



高等教育“十二五”规划教材

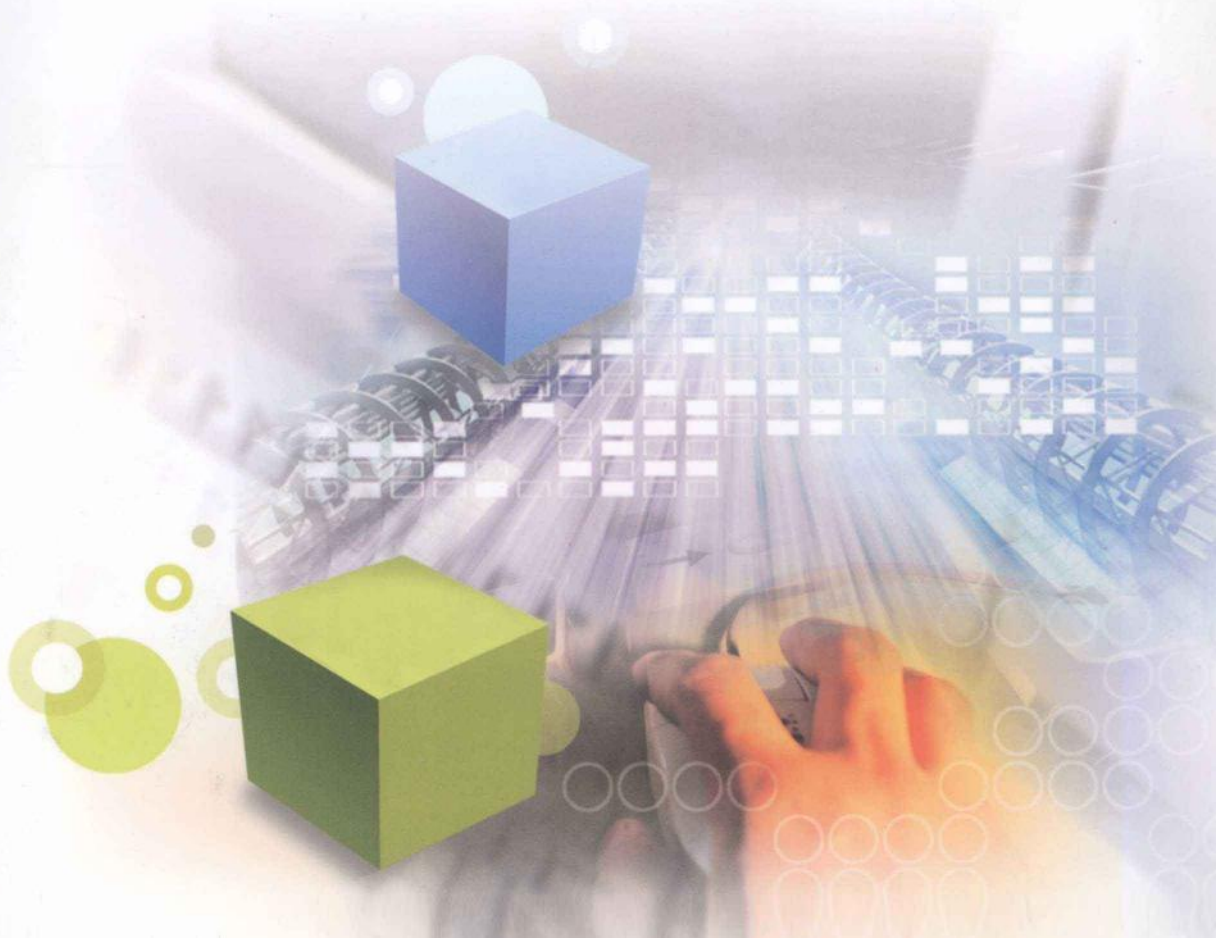
数据库系统与应用

Shujuku Xitong Yu Yingyong

主 编 赵文涛

副主编 张 静 刘艳霞

中国矿业大学出版社



高等

教材

数据库系统与应用

主 编	赵文涛		
副主编	张 静	刘艳霞	
参 编	申志浩	毛 钧	罗军伟
	吴 岩	王凤娥	刘小燕

中国矿业大学出版社

内 容 提 要

本书针对计算机及相近专业的数据库原理课程教学要求,介绍了关系数据库系统的基本原理和数据库应用的开发方法。本书共分13章,重点介绍了数据库理论的基本知识和概念,同时介绍数据库方面的最新研究成果、发展方向和数据库高级技术。

