第 4 章 结构化查询语言	(58)
4.1 结构化查询语言概述	(58)
4.2 数据定义	(59)
4.3 数据查询	(64)
4.4 数据更新	(80)
4.5 视图	(83)
4.6 空值	(84)
4.7 嵌入式 SQL	(85)
4.8 动态 SQL	(88)
4.9 其他 SQL 语句	(88)
习题	(88)
第 5 章 关系模型规范化	(90)
5.1 问题的提出	(90)

5.2	函数依赖	(94)
5.3	模式分解	(102)
5.4	多值依赖	(105)
5.5	其他类型的依赖	(109)
	习题	(110)
第6章	数据库设计	(111)
6.1	数据库设计概述	(111)
6.2	数据库设计的内容	(112)
6.3	数据库系统设计的过程与方法	(117)
6.4	数据建模方法	(125)
6.5	数据建模工具	(143)
	习题	(148)
第7章	存储结构	(150)
7.1	存储介质和文件结构	(150)
7.2	索引	(158)
7.3	创建索引	(169)
	习题	(171)
第8章	存储过程与触发器	(172)
8.1	存储过程	(172)
8.2	触发器	(182)
8.3	函数	(189)
8.4	作业	(194)
	习题	(198)
第9章	事务	(201)
9.1	事务的 ACID 属性	(201)
9.2	事务的并发执行	(205)
	习题	(224)
第10章	查询处理和查询优化	(226)
10.1	查询处理	(226)
10.2	查询优化	(241)
	习题	(250)
第11章	安全性	(251)
11.1	计算机安全性概论	(251)
11.2	数据库安全访问控制机制	(253)
11.3	视图机制	(260)
11.4	审计	(260)
11.5	数据加密	(261)
11.6	统计数据库安全性	(262)
11.7	数据库的恢复	(263)

11.8	数据库的完整性	(276)
11.9	数据库安全新技术	(284)
	习题	(285)
第 12 章	数据库新技术	(286)
12.1	数据自动化管理	(286)
12.2	分布式数据管理	(287)
12.3	C/S 结构与 B/S 结构	(297)
12.4	XML 技术	(300)
12.5	数据抽取、转换和加载工具	(304)
12.6	数据仓库	(311)
12.7	联机分析处理	(316)
12.8	数据挖掘	(323)
	习题	(329)
参考文献	(330)

第 1 章 数据库系统概述

人类社会步入信息时代的今天,信息资源已成为最重要和最宝贵的资源之一,信息资源的开发利用是信息化的重要内容,建立一个能够满足各级部门信息处理要求的、行之有效的信息系统成为一个组织生存和发展的重要条件。数据库技术是数据管理的最新技术,是信息资源开发利用的关键技术之一,是计算机学科中的一个重要分支。1968 年世界上第一个 IMS(Information Management System)系统诞生以来,数据库技术得到了迅猛发展,其应用已渗透到工农业生产、商业、金融、行政、科学研究、工程技术和国防军事等领域的各个部门,伴随着因特网的出现,其遍布社会的各个角落。

本章主要介绍数据库技术的常用术语和基本概念,力求对数据库建立一个粗浅的、轮廓性的认识。

1.1 数据库技术的基本概念

1.1.1 数据和信息

在信息化的社会里,信息以惊人的速度增长,如何有效地组织和利用它们成为急需解决的问题。而信息和数据是分不开的,它们既有联系又有区别。

信息在数据处理领域可以被理解为关于现实世界事物存在方式或运动状态的反映。信息有许多重要特征: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。这些特征构成了信息的最重要的自然属性。同时,信息又是社会上各行各业不可缺少的重要资源,这是信息的社会属性。

为了记载信息,人们使用各种符号和它们的组合来表示信息,这些符号及其组合就是数据。数据是数据库中存储的基本对象。它的表现形式是多样的,除了数值形式外,还包括文字、图形、图像、声音等,都称为数据。

所以数据可以被定义为:描述事物的符号。描述事物的符号可以是数字,也可以是文字、图形、图像、声音和语言等,即数据有多种表现形式,它们都可以经过数字化后存入计算机系统。