本书的特点是紧贴计算机专业本科教学需求、图文并茂,突出基本原理和基本概念的阐述,与实际应用和教学紧密结合,同时简介数据库的高级技术和最新技术成果。本书可作为计算机及相近专业数据库课程的教学用书,也可供其他人员作为自学或参考用书。

图书在版编目(CIP)数据

数据库系统与应用 / 赵文涛主编. —徐州: 中国矿业大学出版社, 2012. 1

ISBN 978 - 7 - 5646 - 1355 - 6

I. ①数… II. ①赵… 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(2011)第 268813 号

书 名 数据库系统与应用
主 编 赵文涛
责任编辑 仓小金
出版发行 中国矿业大学出版社有限责任公司
(江苏省徐州市解放南路 邮编 221008)
营销热线 (0516)83885307 83884995
出版服务 (0516)83885767 83884920
网 址 <http://www.cumtp.com> E-mail: cumtpvip@cumtp.com
印 刷 徐州中矿大印发科技有限公司
开 本 787×1092 1/16 印张 17.75 字数 443 千字
版次印次 2012 年 1 月第 1 版 2012 年 1 月第 1 次印刷
定 价 28.00 元

(图书出现印装质量问题,本社负责调换)

前 言

数据库技术一直是十分活跃的研究领域,是信息基础设施的核心技术和重要基础,同时也是一个极富挑战性的研究领域。从 20 世纪 60 年代中期产生至今已有 40 多年的历史,每年都有数千篇学术论文在重要学术会议和学术刊物上发表,每隔几年就会出现一大批新的挑战性问题,随之又会出现大量解决这些问题的研究成果和新产品。近年来,随着计算机软硬件系统、数据库应用、数据容量和类型的迅速变化,数据库领域的活跃程度和变化速度与日俱增。

本书的编写力争反映当前数据库技术的发展现状和我国数据库的应用水平,结合高等学校数据库系统原理课程教学的实际情况。全书共分 13 章,第一章介绍数据库的基本概念;第二章讲述关系模型和关系运算;第三章讲述关系数据库查询语言 SQL;第四章简述关系系统及其查询优化;第五章讲述关系数据的规范化理论;第六章讲述数据库设计;第七章讲述数据库恢复技术;第八章讲述数据库并发控制;第九章讲述数据库的安全性、完整性;第十章介绍数据库管理系统;第十一章讲述数据库接口技术;第十二章讲述高级数据库技术;第十三章讲述高级数据处理技术。

本书可作为高等学校计算机及相近专业本科数据库课程教材或教学参考书,也可供广大数据库研究与开发人员参考使用。

本书由河南理工大学组织编写,赵文涛任主编,张静、刘艳霞任副主编。编写分工为:申志浩编写第一、二章,刘艳霞编写第三章,张静编写第四、五、十章,吴岩编写第六、九章,毛钧编写第七、八章,罗军伟编写第十一章,赵文涛编写第十二、十三章。在本书的出版得到了中国矿业大学出版社、河南理工大学计算机科学与技术学院的大力支持,在此表示衷心感谢!

在本书的编写过程中,查阅了国内外大量数据库研究成果和文献,力求把数据库领域的新理论、新技术和新方法纳入本书,使之既包括数据库系统的基本理论、概念和技术,也能够反映数据库领域的最新进展。但时间紧迫,书中不足之处在所难免,希望得到广大读者的帮助和批评指正。

编者
2011 年 8 月

目 录

第一章 数据库概论.....	1
第一节 数据库系统概述.....	1
第二节 数据描述.....	8
第三节 数据模型	10
第四节 数据库系统的结构	18
第五节 数据库系统的组成	21
第六节 数据库技术的研究领域	23
第七节 当前流行大型数据库产品简介	23
本章小结	25
习题	25
第二章 关系模型和关系运算	27
第一节 关系模型概述	27
第二节 关系数据结构及形式化定义	28
第三节 关系代数	33
本章小结	39
习题	39
第三章 SQL 语言	41
第一节 概述	41
第二节 数据定义	43
第三节 数据查询	46
第四节 数据操纵	65
第五节 视图	68
第六节 数据控制	74
第七节 嵌入式 SQL	76
本章小结	79
习题	79
第四章 关系系统的查询优化	81
第一节 关系系统	81

第二节 查询优化处理	84
本章小结	90
习题	91
第五章 关系数据库的规范化理论	92
第一节 关系模式设计存在的问题	92
第二节 函数依赖	93
第三节 范式理论	98
* 第四节 模式分解	103
本章小结	107
习题	107
第六章 数据库设计	109
第一节 数据库设计概述	109
第二节 规划和需求分析	113
第三节 概念结构设计	118
第四节 逻辑结构设计	124
第五节 数据库物理设计	127
第六节 数据库的实施和维护	130
本章小结	132
习题	132
第七章 数据库恢复技术	133
第一节 事务的基本概念	133
第二节 故障的种类	136
第三节 恢复的实现技术	137
第四节 恢复策略	140
第五节 具有检查点的恢复技术	142
第六节 数据库镜像	144
第七节 SQLServer 的恢复技术	145
本章小结	147
习题	147
第八章 数据库并发控制	148
第一节 数据库的并发问题	149
第二节 并发控制方法	151
第三节 并发调度的可串行性	155
第四节 两段锁协议	156
第五节 SQLServer 的并发控制	157

本章小结	159
习题	159
第九章 数据库的安全性与完整性	160
第一节 计算机安全性概述	160
第二节 数据库的安全性	163
第三节 统计数据库安全性	181
第四节 数据库完整性	182
本章小结	185
习题	185
第十章 数据库管理系统	186
第一节 DBMS 概述	186
第二节 DBMS 的层次结构	191
本章小结	200
习题	200
第十一章 数据库接口技术	201
第一节 ODBC	201
第二节 ADO, NET	209
第三节 JDBC	217
本章小结	224
习题	224
第十二章 高级数据库技术	226
第一节 分布式数据库技术	226
第二节 面向对象数据库技术	232
第三节 并行数据库技术	234
第四节 多媒体数据库技术	237
第五节 时态数据库技术	242
第六节 空间数据库技术	244
第七节 移动数据库技术	247
第八节 主动数据库技术	249
第九节 数据流数据库技术	252
本章小结	256
习题	256
第十三章 高级数据处理技术	257
第一节 数据仓库与联机分析处理技术	257

第二节 数据挖掘技术.....	262
第三节 Web 信息检索与 Web 数据管理技术	266
本章小结.....	271
习题.....	272
参考文献	273

第一章 数据库概论

【本章重点】 数据库系统的基本概念;数据模型;数据库系统体系结构。

【本章难点】 数据库系统三级模式、二级映象及数据独立性原理;实体—联系模型表示概念模型的方法和实际应用。

数据库技术是计算机科学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。近年来,随着多媒体技术、空间数据库技术和计算机网络的飞速发展,数据库系统的发展十分迅速,应用领域愈来愈广。企事业单位、政府部门的行政管理、办公自动化,企业生产计划、物资管理,财务管理,车票、机票预定系统,铁路车次调度系统,宾馆、酒店管理系统,百货公司订购与销售系统,医院病房、病历管理,教学管理,图书馆管理,气象预报,地理信息系统(GIS),电子商务等都离不开数据库系统。在我国各行业中,数据库技术是应用最为广泛的软件技术之一。数据库的建设规模、数据库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。

第一节 数据库系统概述

一、基本概念

(一) 数据(Data)

数据是指存储在某一种媒体上能够识别的物理符号。数据的概念包括两个方面:其一是描述事物特性的数据内容;其二是存储在某一种媒体上的数据形式。由于描述事物特性必须借助一定的符号,这些符号就是数据形式。数据形式可以是多种多样的,例如某人的出生日期是“1990年2月17日”,当然也可以将该形式改写为“02/17/1990”,但其含义并没有改变。

在数据处理领域数据的概念已经大大地拓宽了,不仅仅指数字、字母、文字和其他特殊字符组成的文本形式的数据,而且还包括图形、图像、动画、影像、声音等多媒体数据。

一般来说,在计算机中处理数据可以采用记录的形式来表示,例如,对一个学生数据来说,可以通过学生姓名、性别、年龄和院系这些信息进行描述。我们可以采用这样的形式进行描述:

学生(王红,女,21,计算机学院)

数据的形式还不能完全表达其内容,需要经过解释。所以数据和关于数据的解释是不可分的,数据的解释是指对数据含义的说明,数据的含义称为数据的语义,数据与其语义是

不可分的。

(二) 数据库(Database, 简称 DB)

数据库可以直观地理解为存放数据的仓库, 只是这个仓库是在计算机的大容量存储器上。数据按一定的格式存放, 它不仅需要存放, 还要便于查找。所以可以认为数据库是被长期存放在计算机内、有组织的、可以表现为多种形式的可共享的数据集合。数据库技术使数据能按一定格式组织、描述和存储, 且具有较小的冗余度、较高的数据独立性和易扩展性, 并可为多个用户所共享。

人们总是尽可能地收集各种各样的数据, 然后对它们进行加工, 目的是要从这些数据中得到有用的信息。数据管理是指对数据的组织、分类、编码、存储、检索和维护。在社会飞速发展的今天, 人们接触的事物越来越多, 反映这些事物的数据量也急剧增加。过去人们手工管理和处理数据, 现在借助计算机来保存和管理复杂的大量数据, 这样就可能方便而充分地利用这些宝贵的数据资源, 数据库技术正是由于这一需求驱动而发展起来的一种计算机软件技术。

(三) 数据库管理系统

数据库管理系统(DataBase Management System, DBMS)是位于用户与操作系统之间的一层数据管理软件, 它的职能是有效地组织和存储数据、获取和管理数据、接受和完成用户提出的访问数据的各种请求。

数据库管理系统主要功能包括以下几个方面:

(1) 数据定义功能

DBMS 提供了数据定义语言 DDL(Data Definition Language), 用户通过它可以方便地对数据库中的相关内容进行定义。例如, 对数据库、表、索引进行定义。

(2) 数据操纵功能

DBMS 提供了数据操纵语言 DML(Data Manipulation Language), 用户通过它可以实现对数据库的基本操作。例如, 对表中数据的查询、插入、删除和修改。

(3) 数据库运行控制功能

对数据库的运行进行管理是 DBMS 的核心部分, 包括对数据库进行并发控制、安全性检查、完整性约束条件的检查和执行、数据库的内部维护(如索引、数据字典的自动维护)等。所有访问数据库的操作都要在这些控制程序的统一管理下进行, 以保证数据的安全性、完整性、一致性以及多用户对数据库的并发使用。

(4) 数据库的建立和维护功能

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, 数据库的重新组织功能和性能监视、分析功能等。这些功能通常是由一些实用程序完成的。它是数据库管理系统的一个重要组成部分。

DBMS 总是基于某种数据模型。现在世界上已经有了很多已经成熟的 DBMS 软件。例如, 大家熟悉的 ACCESS, FOXPRO, dBASE 等小型 DBMS 软件; 大型的 DBMS 软件如 DB2, ORACLE, SQL SERVER, INFORMIX 等。

(四) 数据库系统

数据库系统是指在计算机系统中引入数据库后的系统, 一般由数据库、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。数据库的建立、使用和维护等工作只靠一个 DBMS 远远不够,还要有专门的人员来完成,这些人被称为数据库管理员(Data-Base Administrator,简称 DBA)。