为了了解世界,交流信息,人们需要描述这些事物。在日常生活中直接用自然语言(如汉语)描述。例如可以这样来描述某校一位学生的基本情况:张山,年龄 20 岁,男,甘肃人,2007 年入学。在计算机中,为了存储和处理这些信息,就要抽取对对这些事物感兴趣的特征组成一个记录来描述,所以在计算机中常常这样来描述:(张山,20,男,甘肃,2007),这样的一条记录就是数据。对于这条记录,不能了解其语义的人则无法理解其含义。例如,20 是一个数据,可以是该学生的年龄,也可以是某门课的成绩,还可以是某个班的人数,可见,

数据的形式还不能完全表达其内容,需要通过语义解释。数据的解释是指对数据含义的说明,数据的含义称为语义,数据与其语义是不可分的。

信息和数据的关系可以归纳为:数据是信息的载体;信息是数据的内涵,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,但通常使用时并不严格区分它们。

1.1.2 数据库

数据库(Database,简称为DB),可以直观地理解为存放数据的仓库,只不过这个仓库是在计算机的大容量存储器上,如硬盘就是一类最常见的计算机大容量存储设备,而且数据必须按照一定的格式存放。可以将数据库简单地归纳为:数据库是按一定结构组织并长期存储在计算机内的、可共享的大量数据的有机集合。

1.1.3 数据库管理系统

了解了数据和数据库的概念,下一个问题就是如何科学地组织和存储数据,如何高效地存取和维护数据。数据库管理系统(Database Management System,简称为DBMS)便是完成这一任务的系统软件,它是数据库和用户之间的一个接口,其主要作用是在数据库建立、运行和维护时对数据库进行统一管理和控制。可以从以下几个方面来理解DBMS:

(1)从操作系统角度看,DBMS是操作系统的使用者,它建立在操作系统的基础之上,需要操作系统提供底层服务,如创建、撤销进程,读写磁盘文件,管理处理机和管理内存等。

(2)从数据库角度看,DBMS是管理者,是数据库系统的核心,是为数据库的管理、运行、控制和维护而配置的系统软件,负责对数据库进行统一的管理和控制。

(3)从用户角度看,DBMS是工具或桥梁,是位于操作系统与用户之间的一层数据管理软件。用户发出的或应用程序中的各种操作数据库的命令,都要通过DBMS来执行。

DBMS的主要功能包括以下几个方面:

(1)数据定义功能

DBMS提供数据定义语言(Data Definition Language,简称DDL),用户使用该语言可以方便地对数据库中的数据对象进行定义。

(2)数据操作功能

DBMS还提供数据操作语言(Data Manipulation Language,简称DML),用户可以使用DML操作数据实现对数据库的基本操作,如查询、插入、删除和修改等。

(3)数据库的运行管理功能

数据库在建立、运用和维护时由数据库管理系统统一管理、统一控制,以保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

(4)数据库的建立和维护功能

包括数据库初始数据的输入、转换功能,数据库的转储、恢复功能,数据库的重组功能和性能监视、分析功能等,这些功能通常是由DBMS提供的一些实用程序完成的。

数据库管理系统是数据库系统的一个重要组成部分,常用的DBMS有:Oracle、SQL Server、Informix、Sybase、FoxPro、Access等。

1.1.4 数据库系统

数据库系统(Database System,简称 DBS)是一个复杂的系统,它是在计算机系统中引入数据库后的系统。数据库系统是存储介质、处理对象和管理系统的集合,一般由数据库、数据库管理系统(及其开发工具)、软件系统、硬件系统、数据库管理员和用户构成。它们之间的关系如图1-1所示。有时人们也把数据库系统广义地定义为“数据库+数据库管理系统+数据管理+应用程序+用户”。数据库系统具有以下特点:信息完整和功能通用、程序与数据独立、数据抽象、支持数据的不同视图、控制数据冗余、支持数据共享、限制非授权的存取、提供多种用户界面、表示数据之间的复杂联系、完整性约束、数据恢复等。

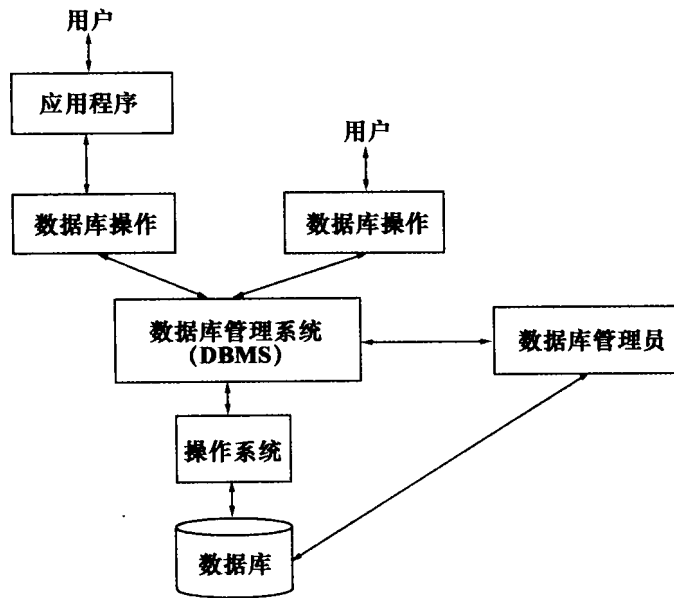


图1-1 数据库系统

1.1.5 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学科。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研究内容。程序是使用数据库系统最基本的方式,因为系统中大量的应用程序都是用高级语言加上数据库的操作语言编写的。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很强的学科。

1.1.6 数据管理技术的发展

数据库管理技术经历了手工管理、文件管理和数据库技术三个发展阶段。数据库的出现改变了传统的信息管理模式,扩大了信息管理的规模,提高了信息的利用能力,缩短了信息传播的过程,实现了世界信息一体化的目标。如今,数据库技术仍在快速发展,其应用仍

在深入。

按照数据模型,数据库技术的发展也经历了三个阶段:网状模型与层次模型、关系模型、面向对象模型。这四种模型中,由于关系模型有着坚实的理论基础(离散数学)、成熟的DBMS产品而广受应用,当前已建立的绝大多数数据库系统都是基于关系模型的,本书主要讲述关系模型。

1.2 数据库管理系统的特点

1. 数据结构化

数据结构化是数据库和文件系统的本质区别。数据结构化是按照一定的数据模型来组织和存放数据,也就是采用复杂的数据模型表示数据结构。数据模型不仅描述数据本身的特点,还描述数据之间的联系。这种结构化的数据反映了数据之间的自然联系,是实现数据的集中控制和减少数据冗余的前提和保证。

由于数据库是从一个企事业单位的总体应用来全盘考虑并集成数据结构的,所以数据库中的数据不再是面向个别应用而是面向系统的。各个不同的应用系统所需的数据只是整体模型的一个子集。数据库设计的基础是数据模型。在进行数据库设计时,要站在全局需要的角度抽象和组织数据,要完整地、准确地描述数据自身和数据之间联系的情况,建立适合总体需要的数据模型。数据库系统是以数据库为基础的,各种应用程序应建立在数据库之上。数据库系统的这种特点决定了它的设计方法,即系统设计时应先设计数据库,再设计功能程序,而不能像文件系统那样,先设计程序,再考虑程序需要的数据。

2. 有较高的数据独立性

独立性就是数据与应用程序之间不存在紧密的相互依赖关系,二者之间是一种松散的耦合,也就是数据的逻辑结构、存储结构和存储方法等不因应用程序的修改而修改,反之亦然。数据独立性分为物理独立性和逻辑独立性两种。

(1) 物理独立性是指数据的物理结构(或存储结构)的改变,如物理存储设备的更换、物理存储位置的变更、存取方法的改变等,不影响数据库的逻辑结构,不会引起应用程序的修改。