图 1-1 和图 1-2 反映了数据库系统组成及其在整个计算机系统中的地位。

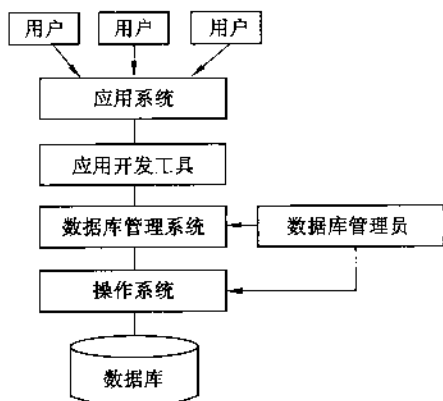


图 1-1 数据库系统

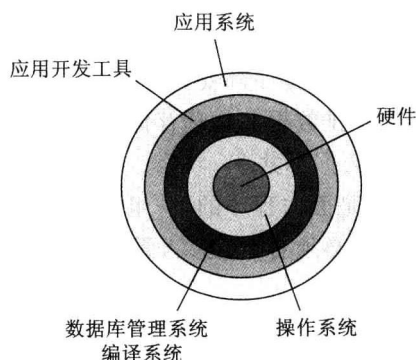


图 1-2 数据库在计算机系统中的地位

二、数据管理技术的产生和发展

计算机对数据的管理是指对数据的组织、分类、编码、存储、检索和维护。与其他技术的发展一样,计算机数据管理也经历了由低级到高级的发展过程,大致经历了如下三个阶段:人工管理、文件系统、数据库系统。这三个阶段特点及其比较如表 1-1 所示。

表 1-1 数据管理三个阶段的比较

		人工管理	文件系统	数据库系统
背景	应用背景	科学计算	科学计算、管理	大规模管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘
	软件背景	没有操作系统	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理,批处理	联机实时处理,分布处理批处理
特点	数据的管理者	人	文件系统	数据库管理系统
	数据面向的对象	某一应用程序 无共享,冗余度极大	某一应用程序 共享性差,冗余度大	整个应用系统 共享性高,冗余度小
	数据的独立性	不独立,完全依赖于程序	独立性差	具有高度的物理独立性和逻辑独立性
	数据的结构化	无结构	记录内有结构,整体无结构	整体结构化,用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

(一) 人工管理阶段

20 世纪 50 年代以前,计算机主要用于科学计算。当时没有管理数据的软件,数据处理方式是批处理。人工管理数据具有如下特点:

(1) 数据不保存在计算机内

因为当时计算机主要用于科学计算,对于数据保存的需求尚不迫切。

(2) 应用程序管理数据

每个应用程序都要包括数据的存储结构、存取方法、输入方式等,程序员编写应用程序时,还要安排数据的物理存储,因此程序员负担很重。

(3) 数据不共享

数据是面向程序的,一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时,由于必须各自定义,无法互相利用、互相参照,因此程序与程序之间有大量的冗余数据。

(4) 数据不具有独立性

程序依赖于数据,如果数据的类型、格式或输入输出方式等逻辑结构或物理结构发生变化,必须对应用程序做出相应的修改。程序与数据的对应关系如图 1-3 所示。

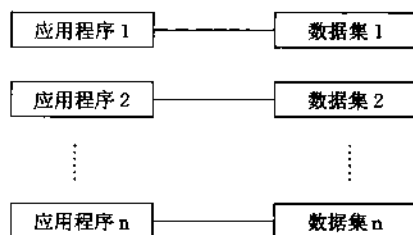


图 1-3 人工管理阶段应用程序与数据的对应关系

(二) 文件系统阶段

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,还用于信息管理。基于计算机的数据处理系统迅速发展起来。随着操作系统的出现,由操作系统中的文件系统专门管理外存中的数据。处理方式上不仅有了批处理,而且能够联机实时处理。

文件系统管理数据有如下特点:

(1) 数据可以长期保存

用户可随时对文件进行查询、修改和增删等处理。

(2) 文件系统管理数据

文件系统把数据组织成相互独立的数据文件,可以对文件进行增删改等操作。文件系统实现了记录内的结构性,但整体无结构。程序和数据之间由文件系统提供存取方法进行转换,使应用程序与数据之间有了一定的独立性。程序员只与文件名打交道,不必明确数据的物理存储,大大减轻了其工作的负担。

但文件系统仍存在以下缺点:

(1) 数据共享性差,冗余度大

文件系统中,一个文件基本上对应一个应用程序,不同的应用程序具有部分相同的数据

时,也必须建立各自的文件,而不能共享相同的数据,因此数据的冗余度大,浪费存储空间。容易造成数据的不一致性,给数据的修改和维护带来了困难。

(2) 数据独立性差

一旦数据的逻辑结构改变,必须修改应用程序,修改文件结构的定义。应用程序的改变,也将引起文件的数据结构的改变,因此数据与程序之间仍缺乏独立性。程序与数据之间的关系如图 1-4 所示。