(2) 逻辑独立性是指数据库总体逻辑结构的改变,如修改数据的定义、增加新的数据项及数据类型、改变数据间的联系等等,无需修改原来的应用程序。

总之,数据独立性就是数据与应用程序之间的互不依赖性。一个具有数据独立性的系统称为以数据为中心的系统或面向数据的系统。显然,数据库系统是以数据为中心的系统。

3. 数据冗余度小、数据共享度高

数据冗余度小是指存储在数据库中的重复数据少。在非数据库系统中,每个应用程序有它自己的数据文件,从而造成存储数据的大量重复。由于在数据库系统方式下,数据不再是面向某个应用,而是面向整个系统,这就使得数据库中的数据冗余度小,从而避免了由于数据大量冗余带来的数据冲突问题,也避免了由此产生的数据维护麻烦和数据统计错误问题,节约了存储空间。

数据库系统通过数据模型和数据控制机制提高数据的共享性。数据共享度高会提高数据的利用率,使得数据更有价值,能够更容易、更方便地使用。数据高度共享使得数据库具有以下几个方面的优点:

- (1) 系统现有用户或程序可以共享数据库中的数据；
- (2) 当系统需要扩充时,再开发的新用户或新程序还可以共享原有的数据资源；
- (3) 多用户或多程序可以在同一时刻共同使用同一数据。

4. 避免了数据的不一致性

当本应相同的数据项在不同的应用中出现不同的数据值时,便出现了数据的不一致性。数据库在理论上可以避免数据冗余,因而也可以避免数据的不一致性。即使存在某些冗余,数据库系统也提供对数据的各种控制和检查,保证在更新数据时更新所有的副本,从而保证数据的一致性。

5. 完善的数据控制功能

(1) 并发控制

避免并发程序之间相互干扰,防止数据库被破坏,避免给用户提供不正确的数据,数据库系统对多用户的并发操作加以控制和协调,也就是避免当多个用户的并发进程同时存取、修改数据库时,可能会发生相互干扰而得到错误的结果或使数据库的完整性遭到破坏。

(2) 实施安全性保护

数据的安全性主要指数据保密,防止数据被非法使用、越权使用。只有数据库管理员(Database Administrator, DBA)拥有完全的操作权限,DBA 可以规定各用户的合理权限。数据库系统保证对数据的存取方法是唯一的。每当用户企图存取敏感数据时,数据库系统就进行安全性检查。在数据库中,对数据进行各种类型的操作(检索、修改、删除等),数据库系统都可以实施不同的安全检查。

(3) 保证数据的完整性

数据的完整性也就是数据的正确性、有效性和相容性,数据的不一致性是失去完整性的例子。数据冗余可能引起数据的不完整性、不一致性,但没有数据冗余,同样可能出现不正确的数据而使数据库失去数据完整性。例如,在一个数值型数据中出现了字母、特殊符号,或一个人的年龄小于0等,都是失去完整性的例子,与是否存在数据冗余无关。数据库系统的集中控制可以避免这些情况的出现,通过由 DBA 定义的完整性检查,对每一次更新操作实施完整性检查,保证数据的完整性。

(4) 发现故障与恢复

数据库在运行过程中很难保证不受破坏,硬件或软件的故障及用户操作的失误,都可能使数据库遭到不同程度的破坏。数据库系统有一套及时发现故障,并迅速地把数据库恢复到故障以前正确状态的措施,如转储、日志、检查点等方法。

6. 事务

当现实世界中的事件发生改变时,在联机 DBMS 中,这些改变是通过一段叫做事务(Transaction)的程序来实时完成的。所以,事务是用户定义的一个数据库序列,这些操作要么全做要么全不做,是一个不可分割的工作单位。它与程序是有区别的,一个程序可以包含多个事务。

事务处理系统(Transaction Processing System, TPS)包括一个或多个存储企业状态的数据库、用以操作企业的管理事务软件,以及组成应用代码的事务本身。一般地,最简单的事务处理系统只有一个 DBMS,这个 DBMS 包含了处理事务的软件。更复杂一些的系统会包含多个 DBMS。

事务处理系统要求:(1)可用性高(High Availability);(2)可靠性高(High Reliability);(3)吞吐量高(High Throughput);(4)响应时间短(Low Response Time);(5)安全性高(High Security)。