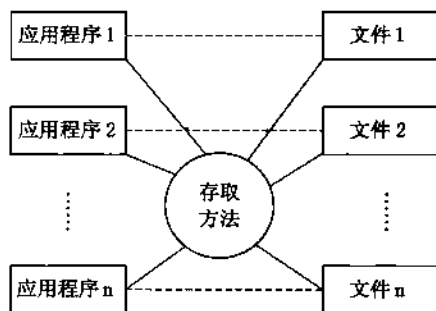


图 1-4 文件系统阶段应用程序与数据之间的对应关系

(三) 数据库管理阶段

20 世纪 60 年代后期以来,计算机应用规模越来越大,数据量急剧增长。文件系统管理数据已经不能满足应用的需求,为了解决多用户、多应用共享数据的要求,数据库技术应运而生,出现了统一管理数据的专门软件系统——数据库管理系统。数据库的特点是数据不再只针对某一特定应用,而是面向全组织,具有整体的结构性,共享性高,因此冗余度小,具有一定的程序与数据间的独立性,并且实现了对数据进行统一的控制。

从文件系统到数据库系统,标志着数据管理技术质的飞跃。20 世纪 80 年代后不仅在大、中型机上实现并应用了数据库管理系统,即使在微型计算机上也配置了经过功能简化的数据库管理系统,使数据库技术得到广泛的应用和普及。数据库系统阶段,程序与数据之间的关系可用图 1-5 表示。

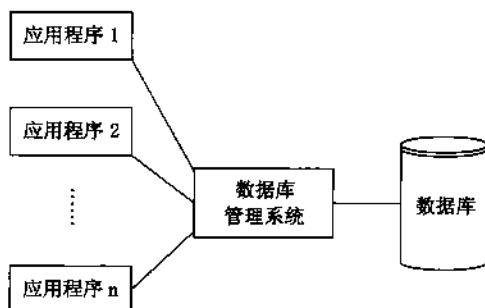


图 1-5 数据库管理阶段应用程序与数据之间的对应关系

(四) 分布式数据库管理阶段

分布式数据库系统是数据库技术、通信技术和网络技术相结合的产物,并随着这三种技术的发展而得到不断地发展和进步。传统的数据库管理属于单机式、主从式或客户服务器

模式,数据库和数据库管理系统集中安装在主机或服务器上,终端(客户)机访问主机(服务器)上的数据库,在终端(客户)机上完成数据输入输出或处理任务。此时,若主机(服务器)发生故障,则整个系统都处于瘫痪状态。因此,这种数据库系统的可靠性不高。

分布式数据库系统通过计算机网络和通信线路可以把分布在不同地域的、不同局域网环境下的、不同类型的数据库系统连接和统一管理起来。分布式数据库系统按照“就近存放,就近使用”原则,将数据库中的数据分别存储在计算机网络的各个节点上。分布式数据库系统既支持客户的局部应用,又支持客户的全局应用。具体地说,客户既可以访问与之连接的本地数据库,又可以自动访问到分布式数据库系统内的非本地数据库中的数据。对于系统内重要的数据,可以在不同地点存储多个副本,当本地的局部数据库系统发生故障时,可以自动到外地数据库系统中存取副本数据,从而提高了整个数据库系统的可靠性和稳定性。

分布式数据库除了具有一般数据库的优点之外,还具有系统的兼容性强、可靠性高、地域范围广、数据量大、客户数多等优点,缺点是由提高可靠性而带来的数据冗余,由不同局部数据库使用平台相连接和统一管理所带来的系统复杂性。

对于一般用户来说,在终端(客户)机上通过数据库应用软件使用分布式数据库和使用集中式数据库一样方便,感觉不到有什么差别,数据存取好像都是在本地进行的,系统内各站点之间的数据传输都是由分布式数据库管理系统在必要时自动完成的。

三、数据库系统的特点

(一) 数据结构化

文件系统中,相互独立的文件的记录内部是有结构的,但记录之间没有联系。数据库系统实现整体数据的结构化,是数据库的主要特征之一,也是数据库系统与文件系统的本质区别。数据库的数据组织方式使数据结构化,即在描述数据时不仅要描述数据本身,还要提供数据之间的联系。

例如,要建立学生成绩管理系统,系统包含学生(学号、姓名、性别、系别、年龄)、课程(课程号、课程名)、成绩(学号、课程号、成绩)等数据,分别对应三个文件。

若采用文件处理方式,因为文件系统只表示记录内部的联系,而不涉及不同文件记录之间的联系,要想查找某个学生的学号、姓名、所选课程的名称和成绩,必须编写一段较复杂的程序来实现。

而采用数据库方式,数据库系统不仅描述数据本身,还描述数据之间的联系,上述查询可以非常容易地联机查到。

(二) 数据的共享性高,冗余度低,易扩充

数据库系统从整体角度看待和描述数据,使得数据不再面向某个应用而是面向整个系统,因此数据可以被多个用户、多个应用共享使用。数据共享可以减少数据冗余,节约存储空间。数据共享还能够避免数据之间的不相容性与不一致性。下面介绍刚才所提到的几个概念。

① 数据冗余度:指同一数据重复存储时的重复程度。

② 数据的一致性:指同一数据不同拷贝的值一样。

③ 扩充性:由于数据面向整个系统,可以很容易增加新的应用,使得数据库的弹性大,

易于扩充。

④ 数据的不一致性:指同一数据不同拷贝的值不一样。

(三) 数据独立性高

数据的物理存储结构是指数据在计算机物理存储设备上的存储结构。在数据库中,数据在磁盘上的存储结构是由 DBMS 来管理和实现的,用户或应用程序不必关心。应用程序直接与数据的逻辑结构相关。数据的逻辑结构又分为局部逻辑结构和全局逻辑结构。不同的应用程序只与自己局部数据的逻辑结构相关。例如,学生的人事部门的应用程序与学生的人事数据相关,健康部门只关心学生的健康数据,教务部门只关心学生的学习成绩和选课数据。

数据库中数据的高度独立性,是指物理数据的独立性和逻辑数据的独立性两个方面。如图 1-6 所示,应用程序与数据的逻辑结构和物理存储结构之间的映射关系由 DBMS 完成。

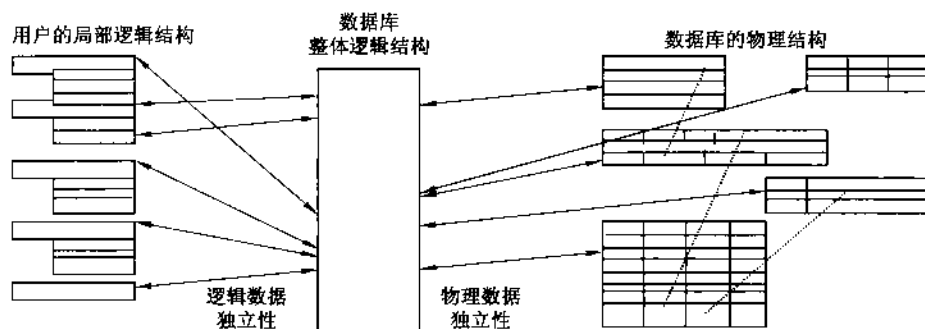


图 1-6 数据库逻辑结构和物理存储结构之间的映射关系

物理数据的独立性:全局逻辑数据结构独立于物理数据结构。即用户的应用程序与数据在数据库中的物理存储结构相互独立。

当数据的物理存储结构改变时,只要数据的逻辑结构不改变,用户的应用程序就不用改变。反之,如果应用程序改变了,数据的逻辑结构不变,数据的物理存储结构也不用改变。

逻辑结构的独立性:指数据的全局逻辑结构独立于局部逻辑结构。即用户的应用程序与数据的全局逻辑结构的相互独立性。

数据的独立性是由 DBMS 实现的两级映射关系来完成的。

(四) 数据由 DBMS 统一管理和控制

数据库的共享是并发的共享,即多个用户可以同时存取数据库中的数据甚至可以同时存取数据库中同一数据。为此,DBMS 还必须提供以下几方面的数据控制功能:

(1) 数据的安全性保护

数据的安全性是指保护数据以防止不合法的使用造成数据的泄密和损坏。使每个用户只能按规定对某些数据以某些方式进行使用和处理。

(2) 数据的完整性检查

数据的完整性指数据的正确性、有效性和相容性。

(3) 并发控制

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得数据库的完整性遭到破坏,因此必须对多用户的并发操作加以控制和协调。

例如,对于一个订票系统,如果只剩下一张票,但同时有两个用户要订票。由于同时操作,那么这两个进程将都看到还剩下--张票,会产生错误。因此,必须对并发操作加以控制和协调。