1.3 数据库的三级模式结构

考察数据库系统的结构,可以选择多种不同的层次或不同的角度。从数据库管理系统的角度看,数据库系统通常采用三级模式结构。

1.3.1 数据库管理系统中数据的抽象级别

1. 概念模式

在数据模型中有“型”(Type)和“值”(Value)的概念。型是指对某一类数据的结构和属性的说明,值是型的一个具体赋值。例如:学生记录定义为(学号,姓名,性别,年龄,籍贯)这样的记录型,而(200701,张三,男,20,甘肃)则是该记录型的一个记录值。

概念模式也称为模式(Schema)或逻辑模式(Logical Schema),这是对数据库的数据的整体逻辑结构和特征的描述。它使用 DDL 进行定义。定义的内容包括数据库的记录型、数据型、数据项的型、记录间的联系等,以及数据的安全性定义(保密方式、保密级别和数据使用权)、数据应满足的完整性条件和数据寻址方式的说明。模式的一个具体值称为模式的一个实例(Instance),同一个模式可以有很多实例。模式是相对稳定的,而实例反映数据库某一时刻的状态,所以是相对变动的。

概念模式是数据库系统模式结构的中间层,既不涉及数据的物理存储细节和硬件环境,也与具体的应用程序、与所使用的应用开发工具及高级程序设计语言无关。

大多数数据模型都有用图表显示数据库模式的某些约定。使用图表显示的模式也被称为模式图。图 1-2 就是一个数据库的模式图,这个图显示的是每个记录类型的结构,而不是每个记录的具体值。

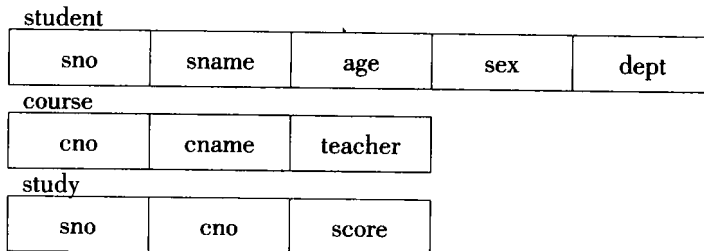


图 1-2 数据库模式图

2. 物理模式

物理模式也称为内模式(Internal Schema)或存储模式(Access Schema),内模式是数据库在物理存储方面的描述,是数据在数据库内部的表示方式,一个数据库只有一个内模式。它定义所有的内部记录类型、索引和文件的组织方式,以及数据控制方面的细节,例如,记录的存储方式是顺序存储还是链接存储;索引按照什么方式组织,是按照 B⁺ 树索引,还是哈希索引;数据是否压缩存储,是否加密;数据的存储记录结构有何规定等。

3. 外模式

外模式(External Schema)也称为子模式(Subschema)或用户模式(User Schema),是用户与数据库系统的接口,它是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述,是数据库用户的数据视图,是与某一应用有关的数据的逻辑表示,是用户用到的那部分数据的描述。外模式由若干个外部记录类型组成,用户使用数据操作语言 DML 对数据库进行操作,实际上是对外模式的外部记录进行操作。

总之,按照外模式的描述提供给用户,按照内模式的描述存储在磁盘上,而概念模式提供了连接这两级模式的相对稳定的中间点,并使得两级的任意一级的改变都不受另一级的牵制。

1.3.2 两层映像及数据独立性

数据库系统的三级模式是对数据的三个抽象级别,把数据的具体组织留给 DBMS 管理,使用户能够逻辑地、抽象地处理数据,而不必关心数据在计算机中的具体表示方式与存储方式。为了能够在内部实现这三个抽象层次的联系和转换,数据库管理系统在这三个模式之间提供了两层映像,图 1-3 所示。

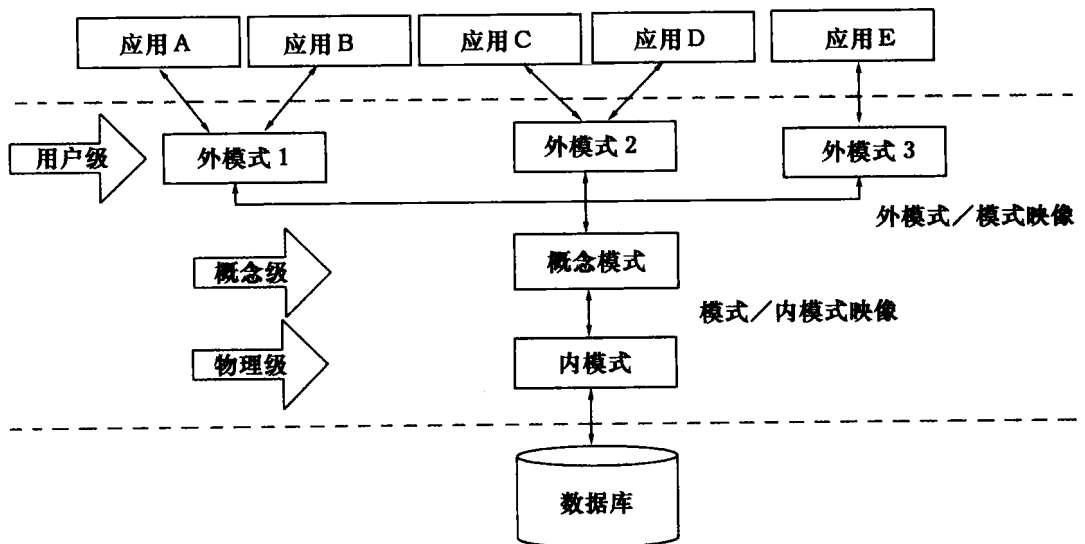


图 1-3 三级模式结构与两层映射

1. 外模式/模式映像

外模式/模式之间的映像,定义并保证了外模式与数据模式之间的关系。外模式/模式的映像定义通常在模式变化时,DBA 可以通过修改映像的方法使外模式不变;由于应用程序是根据外模式进行设计的,只要外模式不改变,应用程序就不需要修改。显然,数据库系统中的外模式与模式之间的映像不仅建立了用户数据库与逻辑数据库之间的对应关系,使得用户能够按子模式进行程序设计,同时也保证了数据的逻辑独立性。

2. 模式/内模式映像

模式/内模式之间的映像,定义并保证了数据的逻辑模式与内模式之间的对应关系。这说明数据的记录、数据项在计算机内部是如何组织和表示的。当数据库的存储结构改变时,

DBA 可以通过修改模式/内模式之间的映像使数据模式不变化。由于用户或程序是按数据的逻辑模式使用数据的,所以只要数据模式不变,用户仍可以按原来的方式使用数据,程序也不需要修改。模式/内模式的映像技术不仅使用户或程序能够按数据的逻辑结构使用数据,还提供了内模式变化而程序不变的方法,从而保证了数据的物理独立性。

1.4 数据模型

数据模型是一种表示数据及其联系的模型,是对现实世界数据特征与联系的抽象反映。通俗地讲,数据模型就是现实世界的模拟,数据模型是数据库系统的核心和基础。现有的数据库系统均是基于某种数据模型的。现在使用最多的是关系模型:以表的集合来表示数据和数据之间的联系。其他模型还包括:层次模型、网状模型、面向对象模型等。

1.4.1 数据模型的组成要素

一般来讲,数据模型是严格定义的一组概念的集合。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。因此,数据模型通常由数据结构、数据操作和数据的约束条件三部分组成。

1. 数据结构

数据库是由对象组成的,包括两类:一类是与数据类型、内容、性质有关的对象,例如网状模型中的数据项、记录,关系模型中的域、属性、关系等;一类是与数据之间联系有关的对象,例如网状模型中的系型(Set Type)。这些对象的集合便构成数据结构。它是一个刻画数据模型性质最重要的方面,描述的是系统的静态特性。

2. 数据操作

数据操作是指对数据库中各类对象(型)的实例(值)允许执行的操作的集合,包括操作及有关的操作规则。数据库主要有检索和更新(包括插入、删除、修改)两大类操作。数据模型必须定义这些操作的确切含义、操作符号、操作规则(如优先级)以及实现操作的语言。数据操作是对系统动态特性的描述。