并发控制的种类和涉及的概念很多,在这里不作详细介绍,在第八章将详细介绍。

(4) 数据库恢复机制

计算机系统的硬件故障、软件故障、操作员的失误以及恶意的破坏也会影响数据库中数据的正确性,甚至造成数据库部分或全部数据的丢失。DBMS 必须具有将数据库从错误状态恢复到某一已知的正确状态(亦称为完整状态或一致状态)的功能。当软件、硬件或系统运行出现各种故障时,要确保存储在数据库的数据不被丢失和破坏,使数据库中存储的数据是永久性的数据。

第二节 数据描述

如何将复杂的各种相关联的事物以计算机和数据库技术所允许的形式存储在数据库中呢?这大体上分为三个步骤,即经历了现实世界(Realistic World)、信息世界(Information World)和计算机世界(Computer World)三个不同的世界的抽象和转换。利用抽象知识,使一个世界向另一个世界过渡,实现在数据库中反映事物的各种信息。这一过程如图 1-7 所示。

一、概念设计中的数据描述

存在于人们头脑之外的客观世界,称为现实世界。例如仓库管理中涉及的货物管理,包括货物的存放、进出和监察等。而信息世界是指现实世界在人们头脑中的反映,人们把它用文字和符号记录下来。信息世界涉及的概念主要有:

(1) 实体(Entity)

是指客观存在、并且可以相互区别的东西。实体可以是具体的对象,例如一名学生、一辆汽车等,也可以是抽象的事件,例如一次足球比赛,一次借书等。

(2) 实体集(Entity Set)

性质相同的同类实体的集合。所有学生、全国足球锦标赛的所有比赛等,都是实体集。

(3) 属性(Attribute)

实体的每一个特性称为属性,每个属性有一个值域,值域的类型可以是整数型、实数型或字符串型。

例如学生有学号、姓名、年龄和性别等属性,相应的值域可以为字符串、整数和字符串型。

(4) 键(Key)

是指能唯一标识每个实体的属性或属性集,有时也称作实体标识符。例如学生的学号可以作为学生实体的键。

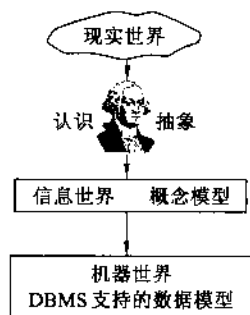


图 1-7 现实世界中客观对象的抽象过程

(5) 实体型(Entity Type)

具有相同属性的实体必然具有共同的特征和性质。用实体名及其属性名集合来抽象和刻画同类实体,称为实体型。例如:学生(学号,姓名,性别,出生年份,系,入学时间)就是一个实体型。

二、逻辑设计中的数据描述

信息世界的信息在机器世界中以数据形式存储。

机器世界涉及的概念主要有:

(1) 字段(Field)

标记实体属性的命名单位称为字段或数据项。字段是可以命名的最小信息单位,字段的命名往往和属性的命名相同,例如学生有学号、姓名、年龄、性别等字段。

(2) 记录(Record)

字段的有序集合称为记录,一般用一个记录描述一个实体。例如一个学生记录,由有序的字集学号、姓名、年龄、性别组成。

(3) 文件(File)

同一类型记录的集合称为文件,文件又可以定义为描述一个实体集的所有记录集。例如所有的学生记录组成了一个学生文件。

(4) 关键码(Key)

能唯一标识文件中每个记录的字段或字段集,称为记录的关键码(简称为键)。

机器世界与信息世界的术语对应关系如表 1-2 所示。

表 1-2 机器世界与信息世界的术语对应关系

信息世界	机器世界
实体	记录
属性	字段或数据项
实体集	文件
实体键	记录键

三、物理存储中的数据描述

计算机存储器中的概念主要有:

(1) 位(Bit)

一个二进制位称为“位”,一位只能取“0”或“1”状态。

(2) 字节(Byte)

8 个二进制位称为一个字节。

(3) 字(Word)

若干个字节组成一个字,一个字所含的二进制位的位数称为字长(有 8 位、16 位、24 位、32 位等)。

(4) 块(Block)或物理块、物理记录

每块的大小为 28~212 字节,块是内存和外存交换信息的最小单位。