3. 数据的约束条件

数据的约束条件是一组完整性规则的集合。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,用以限定符合数据模型的数据库状态以及状态的变化,以保证数据的正确、有效、相容。数据模型应该反映和规定本数据模型必须遵守的基本的完整性约束条件。例如,在关系模型中,任何关系必须满足实体完整性和参照完整性两个条件。

此外,数据模型还应该提供定义完整性约束条件的机制,以反映具体应用所涉及的数据必须遵守的特定的语义约束条件。例如,在学校的数据库中规定大学生入学年龄不得超过30岁等。

1.4.2 关系模型

关系模型(Relational Model)用二维表表示数据和数据之间的联系。每个表有多个列,每列有唯一的列名。

关系模型是基于记录模型的一种。基于记录模型的名称的由来是它用一些固定格式的

记录来描述数据库结构。每张表包含某种特定类型的记录,每个记录类型定义了固定数目的字段(或属性)。表格的列对应于记录类型的属性。

关系数据模型是使用最广泛的数据类型。当今大量的数据库系统都是基于这种关系模型的。

1.4.3 层次模型与网状模型

层次模型、网状模型是 20 世纪 70 年代至 80 年代初期广泛流行的逻辑数据模型。层次模型和网状模型统称为非关系模型。关系模型的数据库系统在 20 世纪 70 年代开始出现,之后发展迅速,并逐步取代了非关系模型数据系统的统治地位。

1. 层次模型

层次模型是数据库系统中最早出现的数据模型,层次数据库系统采用层次模型作为数据的组织方式。层次数据库系统的典型代表是 IBM 公司的 IMS (Information Management System) 数据库管理系统。层次模型用树形结构来表示各类实体以及实体间的联系。

2. 网状模型

用有向图结构来组织数据的数据模型称为网状模型。这种有向图结构也称为网状结构。网状数据库系统采用网状模型作为数据的组织方式。网状数据模型的典型代表是 DBTG 系统,是 20 世纪 70 年代数据系统语言研究会下属的数据库任务组 DBTG (DataBase Task Group) 提出的一个系统方案。DBTG 系统虽然不是实际的软件系统,但是它提出的基本概念、方法和技术具有普遍意义。它对于网状数据库系统的研制和发展产生了重大的影响。后来不少的系统都采用 DBTG 模型或者简化的 DBTG 模型。

1.4.4 面向对象数据模型

由于关系模型比层次模型、网状模型简单灵活,因此,在数据处理领域中,关系数据库的使用已相当普遍。但是,现实世界存在着许多含有复杂数据结构(例如 CAD 数据、图形数据等)的应用领域,它们需要更高级的数据库技术表达这类信息。面向对象的概念最早出现在程序设计语言中,随后迅速渗透到计算机领域的每一个分支,现已使用在数据库技术中。面向对象数据库是面向对象技术与数据库技术相结合的产物,以满足一些新的应用需要,例如面向对象的程序设计环境(CASE)、计算机辅助设计与制造(CAD/CAM)、地理信息系统(GIS)、多媒体应用、基于 Web 的电子商务以及其他非商用领域中的应用。面向对象数据库系统支持的数据模型称为面向对象数据模型,它包括以下几个方面:

1. 对象(Object)

面向对象模型中最基本的概念是对象和类。对象是现实世界中实体的模型化。一切概念上的实体都可以称作对象,如一个数字、一个人、一本书等。对象是由一组数据结构和在这组数据结构上的操作的程序代码封装起来的基本单位。对象之间的接口由一组消息定义。

2. 对象标识 OID(Object Identifier)

面向对象数据库中的每个对象都有一个唯一、不变的标识,称为对象标识。对象通常与实际领域的实体对应。在现实世界中,实体中的属性值可能随着时间的推移会发生改变,但是每个实体的标识始终保持不变。相应地,对象的部分属性、对象的方法会随着时间的推移发生变化,但对对象标识不会改变。